CHAPTER 4

A Recurrent, Excitatory-Inhibitory Neural Network Model of Morphological Processing

Empirical evidence from a variety of sources suggest that the role of morphological information changes over the time-course of word processing (e.g, see Rastle et al., 2000; Longtin et al., 2003; Feldman et al., 2004; Quémart et al., 2011). In particular, orthographic characteristics appear to more strongly impact what information is represented early in processing, just after a word is seen, whereas semantic information becomes more dominant if the input is sufficiently strong or long-lasting. The modeling efforts with a feed-forward neural network described in the previous chapter fell short of capturing how word recognition unfolds over time, as weights were learned in the context of generating a single, static pattern of activation in the semantic output layer. If, instead, a recurrent model is used, the inherently cyclical nature of activation propagation requires that learned weights accommodate multiple phases of information processing, rather than a single passage of information from input to output.

Prior efforts to simulate the dynamics of cognitive processing have found greater success when connectivity is constrained in a biologically plausible manner. Laszlo & Plaut (2012) successfully captured observed dynamics of the N400, an ERP component associated with semantic processing, with a recurrent neural network of written word comprehension by constraining connectivity among units to imitate the configuration of connectivity among actual neurons in the cortex. More specifically, weight values were constrained such that no unit can have both positive and negative outgoing connections (i.e., each unit is either excitatory or inhibitory) and negative weights can only occur within layers, not between layers (also see Cheyette & Plaut, 2017). Given prior neural-network simulations' success in capturing the dynamics of lexical processing, the separation of excitatory and inhibitory weights was also instituted in the current simulation work.

Recurrent networks take longer to learn than feedforward networks, and the added complexity of constraining weights to be either excitatory or inhibitory introduces further challenges to the training process. Thus, in the exploratory simulations presented in this chapter, neural networks were trained to map from orthography to semantics for words from small, artificial languages designed to vary in their morphological richness, comparable to those used in Plaut & Gonnerman (2000) (see Chapter 1 for a detailed summary of these simulations).

The current modeling work aimed to capture several general characteristics of complex word processing: an increasing importance of semantic characteristics and a diminishing role of orthographic characteristics for longer prime durations, as well as an overall advantage of morphologically related primes relative to other types of priming. Given that opaque priming is often cited as primary evidence of decomposition mechanism in morphological processing, the ability of recurrent, excitatory-inhibitory (E-I) networks to capture how opaque priming effects change with increasing prime duration, and for more or less morphologically rich languages, was also of particular interest. Finally, the current models' ability to account for developmental trends in morphological processing research, particularly how opaque priming emerges with reading experience, was assessed by testing the models at two earlier time points in training as well as after training was complete.

4.1 SIMULATION

As in Chapter 3, the LENS neural network simulator (Rohde, 2003) was used to build, train and test the network.

4.1.1 Network Architecture

The E-I recurrent network consisted of 65 input units, two 300-unit hidden layers, and a 200-unit output layer. Additionally, the hidden and output layers each had a corresponding inhibitory layer: 90 inhibitory units for each hidden layer, 50 inhibitory units for the

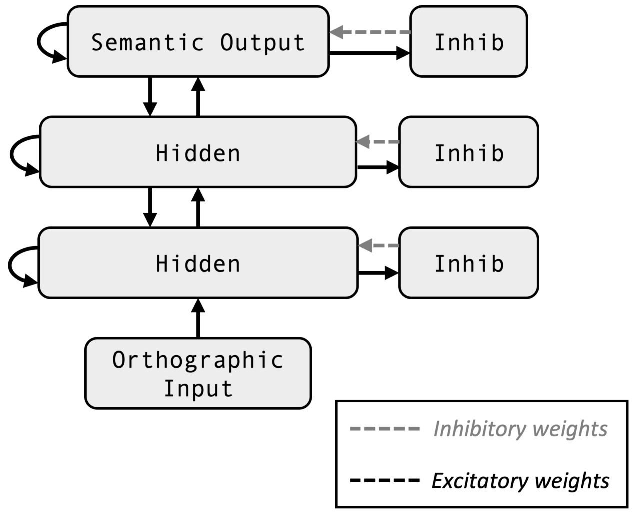

Figure 4.1: The network architecture trained to map from orthographic to semantic representations for morphologically impoverished and morphologically rich languages.

output layer (see Figure 4.1). All units except those in the orthographic layer also received input from a bias unit. All units used a rectified linear (ReLU) activation function, and the network was set to process 5 ticks (activation updates) per time interval.

Due to difficulties in training the network without it becoming stuck in a local minimum of error, only connections to and from the inhibitory layers were strictly enforced as positive or negative (positive if feeding into, negative if issuing out of, the inhibitory layer). All other weights were simply initialized on their respective side of 0, and over training the vast majority of weights remained on their initial side of 0.

