Modeling morphological processing in human magnetoencephalography

Anonymous SCiL submission

Abstract

In this paper, we conduct a magnetoencephalography (MEG) lexical decision experiment and model morphological processing in the human brain, especially the Visual Word Form Area (VWFA) in the ventral visual stream. Five computational models of morphological processing are constructed and evaluated against human brain activities: Letter Markov Model and Syllable Markov Model as “amorphous” models without linguistically defined morphemes, and Morpeme Markov Model, Hidden Markov Model (HMM), and Probabilistic Context-Free Grammar (PCFG) as “morphous” models with different amounts of morphological supervision. The present experiment demonstrates that “amorphous” models underperformed relative to “morphous” models, PCFG was more neurologically accurate than sequential models, and PCFG better explained nested words with non-local dependencies between prefixes and suffixes. These results strongly suggest that morphemes are represented in the VWFA and parsed into hierarchical morphological structures.

1 Introduction

Under the single-route decomposition model of morphologically complex visual word recognition (Taft, 1979, 2004; Taft and Forster, 1975), there are three functionally different stages of morphological processing: morphological decomposition, lexical access, and morphological recombination. In the first stage of morphological decomposition, morphologically complex words are visually decomposed into component morphemes. In the second stage of lexical access, meanings of decomposed morphemes are lexically retrieved. In the third and last stage of morphological recombination, retrieved meanings of decomposed morphemes are semantically composed.

In the cognitive neuroscience literature, Fruchter and Marantz (2015) employed magnetoencephalography (MEG) to spatiotemporally dissociate those stages of morphological processing. Specifically, the first stage of morphological decomposition has been indexed by M170 (Solomyak and Marantz, 2009, 2010; Zweig and Pylkkänen, 2009; Lewis et al., 2011; Fruchter et al., 2013; Gwilliams et al., 2016) or Type II (Tarkiainen et al., 1999; Helenius et al., 1999) evoked response components in the ventral visual stream (Pylkkänen and Marantz, 2003; Hickok and Poeppel, 2007). Dehaene et al. (2005) proposed local combination detectors (LCDs) where linguistic units such as letters, syllables, and morphemes are convolutionally represented and processed in the ventral visual stream from posterior occipital to anterior inferior temporal cortices and, importantly, morphological decomposition has been localized to the left fusiform gyrus known as the Visual Word Form Area (VWFA; Cohen et al., 2000, 2002; Dehaene et al., 2001, 2002). For example, Solomyak and Marantz (2010) and Lewis et al. (2011) computed transition probabilities between linguistically defined morphemes (e.g. \( P(\text{Affix}|\text{Stem}) \)) to successfully predict VWFA activities associated with real (e.g. teach-er) and pseudo (e.g. corn-er) bimorphic words, respectively. These results have strongly shown that morphemes are represented and processed in the VWFA.

However, “amorphous” models without linguistically defined morphemes have been recently proposed in morphological processing (Baayen et al., 2011; Milin et al., 2017; Virpioja et al., 2017; Hakala et al., 2018). For instance, Virpioja et al. (2017) and Hakala et al. (2018) employed Morfessor, an unsupervised model of morphological processing with statistically induced “morphs” (Creutz and Lagus, 2007), to predict human re-
action times and brain activities. Furthermore, as correctly pointed out by Libben (2003, 2006); de Almeida and Libben (2005); Pollatsek et al. (2010), bimorphemic words exclusively tested in the previous literature (Solomyak and Marantz, 2009, 2010; Lewis et al., 2011; Fruchter et al., 2013; Gwilliams et al., 2016) cannot distinguish sequential morphological decomposition from hierarchically morphological parsing of morphologically complex words into hierarchical morphological structures (Song et al., 2019; Oseki et al., 2019). Therefore, whether morphemes are represented in the VWFA and, if so, processed sequentially or hierarchically remains to be an empirical question.

In this paper, we conduct an magnetoencephalography (MEG) experiment where participants perform visual lexical decision on morphologically complex words and, generalizing the computational modeling methodology developed in sentence processing (Brennan et al., 2016; Nelson et al., 2017) to morphological processing, model morphological processing in the human brain, with special focus on the VWFA in the ventral visual stream. Specifically, we construct five computational models of morphological processing, Letter Markov Model and Syllable Markov Model as “amorphous” models without linguistically defined morphemes and Morpheme Markov Model, Hidden Markov Model (HMM), and Probabilistic Context-Free Grammar (PCFG) as “morphous” models with different amounts of morphological supervision, and evaluate them against human brain activities in the VWFA to investigate which model is most neurologically accurate.

