Wray, S., Stockall, L. & Marantz A. (2021). Early Form Based Morphological Decomposition in Tagalog. Neurobiology of Language, Advance publication. https://doi.org/10.1162/nol_a_00062.

1. Early Form Based Morphological Decomposition in Tagalog: MEG evidence from Reduplication, Infixation and Circumfixation
2. Decomposition in Tagalog: MEG evidence
3.

Samantha Wray – Dartmouth College; New York University Abu Dhabi

Linnaea Stockall – Queen Mary University of London

Alec Marantz – New York University; New York University Abu Dhabi

4. Samantha Wray samantha.wray@nyu.edu
5. The authors would like to extend sincere thanks to Jianjun Hua for statistical consultation, M. Julieta Guzman and Nathan Quimpo for their assistance in stimuli creation, and to the members of the New York University and New York University Abu Dhabi Neuroscience of Language Lab.
6. Alec Marantz serves on the editorial team of the Neurobiology of Language Journal as a senior editor
7. New York University Abu Dhabi Institute under Grant G1001

Abstract

Neuro- and psycholinguistic experimentation supports the early decomposition of morphologically complex words within the ventral processing stream, which MEG has localized to the M170 response in the (left) visual word form area (VWFA). Decomposition into an exhaustive parse of visual morpheme forms extends beyond words like “farmer” to those imitating complexity (e.g. “brother”, Lewis et al. 2011), and to “unique” stems occurring in only one word but following the syntax and semantics of their affix (e.g. “vulnerable”, Gwilliams & Marantz 2018). Evidence comes primarily from suffixation; other morphological processes have been under-investigated. This study explores circumfixation, infixation, and reduplication in Tagalog. In addition to investigating whether these are parsed like suffixation, we address an outstanding question concerning semantically empty morphemes. Some words in Tagalog resemble English “winter” as decomposition is not supported (wint-er); these apparently reduplicated pseudoreduplicates lack the syntactic and semantic features of reduplicated forms. However, unlike “winter,” these words exhibit phonological behavior predicted only if they involve a reduplicating morpheme. If these are decomposed, this provides evidence that words are analyzed as complex, like English “vulnerable”, when the grammar demands it. In a lexical decision task with MEG, we find that VWFA activity correlates with stem:word transition probability for circumfixed, infixed and reduplicated words. Furthermore, a Bayesian analysis suggests that pseudoreduplicates with reduplicate-like phonology are also decomposed; other pseudoreduplicates are not. These findings are consistent with an interpretation that decomposition is modulated by phonology in addition to syntax and semantics.

1. Introduction

The process of word recognition is necessarily complicated for words composed of multiple morphemic constituents. Are morphologically complex words decomposed during lexical access? Does this decomposition occur early in the word recognition pipeline before meaning is associated with morphemic units, and what aspects of a word’s internal structure determines this? The current study aims to contribute unstudied morphological phenomena to the growing body of literature focused on early form-based morphemic decomposition.

1.1 Visual word recognition

Full decomposition models (contra non-decompositional models, i.e. Giraudo & Grainger 2000) posit an early automatic form-based decomposition of complex words into the orthographic forms of their constituent morphemes during visual lexical access (including Taft & Forster, 1975; Taft, 1979; Taft, 2004; Crepaldi, Rastle, Coltheart, and Nickels 2010).

Much evidence delineating the discriminatory nature of this morphological parser has emerged. In masked priming studies, "teacher" primes "TEACH" but "brother" also primes "BROTH", despite the fact that the orthographic –er is not an affix in that word (Rastle, Davis, and New 2004; Rastle & Davis 2008). This contrasts with the lack of priming between "brothel" and "BROTH" (Rastle et al. 2004), where -el is not a visual form of an English morpheme. Neural evidence from magnetic resonance imaging (MRI; Gold & Rastle 2007), magnetoencephalography (MEG; Lehtonen, Monahan, and Poeppel 2011; Lewis, Solomyak, and Marantz 2011; Fruchter and Marantz 2015; Cavalli, Colé, Badier and Ziegler 2016) and electroencephalography (EEG; Lavric, Clapp, and Rastle 2007; Morris, Frank, Grainger, & Holcomb 2007; Morris, Grainger, and Holcomb 2008, Royle, Drury, Bourguignon, and Steinhauer 2010; Morris & Stockall 2012; Beyersmann, Iakimova, Ziegler, & Colé 2014) further support a semantics-independent morphological parser as the responsible mechanism for this phenomenon. MEG research by Tarkiainen, Helenius, Hansen, Cornelissen and Salmelin (1999), and fMRI studies by Dehaene, Le Clec, Poline, Le Bihan, and Cohen (2002) localized a possible neural basis for character string processing to the fusiform gyrus, specifically the visual word form area (VWFA). In MEG, this region has been shown to be a generator of a visually-evoked response component peaking approximately 170 ms after stimulus onset (the M170) that was originally targeted for possible relevance for morphology as a bilateral component sensitive to a word's exhaustive parsability (Zweig & Pylkkänen 2009). In subsequent studies, the left M170 was found to index several lexical variables associated with morphological parsing, including affix frequency and the transition probability from a stem to the whole word, both for bound stems and free stems (Solomyak & Marantz 2010). The ERP analog to the M170 response appears to be the N250, which consistently shows effects of morphological priming but not semantic priming in the studies cited above (see Morris & Stockall 2012, and Royle & Steinhauer 2021 for reviews and discussion of this literature).

M170 activity elicited by "brother" words correlates with the stem:whole word transition probability (often abbreviated as TP or TPL in the literature) given a stem of "broth", just as the M170 evoked by genuinely complex words like "teacher" correlates with the stem:whole word transition probability given the stem "teach" ; this is not true for "brothel" words (Lewis et al. 2011). In addition to this dependence of decomposition on the presence of an affix, a viable stem must result from the parse stripping the suffix, as evidenced by the comparison between "brother" and "winter" (Zweig & Pylkkänen 2009), where "winter" patterns with the morphologically simple words given the non-existence of a stem "wint." The stem involved in an exhaustive morphological parse may be bound, provided the word follows morphosyntactic rules associated with its suffix. Thus, M170 activity is predicted by a model computing the M170 from transition probability (and other variables) for "vulnerable" (from the unique bound stem "vulner" to the suffix -able – a transition probability of 1) as it is morphosyntactically and semantically congruent with other adjectives with the –able affix. This is not the case for e.g., "sausage" (from "saus" to "age," also a transition probability of 1), since the combination of "saus(e)" and "age" would not conform to any rule in English, given the meaning of "sausage" (Gwilliams & Marantz 2018).

A summary of the previous results in the literature on morphological processing in occipito-temporal regions is presented in Table 1.

Study	**Morphological Variable**	**Timing and laterality**	**Morphological type**	**Language**
Zweig & Pylkkänen (2009)	complexity	prefix: 174-182 ms bilateral suffix: 170-186 ms right hemisphere	prefixation, suffixation	English
Vartiainen et al. (2009)	complexity	200–800 ms left hemisphere (temporal)	suffixation	Finnish
Solomyak & Marantz (2010)	stem:whole word TP	178–214 ms left hemisphere	suffixation	English
Lewis et al. (2011)	stem:whole word TP	164–208 ms left hemisphere	suffixation	English
Lehtonen et al. (2011)	stem:suffix TP for low semantic opacity	220 ms left hemisphere	suffixation	English

Fruchter et al. (2013)	morphophonological congruency	158–183 ms left hemisphere	irregular	English
Gwilliams & Marantz (2018)	stem:whole word TP	150-180 ms left hemisphere	suffixation	English
Neophytou et al. (2018)	stem:whole word TP	100-200 ms left hemisphere	suffixation	Greek
Hakala et al. (2018)	Morfessor (minimum description length) (Rissanen, 1978 Creutz & Lagus 2007)	140ms-200ms bilateral	suffixation	Finnish
Ohta et al. (2019)	root:affix TP	150ms-200ms left hemisphere	suffixation	Japanese
Stockall et al. (2019)	stem:whole word TP	200-220 ms right hemisphere	prefixation	English

Table 1: A summary of MEG studies demonstrating correlation of morphological variables, including transition probability (TP), with activity in occipito-temporal regions.

The current study expands upon these studies typologically, and more generally informs our knowledge of automatic decomposition during early visual word recognition. The study allows us to determine if previously attested automatic decomposition effects and their accompanying theories extend from languages with relatively more simplistic morphological processes to those with more complicated processes. Moreover, Tagalog exhibits morphologically triggered phonological phenomena that allow us to determine whether phonological cues to morphological complexity are attended to in early visual processing. The results of the current study are consistent with those in Table 1 which demonstrate the correlation of M170 activity with morphological measures, suggesting that the effects of a complex word's internal structure modulate activity in anterior fusiform gyrus regardless of the morphological process underlying that word's complexity. Support for this conclusion is comprised of results from seven word types: (i) reduplicated; pseudoreduplicated of two types: (ii) those exhibiting phonological behavior indicative of morphological complexity; and (iii) those which do not; (iv) infixed; and (v) non-infixed but with a phono-orthographic string that could be an infix (a "winter" type); (vi) circumfixed; (vii) unambiguously morphologically simple words not imitative of complexity. Relevant morphophonological details are reviewed in the sections which immediately follow.

1.2 Reduplication in Tagalog

The current study includes a focus on phonological transparency as a perceptual cue to morphological complexity.

Reduplication in Tagalog can feed transparently applied phonological rules[1], creating phonological non-identity between the base and copy (reduplicant). However, reduplicates in Tagalog can also exhibit a non-transparent application of phonological rules, keeping base and copy more similar phonologically than they would be if the rules applied normally. In non-transparent application, phonological rules apply to both the base and the reduplicant despite the fact that only one of the segments fulfills the environmental requirements for application of the rule, or fail to apply even though one of the segments falls into the usual triggering environment. (Wilbur 1973, Carrier 1979; Marantz 1982; McCarthy & Prince 1995). An example of failure to apply a rule governing the raising of the vowel /o/ to /u/ in reduplication is shown in (1b). Contrast this with transparent application in suffixation in (1a).