Initial values for feedforward weights between non-inhibitory layers were positive, ranging from 0.04 to 0.35 in magnitude. Weights from a non-inhibitory layer to itself or to a previous layer were more weakly positive, ranging from 0.01 to 0.15. Weights from each hidden and output layer to their respective inhibitory layer ranged from 0.06 to 0.5 in magnitude, and were constrained to stay positive over training. Weights from inhibitory layers back to their primary layer had strong negative values (-0.4 to -0.5) and were constrained to stay negative. The hidden, output, and inhibitory layers also received strong inhibition from a bias unit (bias weights ranged from -1.6 to -2.4 for hidden and output layers, exactly -3.0 for inhibitory layers), and these connections were frozen such that they could not be updated during training.

The initial network structure and weight values were tailored to meet the following criteria: First, when the network received input, activation needed to reach the final output layer before being suppressed by the inhibitory units. Second, once hidden and output layers had turned off, inhibitory units needed to deactivate prior to the next example's input onset as well, so as to not suppress propagation of activation in a subsequent example.

4.1.2 Training

The network was trained to map from orthographic to semantic representations of words, as in Chapter 3. However, the words presented in this simulation were constructed specifically to contain varying degrees of systematicity in the mapping from orthography to semantics.

4.1.2.1 Artificial Languages

As in the Plaut & Gonnerman (2000) simulations, the network was trained on two artificial languages: one morphologically rich, one morphologically impoverished. Both languages included 15 possible letters, used to construct thirty 4-letter stems and twenty 2-letter second syllables. Eight of the second syllables were designated as affixes. Stems were designated as *transparent*, *quasi-transparent*, *opaque*, or experimental, and this designation informed how the meanings of words containing that stem were constructed (see the description of semantic representations below). In both languages, words constructed from the experimental stems consisted of two words ending in non-affix endings, and six words ending in affixes (two semantically transparent, two intermediate, and two semantically opaque). Words containing experimental stems were used exclusively for network testing, to ensure differences in effects were due to language differences and not word-specific or stem-specific differences.

In the morphologically impoverished language, 8 of the non-experimental stems were transparent, 6 were quasi-transparent, and 6 were opaque. In the morphologically rich language, 14 were transparent, 4 were quasi-transparent, and 2 were opaque. In each language, 240 words were constructed, eight words per stem. For transparent and quasi-transparent stems, 6 words were formed by combining the stem with an affix (all with meanings transparently or quasi-transparently related to stem meanings, respectively), and 2 were formed by combining the stem with a non-affix second syllable. For words formed from opaque stems, 2 were made by combining the stem with an affix (all with meanings unrelated to stem meanings), and 6 by combining the stem with a non-affix second syllable. Nonwords were random combinations of 6 letters of the 15 used in the language, removing any resulting strings that existed as words in the language or that were one letter off from matching a word in the language.

4.1.2.2 Representations

Orthographic representations used in the orthography-to-semantics training were learned from position-specific letter strings by a separate auto-encoder. Stems and affixes were constructed as random strings of length 4 (stem) or 2 (affix), selected from a set of 15 possible letters. The input (and output) of the auto-encoder consisted of 15 letter units at each of 8 positions. During training, a given 6-letter word was presented at each three positions: centered, or shifted one position to the left or right. For each of these presentations, the network learned a hidden representation (of 65 units) that could generate a corresponding position-specific representation at each of 3 possible output positions (indicated by the activation of one of three "position" units which projected only to the output side of the auto-encoder). For example, given the word GNPRNL starting from position 2 in the input layer (_GNPRNL_) it might be reproduced on a given trial starting from position 1 in the output layer (GNPRNL_ _). Given this training, the representations in the autoencoder's hidden layer learned to maintain information about letter identity and relative letter positions, while being consistent across 3 possible word positions. The 65-unit encoded representations for all three position presentations were used for training the E-I network, all mapping to the same semantic representation so that the network would be robust to variability in orthographic input. Nonword orthographic representations were generated by presenting nonword strings in a random starting position to the same autoencoder and saving the resulting hidden layer activation pattern.

Semantic representations were generated by randomly turning on 4% of 200 units for stem meanings, and 1% for affix meanings. The meanings of a given stem and affix were superimposed to make the meaning of the word they formed, and then transformed to a degree corresponding to the word's assigned transparency. For words formed from transparent stems, 6 words ended in affixes and no transformation of their meanings occurred after combining stem and affix meanings, while 2 words ending in non-affix word endings and had meanings completely unrelated to the stem (meanings were intentionally generating such that there was no overlap in units with a value of 1 between word meaning and stem meaning). The same was true for quasi-transparent stems, except the 6 words with affixes underwent partial transformation of the cumulative meaning (half of "on" units were swapped for other units). Opaque stems formed 6 words with non-affix endings and 2 with affix endings; all meanings were completely unrelated to the stem meaning. For "nonword" examples, all semantic targets were 0.