2 Methods

2.1 Participants

The participants were 26 native English speakers recruited in New York City. All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) and with normal or corrected-to-normal vision. They provided written informed consent and were paid $15/hour for their participation. We excluded 6 participants based on their behavioral performance: 3 participants excluded due to low accuracy (< 75%) and 3 participants excluded due to slow (> 2000 ms) or fast mean reaction times (< 500 ms). Thus, 20 participants were included in the statistical analyses (10 males and 10 females, $M = 28.4, SD = 9.27$).

2.2 Stimuli

The stimuli were 800 morphologically complex trimorphemic words and nonwords. The stimuli selection procedure consisted of several steps. First, 600 trimorphemic words were created based on CELEX (Baayen et al., 1995) in accordance with syntactic, morphological, and phonological selectional restrictions of outer derivational affixes, but without semantic selections explicitly taken into consideration. In this sense, these trimorphemic words constitute the potential lexicon (Halle, 1973) of grammatical (“possible”) but not necessarily acceptable (“actual”) words (cf. Bauer, 2014; Embick, 2016). These 600 trimorphemic words were subcategorized into 300 linear words \([x \ [y \ [z \ \sqrt{\text{Root}} \ \text{Suffix}] \ \text{Suffix}] \text{with productive derivational suffixes (Plag and Baayen, 2009)}\) and 300 nested words \([x \ \text{Prefix} \ [y \ [z \ \sqrt{\text{Root}} \ \text{Suffix}]] \text{ with productive derivational prefixes (Zirkel, 2010)}\). Furthermore, these trimorphemic words have zero surface frequencies in CELEX, thereby encouraging participants to actively perform morphological parsing, according to the Relative Frequency Hypothesis (Hay, 2001, 2002). Second, in order to weed out extremely implausible words, 600 trimorphemic words were normed with the crowdsourced acceptability judgment experiment, where participants rated them on 1~7 Likert scale. Third, 500 trimorphemic words (250 linear and 250 nested) with higher acceptability judgments (> 3.5) and lower standard deviations (< 2.5) were selected and 500 trimorphemic nonwords (250 linear and 250 nested) were also created based on CELEX in violation of syntactic selectional restrictions of inner derivational suffixes, resulting in 1000 trimorphemic stimuli. Fourth, in order to ensure that words and nonwords are correctly judged as such, 1000 trimorphemic stimuli were further normed with the crowdsourced lexical decision experiment, where participants judged whether or not presented stimuli were possible English words as quickly and accurately as possible. Finally, 400 trimorphemic words (200 linear and 200 nested) and 400 trimorphemic nonwords (200 linear and 200 nested) with higher accuracies (> 75%) were selected, resulting in the balanced and extensively normed set of 800 trimorphemic stimuli to be tested in this experiment. The stimuli are summarized in Table 1.
was implemented with the
module. The
model module. The
Markov Model can be defined by an n-order
Markov process that computes the transition prob-
abilities of morphemes at position i given the i–n
case context, e.g. \( P(m_i|m_{i-n}, m_{i-1}) \). When \( n = 1 \), the
1st-order Markov Model (i.e. bigram model) com-
putes the transition probabilities of morphemes at
position i given the i–1 context, e.g. \( P(m_i|m_{i-1}) \).
When \( n = 2 \), the 2nd-order Markov Model (i.e.
trigram model) computes the transition probabil-
ities of morphemes at position i given the i–2 con-
text, e.g. \( P(m_i|m_{i-1}, m_{i-2}) \). Given the Markov
assumption, the local probabilities of component
morphemes in morphologically complex words are
merely their transition probabilities.

The transition probabilities are the model pa-
rameters empirically estimated from morpheme
sequences in the training data via Maximum Like-
lihood Estimation with token weighting and Lid-
stone smoothing at \( \alpha = 0.1 \). The Markov Model
is linear and string-oriented in that the transition
probabilities merely track morphemes from left to
right, which should effectively capture lexically
specific dependencies among morphemes.

