

## Syntactic and semantic restrictions on morphological recomposition: MEG evidence from Greek



K. Neophytou<sup>a,\*</sup>, C. Manouilidou<sup>b</sup>, L. Stockall<sup>c,a</sup>, A. Marantz<sup>a,d,e</sup>

<sup>a</sup> NYUAD Institute, New York University Abu Dhabi, United Arab Emirates

<sup>b</sup> Department of Comparative and General Linguistics, University of Ljubljana, Slovenia

<sup>c</sup> Department of Linguistics, Queen Mary University of London, United Kingdom

<sup>d</sup> Department of Psychology, New York University, USA

<sup>e</sup> Department of Linguistics, New York University, USA

### ARTICLE INFO

#### Keywords:

Morphology

Full Decomposition model

Greek

Visual word processing

Magnetoencephalography

### ABSTRACT

Complex morphological processing has been extensively studied in the past decades. However, most of this work has either focused on only certain steps involved in this process, or it has been conducted on a few languages, like English. The purpose of the present study is to investigate the spatiotemporal cortical processing profile of the distinct steps previously reported in the literature, from decomposition to re-composition of morphologically complex items, in a relatively understudied language, Greek. Using magnetoencephalography, we confirm the role of the fusiform gyrus in early, form-based morphological decomposition, we relate the syntactic licensing of stem-suffix combinations to the ventral visual processing stream, somewhat independent from lexical access for the stem, and we further elucidate the role of orbitofrontal regions in semantic composition. Thus, the current study offers the most comprehensive test to date of visual morphological processing and additional, cross-linguistic validation of the steps involved in it.

## 1. Introduction

### 1.1. Processing visually presented, morphologically complex words

The influence of morphological structure on visual word recognition has occupied the psycholinguistic literature for decades. The various models that have been proposed to describe the processing of visually presented, morphologically complex words make claims regarding the way they are accessed as well as the organization of the mental lexicon. For instance, the Full Listing model (e.g. Butterworth, 1983) assumes that complex words are stored as whole units and, consequently, are accessed as such. In contrast, the Full Decomposition model (Rastle & Davis, 2008; Stockall & Marantz, 2006; Taft & Forster, 1975; Taft, 2004) suggests that any visually presented word is first decomposed into its constituent morphemes (bound or free), and only those constituent morphemes are stored in the lexicon. For example, “the prefix *un-* and the suffix *-y* must be stripped off before the lexical representation of *unlucky* (namely *luck*) can be accessed” (Taft & Forster, 1975, p. 638).

A substantial body of both behavioral and neurocognitive evidence provides support for this latter model. Decades of behavioral work, especially work using the masked priming paradigm to target early

stages of form-based processing (see Rastle & Davis, 2008 for a review), and more recent work using EEG and MEG to directly measure those early stages (see below, Section 1.2.) have provided very strong and consistent support for form-based morphological decomposition, operating across all potentially morphologically complex strings (even opaque and irregular words), occurring between ~130–210 ms post stimulus onset (PSO) and localized to left anterior fusiform gyrus (the Visual Word Form Area - VWFA, Cohen 2000).

Further support for the Full Decomposition model can be found by investigating the predictions of the model about the processing that occurs *after* this early form-based decomposition. The Full Decomposition model argues that initial lexical access operates over morphological constituents (stems and affixes) rather than whole words, so that effects of stem properties (such as lemma frequency and variables related to morphological family size, density, etc.) should be observed earlier than effects of whole-word properties (like surface frequency). The model also proposes that initial decomposition and constituent activation are followed by a later stage of recombination, in which the pieces are assembled and the meaning of the whole word is determined, allowing for whole word frequency effects to be found. Schreuder and Baayen (1995) argue that this recombination should be further sub-divided into separate processes: licensing, where the

\* Corresponding author at: Department of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218, USA.  
E-mail address: [kneophy1@jhu.edu](mailto:kneophy1@jhu.edu) (K. Neophytou).

grammatical combinability of the stem and the affix(es) are checked, and *composition*,<sup>1</sup> where the semantic and syntactic properties of the stem + affix(es) combination are computed. For a transparent word like *pushable*, lexeme lookup would involve activation of the stem *push* (to press against with force) and the suffix *-able* (able to be X). During *licensing*, *push* is identified as having a verbal use, and *-able* as a suffix that can attach to verbs to make adjectives, so the combination is licensed. *Composition* then involves computing the meaning and syntactic properties from the parts (an adjective meaning “able to be pressed against with force”). These two distinct processes should rely on distinct neurocognitive resources and thus be associated with distinct spatial and temporal profiles.