4.1.2.3 Procedure

Examples were presented to the network in a random order, and unit activations were not reset between examples, to imitate the demands of moving from recognizing one word to recognizing the next. Input units were soft-clamped (activations took on an intermediate value between activation at the previous tick and the unit's input), making input values change gradually over multiple *ticks*, or simulated time steps. At the beginning of each training example, no input was presented for a random duration between 2 and 15 ticks, making gaps between stimulus presentations variable. Orthographic inputs were turned on (gradually) for between 25 and 35 ticks, and targets were turned on (i.e., error was accumulated for output activations' distance from target values) for the final 20 ticks that target inputs were on. Thus, the network had to optimize performance for an unknown gap between words, and an unknown but occasionally brief amount of time between orthographic presentation and having to generate correct semantic activation.

Networks were trained using back-propagation, implemented with a bounded variation of momentum descent (`dougsMomentum` in Lens) with a learning rate of 0.05. A batch size of 2000 was used, in which approximately 1 in 5 examples was a nonword (drawn from 4,300 unique options) and the remaining examples were sampled with replacement from the 720 word inputs (240 words in 3 different positions). Weight values were saved every 100 epochs up to 1000 epochs, and every 1000 epochs subsequently, to allow analysis during multiple phases of training. Training ended when all output units across all examples were within 0.5 of their target value (during all ticks of the example for which target penalties were applied).

4.1.3 Testing

Networks were tested at two points during training, as well as at the end of training, via a procedure comparable to primed lexical decision. The actual process of deciding whether an input was a word or nonword was not directly simulated (unlike in, for example, Cheyette & Plaut, 2017). Instead, the decision was approximated as the moment when the overall activation for semantic units meant to be on for the target word passed a certain threshold activation.

4.1.3.1 *Examples*

Related prime-target pairs were selected using words built from the experimental stems, such that the two transparent words for each experimental stem were always used as targets, and paired with themselves (identity priming), each other (transparent morphological priming), the intermediate words for that stem (quasitransparent morphological priming), the opaque words for that stem (opaque morphological priming) and the unaffixed words for that stem (orthographic priming). Thus, the same 20 targets were used across all priming conditions. Unrelated primes were selected for each related prime-target pair to minimize orthographic and semantic relatedness.

4.1.3.2 *Procedure*

Prime-target pairs were presented to the network in a random order, without resetting units' output values between examples. The example began with 20 ticks of no input, followed by a prime whose duration varied from 1 to 10 ticks in length. The orthographic representation of the prime was presented immediately following that of the target (although, given the soft-clamping of input units, the transition from prime to target occurred smoothly over about 5 ticks) and was presented for 30 ticks.

The number of ticks it took for units in the output layer of the network with target values of 1 (corresponding to which units were activated in the semantic representation of the target word) to reach an average activation of 0.5 was used as a proxy for reaction time (see Figure 4.2).

4.2 RESULTS

4.2.1 Training

Training the network to meet the performance criterion required 32,900 and 17,200 epochs for the morphologically impoverished and rich languages, respectively. The greater time

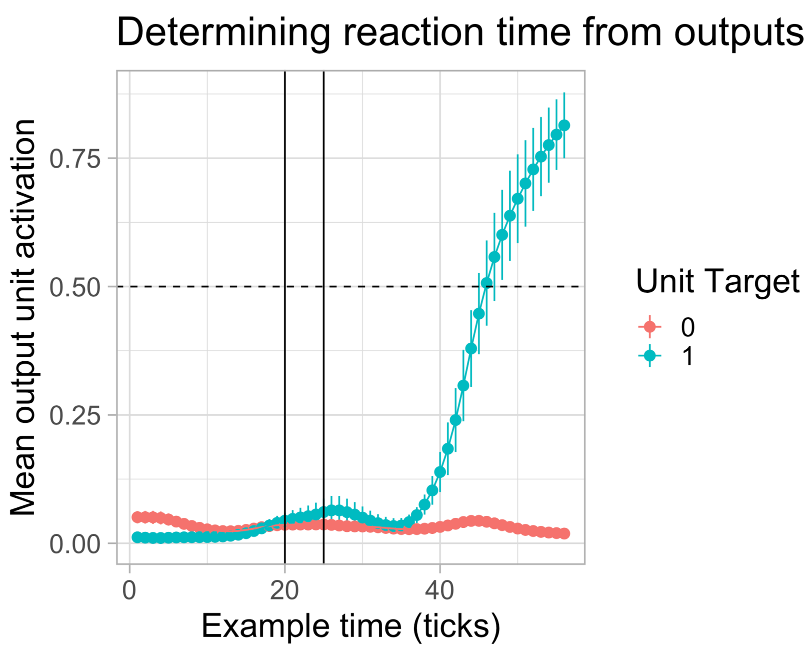

Figure 4.2: Mean activations of output layer for a single testing example with a prime duration of 5 ticks, separately for units whose target values are 1 or 0 for the primed word's semantic representation. Reaction times are estimated as the tick at which mean activation of output units with a target value of 1 surpass 0.5 (dotted line). The leftmost black line denotes onset of prime word input; the rightmost denotes onset of target word input.

needed to train the morphologically impoverished language is unsurprising given the network learned a less systematic mapping from orthography to semantics in this case. Two time points in training prior to reaching the criterion, one with overall network error closest to 20,000 and one with overall network error closest to 5,000, were selected for each language. Testing was conducted at these time points (15,000 and 23,000 epochs for the impoverished language; 5,000 and 10,000 epochs for the rich language) as well as the endpoint of training, to provide insight into patterns of effects during acquisition.