### 2.3.2 Hidden Markov Model

A HMM was implemented with the \texttt{hmm} mod-
ule. A HMM generalizes the Markov Model by
hypothesizing “hidden” structures behind visible
strings. The HMM computes the transition proba-
bilities of POS tags at position i given the i–1 con-
text, e.g. \( P(t_i|t_{i-1}) \), and the emission probabili-
ties of morphemes at position i given POS tags at
the same position, e.g. \( P(m_i|t_i) \). Although the
HMM, like the Markov Model, can be defined by
an n-order Markov process over POS tags, only the
Bigram HMM is investigated in this paper. The lo-
cal probabilities of component morphemes in mor-
phologically complex words are the ratio of prefix
probabilities at position k to position k–1, where
prefix probabilities are the sum of path probabili-
ties compatible with morphemes until position k
(Rabinar, 1989).

While the local probabilities of component mor-
phemes in structurally ambiguous words can be
computed via a forward algorithm (sum of all
paths) or a Viterbi algorithm (max of all paths),
given that most probability mass was allocated to
the best path and thus there were no sub-
stantial differences between forward and Viterbi
algorithms, we adopted the forward algorithm.
Both transition and emission probabilities are

|   | Linear | Nested |
|---|---|---|
| Word | [X](n=200) Y  | [X](n=200) Y |
|   | Z al   | Z al   |
|   | \(\sqrt{Digit}\) | \(\sqrt{Culture}\) |
| Nonword | [X](n=200) Y  | [X](n=200) Y |
|   | Z ion | Z ion |
|   | \(\sqrt{Gulf}\) | \(\sqrt{Kid}\) |

Table 1: Stimuli. The horizontal dimension is syntactic structure: linear vs. nested. The vertical dimension is
lexical status: word vs. nonword. Therefore, there are 4 stimulus types, 200 stimuli for each type, 800 stimuli
in total. The asterisk (*) on subtrees (Y) in nonwords indicates that inner derivational suffixes violate syntactic
selectional restrictions on roots.
the model parameters empirically estimated from tagged morpheme sequences in the training data via Maximum Likelihood Estimation with token weighting and Lidstone smoothing at $\alpha = 0.1$. The HMM is structure-oriented in that hidden structures of POS tags are hypothesized behind visible strings, but still linear because the transition probabilities track POS tags from left to right.

2.3.3 Probabilistic Context-Free Grammar

A PCFG was implemented with the grammar module. A PCFG is most representationally sophisticated among three types of computational models investigated in this paper and, crucially, can model hierarchical structures. The PCFG computes nonterminal production probabilities of right-hand sides given left-hand side nonterminals, e.g. $P(rhs|lhs)$, and terminal production probabilities of right-hand side terminals given left-hand side nonterminals, e.g. $P(m_i|t_i)$, corresponding to HMM emission probabilities. The local probabilities of component morphemes in morphologically complex words are the ratio of prefix probabilities at position $k$ to position $k-1$, where prefix probabilities are the sum of tree probabilities compatible with morphemes until position $k$ (Earley, 1970; Stolcke, 1995). Note that HMMs and PCFGs make different predictions even for bi-morphemic words because derivational affixes are head-lexicalized in PCFGs (e.g. N → V er), while “emitted” from POS tags in HMMs.

Just like HMMs, while the local probabilities of component morphemes in structurally ambiguous words can be computed via an Earley algorithm (sum of all trees) or a Viterbi algorithm (max of all trees), we employed the Earley algorithm which may have interesting consequences for the incremental nature of morphological processing. Both nonterminal and terminal production probabilities are the model parameters empirically estimated from morphological tree structures in the training data via Maximum Likelihood Estimation with token weighting and Lidstone smoothing at $\alpha = 0.1$. The PCFG is hierarchical and structure-oriented in that the probabilities are defined over hierarchical structures permitted by the grammar.

2.4 Evaluation metrics

The information-theoretic complexity metric, surprisal, was employed as linking hypothesis that bridges the gap between representation and processing (Hale, 2001; Levy, 2008). Surprisal of morpheme $m$, $I(m)$, is defined as Equation (1):

$$I(m) = \log_2 \frac{1}{P(m)} = -\log_2 P(m) \quad (1)$$

Surprisal was originally proposed to explain behavioral measures such as reading times in self-paced reading experiments and fixation durations in eye-tracking experiments (Boston et al., 2008; Demberg and Keller, 2008; Roark et al., 2009; Frank and Bod, 2011; Fossum and Levy, 2012). Recently, surprisal has also been used to successfully predict neural measures like N400 components in EEG experiments and BOLD signals in fMRI experiments (Frank et al., 2015; Brennan et al., 2016; Willems et al., 2016; Henderson et al., 2016; Nelson et al., 2017; Lopopolo et al., 2017).