Fruchter and Marantz (2015) were the first to simultaneously test two of the predictions of this model. Using MEG to investigate the processing of morphologically complex words in English they were able to distinguish between two post-decomposition processing responses: a region in Middle Temporal Gyrus (MTG) that showed earlier sensitivity to a variable indexing initial lexical access (derivational family entropy, Moscoso del Prado Martín et al., 2004) from 241 to 387 ms PSO and later sensitivity to whole word surface frequency from 431 to 500 ms PSO; and a region in orbitofrontal cortex (OF) that showed sensitivity to derived semantic coherence in a more sustained response (354–500 ms). These results are clearly consistent with the Full Decomposition model as outlined above, and this study provides an important template for further investigation. However, to date no single study has investigated both decomposition and recomposition of the same items, or tried to test both the prediction that stem frequency effects should precede whole word frequency effects and the prediction that *licensing* and *composition* should be dissociable processes. The current study aims to do exactly that, offering the most comprehensive test to date of the Full Decomposition model, as well as of the neurobiological profile of the different responses associated with complex morphological processing.

## 1.2. From decomposition to recombination

### 1.2.1. Early, obligatory, form-based decomposition

The most compelling behavioral evidence that morphological decomposition is an early, obligatory, form-based process, which applies to all potentially morphologically complex strings regardless of semantic transparency, comes from a sizeable body of studies using the masked priming paradigm and investigating the processing of *pseudo-complex* words like 'brother', which can be parsed into a stem + affix (broth + er), even though they are not morphologically complex. Rastle and Davis (2008) review 19 such studies, reporting data from English, French, Dutch and Russian. All but two of these studies find significant lexical decision priming for pseudo-complexity, and in nearly every study, the magnitude of this priming was similar to the magnitude of priming for transparent morphologically related prime-target pairs such as 'teacher'-'teach'. The masked priming effect for the transparent and pseudo-complex conditions stands in clear contrast to the lack of any priming for orthographically similar pairs like 'brothel'-'broth', which can not be parsed into a stem + affix.

Several ERP studies have adopted this masked priming paradigm and confirmed the relative time course the behavioral studies suggest: Lavric, Clapp, and Rastle (2007), Morris, Frank, Grainger, and Holcomb (2007), Morris, Grainger, and Holcomb (2008) and Royle, Drury, Bourguignon, and Steinhauer (2010) all find equivalent amplitude modulation of the early N250<sup>2</sup> response for both pseudo-complex and

transparently complex priming conditions, but not for orthographic priming conditions.

A separate line of research has used MEG to investigate the timing of morphological decomposition, and established that the left fusiform gyrus, which includes the VWFA, is highly sensitive to possible morphological form complexity between 130 and 210 ms after the onset of a visually presented word. Zweig and Pykkänen (2009) compared the processing of morphologically complex words like *teacher* and *refill*, with the processing of monomorphemic words like *stretch* and *throng*, and found a significant increase in activity peaking ~170 ms PSO for the complex words (the M170 response). Zweig and Pykkänen (2009) argue that this increased activation for morphologically complex words indexes precisely the early stage of visual word form based morphological segmentation argued for from the behavioral masked morphological priming literature.