4.2.2 Testing

Examples for which the reaction time threshold was never reached, or for which the threshold was already surpassed prior to target presentation[1], were removed from analysis, resulting in the removal of 10,544, 6,417, and 4,030 out of 19,800 examples for the morphologically impoverished language at time points 1, 2 and 3, respectively. For the morphologically rich language, 1,913, 680, and 2,464 examples were removed at time points 1, 2, and 3, respectively. Simulated reaction times (number of ticks to cross the mean output activation threshold) for prime-target testing examples were fitted with linear mixed effect models, including random intercepts for target word. Changes in relative simulated priming magnitudes with increased prime duration were not monotonic; For instance, opaque priming appears to grow stronger and subsequently weaker relative to orthographic priming as prime duration increases, for the network fully trained on the morphologically rich language. To allow ease of model interpretation given this feature of model behavior, only the subset of data for which prime duration was 1, 4, 7, or 10 ticks was included, and prime duration was treated as categorical rather than a continuous.

Assessing model fit with an ANOVA revealed a significant four-way interaction of training language, priming condition, relatedness, and prime duration for all three points in acquisition ($F_{T1} = 1.74$, $p = 0.037$; $F_{T2} = 5.93$, $p < 0.0001$; $F_{T3} = 7.98$, $p < 0.0001$).

[1]The semantic output threshold was occasionally surpassed prior to presentation of the target if the prime was sufficiently semantically related to the target and presented for a sufficiently long duration.