(1) Phonological rule application and suffixation/reduplication

Stem		Complex form	
a. tapos	"ending"	tapusin	"to be finished" (Zuraw 2009)
b. b**o**to	"vote"	b**o**b**o**to	"will vote"

1.3 Pseudoreduplication in Tagalog

There is a class of Tagalog words that superficially appear to be reduplicated but do not have an independent stem and lack the morphosyntax of a reduplicated word (termed "pseudoreduplicates" by Zuraw (2002)). Attempts to reduce the repeated orthophonological material to a base and reduplicating morpheme both violate stem minimality constraints in Tagalog (stems are generally bi-syllabic) and are rejected by native speakers as words of the language. Examples of pseudoreduplicates are shown in (2):

[1] We use the term "rule" to refer to emergence of phonological phenomenon. Whether this occurs in a serial application, or as Zuraw (2002) suggests, via the ranking of Optimality Theoretic constraints, is beyond the scope of the current study and has no bearing on the results discussed within.

(2) Pseudoreduplicated words (Zuraw 2002)

a.	mismis	“scraps”	*mis
b.	luloŋ	“swallowing”	*loŋ
c.	ŋasŋas	“scandal”	*ŋas

For a subset of these pseudoreduplicated words, phonological rules are applied transparently with no exceptions for identity between the ‘base’ and ‘reduplicant’, consistent with the word being morphologically simple. For a minority of the pseudoreduplicated words, however, a rule is over/under applied, much as it would be for a true reduplicated word. Examples of pseudoreduplicants exhibiting transparent and non-transparent application are shown in (3). Pseudoreduplicated words which exhibit non-transparent application of phonological rules are marked with [+i] as they phonologically *imitate* true reduplicates; those which transparently apply phonological rules as expected of morphologically simple words are marked with [-i].

(3) Transparent and non-transparent phonology in pseudoreduplicates (Zuraw 2002)

a. dubdob[2]	“vehemence”	Transparent application [-i]
b. gonggong	“grunt fish”	Non-transparent application [+i]

The current study aimed to answer the question: are [-i] pseudoreduplicated words which transparently apply rules processed differently than those [+i] pseudoreduplicated words which do not? Specifically, given that non-transparent application makes a pseudoreduplicated word appear more like a product of morphological reduplication, are these [+i] pseudoreduplicated words processed like reduplicated words? If pseudoreduplicated words are decomposed in parallel to truly reduplicated words, the neurolinguistic evidence would support Zuraw’s (2002) hypothesis that these words are represented with a syntactically and semantically null reduplicating morpheme.

[2] Native speaker judgment for items in the current study placed a certain degree of variability on non-transparent application of the vowel height rule for pseudoreduplicated words, in addition to the variability noted by Zuraw (2002). If the underapplication of the vowel height rule was acceptable, the word was considered to have non-transparent application, even if the transparent form was also considered acceptable.

1.4 Infixation in Tagalog

In Tagalog, an infix follows the first consonant of the base (Schachter & Otanes 1983). Tagalog utilizes several infixes, including -*in*- which marks perfective aspect[3]. Examples of this infix are shown in (4):

(4) -in- Infixation

	Stem	Infixed
a.	subok "try"	sinubok "tried"
b.	gapos "cord"	ginapos "tied/banned"
c.	gulat "surprise"	ginulat "shocked someone"

Tagalog also has words with initial syllables ending in /in/ which are not morphologically complex. In this way, these words are analogous to previously-studied word types in English discussed in detail above that contain phono-orthographic strings consistent with an affix but that are not treated as morphologically complex by visual perception areas in the brain sensitive to relations between morphemes. Specifically, much like "winter" or "sausage," the stripping of the affix does not result in a viable stem, and furthermore the word is not morphosyntactically congruent with words that contain the affix (Zweig & Pylkkänen 2009, Gwilliams & Marantz 2018). Examples of words with initial syllables ending in /in/ that are morphologically simple appear in (5). Note that there is no isolable stem in these words, and they do not exhibit the morphosyntax indicative of –in– infixed words (namely, the words are not perfective verbs). We term these words pseudo-infixed.

(5) Pseudo-infixed /in/

a.	ministro "ministry"	*mistro
b.	ninoŋ "godfather"	*noŋ
c.	pinsaŋ "cousin"	*pisaŋ

[3] Although the current study is comprised of -in- infixed words which are completive, when -in- appears with reduplication, it indicates imperfect patient focus.

The current study then aims to discover if pseudo-infixed words are processed as the evidence from English processing predicts (i.e. broth-er vs. winter (Zweig & Pylkkänen 2009); excurs-ion vs. sausage (Gwilliams & Marantz 2018)). If morphosyntactic indexing and stem viability are coded for Tagalog infixes much the same way they are for English suffixes, we expect that the pseudo-infixes will not be automatically stripped during the word recognition process.

1.5 Predictions and Design

The present study aims to explore the implications of Tagalog morphology, including reduplication, infixation, and circumfixation, for the early evoked activity in occipito-temporal cortex associated automatic decomposition in visual word recognition models. Furthermore, the study aims to determine whether words that appear to be reduplicated or infixed based on their written form are automatically decomposed, and what modulates this decomposition. The study includes two blocks, run in the same experimental session. Block 1 investigates processing of words formed through reduplication and words with circumfixes. Block 1 also compares real reduplicated words to [-i] pseudoreduplicated words which transparently apply phonological rules and [+i] pseudoreduplicated words which non-transparently apply rules (i.e. are reduplicate-like). Block 2 compares processing of infixed words to pseudo-infixed words which superficially appear to have an infix but which are morphologically simple.

A summary of the design of the two blocks with accompanying hypotheses about decomposition for each word type is presented in Table 2:

Condition	Sample Item	Prediction for decomposition	Results for decomposition
Block 1			
simple	aberya “flawed”	✗	✗
reduplicated	araw-araw “everyday”	✓	✓
[-i] pseudoreduplicated – transparent phonology	musmos “naïve”	✗	✗
[+i] pseudoreduplicated – non-transparent phonology	gonggong “grunt fish”	✓	✓
circumfixed	ka-ruwag-an “cowardice”	✓	✓
Block 2			
simple	lungkot “sadness”	✗	✗
infixed –in-	t-in-awag “called”	✓	✓
pseudo-infixed /in/	bintang “accusation”	✗	✓
circumfixed	ka-bayar-an “payment”	✓	✓

Table 2: Conditions of MEG experiment investigating the processing of reduplicated and infixed forms, and words which orthographically appear to be reduplicated or infixed but are morphologically simple. The Simple condition contains unambiguously simple words which have no orthographic imitation of complexity. Hyphens are included to indicate morpheme boundaries.[4]

This experiment tests several hypotheses about what information is used in early, automatic morpheme segmentation by the visual system, and from which morphemes this information is accessible. First, we address the hypothesis that circumfixed, infixed, and

[4] Note that there is an inconsistent distribution of parts of speech across conditions, as words which have reduplication or circumfixation as their only means of varying morphological complexity tend to be nouns, whereas infixed words tend to be verbs. However, transition probability is the feature of interest, and it has been demonstrated to influence the processing of both nouns and verbs, even within the same experiment (Lewis et al. (2011)).

reduplicated words will be processed as a function of their morphemic transition probability, as has been attested for English, Greek, and Finnish suffixes. Under this hypothesis, pseudo-infixed words will not be automatically parsed. Furthermore, we hypothesize that the decomposition of pseudoreduplicated words will be modulated by phonological transparency, as those which imitate reduplicated words by virtue of their nontransparent application of phonological rules will be processed as if they are reduplicated.

2. Methodology

2.1 Participants

Twenty right-handed participants took part in the study (13 females, ages 24-46, mean age = 33). A language history was collected, and speakers who self-reported being native speakers of Tagalog were retained in the study; speakers who self-reported their native language as another Filipino language such as Cebuano/ Bisaya were not retained. All participants reported normal or corrected-to-normal vision. Written informed consent was obtained from all individuals prior to participation in the experiment.

2.2 Materials

Stimuli were selected from a Tagalog dictionary (English 1965), in addition to words identified by Zuraw (2002). Frequency counts were taken from a 5-million word Wikipedia corpus (Oco & Roxas 2012). Finally, the stimuli were vetted by a native speaker for lexicality and decomposability (defined as ability to isolate a definable stem). To determine whether or not each word transparently applied phonological rules, the native speaker also provided judgments on forms which incorporated additional affixation not utilized in the experiment. A summary of the properties of the stimuli is in Table 3:

Condition	Average frequency in parts per million (SD)	Average length in letters (SD)
Block 1		
reduplicated	1.11 (±.85)	7.5 (±1.46)
pseudoreduplicated – transparent application	1.19 (±1.17)	5.4 (±.61)
pseudoreduplicated – non-transparent application	1.03 (±2.51)	6.3 (±.87)
circumfixed	1.06 (±.76)	9.5 (±.97)
Block 2		
infixed –in-	18.9 (±26.22)	7.4 (±1.07)
pseudo-infixed /in/	21.1 (±29.47)	6.5 (±1.54)
circumfixed	17.4 (±24.13)	9.1 (±.96)

Table 3: Properties of items included as visual lexical decision stimuli in experiments with concurrent MEG.

Nonwords in both blocks were created using the nonce word generator toolkit Wuggy by scrambling possible syllables using real Tagalog words as training input (Keuleers & Brysbaert 2010). Then, an appropriate number of the nonce stems underwent the morphological processes in Table 3. For example, an equal number of nonce stems was "reduplicated" to the reduplicated items included as target items in the experiment. This was simply to ensure that participants did not develop a strategy for decision that obscured the desired results.

Although circumfixed items were consistently the longest items in length of letters, and frequency was only matched within block and not across blocks, both length and frequency were added as fixed effects in the linear mixed effects model (described in detail in section 2.4) so that they did not confound an analysis focusing on Condition.

2.3 Procedure

Data were collected at New York University Abu Dhabi overseen by New York University Abu Dhabi's Institutional Review Board. Before beginning, all participants provided informed, written consent. Participants lay supine in a dimly-lit magnetically shielded room while stimuli were presented on a screen suspended 85 cm above the head. Stimuli were presented in black Times New Roman font (corresponding to a display size of ¾ inch/ 2 cm) against a grey background using the experiment control software Presentation (Neurobehavioral Systems). Prestimulus presentation of a fixation cross in the middle of the screen lasted for 50 ms. Stimulus order was fully randomized across and between 5 sets for each blocks, and participants were directed to indicate via button press with the non-dominant (left) hand whether they recognized each word as a word of their language or not. Participants were instructed to answer as quickly and as accurately as possible. After each block, participants could take a self-timed break during which they could perform small movements to remain comfortable. A short break also occurred between blocks 1 and 2. The total time for the experiment averaged 20 minutes.

