Modeling Morphological Processing in Human Magnetoencephalography

Yohei Oseki  
*Waseda University*, oseki@aoni.waseda.jp

Alec Marantz  
*New York University*, marantz@nyu.edu

Follow this and additional works at: [https://scholarworks.umass.edu/scil](https://scholarworks.umass.edu/scil)

Part of the Computational Linguistics Commons, Morphology Commons, and the Psycholinguistics and Neurolinguistics Commons

**Recommended Citation**  
DOI: [https://doi.org/10.7275/7tak-9s75](https://doi.org/10.7275/7tak-9s75)  
Available at: [https://scholarworks.umass.edu/scil/vol3/iss1/22](https://scholarworks.umass.edu/scil/vol3/iss1/22)

This Paper is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Proceedings of the Society for Computation in Linguistics by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
Modeling Morphological Processing in Human Magnetoencephalography

Yohei Oseki
Faculty of Science & Engineering
Waseda University
oseki@aoni.waseda.jp

Abstract

In this paper, we conduct a magnetoencephalography (MEG) lexical decision experiment and computationally model morphological processing in the human brain, especially the Visual Word Form Area (VWFA) in the visual ventral stream. Five neurocomputational models of morphological processing are constructed and evaluated against human neural activities: Character Markov Model and Syllable Markov Model as “amorphous” models without morpheme units, and Morpheme Markov Model, Hidden Markov Model (HMM), and Probabilistic Context-Free Grammar (PCFG) as “morphous” models with morpheme units structured linearly or hierarchically. Our MEG experiment and computational modeling demonstrate that “morphous” models outperformed “amorphous” 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 human brain 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 from the mental lexicon. In the last stage of morphological recombination, retrieved meanings of decomposed morphemes are semantically composed.

In the cognitive neuroscience literature, Fruecht 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 evoked response components such as M170 (Zweig and Pylkkanen, 2009; Solomyak and Marantz, 2010; Lewis et al., 2011; Fruecht et al., 2013; Gwilliams et al., 2016) or Type II (Tarkiainen et al., 1999; Helenius et al., 1999) in the visual ventral stream of the human brain (Pylkkänen and Marantz, 2003; Hickok and Poeppel, 2007). Moreover, Dehaene et al. (2005) proposed local combination detectors (LCDs) where linguistic units such as characters, syllables, and morphemes are convolutionally represented and processed in the visual ventral stream from posterior occipital to anterior temporal cortices and, importantly, morphemes have 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 from stems to suffixes (e.g. P(Suffix|Stem)) to successfully predict neural responses to real (e.g. teach-er) and pseudo (e.g. corn-er) bimorphemic words, respectively. These results have suggested that morphemes may be neurologically real in the human brain.

However, “amorphous” models without morpheme units have recently been proposed in the morphological processing literature (Baayen et al., 2011; Virpioja et al., 2017). For instance, Baayen et al. (2011) and Milin et al. (2017) proposed Naive Discriminative Learning (NDL), a connectionist model with direct mappings from forms to meanings, to explain morphological processing without morpheme units. In addition, Virpioja et al. (2017) and Hakala et al. (2018) employed...
Morfessor, an unsupervised finite-state model with statistically induced “morphs" (Creutz and Lagus, 2007), to predict human reaction times and neural responses without linguistically defined morphemes. Furthermore, as correctly pointed out by Libben (2003, 2006), bimorphic words exclusively tested in the previous literature (Zweig and Pyllkänen, 2009; Solomyak and Marantz, 2010; Lewis et al., 2011) cannot distinguish linear morphological decomposition from hierarchical morphological parsing (cf. Song et al., 2019; Oseki et al., 2019). Therefore, whether morphemes are represented in the human brain and, if so, processed linearly or hierarchically remains to be empirically investigated.

In this paper, we conduct an magnetoencephalography (MEG) experiment where participants perform visual lexical decision on morphologically complex words and, generalizing the computational modeling technique developed in the sentence processing literature (Frank et al., 2015; Brennan et al., 2016), computationally model morphological processing in the human brain, with special focus on the VWFA in the visual ventral stream. Specifically, five neurocomputational models of morphological processing are constructed and evaluated against human neural activities: Character Markov Model and Syllable Markov Model as “amorphous” models without morpheme units, and Morpheme Markov Model, Hidden Markov Model (HMM), and Probabilistic Context-Free Grammar (PCFG) as “morphous” models with morpheme units structured linearly or hierarchically.