Solomyak and Marantz (2010) and Lewis, Solomyak, and Marantz (2011) show that the magnitude of the M170 response is, in fact, highly correlated with the stem:whole-word transition probability (TP) of a potentially morphologically complex word. Words with low TPs such as *taxable* are associated with greater M170 amplitudes than words with higher TPs such as *tolerable*. Solomyak and Marantz (2010) show that while the M170 response is sensitive to morphological TP and affix frequency, it is not modulated by orthographic form features (e.g. string frequency, bigram transition probability). Lewis et al. (2011) investigate pseudo-complex items (e.g. *brother*, *sandal*), and find that pseudo-stem:whole-word TP also significantly modulated M170 activity for these words, consistent with the evidence from the masked-priming experiments. Lewis and colleagues conclude that their findings “support an obligatory decomposition model of complex word recognition whereby the VWFA decomposes all words based on their visual word forms and on the statistical correlations between forms of affixes and stems” (p. 126).

Solomyak and Marantz (2010) found a distinction between the orthographic and morphological properties of the stimulus. The orthographic properties of words, such as the positional letter frequency of the affix and the transition probability from the last letters of the stem to the first letters of the affix, were found to modulate brain activity around 130 ms, but had no effect on the later M170. This distinction builds on the finding by Tarkiainen, Helenius, Hansen, Cornelissen, and Salmelin (1999) that used MEG to identify two occipitotemporal responses evoked in the processing of visually presented word or symbol strings: an early response originating bilaterally in occipital cortex (V2-V4v) sensitive to visual noise and string length, but not to whether the string carried linguistic information (Type I response), which peaked around 100 ms PSO, and a later, letter-specific response around 130–170 ms PSO in left inferior temporal (IT) cortex, including areas identified as the VWFA (Cohen et al., 2000). This Type II/M170 response has since been found to be sensitive to a number of visual word form features. For instance, Lewis et al. (2011) report that the posterior portions of the fusiform gyrus were affected by orthographic properties of the stimulus, whereas the more anterior portions were modulated by particular morphological properties of the stimulus.

In an attempt to develop a coherent account of the temporal and the spatial resolution of these responses, as well as the lexical variables associated with each of them, Gwilliams, Lewis, and Marantz (2016) revisited the original study by Tarkiainen et al. (1999). Using a distributed source analysis instead of dipole modeling, and utilizing the directionality of activity with respect to the cortical surface, they deconstructed the Type II response into two distinct responses: an early, posterior response with negative polarity peaking around 130 ms PSO, the Type II-Noise response, and a later, more anterior response with positive polarity around 170 ms PSO, the Type II-Letter response. The orthographic properties of the stimulus were associated with the Type II-Noise response, whereas the morphological features of the items modulated the Type II-Letter response.

<sup>1</sup> Such a stage has been described by other studies as well under different names: *licensing* in Burani et al. (1999); *recombination* in Taft (2004) and Fruchter and Marantz (2015); *unification* in Hagoort (2005).

<sup>2</sup> Holcomb and Grainger (2007) propose that the amplitude of the N250 may reflect the degree of mismatch between letter and letter-cluster representations that are activated by the prime stimulus, and those representations receiving activation from the target.

### 1.2.2. Steps to recombination

Below we review the evidence for each of the recombination stages, and their spatio-temporal profiles.

**1.2.2.1. Lexeme lookup.** One of the most widely reported evoked responses found in the language processing literature is the N400/M350 component. This response is associated with the lookup of lexemes/morphemes and originates in relatively anterior portions of the temporal lobe, starting at about 200 ms PSO and peaking between 300 and 400 ms. Previous studies have shown family size and family frequency, as well as repetition, to modulate this response (Pylkkänen & Marantz, 2003; Solomyak & Marantz, 2010). Solomyak and Marantz (2010) also report a lemma frequency effect on the M350, an effect dissociating the processing of the stem from the processing of the whole word (see Fruchter & Marantz, 2015; Lau, Phillips & Poeppel, 2008 for a review of further findings associating this response with lexical access).