MEG data were continuously recorded concurrently with accuracy and reaction time (RT) data. MEG data were recorded with a 1000 Hz sample rate on a 208-channel axial gradiometer system (Kanazawa Institute of Technology, Kanazawa, Japan) and went through an online low-pass filter at 200 Hz and high-pass filter at 0.1 Hz.

Participants' head shapes were digitized for source localization and coregistration using a FastSCAN laser scanner (Polhemus, VT, USA). Digitized head shapes were downsampled to create a smoothed surface using the FastSCAN software. Digital fiducial points were marked for each participant across the forehead, the anterior of the left auditory canal, and the anterior of the right auditory canal. Marker coils were taped to each participant's head where the fiducials were recorded. A measurement of marker coil position was taken before and after each block to correct for participant movement post-hoc.

2.4 Analysis

The first step in preprocessing MEG data was noise removal from the raw data using eight reference channels located away from the individual's head using the Continuously Adjusted Least Squares Method (CALM) (Adachi, Shimogawara, Higuchi, Haruta & Ochiai 2001) which was performed using the MEG160 software (Yokohawa Electric Corporation and Eagle Technology Corporation, Tokyo, Japan). Subsequent pre-processing and analysis of MEG data was performed using MNE-Python (Gramfort, Luessi, Engemann, Strohmeier, Brodbeck, Parkkonen, Hämäläinen 2014; Gramfort, Luessi, Engemann, Strohmeier, Brodbeck, Goj, Jas, Brooks, Parkkonen, Hämäläinen 2013) and Eelbrain 0.25.2 (Brodbeck 2017) An Independent Components Analysis (ICA, specifically fast-ica) was performed on the full noise-reduced data to isolate and remove components corresponding to biomagnetic artifacts such as eye movement (blinks, saccades) and pulse. Following ICA, the data went through a low-pass infinite impulse response (IIR) 4th order Butterworth forward-backward filter with an upper cutoff frequency of 40 Hz. The data was epoched from 500 ms preceding stimulus onset to 500 ms following stimulus onset. Manual rejection of epochs to remove those contaminated by motor artifacts as well as those with activity exceeding +/-2,000 fT/cm was performed using Eelbrain, resulting in removal of 1.7 % of trials. Epochs were not baseline corrected. Rather, 50 ms preceding the fixation cross were included as a fixed effect in the linear mixed effects model, following Alday (2019).

MEG data were co-registered with the FreeSurfer average brain (CorTechs Labs Inc, La Jolla, CA, USA) by manually scaling the participants' digitized head shapes and the FreeSurfer average skull. An ico-4 source space was created consisting of 5124 sources using a cortically-constrained minimum norm estimate model (Hämäläinen & Ilmoniemi 1994). Signed minimum estimates were used based on previous research showing their superiority to unsigned estimates in studying orthographic processing (Gwilliams, Lewis & Marantz 2016). For each source, a Boundary Element Model (BEM, see Mosher, Leahy, and Lewis 1999) was used to compute the forward solution. The inverse solution using the forward solution was calculated and subsequently applied to the data with a fixed orientation of the dipole current. A signed fixed orientation for the source estimates was used to calculate the inverse solution, such that the direction of the current was defined and dipoles were perpendicular to the cortical surface. Finally, the data were noise-normalized in the spatial dimension, resulting in a dynamic statistical parameter map (dSPM, see Dale, Liu, Fischl, Buckner, Belliveau, Lewine, Halgren 2000).

Using the anterior fusiform functional region of interest (fROI) defined by Gwilliams et al. (2016), activity averaged across space was plotted using MNE-Python (Gramfort et al. 2013, Gramfort et al. 2014) for the M170 to be manually identified. Further analyses on this data were performed by using activity averaged across space and time as input for linear mixed effects models (using R 3.6.1: R Core Team (2019); lme4 1.1-21: Bates, Maechler, Bolker, and Walker (2015)).

Behavioral data (specifically, RTs and accuracy) were analyzed using linear mixed effects models (also using R: R Core Team, (2019); lme4 Bates et al. (2015)). Items below chance accuracy were excluded from all analyses except the analysis of accuracy.

3. Results

3.1 MEG Data

3.1.1. Complex words

Analyses were focused on activity in the left hemisphere fusiform gyrus (Figure 1), specifically in the anterior region identified by Gwilliams et al. (2016) as a functional ROI, plotted in Figure 1. Gwilliams et al. (2016) identified this fROI by running an English adaptation of the Tarkiainen et al. (1999) study on "Type Two" responses associated with the perception of visible letter strings vs. those obscured with visual noise, which was earlier and more posterior, and the perception of letter strings vs. symbol strings, which was later and more anterior. Crucially, they demonstrated that activity in the anterior region correlated with transition probability from morphologically complex English words (Solomyak & Marantz 2010), and were able to spatiotemporally separate this response from activity associated with the visual noise manipulation. We selected 150 to 200 ms as the time window for analysis and the most likely candidate for the M170. As presented in detail in section 1.1, previous research has variously identified time windows from 100-200 ms (Neophytou, Manouilidou, Stockall, and Marantz 2018, Stockall, Manouilidou, Gwilliams, Neophytou, and Marantz 2019, Fruchter et al. 2013), 130-180 ms (Gwilliams et al. 2016), 150-180ms (Gwilliams & Marantz 2018), 140-220 ms (Lewis et al. 2011). This selection appeared consistent with the wave form morphology; averaged activity from this fROI plotted by condition is shown in Figure 2.

Figure 1: Ventral view of region of interest (ROI) for M170: VWFA (left) using coordinates from Gwilliams et al. (2016), located approximately in anterior fusiform gyrus (right). Shows inflated cortical surface of FreeSurfer average subject (Fischl et al. 1999). Plot created in MNE-Python (Gramfort et al. 2013, Gramfort et al. 2014)

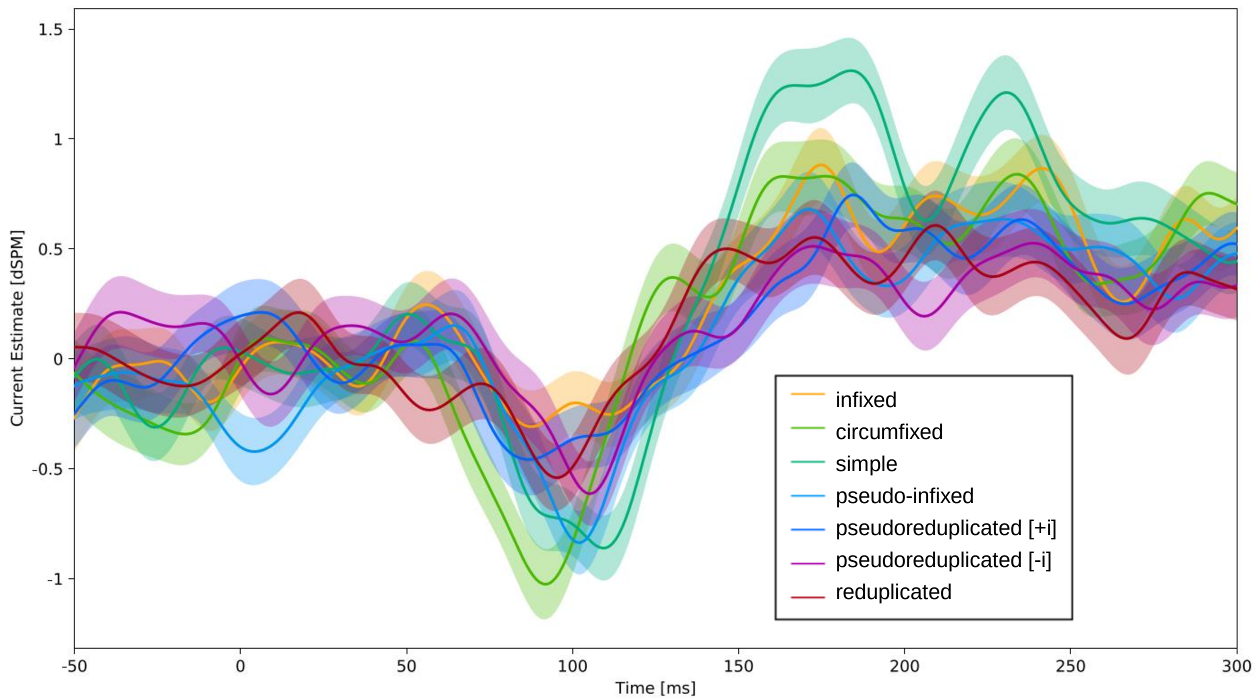


Figure 2: Time course and average activity (current estimates in unitless z) in VWFA from time of stimulus presentation to 300ms after stimulus presentation. Shaded areas represent standard error of the mean. Plot created in Eelbrain (Brodbeck 2017).

Analysis of the neural results was completed in two steps: first, a linear mixed effects regression (LMER) was fit for activity elicited across all word types. Then, activity for simple words that could potentially be parsed ([-i] pseudoreduplicated, [+i] pseudoreduplicated, pseudo-infixed) were compared to their complex counterparts using Bayesian estimation and evaluating the resulting posterior probability distributions.

For the first analysis, we used an LMER to investigate the effects of morphemic transition probability, as well as additional lexical properties, on left hemisphere dSPM averaged across space (the VWFA) as well as averaged across time (from 150 to 200 ms). Fixed effects in the model included the *base dSPM* of 50 ms pre-stimulus period (following Alday 2019) with 50 ms selected as the pre-stimulus baseline time period to mirror the 50 ms time period of interest for post-stimulus dSPM, stem:whole word *transition probability*, word *length* in letters, natural log of stem *frequency* as continuous variables, as well as the fixed effect of the categorical variable *condition* (reduplicated, circumfixed, infixed –in, simple, pseudo-infixed –in, pseudoreduplicated [+i], pseudoreduplicated [-i]). The *interaction* of transition probability and condition was also included in the model. A by-subject intercept and by-subject slope of *length* were also included in the model. The significance of fixed effects was determined using Wald tests on the coefficients using the Satterthwaite approximation for the degrees of freedom (implemented in the lmerTest package, Kuznetsova et al., 2017). Selection of the random effects proceeded via backward selection from the maximal model for both subject and item effects using the lmerTest package 3.1-1 (Kuznetsova, Brockhoff & Christensen 2017) (for discussion, see Barr, Levy, Scheepers, and Tily et al. 2013; Barr 2013; Bates, Kliegl, Vasishth & Baayen, (2015) and Matuschek, Kliegl, Vasishth, Baayen & Bates (2017)). Treatment coding is specified for condition, with the reference level being the reduplicated condition. To check for collinearity, the generalized variance inflation factor (GVIF) was calculated using the car package (Fox & Weisburg 2019); when taking degrees of freedom into account, no GVIF was greater than 2.94. The full model summary after random effect reduction is shown in Table 4.