2 Methods
2.1 Participants
The participants were 26 native English speakers recruited at New York University. 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 trimorphic words and nonwords. The stimuli creation procedure consisted of several steps. First, 600 trimorphic words were created based on the CELEX database (Baayen et al., 1995) in accordance with syntactic (syntactic categories), morphological (affix combinations), and phonological (orthographic adjustments) selectional restrictions of derivational affixes, but without semantic selectional restrictions explicitly taken into consideration. In this sense, these trimorphic words are grammatical (“possible”) but not necessarily acceptable (“actual”) words (cf. Halle, 1973; Bauer, 2014). These 600 trimorphic words were subcategorized into 300 linear words [X Prefix [Y Root] Suffix] with productive derivational suffixes (Plag and Baayen, 2009) and 300 nested words [X Prefix [Y Root] Suffix] with productive derivational prefixes (Zirkel, 2010). Furthermore, these trimorphic words have zero surface frequencies in the CELEX database, thereby enhancing the possibility that those words have never been encountered by participants and stored in the mental lexicon (Hay, 2003). Second, in order to weed out semantically implausible words, 600 trimorphic words were normed with crowdsourced acceptability judgment experiments, where participants judged them on 1~7 Likert scale. Third, 500 trimomorphic words (250 linear and 250 nested) with higher acceptability judgments (> 3.5) and lower standard deviations (< 2.5) were selected and, correspondingly, 500 trimorphic nonwords (250 linear and 250 nested) were also created based on the CELEX database in violation of syntactic selectional restrictions of inner derivational suffixes, resulting in 1000 trimorphic words and nonwords. Fourth, in order to ensure that words and nonwords are correctly judged as such, 1000 trimorphic stimuli were further normed with crowdsourced lexical decision experiments, where participants decided whether presented stimuli were possible English words or not as quickly and accurately as possible. Finally, 400 trimorphic words (200 linear and 200 nested) and 400 trimorphic nonwords (200 linear and 200 nested) with higher accuracies (> 75%) were selected, resulting in the balanced and extensively normed set of 800 trimorphic stimuli to be tested in this experiment. The stimuli are summarized in Table 1:
2.3 Procedure

The experiment was conducted in the Neuroscience of Language Lab at New York University, New York. Before MEG recording, each participant’s head shape was digitized with a Polhemus FastSCAN laser scanner (Polhemus, Vermont, USA) and five fiducial points were marked on his/her forehead, onto which marker coils were attached during the recording. In order to familiarize the participants with visual lexical decision, the participants completed one practice block with 16 practice stimuli, 4 stimuli per each stimulus type, that do not overlap with the target stimuli. The task instructions were exactly the same as the main experiment, but the participants received feedback after each trial (“CORRECT” or “INCORRECT”) during the practice block.

A 157-channel axial gradiometer whole-head MEG system (Kanazawa Institute of Technology, Kanazawa, Japan) recorded the MEG data continuously at a sampling rate of 1000 Hz (1 datapoint per each millisecond), while the participants lay in a dimly lit magnetically shielded room (MSR) and performed visual lexical decision. The MEG data were filtered online between DC and 200 Hz with a notch filter at 60 Hz. Five marker coils were attached to the corresponding fiducial points marked on the forehead and their positions were measured before and after the main experiment, in order to align the MEG data and head shapes and estimate how much the participants moved during the MEG recording. The main experiment itself lasted for about 35 minutes.

The stimuli were presented with PsychoPy package (Peirce, 2007, 2009) in Python. They were projected on the screen approximately 50 cm away from the participants and presented in white 30 lowercase Courier New font on a grey background. The 800 stimuli were randomly distributed into 8 blocks of 100 stimuli with 25 stimuli from each stimulus type. First, the explanation appeared on the screen: “In this experiment, you will read English words and determine whether you think they are possible English words. We are not concerned with whether or not these words are actual English words already listed in a dictionary. Instead, we are interested in whether or not these words could be used by a native speaker of English”. Then, the task instruction appeared on the screen: “The experiment is about to begin. Please fixate on the cross in the center of the screen. Respond with your index finger if the string is word. Respond with your middle finger if it is not a word”. Each trial consisted of the fixation cross (+) for 500 ms, the blank for 300 ms, and the stimulus until the participants respond with their index finger (YES) or middle finger (NO) of their left hand. The inter-stimulus interval (ISI) followed the standard normal distribution with the mean of 400 ms and the standard deviation of 100 ms.