Recent studies include alternative measures of morphological processing, such as derivational family entropy. Entropy describes the uncertainty of choosing one element over another in a given set of candidates (Shannon, 1948). This uncertainty is translated into probability and in the language research domain this probability often relies on corpus-based frequency measurements. In derivational family entropy the set of candidates includes the different derivational forms of the stem. According to Moscoso del Prado Martín et al. (2004), the distribution of frequencies within the derivational family of a stem affects complex word recognition more than mere stem frequency. Fruchter and Marantz (2015) explored the effect of this measure on evoked neural activity, and report a highly significant positive correlation between derivational family entropy and brain activity in the left STG/MTG - areas reliably associated with lexical access (Binder et al., 1997; Friederici, 2012; Hickok & Poeppel, 2007; Indefrey & Levelt, 2004). Fruchter and Marantz (2015) observe these morphological family effects from 240 to 390 ms PSO, more than 200 ms before they detect a surface frequency effect, as predicted by the Full Decomposition model. Several other studies have also shown no effect of surface frequency on this response (King, Linzen, & Marantz, 2018; Solomyak & Marantz, 2009). Left MTG has also been implicated in morphological decomposition in several fMRI studies (Lehtonen, Vorobyev, Hugdahl, Tuokkola, & Laine, 2006; Marslen-Wilson & Tyler, 2007; Tyler, Stamatakis, Post, Randall, & Marslen-Wilson, 2005).

**1.2.2.2. Syntactic licensing.** No previous study has directly addressed the question of where and when our brains determine whether the morphemes in a complex word can be grammatically assembled, a process that would correspond to Schreuder and Baayen's (1995) grammatical *licensing*. However, a number of studies have investigated the temporal and spatial dynamics of early access to grammatical category.

For instance, Friederici's (1995, 2002) serial model of language processing argues that the syntactic category of a word is identified 100–300 ms after stimulus presentation. Support for this model comes from ERP/EEG studies (Hahne & Friederici, 1999; Hahne & Jescheniak, 2001) studying syntactic and morphological processing, which identified an early left-anterior negative ERP response (ELAN) in the inferior portion of the superior temporal gyrus (STG) peaking at about 250 ms. The ELAN response is typically triggered by either a free-standing function word or bound inflectional affix which is inconsistent with the syntactic context, and is interpreted as evidence that an initial syntactic structure has been assigned to this item. MEG studies focusing on word-category violations, as well as morphological priming, also identify an early response, providing further support for a stage whereby the word category is accessed (Dikker, Rabagliati, & Pylkkänen, 2009; Lehtonen, Monahan, & Poeppel, 2011; Monahan, Fiorentino, & Poeppel, 2008).

One of the variables used to identify this response and better understand what triggers it are various forms of entropy. For instance, Linzen, Marantz and Pylkkänen (2013) investigated the effect of

subcategorization frame entropy in the activity in the Anterior Temporal Lobe (ATL). This variable quantifies the uncertainty over the possible syntactic phrases a verb can take as its complement and was found to significantly correlate with activity in the ATL, from 200 to 300 ms. Building on the findings of this study, King et al. (2018) used Noun/Verb (N/V) entropy as a variable to investigate whether the syntactic category features of the word itself also modulate the activity in the ATL in this early time window. N/V entropy is calculated for words like *scoop*, which can be used as a noun or verb. High entropy means that the probability of using *scoop* as a verb is similar to the probability of using it as a noun. Low entropy means that the probabilities of the two entries are very different. King and colleagues report a significant correlation between N/V entropy and activity centered at about 220 ms in the ATL, and argue that it indexes a point in lexical access at which the syntactic properties of the stem are assessed.

However, not all studies on the processing of syntactic category information argue for the involvement of the ATL. Dikker et al. (2009), in a word-category violation study, did not find any evidence that the ATL was sensitive to syntactic category mismatch. Rather, they argue that posterior temporal lobe (PTL) is the area that deals with syntactic category constraints. Recent work by Flick et al. (2018) investigating nominal modification in sentence contexts also identifies the left PTL as the locus of syntactic composition.

Based on these previous findings, then, we would expect effects of syntactic licensing of stem + affix combinations to arise between 200 and 300 ms PSO and rely on left temporal lobe areas. Whether the anterior or the posterior parts of this area are involved in this stage is still unclear.