Formula: dSPM ~ base_dSPM + TP * condition + Length + BaseFreqlog + (1 | Subject) + (Length | Subject) + (BaseFreqlog|Subject)

Fixed effects:

	Estimate	df	t value	Pr(>|t|)
(Intercept)	0.75	408.02	1.635	0.10286
Base dSPM	-0.14	4333.88	-9.545	2e-16 ***
Transition Probability	0.56	4295.51	1.371	0.17035
Condition = simple	0.35	4295.31	0.975	0.32940
Condition = pseudo-infixed	0.78	4295.53	-0.292	0.77000
Condition = pseudoredup [+i]	-0.56	4295.31	-1.456	0.14560
Condition = pseudoredup [-i]	-0.79	4295.48	-2.039	0.04146 *
Condition = circumfix	0.73	4294.6	2.776	0.00554 **
Condition = infixed	0.78	4294.64	2.397	0.01659 *
Length	-0.04	213.07	-0.701	0.48392
log (Base Frequency)	-0.02	36.69	-0.469	0.64187

Interaction, TP:Condition = circumfix	-0.51	4295.16	-0.949	0.34266
Interaction, TP:Condition= infixed	-1.20	4295.12	-2.234	0.02554 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Random effects:

	Variance
Subject	0.399313
Length|Subject	0.009153
Base Frequency|Subject	0.015711
Residual	10.313368

Table 4: Summary of LMER showing correlation coefficients of lexical statistics and word types to source component amplitudes (*left* hemisphere). Treatment coding was used for the categorical predictor condition, with the *reduplicate* condition serving as the reference level. Estimates have been rounded to 2 decimal places. Calculation of p values from t-tests and dfs was performed using Satterthwaite’s method in the lmerTest package (Kuznetsova, Brockhoff & Christensen 2017).

There was a significant interaction between transition probability and the reduplicated and infixed levels of condition indicating that the effect of transition probability on dSPM was not consistent across morphological types. The effect of transition probability for reduplicated words was significantly different than for infixed words (t(4295.12) = -2.23, p = 0.03). There was no significant difference on the effect of transition probability for circumfixed words and reduplicated words (t(4295.16) = -0.95, p = 0.34). This is plotted in Figure 3, which shows that the relationship between transition probability and dSPM is positive for reduplicated and circumfixed words: as it becomes more likely for a whole word to contain its stem, more activity is elicited in the left hemisphere VWFA. This pattern is consistent with those attested for English and Greek suffixes (English: Solomyak & Marantz 2010, Lewis et al. 2011, Gwilliams & Marantz 2018; Greek: Neophytou et al. 2018). However, for infixed words, as it becomes more likely for a whole word to contain its stem, less activity is elicited. The morphologically simple words (conditions: simple, pseudo-infixed, pseudoreduplicated [+i], pseudoreduplicated [-i]) all have Transition Probabilities equal to 1, so there was no corresponding interaction term and the main effects can be interpreted directly. Of most interest are the comparisons between reduplicated and pseudoreduplicated [-i] as well as between reduplicated and pseudoreduplicated [+i]. There was a significant difference between reduplicated and pseudoreduplicated [-i] (t(4295.48) = -2.039. p = 0.04). This is consistent with the hypothesis that pseudoreduplicated [-i] would not be processed like reduplicated words, that is, they would not be automatically decomposed, because they are not phonologically imitative of reduplicated words. In contrast, there was no significant difference between reduplicated and pseudoreduplicated [+i] words ($t(4295.31) = -1.46$. $p = 0.15$). Finally, both length ($t(213.07) = -0.70$, $p = 0.48$) and stem frequency ($t(36.69) = -0.47$, $p = 0.64$) were not significant.

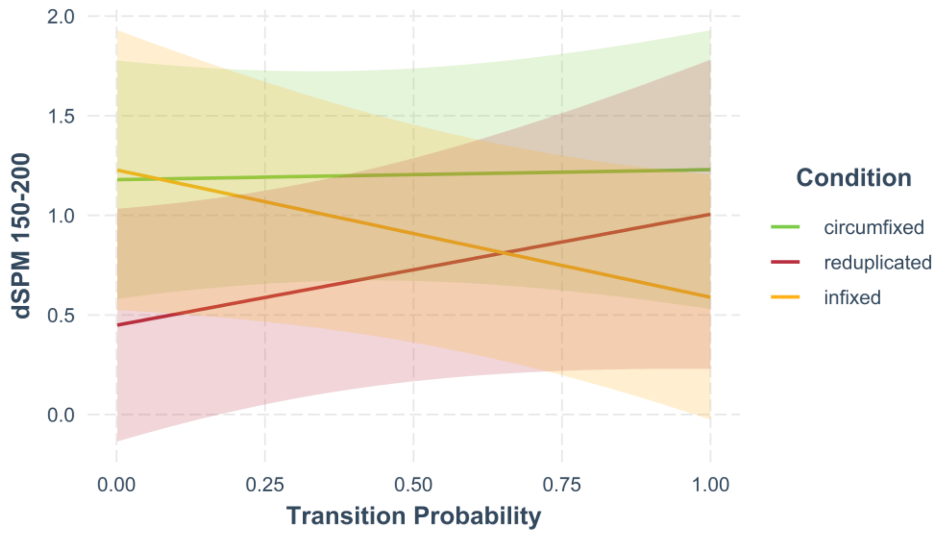
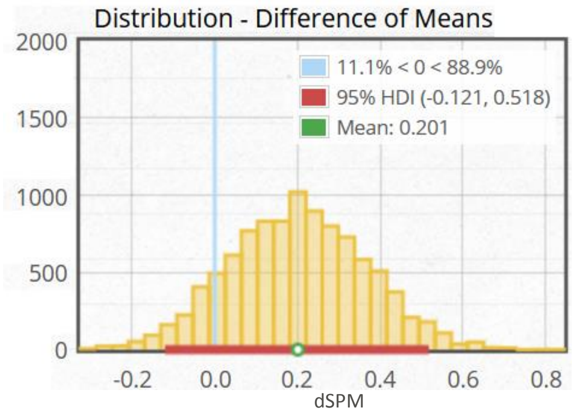


Figure 3: Average activity plotted against stem:whole word transition probability separated by word type. This illustrates an interaction between condition and transition probability. Shaded areas represent 95% confidence interval. Plot created in R (R Core Team 2019) using jtools 2.0.1 (Long 2019).

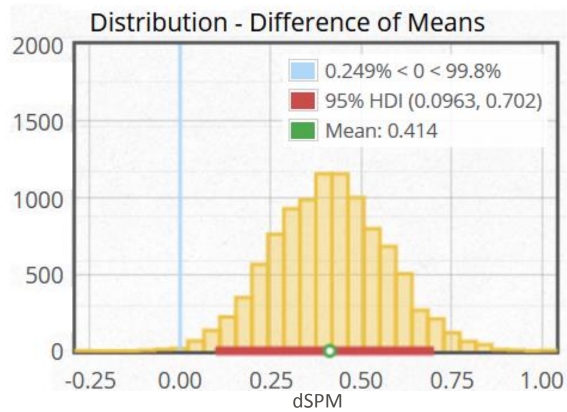
To determine if there was a bilateral effect, the process was repeated for the right-hemisphere homologue to the VWFA. No effect was found, but the results can be found in the supplementary materials.

3.1.2. Comparison between complex and pseudo-complex words

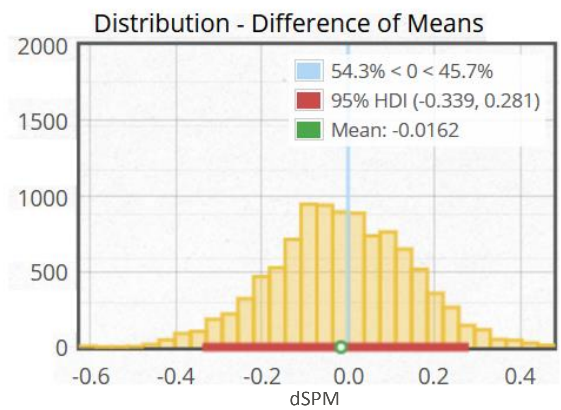
It is possible to evaluate comparisons between word types further by using a Bayesian Parameter Estimation approach. A posterior probability distribution was calculated for the difference in dSPM values between a complex word type (reduplicated and infixed) and its corresponding pseudo- word type ([+i] pseudoreduplicated, [-i] pseudoreduplicated, and pseudo-infixed), using Metropolis-within-Gibbs Markov chain Monte Carlo (MCMC) sampling with 10,000 samples (using the Bååth 2012 implementation of Kruschke 2012, 2013). Based on the posterior probability distribution, shown in the Difference of Means in Figure 4, we quantified the probability that word types elicited similar dSPM values based on comparing observed dSPM from complex and pseudo-complex types.



(a) reduplicated vs. [+i] pseudoreduplicated



(b) reduplicated vs. [-i] pseudoreduplicated



(c) infixed vs. pseudo-infixed

Figure 4: Histograms of differences of means produced by 10,000 MCMC samples per word type. The vertical light blue line marks 0 difference between the predicted means. The horizontal red line indicates the Highest Density Interval (HDI), or 95% of the predicted difference of means. Plots from Bååth (2012) implementation of Kruschke (2013).