Assuming further that morphological processing is incremental, we compute surprisal of morphologically complex words as cumulative surprisal, the cumulative sum of surprisal of their component morphemes. Cumulative surprisal of word $w$, $I(w)$, is defined as Equation (2):

$$I(w) = \sum_{m = m_1, \ldots, m_n} I(m) \quad (2)$$

where $I(m)$ is the surprisal of morpheme $m$ computed by computational models via respective incremental algorithms.

Two evaluation metrics can be derived from cumulative surprisal: linguistic and psychological accuracies (Frank and Bod 2011; Fossum and Levy 2012; Frank et al. 2015; cf. Virpioja et al. 2017). Linguistic accuracy of model $M$, $LA(M)$, is defined as Equation (3):

$$LA(M) = -\frac{1}{n} \sum_{i=1}^{n} I(m_i) \quad (3)$$

where $I(m)$ is the surprisal of morpheme $m$. In particular, we compute linguistic accuracy to evaluate how well computational models can predict words never seen before (cf. Chomsky, 1957), independently of human language processing.

Neurological accuracy of model $M$, $NA(M)$, is defined as Equation (4):

$$NA(M) = \Delta D_B - \Delta D_M \quad (4)$$

where $\Delta D$ is delta deviance defined as $-2 \times$ log-likelihood and $B$ is the baseline model without surprisal included. In effect, neurological accuracy quantifies how well computational models
explain human brain activities above and beyond control predictors included in the baseline model.

Moreover, in order to understand where (i.e. for what types of words) computational models make better or worse predictions (cf. Sprouse et al., 2018), we closely inspect residual errors via the new evaluation metric proposed here as residual score. The residual score of model $M$ for word $w$, $RS_M(w)$, is defined as Equation (5):

$$RS(M) = \sum_{i=1}^{n} |\epsilon_B(w_i)| - |\epsilon_M(w_i)| \quad (5)$$

where $\epsilon_M(w)$ is the residual error of model $M$ for word $w$. Since the globally best model does not necessarily generate perfect fits to every word, residual errors of the globally best model may be higher for some types of words than the alternative models. We compute residual scores of computational models sorted by linear and nested structures to address the question whether hierarchical models make better predictions for nested words than sequential models.

### 2.5 Statistical analyses

We performed mixed-effects regression (Baayen et al., 2008) by averaging source estimates within functionally-defined ROI (fROI) based on spatiotemporal regression. Since the previous literature (Lewis et al., 2011; Simon et al., 2012; Gwilliams et al., 2016) showed that lemma frequency is a significant predictor of the M170, we conducted spatiotemporal regression with lemma frequency as the predictor of interest. Lemma frequency, sometimes called cumulative root frequency (del Prado Martin et al., 2004) or family frequency (Fruchter and Marantz, 2015), is defined as the sum of frequencies of words that share the same lemma. For example, the lemma frequency of globalization is the sum of frequencies of globe, global, globalize, and so on. Spatiotemporal regression in left inferior temporal lobe ROI and 150-200 time window with log-transformed lemma frequency as target predictor and squared length as control predictor; Right: timecourse of neural activity averaged over space and time within the fROI and sorted by linear and nested structures. The $x$-axis is time in milliseconds, while the $y$-axis is source estimates in dSPM (Dale et al., 2000). Color indicates two types of structure: yellow = linear, blue = nested. Pink horizontal span marks 150-200 ms analysis time window.

![Figure 1: fROI for mixed-effects regression. Left: spatial extent of the fROI defined as the significant cluster found via spatiotemporal regression in left inferior temporal lobe ROI and 150-200 time window with log-transformed lemma frequency as target predictor and squared length as control predictor; Right: timecourse of neural activity averaged over source points within the fROI and sorted by linear and nested structures. The x-axis is time in milliseconds, while the y-axis is source estimates in dSPM (Dale et al., 2000). Color indicates two types of structure: yellow = linear, blue = nested. Pink horizontal span marks 150-200 ms analysis time window.](image)
model, the target regression model was then fitted with surprisal included as an additional fixed effect on top of control predictors and random effects held constant. The control predictors were squared length (New et al., 2006) and affix frequency, defined as the sum of frequencies of two affixes in CELEX (two suffixes for linear words; one suffix and one prefix for nested words). Mixed models were fitted via Maximum Likelihood Estimation with \texttt{nlminb} optimizer in \texttt{optimx} package and the maximum number of iterations R permits. Given that the baseline and target models are minimally different only in surprisal, computational models can be evaluated with nested model comparisons via log-likelihood ratio tests based on \( \chi^2 \)-distribution with \( df = 1 \), where \( df \) is the difference in number of parameters between nested models.