**1.2.2.3. Semantic composition.** A number of studies on semantic processing have been conducted on the sentence and phrase levels (Bemis & Pylkkänen, 2011; Pylkkänen, Bemis, & Elorrieta, 2014; Pylkkänen, Olivieri, & Smart, 2009). Pylkkänen, Martin, McElree, and Smart (2009), using a semantic violation paradigm, identified an Anterior Midline Field (AMF) response associated with semantic composition violations for morphologically complex words in sentential context, independently from world knowledge. They located this semantic response in the left ventromedial and inferior prefrontal cortex. Because these areas are otherwise often associated with decision making (see Wallis, 2007 for a review), Pylkkänen et al. (2009) explicitly tested the possibility that the apparent semantic violation effects observed in previous studies were due to decision/task effects. They find that the AMF effect associated with semantic violations and mismatches is elicited even when participants were simply reading for comprehension and had no decision task.

To see if the same response could be identified for words presented in isolation, Fruchter and Marantz (2015) investigated how semantic coherence correlates with brain activity. The semantic coherence measure they created measures the deviation of a complex word's actual frequency from the frequency that would be expected given its stem frequency and the phonological transition probability of the stem/affix combination. Their results showed a highly significant facilitatory effect of semantic coherence in the left lateral OF regions from 300 to 500 ms PSO.

While we acknowledge that fMRI studies of morphologically complex words associate semantic processing with other areas, notably the left inferior frontal gyrus (Friederici, 2012; Lehtonen et al., 2006; Vannest, Newport, Newman, & Bavelier, 2011), the goal of the current project was to see whether the morpho-semantic coherence effect of Fruchter and Marantz (2015) and the morpho-semantic illformedness effect of Pylkkänen et al. (2009) could also be found for an entirely different language, and for pseudoword stimuli, for which no stored whole word representation is available, thus clarifying whether the effect is associated with the actual composition or with comparison between a computed parse and a stored whole word.

### 1.3. Finer grained distinctions: suffixation restrictions in Greek

In order to investigate whether we can observe distinct processing profiles for licensing and composition, we turn to the restrictions on deverbal suffixation in Greek. Manouilidou and colleagues have conducted several studies investigating the processing of familiar and novel deverbal words in Greek (Manouilidou, 2006, 2007; Manouilidou & Stockall, 2014). The aim of this research was to investigate native speakers' processing of pseudowords that violated the selectional restrictions of affixes either in terms of the syntactic category (Cat) or the argument structure (ArgStr) specifications of the verbal stem. A consistent finding across all studies, which utilized the lexical decision paradigm, was that Greek speakers reliably distinguished between the two, with ArgStr violations yielding higher acceptance rates (i.e. higher error rates) and longer reaction times.

Manouilidou (2007) focused on the formation of *deverbal adjectives* using the suffixes *-simos*, *-tos* and *-menos*. The *-simos* and *-tos* derived adjectives behave like *-able* adjectives in English, while *-menos* adjectives function as adjectival passives. The syntactic category restrictions of these suffixes require them to attach to verbal bases, while the argument structure restrictions require these verbs to be transitive. Therefore, pseudowords of the type *kouvertitos* 'blanket-able' (noun + *-able*) are Cat violation items, and pseudowords of the types *gerastos* 'age-able' (intransitive verb + *-able*) are ArgStr violation items. Manouilidou (2007) found that violations of the Cat restriction were more robustly and more quickly rejected than violations of ArgStr restrictions.

Manouilidou and Stockall (2014) extended this line of research to include a comparison with English. Greek deverbal suffixes with Cat and ArgStr restrictions were compared with the English prefix *re-*, which typically requires a verbal stem, and further requires that the verbal stem be able to take a direct object. Cat violation and ArgStr violation items were constructed for each language by violating the respective set of features in each condition, and, as in the previous studies, native speakers judged whether these novel items were possible words or not. Strikingly, the exact same pattern of results was found for both languages: Cat violation items were judged unacceptable significantly more frequently, and more quickly, than ArgStr violation items. This consistent patterning across the two languages, which seems robust to the difference between prefixes and suffixes, and to other differences between the languages (orthographic system, permissiveness of word formation processes), suggests that access to grammatical category features and access to argument structure features are two distinct processes, with the former being faster and more reliable than the latter.