First, we begin with a comparison of reduplicated words and pseudoreduplicated words. Figures 4a and 4b demonstrate a contrast between pseudoreduplicated types. The difference between reduplicated and [+i] pseudoreduplicated, shown in 4a, is estimated to be credibly zero, as indicated by a 0 estimated difference of means being within 95% highest posterior probability density interval.[5] This is indicative of equivalent values. This is consistent with an interpretation that [+i] pseudoreduplicated words and reduplicated words elicit similar dSPM values. In

[5] An alternative approach is to specify a Region of Practical Equivalence (ROPE, for details see Kruschke, 2013) based on effect size and determine if 95% percent of the Difference of Means Distribution falls within this.

contrast, in Figure 4b, the difference between reduplicated words and [-i] pseudoreduplicated words was determined to be non-zero: a 0 estimated difference of means is outside the 95% likelihood density. This is consistent with an interpretation that [-i] pseudoreduplicated words and reduplicated words elicit different dSPM values.

Next, a comparison of infixed words and pseudo-infixed words was undertaken. This difference was also estimated to be credibly zero, as shown in Figure 4c. A 0 estimated difference of means is within 95% likelihood density.

Taken together, these provide evidence that [+i] pseudoreduplicated and pseudo-infixed words are processed like their complex (reduplicated) counterparts, whereas [-i] pseudoreduplicated transparent words are not. This is indicative of decomposition for two of the three pseudo-complex types. Our hypotheses stated that [+i] pseudoreduplicated nontransparent words would be automatically decomposed given that their phonology is imitative of reduplicated words, whereas [-i] pseudoreduplicated transparent words would not be.

3.2 Behavioral Data

3.2.1 Reaction time

RTs for responses to target items were analyzed using two linear mixed-effects regression models, one fit to all words, and one fit to complex words only, to determine a possible effect of transition probability. Before analysis, RT were trimmed to discard responses less than 300 ms or more than 1000 ms from stimulus onset, and RT was log transformed. A graphical summary of RT are shown in Figure 5.

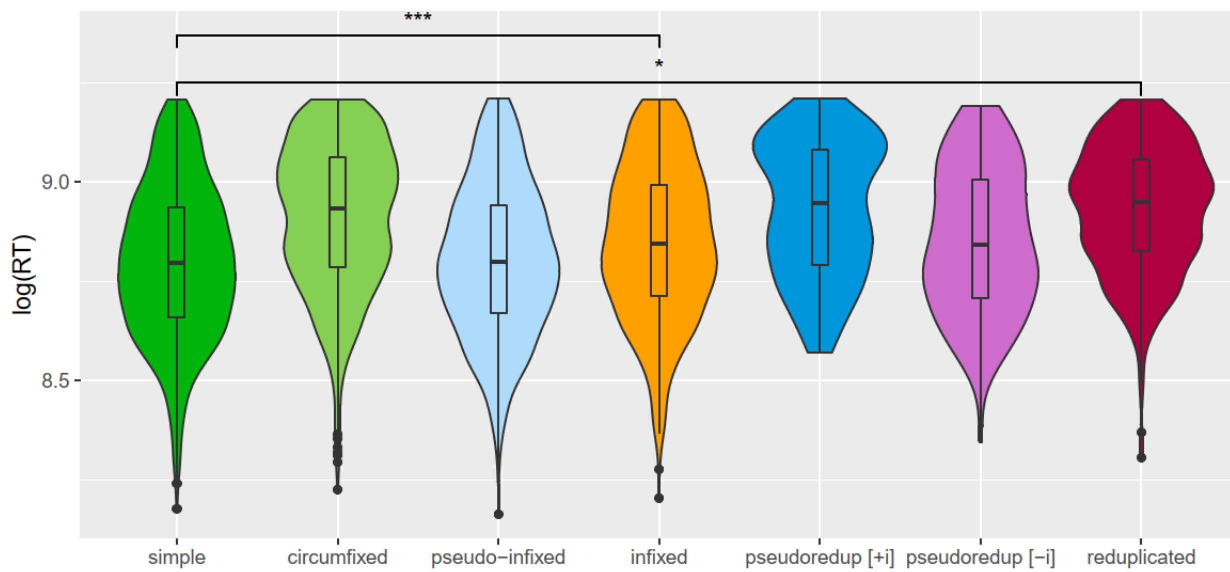


Figure 5: Violin plot showing graphical summary of RTs. Comparisons between morphologically simple and other conditions are from the model in Table 5. Plot created in R 3.6.1 (R Core Team 2019) using ggplot2 3.3.0 (Wickham 2016) and ggsignif 0.6.0 (Ahlmann-Eltze 2019).

Fixed effects included in the full model were: *condition* (morphologically simple, circumfixed, pseudo-infixed, infixed, [+i] nontransparent pseudoreduplicated, [-i] transparent pseudoreduplicated, reduplicated), log-transformed *item frequency*, and item *length* in letters. After reducing from a maximal model, random intercepts for participant and item were also included in the model, as well as a by-subject slope for *item frequency*. GVIF was calculated to check for collinearity, with no GVIF greater than 1.83. Length was correlated with response speed ($t(259) = 6.81$, $p < 0.001$); longer words were responded to more slowly than shorter words. Frequency was also correlated ($t(88) = -4.33$, $p < .001$), with more frequent words being recognized more quickly.

Treatment coding was specified, allowing for a comparison of conditions to the morphologically simple condition. Two of the morphologically complex conditions were significantly different from the morphologically simple condition when controlling for length and frequency (reduplicate $t(247) = 2.16$, $p = .032$; infix $t(239) = 3.61$, $p < .001$). However, despite predictions from the MEG results supporting the automatic decomposition of pseudo-infixed words, there was no significant difference between pseudo-infixed and morphologically simple words ($t(224) = -1.00$, $p = 0.32$). The MEG results also supported automatic decomposition for [+i] nontransparent pseudoreduplicated words. For the behavioral results, the difference between [+i] words and morphologically simple words was not significant (t(254) = 1.80, p = .07). On the other hand, the MEG results do not support the automatic decomposition of [-i] transparent words. In this, the behavioral results agree, since those results are not significant either (t(241) = 0.279, p = .78).

All words:
Formula: RTlog ~ Condition + Freqlog + Length + (1 | Subject) + (1 | Item) + (WordFreq|Subject)

Fixed effects:

	Estimate	df	t value	Pr(>|t|)
(Intercept)	-8.66	175	266.918	< 2e-16 ***
Condition = circumfix	0	250	0.054	0.957
Condition = pseudo-infix	-0.02	224	-0.996	0.32
Condition = infix	0.06	238	3.614	0.000368 ***
Condition = pseudoredup [+i]	0.05	254	1.800	0.0731 .
Condition = pseudoredup [-i]	0.01	241	0.279	0.7807
Condition = reduplicate	0.05	247	2.163	0.0315 *
Length	0.03	259	6.805	6.99e-11 ***
Word Frequency	-0.02	88	-4.331	3.94e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Random effects:

	Variance	Correlation
Subject	3.829e-03	
Word Frequency|Subject	9.685e-05	0.54
Item	4.758e-03	
Residual	2.366e-02	

Table 5: Summary of LMER showing correlation coefficients of RT, lexical statistics and word types to RT. Treatment coding is specified, allowing for a comparison of conditions to the *morphologically simple* condition. Estimates have been rounded to 2 decimal places. Calculation of p values from t-tests and dfs was performed using Satterthwaite's method in the lmerTest package (Kuznetsova, Brockhoff & Christensen 2017).

3.2.2 Accuracy

Overall, accuracy rates were high for both blocks, with an average of 91% accuracy across subjects and items. A binomial logit generalized linear mixed-effects model was fit to analyze accuracy, using *log RT* as a predictor (following Davidson & Martin 2013). In addition to RT, item *condition*, *log frequency,* and item *length* were included in the model. Inclusion of random slopes and intercepts was reduced iteratively starting from a maximal model as described above, resulting in a model with by-subject and by-item intercepts. GVIF was calculated to check for collinearity, and no GVIF was found to be greater than 1.90.

Frequency was found to be a significant predictor of accuracy (z = 2.72, p = .00646) As shown in Table 6, simple words were set as the reference level with treatment coding for levels of condition. Reduplicated words were found to be significantly different from simple words (z = 2.32, p = 0.02044). The summary of the full model is shown in Table 6.

All words

Formula: Accuracy ~ Condition + RTlog + Freqlog + Length + (1 | Subject) + (1|Item)

Fixed effects:

	Estimate	z value	Pr(>|z|)
(Intercept)	-0.40	-0.08	0.93754
Condition = circumfix	1.13	1.44	0.14974
Condition = pseudo-infix	0.81	1.43	0.15234
Condition = infix	0.98	1.81	0.07059 .
Condition = pseudoredup [+i]	-0.50	-0.67	0.50561
Condition = pseudoredup [-i]	-0.65	-0.92	0.35801
Condition = reduplicate	1.66	2.32	0.02044 *
log(RT)	0.60	1.02	0.30701
Length	-0.19	-1.19	0.23285
log(Frequency)	0.30	2.72	0.00646 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Random effects:

	Variance
Subject	0.9919
Item	2.4244

Table 6: Summary of binomial mixed effect logistic regression showing correlation coefficients of RT, lexical statistics and word types to Accuracy. Treatment coding is specified, allowing for a comparison of conditions to the *morphologically simple* condition.

4. Discussion

As outlined in detail in the introduction, the focus of the present study was: Are reduplication, circumfixation, and infixation subject to automatic decomposition by the visual system? Furthermore, are words which superficially appear to be reduplicated or infixed but lack the morphosyntactic and semantic features of these words treated as complex words by the visual system? Finally, is the tendency for a word to be treated by the visual system like a reduplicated word modulated by its conformity to phonological rules?

We addressed these questions by measuring activity elicited in the putative visual word form area in anterior fusiform gyrus. The major findings from the present study are outlined below. In sum, results from the present study are largely consistent with theories of visual word processing that incorporate automatic decomposition of a word into its stem and affixes (Taft & Forster, 1975; Taft, 1979; Taft, 2004; Crepaldi, et al. 2010). The present study makes two novel contributions to the literature concerning this topic: first, it adds typological breadth through the inclusion of the understudied language Tagalog, and second, it demonstrates that words formed via previously unstudied morphological processes are also decomposed during visual word recognition. Furthermore, the current study presents further evidence, previously attested for the English irregular past tense (Fruchter et al. 2013), of a mechanism for early automatic decomposition at the intersection of morphology and phonology: if a pseudo-complex word applies phonological rules analogous to a complex word, it will be decomposed, despite the lack of any morphosyntactic indicators of complexity. However, our current results diverge from previously-attested constraints of morphosyntactic congruency or stem viability as pseudo-infixed words appear also to be automatically decomposed despite a lack of stem viability without the affix.