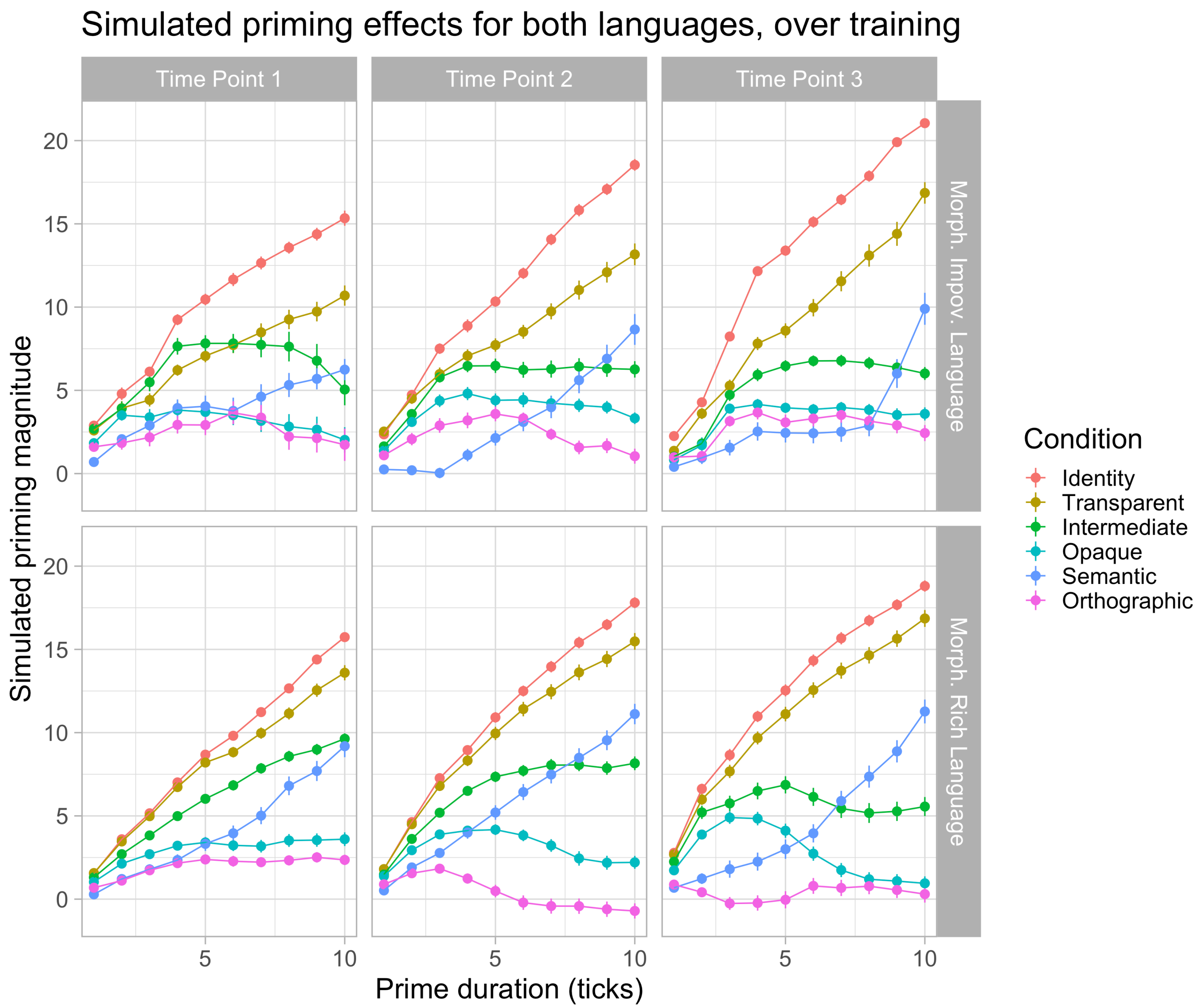

Figure 4.3: Mean simulated priming magnitudes from network testing with varying prime-target relations and prime durations, across languages and training time points. Error bars denote standard error.

Morph. impov.	Time 1			Time 2			Time 3		
prime dur. (ticks)	*4*	*7*	*10*	*4*	*7*	*10*	*4*	*7*	*10*
Transp, Id	-2.35	-3.57*	-4.06*	-2.00	-3.93*	-4.41*	-4.13*	-6.17*	-3.77*
Transp, Sem	1.38	1.39	0.56	3.18*	4.03*	1.90	3.27	5.84*	6.59*
Op, Orth	1.04	-0.24	0.81	1.25	1.29	2.41*	0.50	0.77	0.58
Morph. rich	Time 1			Time 2			Time 3		
prime dur. (ticks)	*4*	*7*	*10*	*4*	*7*	*10*	*4*	*7*	*10*
Transp, Id	-0.41	-1.13	-1.56*	-0.61	-1.33	-1.99*	-1.12	-1.69	-1.81*
Transp, Sem	2.66*	2.87*	2.62*	3.23*	3.99*	3.57*	5.73*	5.66*	3.59*
Op, Orth	0.72	0.82	1.02	2.48*	3.03*	2.24*	3.65*	0.36	-0.18

Table 4.1: Beta values for interaction of relatedness, prime duration, and condition in predicting model RTs for select pairs of conditions, across languages and training time points. Effects for each prime duration are relative to those for a prime duration of 1 tick. Asterisks denote effects with p-values below $\alpha = 0.00093$.

Smaller models were then fitted to investigate planned contrasts of priming conditions for each combination of language and training time point. The contrasts of transparent and identity, transparent and semantic, and opaque and orthographic priming conditions were each assessed in a three-way interaction with relatedness and prime duration[2]. The Bonferroni correction was applied to these analyses, to account for the assessment of 54 separate hypotheses ($\alpha = 0.00093$). Results from these analyses are shown in Table 4.1, but the most notable patterns of effects are discussed below. See Figure 4.3 for simulated priming magnitudes with varying priming condition, prime duration and time point in training, for both languages.

Several overall patterns of effects across both languages, as evident in Figure 4.3, align with prior empirical findings. Identity priming and semantically transparent morphological priming are consistently stronger than the other conditions (for the rich language, final time point: $\beta = 8.01$, $p < 0.0001$; for the impoverished language, final time point: $\beta = 6.75$, $p < 0.0001$). Facilitation by morphological primes with intermediate

[2]The effects of priming for the other three prime durations, 4 ticks, 7 ticks, and 10 ticks were assessed relative to the 1-tick prime duration. The 1-tick prime duration serves as an informative control, as the orthographic representation of the prime word is not presented for long enough in this case to yield sufficient differentiation between priming conditions.

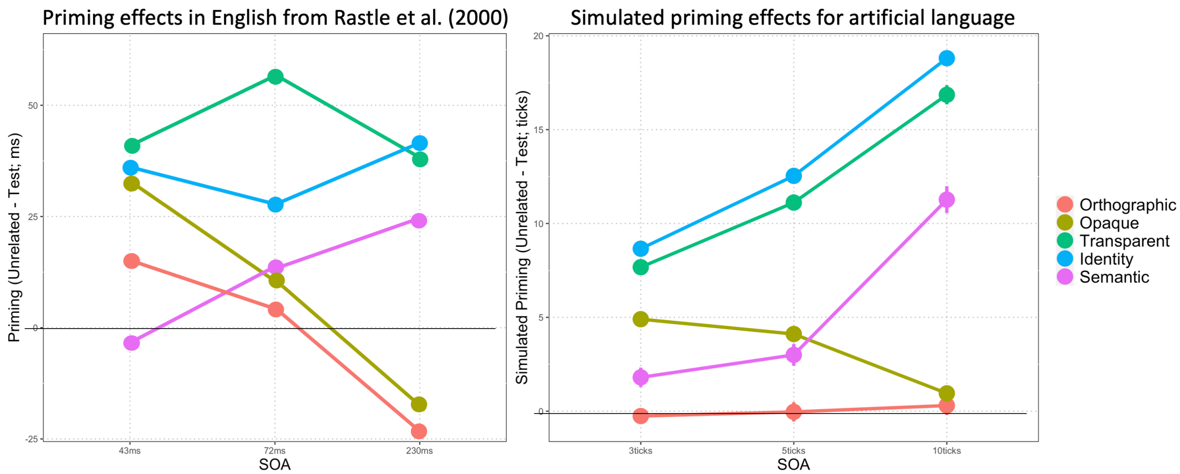

Figure 4.4: Direct comparison of empirical results from Rastle et al. (2000) and simulated results for the network fully trained on a morphologically rich language.