3 Results

3.1 Linguistic and neurological accuracies

Linguistic and neurological accuracies of computational models are summarized in Figure 2, where the \( x \)-axis is linguistic accuracy (negative average surprisal) and the \( y \)-axis is neurological accuracy (decrease in delta deviance). Points represent computational models: emerald = Letter Markov Model, orange = Syllable Markov Model, purple = Morpheme Markov Model, pink = Hidden Markov Model, green = Probabilistic Context-Free Grammar. The horizontal dashed line is \( \chi^2 = 3.84 \), the critical \( \chi^2 \)-statistic at \( p = 0.05 \) with \( df = 1 \). All “morphous” models were statistically significant (\( p < 0.01 \)).

Nested model comparisons via log-likelihood tests revealed that while no “amorphous” models were statistically significant, all “morphous” models were statistically significant (\( p < 0.05 \)). Among these “morphous” models, the PCFG was neurologically most accurate: PCFG (\( \chi^2 = 8.495, p < 0.01 \)) > Markov Model (\( \chi^2 = 7.989, p < 0.01 \)) > HMM (\( \chi^2 = 6.551, p < 0.05 \)) > Syllable (\( \chi^2 = 2.42, ns \)) > Letter (\( \chi^2 = 0.004, ns \)).

3.2 Residual scores

Residual scores are summarized in Figure 3, where the \( x \)-axis is computational models and the \( y \)-axis is residual scores. The residual scores are sorted by two types of syntactic structures (linear vs. nested) and averaged across individual derivational affixes. The horizontal dashed line indicates...
Figure 3: Residual scores, sorted by linear (Left) and nested (Right) structures and averaged across individual derivational affixes. The x-axis is computational models and the y-axis is residual scores. The horizontal dashed line indicates a “tie” borderline where computational models do not diverge from the baseline model, and more positive and negative residual scores mean better and worse predictions made by computational models, respectively.

For linear words, there are no differences among three “morphous” models, despite a weak trend where the sequential models like Markov Model and HMM showed slightly higher residual scores than the hierarchical models like PCFG. For nested words, interestingly, PCFG was the only model which made better predictions and the sequential models were generally mute.

4 Discussion

In summary, the present experiment demonstrated that “amorphous” models of morphological processing underperformed relative to “morphous” models and PCFG was more neurologically accurate than sequential models. We can conclude from these results that morphemes are represented in the VWFA (pace Virpioja et al., 2017; Hakala et al., 2018) and parsed into hierarchical morphological structures (pace Beesley and Karttunen, 2003; Roark and Sproat, 2007). Methodologically, this experiment is one of the first attempts to generalize the computational modeling methodology developed in sentence processing (Brennan et al., 2016; Nelson et al., 2017) to morphological processing.

Moreover, the new evaluation metric named residual scores indicated that PCFG better explained nested words with non-local dependencies between prefixes and suffixes than sequential models such as Markov Model and HMM. This result was exactly predicted because linear and nested words are finite-state and context-free languages in the Chomsky hierarchy (Hopcroft and Ullman, 1979; Partee et al., 1990; Sipser, 1997), only the former of which can be accepted by sequential models, but both of which can be recognized by PCFG. Probabilistically speaking, sequential models have trouble with transition probabilities from prefixes to roots in nested words (e.g. dehumanize) because prefixes (e.g. de-) and roots (e.g. human) never form morphological constituents (e.g. *dehuman) and thus appear in even large corpora.
Now the further question arises why low-level visual evoked response components like M170 in the ventral visual stream “know” high-level linguistic representations like hierarchical morphological structures. One possibility is that, as suggested by Pammer et al. (2004), there is a functional connectivity between left fusiform gyrus and left inferior frontal gyrus in visual word recognition, so that M170 can be modulated in a top-down feedback fashion by “Broca’s area”, the traditional “language” area proposed to process hierarchical syntactic structures (Friederici, 2002, 2012). While this possibility becomes less surprising if even visual cortex can be sensitive to syntax (Dikker et al., 2009), the functional connectivity between the two areas remains to be empirically investigated in future (Carreiras et al., 2014; Woodhead et al., 2014).