4.1 Automatic early decomposition of infixed, reduplicated, and circumfixed words

Segmental information is used by the early visual system to decompose many types of complex words, including those formed by some process other than affixation, namely reduplication. This is evidenced by the effect of stem:whole word transition probability on elicited activity in the left hemisphere. These results are consistent with a robust collection of results from previous studies on suffixation in English (Solomyak & Marantz 2010, Lewis et al. 2011, Gwilliams & Marantz 2018) and Greek (Neophytou et al. 2018). Furthermore, Stockall, Manouilidou, Gwilliams, Neophytou, and Marantz (2019) determined that early automatic form-based decomposition of prefixed English words followed a similar pattern to suffixed words, differing only in hemisphere laterality.

The results of the current study with respect to activity in the left-hemisphere VWFA for morphologically complex words are also noteworthy because of the significant interaction between stem:whole word transition probability and word type. Reduplicated words elicit greater activity for higher values of stem:whole word, which is consistent with both the prefix and suffix literature (Table 1 above). However, infixed words exhibit the opposite pattern. It is possible also that a single stem:whole word transition probability value for infixed words is not sufficient to completely capture their morphological structure, as they have two morpheme boundaries where the infix meets the stem at both its left and right edges. What remains true, despite the direction of the correlation between transition probability and dSPM, is that transition probability for all complex words correlated with activity in left VWFA.

4.2 Decomposition of words with orthophonemic strings which imitate infixes

Our results in support of automatic decomposition of words with pseudo-infixes diverge from results from previous studies on English, which have investigated underlying rules governing visual morpheme representations. Three different kinds of pseudo complex items have been investigated in English: words like *brother*, which contain a viable free stem ‘broth’ as well as the viable affix '-er', words like *winter*, which have the affix, but no viable stem, and words like *vulnerable*, which similarly have no viable free stem, but differ from winter-type words in that the affix makes the same contribution to the syntax and semantics of the whole word as it does in clearly complex words like *workable*[6]. Tagalog pseudo-infixed words are most similar to English winter-type words: removing the infix does not leave a viable stem, and the whole word does not have the grammar that would be expected if it contained the infix -in-. Despite this, we presented results consistent with the hypothesis that pseudo-infixed words are automatically decomposed anyway: values of activity from both pseudo-infixed and infixed words were compared using a Bayesian estimation, indicating the values were probably very similar. However, the behavioral evidence did not show that pseudo-infixed words were processed at a different speed than other morphologically simple words; truly morphologically infixed words were.

[6] The suffix -ble creates adjectives with ‘possibility’ semantics (Oltra-Massuet 2013), in both *workable* and *vulnerable* (compare 'winter' which is neither an agentive nominal nor a comparative adjective).

4.3 Morphologically simple pseudoreduplicated words imitate morphologically complex reduplicated words in their application of phonological rules

The current study compared two types of pseudoreduplicates: those that imitated truly complex reduplicated words in their phonology ([+i]; non-transparent), and those that applied phonological rules as expected for morphologically simple words ([-i]; transparent). The former elicited activity patterns consistent with automatic decomposition as if they were morphologically complex, whereas the latter did not. Therefore, conformity to phonological rules modulates the decomposability of pseudoreduplicated words.

Morphophonological generalizability aiding in the segmentation of complex and pseudo-complex words follows from previous research on English irregular past tense processing. Fruchter et al. (2013) demonstrated that irregular verbs are decomposed into stems and affixes in early written word recognition by correlating priming within the M170 time window to an irregular verb's conformity to a morphophonological rule (formalized computationally by Albright & Hayes 2003).

5. Conclusion

Our results make several important contributions to our understanding of the neural correlates of morphological decomposition. First, reduplication, infixation, and circumfixation are all comparable to prefixation and suffixation in that they are automatically parsed by the ventral visual system during word recognition, as evidenced by stem:whole word transition probability correlations with activity in VWFA. Additionally, we posit that phono-orthographic cues to morpheme boundaries aid in this automatic decomposition process, as words which are not reduplicated but appear to be so superficially due to their under- and over- application of phonological rules are also decomposed. Collectively, these results are consistent with models of visual word recognition that entail automatic decomposition for all morphological processes.

References

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y. and Ochiai, M., 2001. Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. IEEE Transactions on Applied Superconductivity, 11(1), pp.669-672.

Albright, A. and Hayes, B., 2003. Rules vs. analogy in English past tenses: A computational/experimental study. Cognition, 90(2), pp.119-161.

Alday, P. M. 2019. How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. Psychophysiology, 56(12), e13451.

Ahlmann-Eltze, C. 2019. ggsignif: Significance Brackets for ‘ggplot2’. R package version 0.5. 0.

Bååth, R. (2012). Bayesian Estimation Supersedes the t-test (BEST)-online. http://www.sumsar.net/best_online/

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

Bates, D., Maechler, M., Bolker, B., and Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.

Beyersmann, E., Iakimova, G., Ziegler, J. C., & Colé, P. 2014. Semantic processing during morphological priming: An ERP study. Brain Research, 1579, 45–55.

Brodbeck, C. 2017. Eelbrain (Version v0.25). Zenodo. doi:http://doi.org/10.5281/zenodo.438193

Carrier, J.L., 1979. The interaction of morphological and phonological rules in Tagalog: a study in the relationship between rule components in grammar (Doctoral dissertation, Massachusetts Institute of Technology).

Cavalli, E., Colé, P., Badier, J. M., Zielinski, C., Chanoine, V., & Ziegler, J. C. 2016. Spatiotemporal dynamics of morphological processing in visual word recognition. Journal of Cognitive Neuroscience, 28(8), 1228-1242.

CorTechs Labs Inc, La Jolla, CA, USA

Crepaldi, D., Rastle, K., Coltheart, M. and Nickels, L., 2010. ‘Fell’primes ‘fall’, but does ‘bell’prime ‘ball’? Masked priming with irregularly-inflected primes. Journal of memory and language, 63(1), pp.83-99.

Creutz, M., & Lagus, K. 2007. Unsupervised models for morpheme segmentation and morphology learning. ACM Transactions on Speech and Language Processing (TSLP), 4(1), 1-34.

Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D. and Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron, 26(1), pp.55-67.

Davidson, D. J., & Martin, A. E. 2013. Modeling accuracy as a function of response time with the generalized linear mixed effects model. Acta psychologica, 144(1), 83-96.

Dehaene, S., Le Clec'H, G., Poline, J.B., Le Bihan, D. and Cohen, L., 2002. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. Neuroreport, 13(3), pp.321-325.

English, L. 1965. English-Tagalog Dictionary. Manila: Department of Education, Republic of the Philippines

Fox J, Weisberg S. 2019. An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Fruchter, J., & Marantz, A. 2015. Decomposition, lookup, and recombination: MEG evidence for the full decomposition model of complex visual word recognition. Brain and language, 143, 81-96.

Fruchter, J., Stockall, L. and Marantz, A., 2013. MEG masked priming evidence for form-based decomposition of irregular verbs. Frontiers in human neuroscience, 7, p.798

Giraudo, H. and Grainger, J., 2000. Effects of prime word frequency and cumulative root frequency in masked morphological priming. Language and cognitive processes, 15(4-5), pp.421-444.

Gold, B.T. and Rastle, K., 2007. Neural correlates of morphological decomposition during visual word recognition. Journal of Cognitive Neuroscience, 19(12), pp.1983-1993.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L. and Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. Frontiers in neuroscience, 7, p.267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L. and Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. Neuroimage, 86, pp.446-460.

Gwilliams, L. and Marantz, A., 2018. Morphological representations are extrapolated from morpho-syntactic rules. Neuropsychologia, 114, pp.77-87.

Gwilliams, L., Lewis, G.A. and Marantz, A., 2016. Functional characterisation of letter-specific responses in time, space and current polarity using magnetoencephalography. Neuroimage, 132, pp.320-333.

Hakala, T., Hulten, A., Lehtonen, M., Lagus, K. and Salmelin, R., 2018. Information properties of morphologically complex words modulate brain activity during word reading. Human brain mapping, 39(6), pp.2583-2595.

Hämäläinen, M.S. and Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Medical & biological engineering & computing, 32(1), pp.35-42.

Keuleers, E. and Brysbaert, M., 2010. Wuggy: A multilingual pseudoword generator. Behavior research methods, 42(3), pp.627-633.

Lehtonen, M., Monahan, P.J. and Poeppel, D., 2011. Evidence for early morphological decomposition: Combining masked priming with magnetoencephalography. Journal of cognitive neuroscience, 23(11), pp.3366-3379.

Lewis, G., Solomyak, O. and Marantz, A., 2011. The neural basis of obligatory decomposition of suffixed words. Brain and language, 118(3), pp.118-127.

Kanazawa Institute of Technology, Kanazawa, Japan

Kruschke, J. K. 2013. Bayesian estimation supersedes the t test. Journal of Experimental Psychology: General, 142(2), 573.

Kuznetsova A., Brockhoff P.B. and Christensen R.H.B. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." Journal of Statistical Software, 82(13), pp. 1–26. doi: 10.18637/jss.v082.i13.

Lakens, D. 2017. Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. Social Psychological and Personality Science, 8(4), 355-362

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.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. 2019. Package “emmeans”: Estimated marginal means, aka least-squares means. Compr. R Arch. Netw, 1-67.

Long, JA. 2019. jtools: Analysis and Presentation of Social Scientific Data. R package version 2.0. 1, https://cran.r-project.org/package=jtools.

McCarthy, J.J. and Prince, A., 1995. Faithfulness and reduplicative identity. Linguistics Department Faculty Publication Series, p.10.

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

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

Morris, J., & Stockall, L. 2012. Early, equivalent ERP masked priming effects for regular and irregular morphology. Brain and language, 123(2), 81-93.

Mosher, J.C., Leahy, R.M. and Lewis, P.S., 1999. EEG and MEG: forward solutions for inverse methods. IEEE Transactions on Biomedical Engineering, 46(3), pp.245-259.

Neurobehavioral Systems, Version 18.0; www.neurobs.com

Neophytou, K., Manouilidou, C., Stockall, L. and Marantz, A., 2018. Syntactic and semantic restrictions on morphological recomposition: MEG evidence from Greek. Brain and language, 183, pp.11-20.