semantic transparency levels out to be more moderate with longer prime durations (see Gonnerman et al., 2007; Xu & Taft, 2015, for comparable empirical findings with quasi-transparent stems). Semantic priming grows stronger for longer prime durations (comparing 10 ticks with 4 ticks at final time point: $\beta = 9.46$, $p < 0.0001$ for rich language; $\beta = 6.63$, $p < 0.0001$ for impoverished language). More generally, and particularly by the end of training, semantically related primes (those in the identity, transparent, semantic, and intermediate conditions) show the greatest facilitation for speedy recognition of the target word when prime presentation duration is long, whereas opaque and orthographic priming facilitate more weakly if at all.

As evident in Figure 4.4, comparing simulated effects from the morphologically rich language with results from the time-course study reported by Rastle et al. (2000), the most noticeable difference between these two patterns of results is the lack of an inhibitory effect of orthgographic and opaque priming for long prime durations. This could be due to a lack of variety in artificial words' meanings, or a lack of sufficient distance within the semantic space made possible by a mere 200 output units; semantic processing cannot be brought too far off track in such a constrained context.

Some patterns of effects differ notably between the two languages. First, the difference between semantically transparent morphological priming and identity priming is less notable for the rich language than for the impoverished language. This trend is supported by the statistics shown in Table 4.1: in the rich language, transparent priming is only found to be significantly weaker than identity priming for the longest prime duration (relative to the shortest prime duration), while in this effect is larger and more often significant in the case of the impoverished language. Similarly, the difference between semantically transparent morphological priming and semantic priming is greater and more often significant (again, see Table 4.1). These two distinctions between the two languages can be summarized as an overall stronger effect of transparent priming in the morphologically rich language, such that it behaves less like semantic priming and more like identity priming.

The comparison of semantically opaque morphological priming and orthographic priming effects also reveals notable differences between languages. As understanding and contextualizing opaque priming across language, development, and task was a primary goal of this simulation work, these results are described in more detail. For the two fully-trained models, no advantage of opaque priming over orthographic priming is found in the case of the impoverished language, while a significant advantage is found for a 4-tick prime duration (relative to a 1-tick prime duration) in the case of the morphologically rich language. The network trained on the morphologically rich language demonstrates a pattern of priming effects comparable to that of a skilled reader of languages such as English (Rastle et al., 2000, 2004) or French (Longtin et al., 2003).

Turning to these effects' developmental emergence, opaque priming is not found to be significantly stronger than orthographic priming in either language at the earliest time point in training. Thus, the advantage of opaque priming over orthographic priming takes a good deal of training to emerge, in alignment with the findings of Beyersmann et al. (2012a) and Dawson et al. (2021). At the second time point, however, a significant advantage of opaque over orthographic priming is found for all three prime durations for the morphologically rich language, whereas such an advantage is only found for 10-ticks relative to 1-tick for the morphologically impoverished language. It is worth noting that, although opaque priming for the impoverished language did not reach significance with the Bonferroni-corrected α at 4-tick and 7-tick prime durations at time point 2, effects were greater in magnitude for these two prime durations than for any effects at time point 1 or time point 3; that is, the same pattern of opaque priming effects may be manifesting at time point 2 in both languages, to differing degrees. Thus, semantically opaque morphological priming in these networks appears to emerge first as a stronger and longer-lasting effect, before attenuating or being relegated to early processing with additional training. Although unexpected, these results do not necessarily conflict with the currently available results on morphological processing in developing readers (see chapter discussion).

One final noteworthy difference between the morphologically rich and impoverished languages is that, once fully trained, both opaque and orthographic priming appear to be more facilitatory at long prime durations for the impoverished language ($\beta = 2.71$,$p < 0.0001$) relative to the rich language ($\beta = 0.38$, $p = 0.15$). The network trained on the morphologically impoverished language, lacking aid from systematic morphological structure and subject to pressure to activate accurate semantics as quickly as possible, may learn to make use of orthographic similarity regardless of the prime's ending.

4.3 DISCUSSION

The neural network model presented in this chapter captures several general aspects of the time-course of morphological processing in skilled readers. The increasing role of semantics and diminishing aid of orthographic information for longer prime durations in this model echoes the results of morphological priming studies where prime duration has been varied (e.g., Rastle et al., 2000; Longtin et al., 2003; Feldman et al., 2004); to our knowledge, these aspects of complex word processing dynamics have not been simulated previously.