Table 1: Summary of stimuli. The horizontal dimension is morphological structure: linear vs. nested. The vertical dimension is lexicality status: word vs. nonword. The asterisk (*) on subtrees (Y) of nonwords indicates that inner derivational suffixes violate syntactic selectional restrictions on syntactic categories of roots.

<table>
<thead>
<tr>
<th></th>
<th>Linear</th>
<th>Nested</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word</strong></td>
<td>![Diagram](Linear Word)</td>
<td>![Diagram](Nested Word)</td>
</tr>
<tr>
<td></td>
<td>X ∈ 200</td>
<td>X ∈ 200</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Z</td>
</tr>
<tr>
<td></td>
<td>Digit</td>
<td>Culture</td>
</tr>
<tr>
<td><strong>Nonword</strong></td>
<td>![Diagram](Linear Nonword)</td>
<td>![Diagram](Nested Nonword)</td>
</tr>
<tr>
<td></td>
<td>X ∈ 200</td>
<td>X ∈ 200</td>
</tr>
<tr>
<td></td>
<td>*Y</td>
<td>*Y</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>Z</td>
</tr>
<tr>
<td></td>
<td>ion</td>
<td>ion</td>
</tr>
<tr>
<td></td>
<td>√Gulf</td>
<td>√Kid</td>
</tr>
</tbody>
</table>
### 2.4 Computational models

Five computational models were implemented with Natural Language Tool Kit package (Bird et al., 2009) in Python: Character Markov Models (Character), Syllable Markov Models (Syllables), Morpheme Markov Models (Markov), Hidden Markov Model (HMM), and Probabilistic Context-Free Grammar (PCFG). Those models were trained on the entire CELEX database via Maximum Likelihood Estimation with token weighting and Lidstone smoothing at $\alpha = 0.1$. The architectures of Markov Model, HMM, and PCFG are summarized below.

#### 2.4.1 Markov Model

Markov Models (also called $n$-gram models) are defined by $n$-order Markov processes that compute transition probabilities of linguistic units (e.g. characters, syllables, morphemes) at position $i$ given $i-n$ context (e.g. $P(x_i|x_{i-n},x_{i-1})$). Since the length of morphologically complex words is inherently limited relative to syntactically complex sentences, Markov Models were defined with $n = 1$ (i.e. bigram models), which compute transition probabilities of linguistic units at position $i$ given the immediately preceding unit (e.g. $P(x_i|x_{i-1})$). For training, Markov Models were trained on character strings (Character Markov Model), syllable strings (Syllable Markov Model), and morpheme strings (Morpheme Markov Model), respectively, where character and morpheme strings were available from the CELEX database, while syllable strings were generated with syllabify module implemented in Python by Kyle Gorman through ARPABET transcriptions assigned by LOGIOS Lexicon Tool in the Carnegie Mellon University Pronouncing Dictionary. For testing, those trained Markov Models then computed morpheme probabilities of morphologically complex words as the ratio of prefix probabilities at position $k$ to position $k-1$, where prefix probabilities are the sum of path probabilities compatible with morphemes until position $k$ (Rabinar, 1989). HMMs are linear models, which should accurately predict local dependencies of linear words (e.g. $N-1$ for $digitally$), but also non-local dependencies of nested words (e.g. $unpredictable$) if component local dependencies (e.g. $A-V$ for $unpredict$) are attested in the training data.

#### 2.4.2 Hidden Markov Model

HMMs generalize Markov Models with $n$-order Markov processes defined over “hidden” linear strings. HMMs compute transition probabilities of part-of-speech (POS) tags at position $i$ given $i-n$ context (e.g. $P(t_i|t_{i-n},t_{i-1})$), and emission probabilities of morphemes at position $i$ given POS tags at the same position $i$ (e.g. $P(m_i|t_i)$). Like Markov Models, HMMs were also defined with $n = 1$, which compute transition probabilities of POS tags at position $i$ given the immediately preceding POS tag (e.g. $P(t_i|t_{i-1})$). For training, HMMs were supervisedly trained on tagged morpheme strings generated from morphological structures available from the CELEX database (e.g. $[(digit, N), (al, A), (ly, B)]$). For testing, those trained HMMs then computed morpheme probabilities of morphologically complex words as the ratio of prefix probabilities at position $k$ to position $k-1$, where prefix probabilities are the sum of path probabilities compatible with morphemes until position $k$ (Earley, 1970; Stolcke, 1995). PCFGs are hierarchical models, which should accurately predict not only local dependencies of linear words (e.g. $digitally$), but also non-local dependencies of nested words (e.g. $unpredictable$).
2.5 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):