5 Conclusion

In this paper, we conducted a magnetoencephalography (MEG) experiment where participants performed visual lexical decision on morphologically complex words and, generalizing the computational modeling methodology developed in sentence processing (Brennan et al., 2016; Nelson et al., 2017) to morphological processing, modeled morphological processing in the human brain, with special focus on the VWFA in the ventral visual stream. Five computational models of morphological processing were constructed and evaluated against human brain activities in order to investigate whether morphemes are represented in the VWFA and parsed into hierarchical morphological structures: Letter Markov Model and Syllable Markov Model as “amorphous” models without linguistically defined morphemes, and Morpheme Markov Model, Hidden Markov Model (HMM), and Probabilistic Context-Free Grammar (PCFG) as “morphous” models with different amounts of morphological supervision. The present experiment demonstrated that “amorphous” models of morphological processing underperformed relative to “morphous” models. PCFG was more neurologically accurate than sequential models, and PCFG better explained nested words with non-local dependencies between prefixes and suffixes. These results strongly suggest that morphemes are represented in the VWFA and parsed into hierarchical morphological structures.

References

Roberto de Almeida and Gary Libben. 2005. Changing morphological structures: The effect of sentence context on the interpretation of structurally ambiguous English trimorphemic words. Language and Cognitive Processes, 20:373–394.

Harald Baayen, Douglas Davidson, and Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59:390–412.

Harald Baayen, Petar Milin, Dusica Filipovic Durdevic, Peter Hendrix, and Marco Marelli. 2011. An Amorphous Model for Morphological Processing in Visual Comprehension Based on Naive Discriminative Learning. Psychological Review, 118:438–481.

R. H. Baayen, R. Piepenbrock, and L. Gulikers. 1995. The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Douglas Bates, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67:1–48.

Laurie Bauer. 2014. Grammaticality, acceptability, possible words and large corpora. Morphology, 24:83–103.

Kenneth Beesley and Lauri Karttunen. 2003. Finite State Morphology. CSLI Publications, University of Chicago Press.

Steven Bird, Ewan Klein, and Edward Loper. 2009. Natural Language Processing with Python. O’Reilly Media.

Marisa Boston, John Hale, Reinhold Kliegl, Umesh Patil, and Shravan Vasishth. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. Journal of Eye Movement Research, 2:1–12.

Jonathan Brennan, Edward Stabler, Sarah Van Wagenen, Wen-Ming Luh, and John Hale. 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. Brain and Language, 157-158:81–94.

Manuel Carreiras, Blair Armstrong, Manuel Perea, and Ram Frost. 2014. The what, when, where, and how of visual word recognition. Trends in Cognitive Sciences, 18:90–98.

Noam Chomsky. 1957. Syntactic Structures. Mouton, The Hague.

Laurent Cohen, Stanislas Dehaene, Lionel Naccache, Stephane Lehericy, Ghislaine Dehaene-Lambertz, Marie-Anne Henaff, and Francois Michel. 2000. The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain, 123:291–307.
Laurent Cohen, Stephane Lehericy, Florence Chocron, Cathy Lemer, Sohie Rivaud, and Stanislas Dehaene. 2002. Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, 125:1054–1069.

Mathias Creutz and Crista Lagus. 2007. Unsupervised models for morpheme segmentation and morphology learning. *ACM Transactions on Speech and Language Processing*, 4:3.

Anders M. Dale, Arthur K. Liu, Bruce R. Fischl, Randy L. Buckner, John W. Belliveau, Jeffrey D. Lewine, and Eric Halgren. 2000. Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron*, 26:55–67.

Stanislas Dehaene, Gurvan Le Clec’H, Jean-Baptiste Poline, Denis Le Bihan, and Laurent Cohen. 2002. The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *NeuroReport*, 13:321–325.

Stanislas Dehaene, Laurent Cohen, Mariano Sigman, and Fabien Vinckier. 2005. The neural code for written words: a proposal. *Trends in Cognitive Sciences*, 9:335–341.

Stanislas Dehaene, Lionel Naccache, Laurent Cohen, Denis Le Bihan, Jean-Francois Mangin, Jean-Baptiste Poline, and Denis Riviere. 2001. Cerebral mechanisms of word masking and unconscious repetition priming. *Nature NeuroScience*, 4:752–758.