At the final training time point, the finding of significantly greater opaque priming relative to orthographic priming in the morphologically rich language but not in the morphologically impoverished language corresponds with our prediction that greater morphological richness should give rise to stronger opaque priming effects. However, it appears that the morphologically rich language in these simulations may be more likened to English or French (Longtin et al., 2003; Rastle et al., 2004) than to Hebrew, Arabic, or German (Frost et al., 2000; Boudelaa & Marslen-Wilson, 2001; Smolka et al., 2009), as opaque priming is found for the 4-tick prime duration but not for longer ones. These results motivate the examination of an even more morphologically rich language, to determine if overt opaque priming can be detected even in a fully-trained model in that case.

The current model better captures the time-course of complex word recognition by being trained in a dynamic and time-pressured context. The feedforward model (Chapter 3) was trained in a context where all orthographic input is immediately and fully processed. Thus, there was no phase of recognition where making use of information beyond than the full identity of the word was particularly advantageous. However, in the recurrent model presented in this chapter, the network is trained such that weights are pressured to take advantage of information at any level of complexity as it becomes available, such that words can be recognized under increasingly time pressured scenarios. Varying the amount of time before targets turned on among training examples allowed for slower processing techniques to yield some reduction of overall error during early phases of training, while motivating more efficient strategies, such as early sensitivity to orthographic structures that typically denote morphological complexity, to eventually play a more significant role. The developmental trajectory of these effects over training also provides insights regarding how such processes emerge.

Although the lack of opaque priming for both languages at training time point 1 is unsurprising, the stronger and long-lasting effects of opaque priming for both languages at time point 2 was not expected. It seems that at an intermediate phase of training, these networks become attuned to the orthographic structures typically associated with morphological relatedness, and these structures are made use of in a way that impacts how the prime word is represented even once it has been presented for a considerable duration (e.g., 10 ticks). However, processing complex words in this way will lead to new errors, due to the opaque and quasi-transparent words present in both languages. Thus, later in training the network attenuates the use of this type of sensitivity (in the case the impoverished language) or overrides its use with higher-level information as it becomes available (in the case of the rich language).

If this developmental phenomenon reflects the actual emergence of semantically opaque morphological priming, we would expect to see opaque priming emerging for both masked and overt priming contexts simultaneously, followed by a weakening of opaque priming in the overt context in languages in which skilled readers don't show opaque, overt priming. In English, masked priming by semantically opaque words does not typically emerge until developing readers have gained quite a bit of reading experience (Beyersmann et al., 2012a; Dawson et al., 2021); however, no developmental studies of morphological processing in English have looked at the emergence of sensitivity to semantically opaque words in an overt priming context.

To our knowledge, only one developmental study of morphological processing, conducted in French, has systematically manipulated prime duration: Quémart et al. (2011) investigated sensitivity to transparent, opaque, orthographic, and semantic priming across three prime durations (60 ms, 250 ms, and 800 ms) and two age groups (3rd to 7th graders, and adults). For the 60 ms prime, semantically transparent and opaque primes significantly and equivalently facilitated lexical decision for both the children and the adults, whereas semantic and orthographic priming were not significant for either group. However, for the 250 ms prime, transparent and opaque priming were significant and equivalent for children, whereas adults showed only significant transparent priming and marginal semantic priming. These results have not been replicated, in French or any other language, but if they are robust they may mirror the simulated pattern of opaque priming emergence found here. See Chapter 8 for further discussion and investigation of longer-lasting opaque priming in early readers.

Although the exploratory modeling efforts described in this chapter are a step forward in our understanding of morphological processing dynamics, a good deal remains to be done. First, these networks do not entirely abide by the constraints of a true excitatory-inhibitory network, as the constraint of only positive weights within layers was relaxed in order to facilitate training. It is unclear whether enforcing such constraints more strictly would notably impact the patterns of effects reported here. Second, all words in the artificial languages presented to the network were complex, and all prime-target pairs consisted of two complex words, whereas the majority of empirical results to which these simulated findings are compared were conducted with a complex word priming a simple target. Whether such a distinction impacts the simulated pattern of effects found here must be explored in future work. Additionally, the process of lexical decision was not suitably captured in the current model. A large number of examples had to be removed from testing due to the nature of the semantic activation threshold used to approximate lexical decision, leading to an imbalance in number of testing examples across the two languages. This issue could be addressed by implementing a leaky competing integrator model of decision making (see Cheyette & Plaut, 2017; Usher & McClelland, 2001). Finally, and most notably, the languages presented to these networks were exceedingly small (only 240 words, each presented at three different positions) and obviously do not capture the true complexity of real language or the contexts in which it is encountered. Expanding the current model and speeding up training procedures such that the model can be trained on authentic language input, as done with the feedforward network in Chapter 3, would greatly enhance our ability to interpret the implications of simulated results for real word processing. Presenting realistically ordered examples (e.g., THE, DOG, BARKED instead of DOG, TREE, COULD) or incorporating phonological information (see Plaut et al., 1996) would also be worthwhile steps towards a more ecologically plausible future model.

BIBLIOGRAPHY

Alegre, M., & Gordon, P. (1999). Frequency Effects and the Representational Status of Regular Inflections. *Journal of Memory and Language*, *40*(1), 41–61.