Oco, N. and Roxas, R.E., 2012. Pattern matching refinements to dictionary-based code-switching point detection. In Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation (pp. 229-236).

Ohta, S., Osekai, Y., and Marantz, A., 2019. Dissociating the effects of morphemes and letters in visual word recognition: An MEG study of Japanese verbs. Presented at the 25th meeting of Architectures and Mechanisms for Language Processing, Moscow, Russia.

Oltra-Massuet, Isabel. 2013. Deverbal Adjectives at the interface: a crosslinguistic investigation into the morphology, syntax and semantics of-ble. Vol. 28. Walter de Gruyter.

Polhemus, VT, USA

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

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

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

Rissanen, J. 1978. Modeling by shortest data description. Automatica, 14(5), 465-471.

Royle, P., & Steinhauer, K. 2021. Neural correlates of morphology computation and representation. in Grimaldi, M., Y. Shtyrov & E. Brattico, (eds.), Language Electrified. Techniques, Methods, Applications, and Future Perspectives in the Neurophysiological Investigation of Language, Springer.

Sawilowsky, S.S., 2009. New effect size rules of thumb. Journal of Modern Applied Statistical Methods, 8(2), p.26.

Schachter, P. and Otanes, F.T., 1983. Tagalog reference grammar. University of California Press.

Schuirmann D. J. 1987. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics, 15, 657–680.

Solomyak, O. and Marantz, A., 2010. Evidence for early morphological decomposition in visual word recognition. Journal of Cognitive Neuroscience, 22(9), pp.2042-2057.

Stockall, L., Manouilidou, C., Gwilliams, L., Neophytou, K. and Marantz, A., 2019. Prefix Stripping Re-Re-Revisited: MEG Investigations of Morphological Decomposition and Recomposition. Frontiers in psychology, 10, p.1964.

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

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

Taft, M., 2004. Morphological decomposition and the reverse base frequency effect. The Quarterly Journal of Experimental Psychology Section A, 57(4), pp.745-765.

Tarkiainen, A., Helenius, P., Hansen, P.C., Cornelissen, P.L. and Salmelin, R., 1999. Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, *122*(11), pp.2119-2132.

Vartiainen, J., Aggujaro, S., Lehtonen, M., Hultén, A., Laine, M. and Salmelin, R., 2009. Neural dynamics of reading morphologically complex words. Neuroimage, 47(4), pp.2064-2072.

Fox, J. and Weisberg, S., 2019. An R companion to applied regression. Sage publications.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York

Wilbur, R.B., 1973. The phonology of reduplication. Bloomington: Indiana University Linguistics Club.

Yokohawa Electric Corporation and Eagle Technology Corporation, Tokyo, Japan

Zuraw, K., 2002. Aggressive reduplication. Phonology, 19(3), pp.395-439.

Zuraw, K. 2009. Frequency influences on rule application within and across words. Proceedings of CLS (Chicago Linguistic Society) 43.

Zweig, E. and Pylkkänen, L., 2009. A visual M170 effect of morphological complexity. Language and Cognitive Processes, 24(3), pp.412-439.

Supplementary Material A: Right-hemisphere analysis

Formula: dSPM ~ base_dSPM + TP * condition + Length + Freqlog + (1 | Subject)

Fixed effects:

	Estimate	df	t value	Pr(>|t|)	
(Intercept)	-0.20	519.6	-0.503	0.615	
Base dSPM	-0.14	4332	-9.383	2e-16	***
Transition Probability	1.27e-03	4315	0.004	0.997	
Condition = simple	0.12	4315	0.405	0.685	
Condition = pseudo-infixed	-0.17	4315	-0.599	0.549	
Condition = pseudoredup [+i]	-0.06	4315	-0.197	0.844	
Condition = pseudoredup [-i]	0.26	4315	0.828	0.408	
Condition = circumfix	-0.21	4315	-0.968	0.333	
Condition = infixed	-0.12	4315	-0.449	0.653	
Length	0.05	4315	1.374	0.169	
Base Frequency	-8.79e-04	4315	-0.031	0.975	
Interaction, TP:Condition = circumfix	0.21	4315	0.473	0.636	
Interaction, TP:Condition= infixed	-0.33	4315	-0.765	0.444	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Random effects:

	Variance
Subject	0.5413
Residual	6.7971

Table 7: Summary of LMER showing correlation coefficients of lexical statistics and word types to source component amplitudes (*right* hemisphere). Treatment coding was used for the categorical predictor condition, with the *reduplicate* condition serving as the reference level. Estimates have been rounded to 2 decimal places. Calculation of p values from t-tests and dfs was performed using Satterthwaite’s method in the lmerTest package (Kuznetsova, Brockhoff & Christensen 2017).

Figure 1

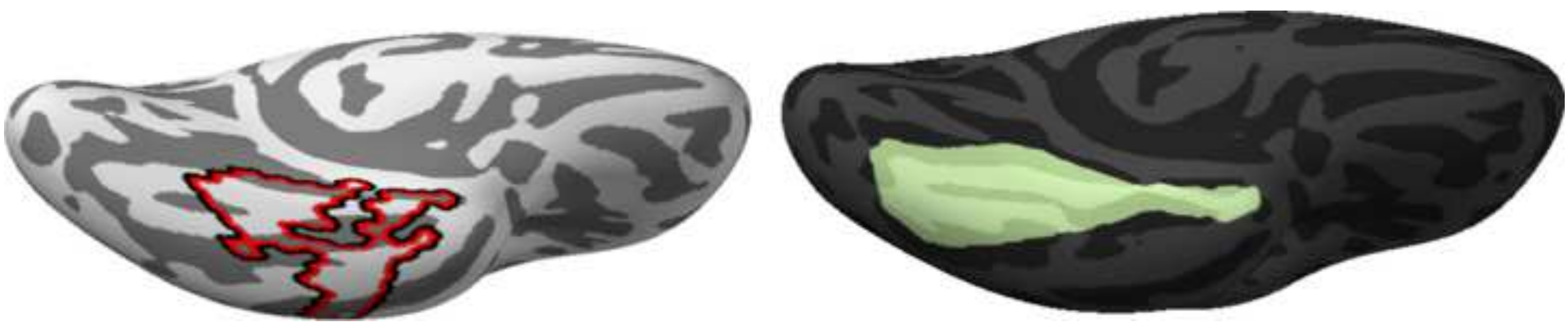


Figure 2

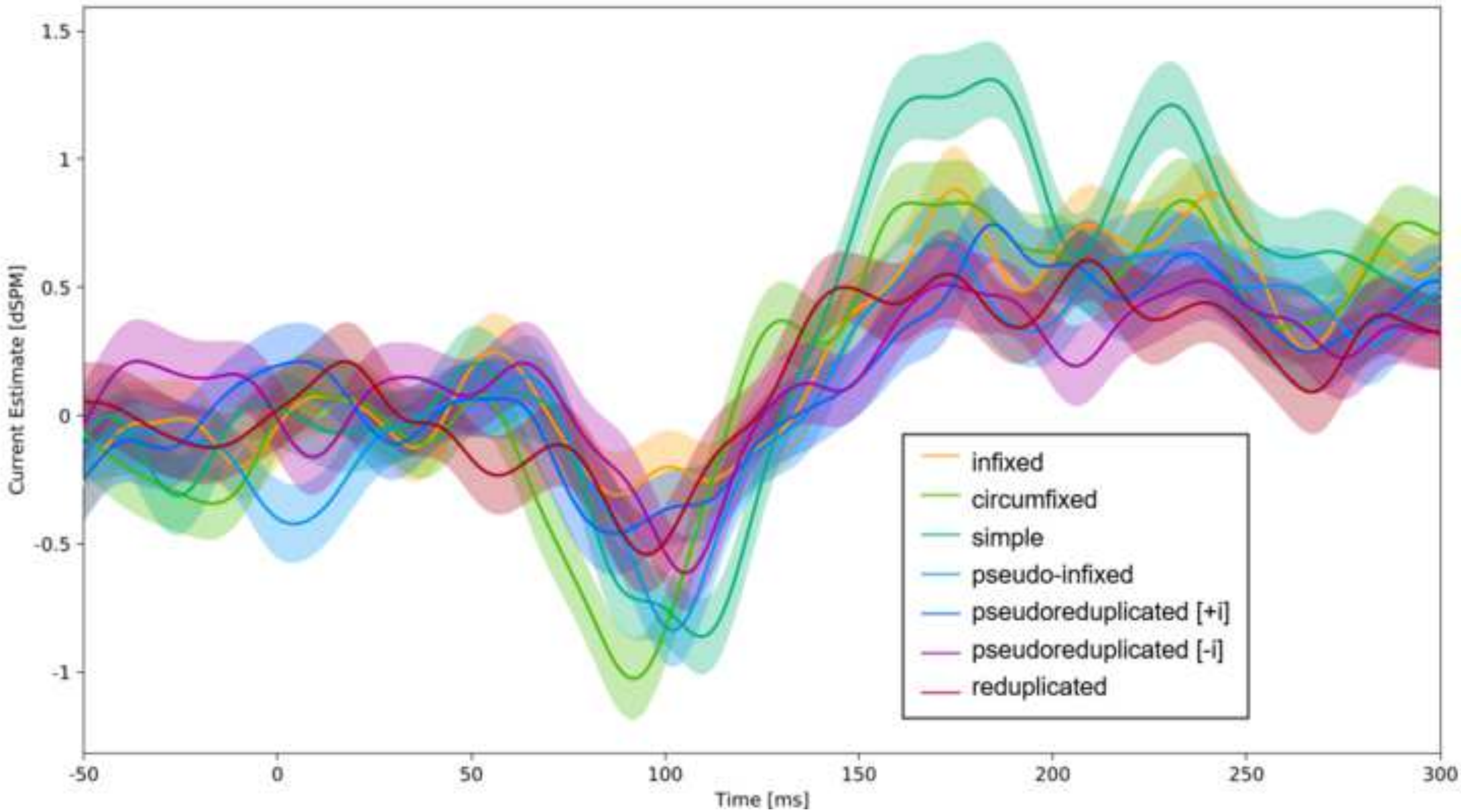

1.5
1
0.5
0
-0.5
-1
Current Estimate [dSPM]
-50
0
50
100
150
200
250
300
Time [ms]
infixed
circumfixed
simple
pseudo-infixed
pseudoreduplicated [+i]
pseudoreduplicated [-i]
reduplicated