\[
I(m) = \log_2 \frac{1}{P(m)} = -\log_2 P(m)
\]

where \( P(m) \) is the probability of morpheme \( m \) computed by computational models via respective incremental algorithms. 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; Fossom and Levy, 2012). Recently, surprisal has also been extended to 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 (cf. prefix stripping; Taft and Forster, 1975; Stockall et al., 2019), we compute surprisal of morphologically complex words as cumulative surprisal, the cumulative sum of surprisal of component morphemes. Cumulative surprisal of word \( w \), \( I(w) \), is defined as Equation (2):

\[
I(w) = I(m_1, \ldots, m_n) = \sum_{i=1}^{n} I(m_i)
\]

where \( I(m) \) is the surprisal of morpheme \( m \) computed by computational models.

Two evaluation metrics are then derived from cumulative surprisal: neurological and error accuracies (cf. Frank et al., 2015; Sprouse et al., 2018). The neurological accuracy of model \( M \), \( NA(M) \), is defined as Equation (3):

\[
NA(M) = DB - DM
\]

where \( DB \) and \( DM \) are deviance defined as \(-2\) times log-likelihoods of baseline and target models, respectively. Neurological accuracy quantifies decreases in deviance \((-\Delta D)\) and evaluates how well computational models explain human neural activities beyond control predictors included in the baseline model (cf. Frank et al., 2015).

The error accuracy of model \( M \), \( EA(M) \), is defined as Equation (4):

\[
EA(M) = \sum_{i=1}^{n} |\epsilon_B(w_i) - \epsilon_M(w_i)|
\]

where \( \epsilon_B(w) \) and \( \epsilon_M(w) \) are residual errors of baseline and target models for word \( w \), respectively. Error accuracy quantifies decreases in absolute residual errors \((-\Delta\epsilon)\) and evaluates cost-benefit tradeoffs of computational models (cf. Sprouse et al., 2018). We compute error accuracies of computational models with respect to linear and nested morphological structures to address the question whether hierarchical models make better predictions for nested words than linear models.

2.6 Statistical analyses

We performed linear mixed-effects regression (Baayen et al., 2008) by averaging neural activities within the functionally defined region of interest (fROI) based on spatiotemporal cluster permutation regression (Maris and Oostenveld, 2007). In the previous literature (cf. Williams et al., 2016), lemma frequency has been proposed as a significant predictor of the M170 and, thus, employed as the predictor of interest for spatiotemporal regression. Lemma frequency (cf. del Prado Martin et al., 2004) 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 the left inferior temporal lobe and the 150-200 time window with log-transformed lemma frequency as target predictor and squared length as control predictor identified the significant cluster where the clear M170 peak can be observed, as shown in Figure 1. Finally, the neural activities were averaged over space and time within the fROI to compute by-trial dSPMs (Dale et al., 2000), which were then exported to R for mixed-effects regression.

Linear mixed-effects regression was implemented with lme4 package (Bates et al., 2015) in R. The baseline regression model was first fitted with by-trial dSPMs as the dependent variable, control predictors as fixed effects, and by-subject and by-word random intercepts as random effects. For each computational model, the target regression model was then fitted with cumulative surprisal included as an additional fixed effect on top.
of control predictors and random effects held constant. The control predictor was squared length (New et al., 2006) also included to functionally define the ROI. Mixed-effects models were fitted via Maximum Likelihood Estimation with nlmixr optimizer in optimx package and the maximum number of iterations R permits. Given that the baseline and target models are minimally different only in cumulative 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 Neurological accuracy

Neurological accuracies of computational models are summarized in Figure 3, where the $x$-axis is computational models and the $y$-axis is neurological accuracies (i.e. decreases in absolute residual errors), categorized into linear and nested morphological structures and averaged across individual derivational affixes. The horizontal dashed line indicates a “tie” borderline where computational models do not diverge from the baseline model. More positive and negative error accuracies mean better and worse predictions relative to the baseline model.

For linear words, all neurologically accurate “morphous” models made significant contributions, among which Markov Model made best predictions relative to the baseline model. For nested words, interestingly, PCFG was the only computational model which reduced residual errors, while linear models such as HMM and Markov Model made only slight or even worse predictions relative to the baseline model, respectively.

4 Discussion

In summary, our MEG experiment and computational modeling demonstrated that “morphous” models of morphological processing outperformed “amorphous” models and, importantly, PCFG was most neurologically accurate among those “morphous” models. We can conclude from these results that morphemes are neurologically represented in the human brain (pace Baayen 2014).
et al., 2011; Milin et al., 2017) and parsed into hierarchical morphological structures (pace Virpioja et al., 2017; Hakala et al., 2018). In addition, this paper successfully generalized the computational modeling technique developed in the sentence processing literature (Frank et al., 2015; Brennan et al., 2016) to morphological processing.