Vera Demberg and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109:193–210.

Suzanne Dikker, Hugh Rabagliati, and Liina Pylkkänen. 2009. Sensitivity to syntax in visual cortex. *Cognition*, 110:293–321.

Jay Earley. 1970. An ecent context-free parsing algorithm. *Communications of the Association for Computing Machinery*, 13:94–102.

David Embick. 2016. Approaching Polymorphy. Unpublished manuscript.

Victoria Fossum and Roger Levy. 2012. Sequential vs. hierarchical syntactic models of human incremental sentence processing. *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics*, pages 61–69.

Stefan Frank and Rens Bod. 2011. Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, 22:829–834.

Stefan L Frank, Leun J Otten, Giulia Galli, and Gabriella Vigliocco. 2015. The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140:1–11.

Angela Friederici. 2002. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6:78–84.

Angela Friederici. 2012. The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16:262–268.

Joseph Fruchter and Alec Marantz. 2015. Decomposition, lookup, and recombination: MEG evidence for the Full Decomposition model of complex visual word recognition. *Brain and Language*, 143:81–96.

Joseph Fruchter, Linnaea Stockall, and Alec Marantz. 2013. MEG masked priming evidence for form-based decomposition of irregular verbs. *Frontiers in Human Neuroscience*, 7:798.

Laura Gwilliams, Gwyneth Lewis, and Alec Marantz. 2016. Functional characterisation of letter-specific responses in time, space and current polarity using magnetoencephalography. *Neuroimage*, 132:320–333.

Tero Hakala, Annika Hulten, Minna Lehtonen, Krista Lagus, and Riitta Salmelin. 2018. Information properties of morphologically complex words modulate brain activity during word reading. *Human Brain Mapping*, 39:2583–2595.

John Hale. 2001. A Probabilistic Earley Parser as a Psycholinguistic Model. *Proceedings of NAACL-2001*, pages 159–166.

Morris Halle. 1973. Prolegomena to a Theory of Word Formation. *Linguistic Inquiry*, 4:3–16.

Jennifer Hay. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics*, 39:1041–1070.

Jennifer Hay. 2002. From speech perception to morphology: Affix ordering revisited. *Language*, 78:517–555.

P. Helenius, A. Tarkiainen, P. Cornelissen, P.C. Hansen, and R. Salmelin. 1999. Dissociation of Normal Features of Normal Feature Analysis and Deficient Processing of Letterstrings in Dyslexic Adults. *Cerebral Cortex*, 9:476–483.

John M. Henderson, Wonil Choi, Matthew W. Lowder, and Fernanda Ferreira. 2016. Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. *NeuroImage*, 132:291–300.

Gregory Hickok and David Poeppel. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8:393–402.

John Hopcroft and Jeffrey Ullman. 1979. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading, MA.

Roger Levy. 2008. Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177.
Gwyneth Lewis, Olla Solomyak, and Alec Marantz. 2011. The Neural Basis of Obligatory Decomposition of Suffix Words. *Brain and Language*, 118:118–127.

Gary Libben. 2003. Morphological parsing and morphological structure. In Egbert Assink and Dominiek Sandra, editors, *Reading Complex Words*, pages 221–239. Kluwer, New York.

Gary Libben. 2006. Getting at psychological reality: On- and off-line tasks in the investigation of hierarchical morphological structure. In G. Wiebe, G. Libben, T. Priestly, R. Smyth, and S. Wang, editors, *Phonology, Morphology, and the Empirical Imperative*, pages 349–369. Crane, Taipei.

Alessandro Lopopolo, Stefan L. Frank, Antal van den Bosch, and Roel M. Willems. 2017. Using stochastic language models (SLM) to map lexical, syntactic, and phonological information processing in the brain. *PLoS ONE*, 12:e0177974.

Petar Milin, Laurie Feldman, Michael Ramscar, Peter Hendrix, and Harald Baayen. 2017. Discrimination in lexical decision. *PLoS ONE*, 12.

Matthew J Nelson, Imen El Karoui, Kristof Giber, Xiaofang Yang, Laurent Cohen, Hilda Koopman, Sydney S Cash, Lionel Naccache, John T Hale, Christophe Pallier, et al. 2017. Neuropsychological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 114:3669–3678.