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*(JUL), 1–12.

Amenta, S., Crepaldi, D., & Marelli, M. (2020a). 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. (2020b). 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*. 62. 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.

Armstrong, B. C., Watson, C. E., & Plaut, D. C. (2012). Sos! an algorithm and software for the stochastic optimization of stimuli. *Behavior research methods*, *44*(3), 675–705.

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, MA: 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. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Baayen, R. H., Milin, P., Ević, 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., & Smolka, E. (2020). Modeling Morphological Priming in German With Naive Discriminative Learning. *Frontiers in Communication*, *5*(April).

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.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The english lexicon project. *Behavior research methods*, *39*(3), 445–459.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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. (2012a). 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. (2016a). Embedded stem priming effects in prefixed and suffixed pseudowords. *Scientific Studies of Reading*, *20*(3), 220–230.

Beyersmann, E., Coltheart, M., & Castles, A. (2012b). 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*, *208*.

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

Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., & Grainger, J. (2016c). Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin and 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.

Blevins, J. P. (2016). *Word and paradigm morphology*. 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. Aronoff, & M. L. Kean (Eds.) *Juncture*. Cambridge, Mass: Academic Press.

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389.

Brontë, C. (2002). *Jane Eyre*. Mineola, NY: Dover Publications, Inc. (Original work published 1847).

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Brysbaert, M., & New, B. (2009). Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior research methods*, *41*(4), 977–990.

Burani, C., 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.

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.

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*.

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*, *0*(0), 1–15.

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. (2010a). '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. (2010b). 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. (2019). Individual differences in morphological processing in developing and skilled readers. *Society for the Scientific Study of Reading*, *116*, 1–3.

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

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.

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods*, *47*(1), 1–12.

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

Deacon, H., Tong, X., & Mimeau, C. (2019). Morphological and semantic processing in developmental dyslexia across languages: 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.

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

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). Derivational morphology and skilled reading. In *The Cambridge Handbook of Psycholinguistics*, 30, (pp. 311–350). Cambridge University Press.

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.

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. MIT press.

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

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., & Kostić, 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*(March), 1–19.

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.

Feldman, L. B., Soltano, E. G., Pastizzo, M. J., & Francis, S. E. (2004). What do graded effects of semantic transparency reveal about morphological processing? *Brain and Language*, *90*(1-3), 17–30.

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 decomposition supports word recognition in primary school children learning to read: Evidence from masked priming of german derived words. *Journal of Research in Reading*.

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.

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

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.

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*.

Hasenäcker, J., Solaja, O., & Crepaldi, D. (2021). Does morphological structure modulate access to embedded word meaning in child readers? *Memory & Cognition*, *49*(7), 1334–1347.

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.

Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a foreign language*, *23*, 403–430.

Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural networks*, *1*(4), 295–307.

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.

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.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The british lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic english words. *Behavior research methods*, *44*(1), 287–304.

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.

Kuroki, D. (2021). A new jspsych plugin for psychophysics, providing accurate display duration and stimulus onset asynchrony. *Behavior Research Methods*, *53*(1), 301–310.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

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.

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*.

Lei Ba, J., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. *ArXiv e-prints*, (pp. arXiv–1607).

Leinonen, A., Grønholm-Nyman, P., Järvenpää, M., Søderholm, C., Lappia, O., Laine, M., & Krause, C. M. (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.

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific reports*, *10*(1), 1–11.

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 (Volume 1: Long Papers)*, (pp. 2059–2068).

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in psychology*, *6*, 1171.

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*, (pp. 315–333).

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., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.

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), e0171935.

Milin, P., Filipović Durdević, 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). John Wiley & Sons, Inc.

Morris, D. (2013). *Diagnosis and correction of reading problems*. Guilford Publications.

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

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). Springer, Boston, MA.

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, The Netherlands*.

Moscoso Del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Harald Baayen, R. (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., 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.

Moscoso Del Prado Martín, F., Kostić, 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.

Nagy, W., Berninger, V. W., & Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of educational psychology*, *98*(1), 134.

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

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*, (pp. 1532–1543).

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, *11*(4), 357–383.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and cognitive processes*, *12*(5-6), 765–806.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: empirical and computational support for a single-mechanism account of lexical processing. *Psychological review*, *107*(4), 786.

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*, (pp. 824–829).

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.

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.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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.

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*(7–8), 972–1001.

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

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, MA: 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, MA: 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.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26–43.

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

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 individual-level 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*(FEB), 1–21.

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.

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*(August), 1–21.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, *108*(3), 550.

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., Kaiser, L., & Polosukhin, I. (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*.

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*(3), 904–910.

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

Zeno, S. M., Ivens, S. H., Millart, R. T., & Duwuri, R. (Eds.) (1995). *The educators word frequency guide*. Brewster, NY: Touchstone Applied Science Associate, Inc.

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.

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.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of dutch compounds. *Language and cognitive processes*, *9*(3), 341–368.