Figure 3

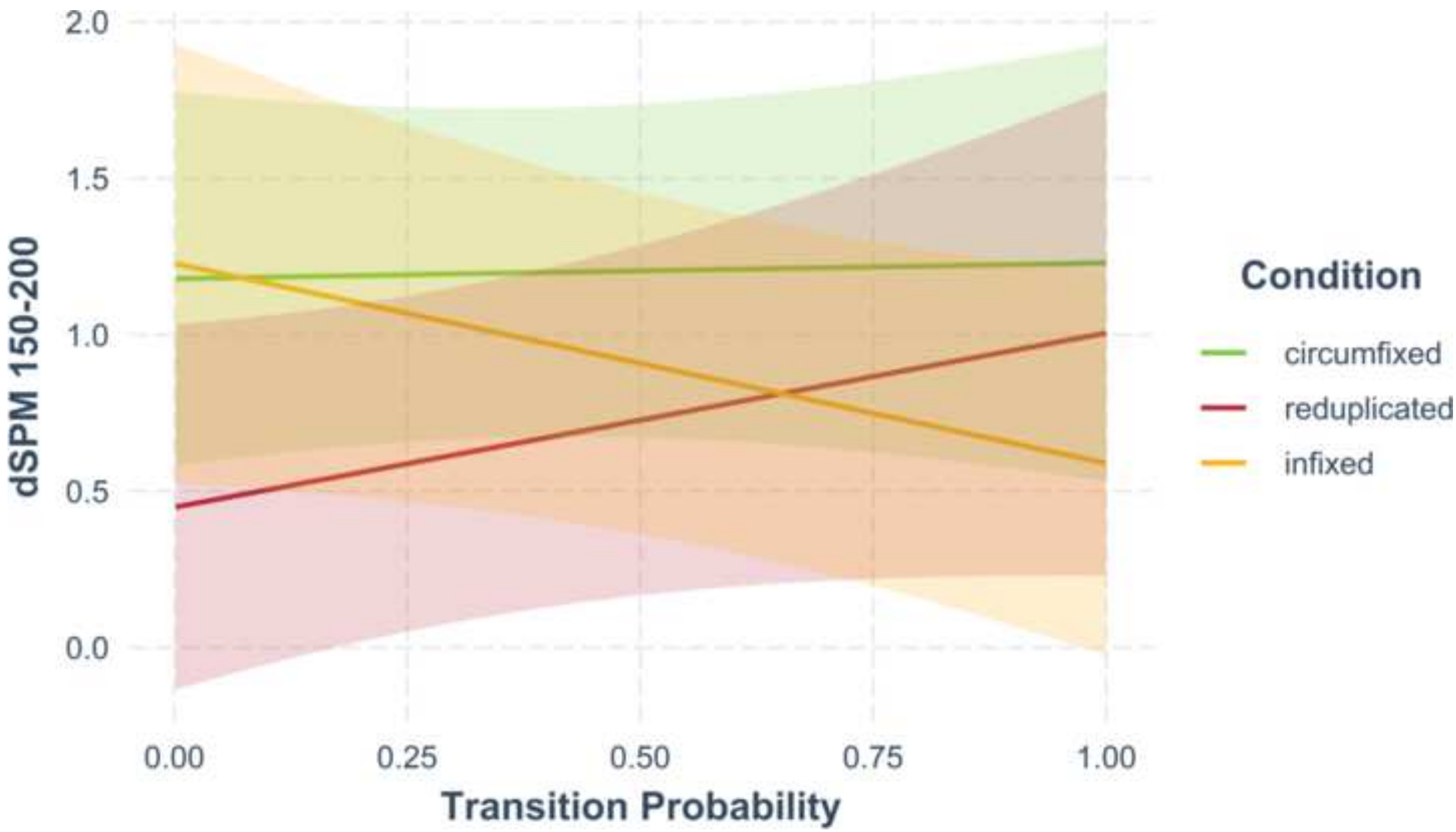
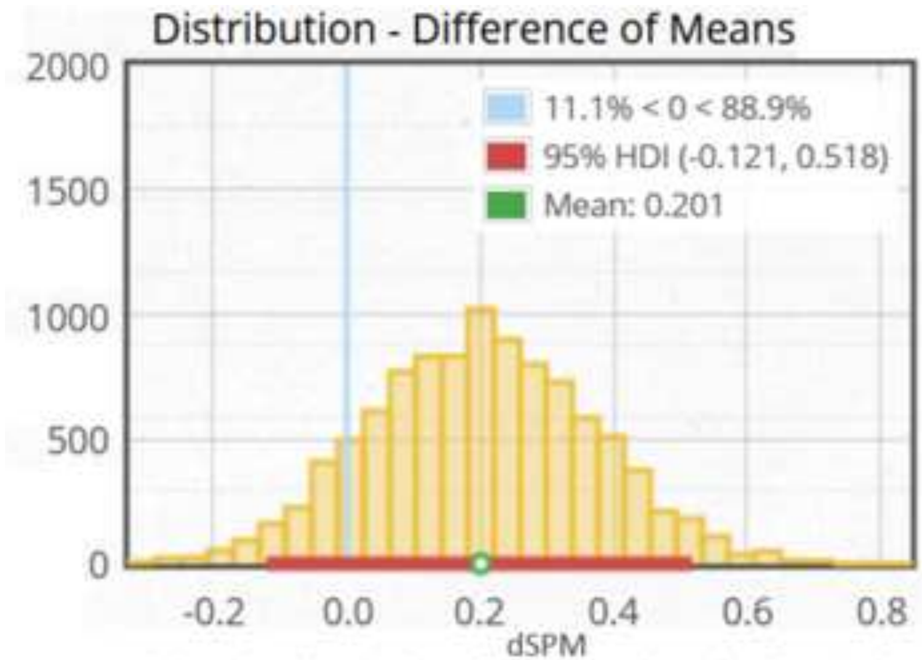
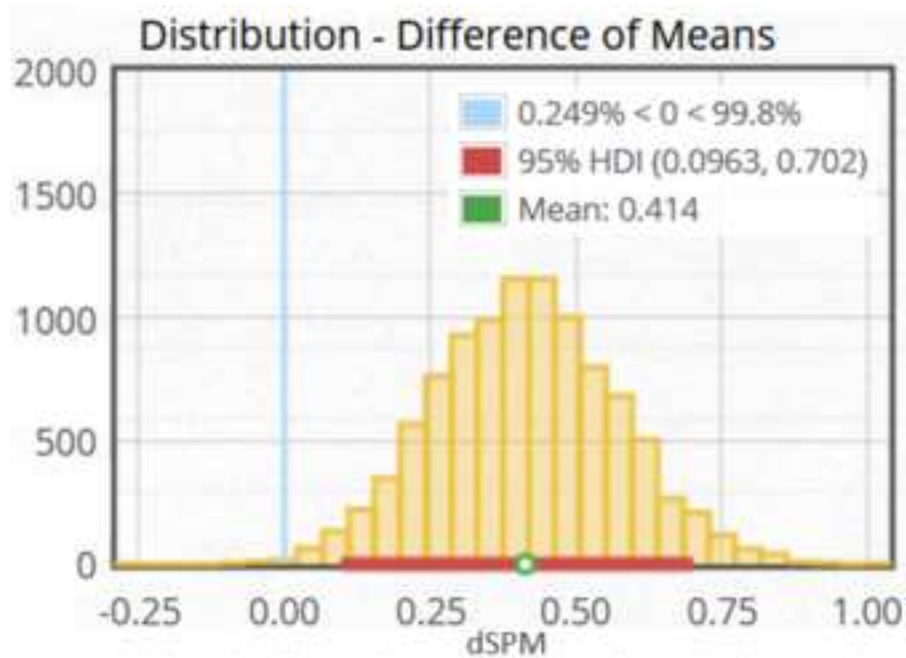

2.0
1.5
1.0
0.5
0.0
dSPM 150-200
0.00
0.25
0.50
0.75
1.00
Transition Probability
Condition
circumfixed
reduplicated
infixed

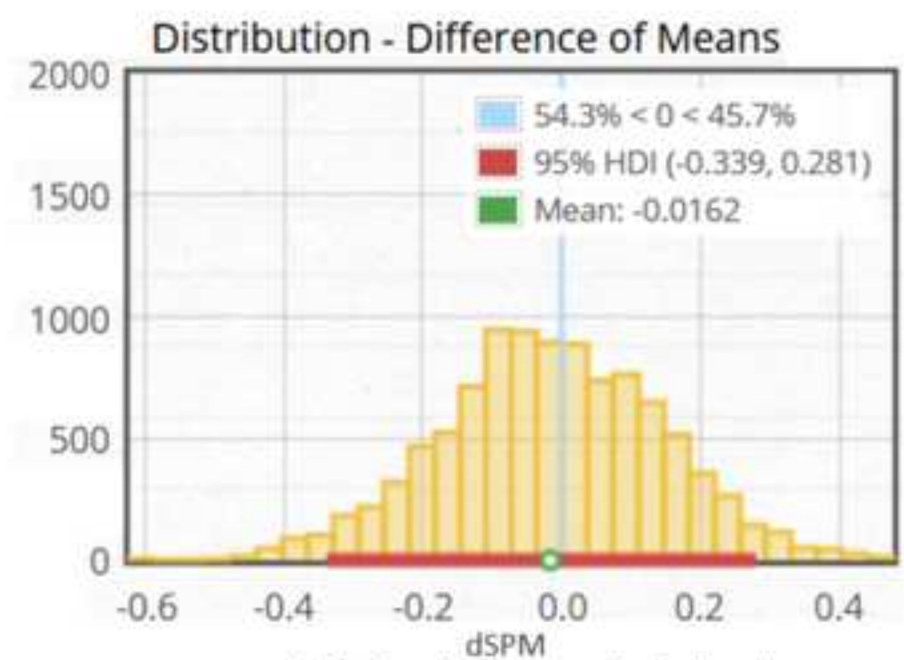
Figure 4



(a) reduplicated vs. [+i] pseudoreduplicated



(b) reduplicated vs. [-i] pseudoreduplicated



(c) infixed vs. pseudo-infixed

Figure 5

Click here to access/download;Figure;figure_5.tiff