Boris New, Ludovic Ferrand, Christophe Pallier, and Marc Brysbaert. 2006. Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin and Review*, 13:45–52.

Carolus Oldfield. 1971. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9:97–113.

Yohei Oseki, Charles Yang, and Alec Marantz. 2019. Modeling Hierarchical Syntactic Structures in Morphological Processing. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 43–52.

Kristen Pammer, Peter Hansen, Morten Kringelbach, Ian Holliday, Gareth Barnes, Arjan Hillebrand, Krish Singh, and Piers Cornelissen. 2004. Visual word recognition: the first half second. *Neurolmage*, 22:1819–1825.

Barbara Partee, Alice ter Meulen, and Robert Wall. 1990. *Mathematical Methods in Linguistics*. Springer, Dordrecht.

Ingo Plag and Harald Baayen. 2009. Suffix Ordering and Morphological Processing. *Language*, 85:109–152.

Alexander Pollatsek, Denis Drieghe, Linnnea Stockall, and Roberto de Almeida. 2010. The interpretation of ambiguous trimorphemic words in sentence context. *Psychonomic Bulletin and Review*, 17:88–94.

Fermin Moscoso del Prado Martín, Aleksandar Kostic, and Harald Baayen. 2004. Putting the bits together: an information theoretical perspective on morphological processing. *Cognition*, 94:1–18.

Liina Pylkkänen and Alec Marantz. 2003. Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, 7:187–189.

Lawrence Rabinar. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77:257–286.

Brian Roark, Asaf Bachrach, Carlos Cardenas, and Christophe Pallier. 2009. Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 324–333.

Brian Roark and Richard Sproat. 2007. *Computational Approaches to Morphology and Syntax*. Oxford University Press, Oxford.

Dylan Simon, Gwyneth Lewis, and Alec Marantz. 2012. Disambiguating form and lexical frequency effects in MEG responses using homonyms. *Language and Cognitive Processes*, 27:275–287.

Michael Sipser. 1997. *Introduction to the Theory of Computation*. PWS Publishing, Boston, MA.

Olla Solomyak and Alec Marantz. 2009. Lexical access in early stages of visual word processing: A single-trial correlational MEG study of heteronym recognition. *Brain and Language*, 108:191–196.

Olla Solomyak and Alec Marantz. 2010. Evidence for Early Morphological Decomposition in Visual Word Recognition: A Single-Trial Correlational MEG Study. *Journal of Cognitive Neuroscience*, 22:2042–2057.

Yoonsang Song, Youngah Do, Jongbong Lee, Arthur Thompson, and Eileen Waegemaekers. 2019. The reality of hierarchical morphological structure in multimorphemic words. *Cognition*, 183:269–276.

Jon Sprouse, Sagar Indurkhya, Beracah Yankama, Sandiway Fong, and Robert C. Berwick. 2018. Colorless green ideas do sleep furiously: gradient acceptability and the nature of the grammar. *Linguistic Review*, 35:575–599.

Andreas Stolcke. 1995. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, 21:165–201.
M. Taft. 1979. Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7:263–272.

M. Taft. 2004. Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology*, 57:745–765.

M. Taft and K. I. Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14:638–647.

A. Tarkiainen, P. Helenius, P. C. Hansen, P. L. Cornelissen, and R Salmelin. 1999. Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, 122:2119–2132.

Sami Virpioja, Minna Lehtonen, Annika Hulten, Henna Kivikari, Riitta Salmelin, and Krista Lagus. 2017. Using Statistical Models of Morphology in the Search for Optimal Units of Representation in the Human Mental Lexicon. *Cognitive Science*, pages 1–35.

Roel M. Willems, Stefan L. Frank, Annabel D. Nijhof, Peter Hagoort, and Antal van den Bosch. 2016. Prediction During Natural Language Comprehension. *Cerebral Cortex*, 26:2506–2516.

Z.V. J. Woodhead, G.R. Barnes, W. Penny, R. Moran, S. Teki, C.J. Price, and A.P. Leff. 2014. Reading Front to Back: MEG Evidence for Early Feedback Effects During Word Recognition. *Cerebral Cortex*, 24:817–825.

Linda Zirkel. 2010. Prefix combinations in English: structural and processing factors. *Morphology*, 20:239–266.

Etyan Zweig and Liina Pylkkänen. 2009. A Visual M170 Effect of Morphological Complexity. *Language and Cognitive Processes*, 24:412–439.