Moreover, error accuracies of computational models indicated that PCFG better explained nested words with non-local dependencies between prefixes and suffixes than linear models such as Markov Model and HMM. This result follows straightforwardly from formal language theory, where 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), the former of which can be modeled by both linear and hierarchical models, but the latter of which can only be parsed by hierarchical models like PCFG. Furthermore, from the probabilistic perspective, linear models have trouble with transition probabilities from prefixes to roots in nested words (e.g. *unpredictable) because prefixes (e.g. un-) and roots (e.g. predict) form no morphological constitutes (e.g. *unpredict) and thus never appear in the training data.

Now the theoretical question arises why low-level visual evoked response components like M170 in the visual ventral stream “know” high-level linguistic representations like abstract hierarchical structures. One possibility is that, given the functional connectivity between the left fusiform gyrus and the left inferior frontal gyrus in visual word recognition (Pammer et al., 2004), M170 can be modulated in a top-down feedback manner by “Broca’s area”, the traditional “language” area proposed to process abstract hierarchical structures (Friederici, 2002, 2012). This possibility becomes even less surprising if visual cortex can be sensitive to abstract hierarchical structures (Dikker et al., 2009). Therefore, the functional connectivity between the left fusiform and inferior frontal gyri remains to be empirically investigated in the future research (Carreiras et al., 2014; Woodhead et al., 2014).

Nevertheless, there are several limitations with our computational modeling. One of the several important issues is that “amorphous” models in-

Figure 2: Neurological accuracies of computational models. The x-axis is computational models, while the y-axis is neurological accuracies (i.e. decreases in deviance). Points represent computational models: blue = Character Markov Model, orange = Syllable Markov Model, yellow = Morpheme Markov Model, green = Hidden Markov Model, purple = 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 ($p < 0.01$).
Figure 3: Error accuracies of computational models. The x-axis is computational models, while the y-axis is error accuracies (i.e. decreases in absolute residual errors), categorized into linear (Left) and nested (Right) morphological structures and averaged across individual derivational affixes. The horizontal dashed line indicates a “tie” borderline where computational models do not diverge from the baseline model, and more positive and negative error accuracies mean better and worse predictions relative to the baseline model.

Investigated in this paper are too simplistic as compared to computational models recently proposed in the morphological processing literature such as Naive Discriminative Learning (Baayen et al., 2011; Milin et al., 2017) or Linear Discriminative Learning (Baayen et al., 2018, 2019). Those state-of-the-art computational models of morphological processing remain to be constructed and evaluated against human neural activities and computational models investigated in this paper.

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 technique developed in the sentence processing literature (Frank et al., 2015; Brennan et al., 2016), computationally modeled morphological processing in the human brain, with special focus on the VWFA in the visual ventral stream. Five neuro-computational models of morphological processing were constructed and evaluated against human neural activities in order to investigate whether morphemes are neurologically represented in the human brain and parsed into hierarchical morphological structures: Character Markov Model and Syllable Markov Model as “amorphous” models without morpheme units, and Morpheme Markov Model, Hidden Markov Model (HMM), and Probabilistic Context-Free Grammar (PCFG) as “morphous” models with morpheme units structured linearly or hierarchically. Our MEG experiment and computational modeling demonstrated that “morphous” models of morphological processing outperformed “amorphous” models, PCFG was most neurologically accurate among those “morphous” models, and PCFG better explained nested words with non-local dependencies between prefixes and suffixes. These results strongly suggest that morphemes are neurologically represented in the human brain and parsed into hierarchical morphological structures. In conclusion, neuro-computational modeling of natural language must be a promising future direction in the cognitive computational neuroscience of language (Kriegeskorte and Douglas, 2018; Naselaris et al., 2018).
Acknowledgments

We would like to thank three anonymous reviewers of the Society for Computation in Linguistics and the members of the Neuroscience of Language Lab at New York University for valuable comments and suggestions. This work was supported by JSPS KAKENHI Grant Numbers JP18H05589 and JP19H04990 (YO) and the NYU Abu Dhabi Institute Grant Number G1001 (AM).

References


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


Victoria Fossum and Roger Levy. 2012. Sequential vs. hierarchical syntactic models of human incremental sentence processing. Proceedings of the