### 1.4. The current study

#### 1.4.1. The morphological processing experiment

The present study combines the Greek stimuli and judgment task from Manouilidou and Stockall (2014) with MEG recording, in order to directly observe the processing of affixation restrictions in the context of complex morphological processing, and determine whether the conclusions of the behavioral studies (e.g. Manouilidou, 2007; Manouilidou & Stockall, 2014) are justified. More specifically, we expect to find:

- (a) For the grammatical words:
  - a. an early effect of decomposition within the first 200 ms PSO in the left fusiform gyrus
  - b. a dissociation between early (200–400 ms) stem frequency effects and later (> 400 ms) surface frequency effects in left temporal lobe
- (b) For the pseudowords:
  - a. licensing effects (effects of the mismatch between the stem and affix in the Cat violation items) in the temporal lobe from 200 to

300 ms

- b. composition effects (effects of the mismatch between the stem and affix in the ArgStr violation items), are expected between 300 and 500 ms in the OF.

Equally importantly, since most of the literature on the neurocognition of morphological processing of complex lexical items focuses on English, running this study in Greek provides cross-linguistic evidence for complex morphological processing. Greek is a relatively understudied language and, to the best of our knowledge, no other MEG study has been conducted in Greek.

#### 1.4.2. The Tarkiainen localizer

Along with the main experiment, the present study included a second task, a functional localizer. Functional localizers are used to identify functionally defined regions of interest (fROIs) within individual subjects, and they rely on a very simple, short task that taps into the same, or similar, neural networks as the main task. The functional regions identified in the localizer data can be extracted and used for the analysis of the main task data. Anatomical regions have quite abstract borders and are relatively large; therefore, by using fROIs, we achieve a more focal analysis of the data.

Gwilliams et al. (2016) created an abridged version of the original Tarkiainen et al. (1999) task to use as a functional localizer for the early processing of visually presented words. The localizer included less than one third of the items used in the full paradigm and lasted for about 6 min (see Section 2). Unlike the original task, participants did not have to actively respond to stimuli. Gwilliams et al. (2016) ran an English adaptation of both the full paradigm and the abridged localizer, and compared the findings from the two tasks against each other. The results showed that both tasks evoked similar responses, suggesting that the abridged paradigm can be used as a functional localizer to identify the brain regions associated with the M100, the M130 and the M170 responses, also referred to as Type I, Type II-Noise, and Type II-Letter responses respectively.

For the current study we ran a Greek adaptation of this abridged task as the functional localizer. The purpose of the localizer in this study was to facilitate the analysis of the MEG data of the main experiment by having smaller and well-targeted brain regions to analyze for the effects of early, obligatory, form-based decomposition.

## 2. Material & methods

### 2.1. Design and stimuli

#### 2.1.1. The Tarkiainen localizer

The stimuli consisted of 6 different conditions, varying in length and Gaussian noise level. In terms of length, the stimuli were divided into 1-element items and 4-element items. The 1-element items included 4 one-symbol items presented only at the lowest Gaussian noise level (level 1), and 23 one-letter items presented both at the lowest Gaussian noise level and the highest Gaussian noise level (level 24). The one-symbol items included a triangle, a square, a diamond and a circle, and the one-letter items consisted of 23 of the 24 letters of the Greek alphabet (O 'omikron', was excluded due to similarity with the circle). The 4-element items included 4 four-symbol items presented only at the lowest Gaussian noise level, and 50 four-letter Greek words, all of them presented both at noise level 1 and noise level 24. The four symbols used were again triangle, square, diamond and circle, and were presented in four different permutations. All words were frequent Standard Modern Greek singular nouns in nominative case. In total we had 154 items, but since they were not equally divided into the six conditions, randomly selected items from each condition were repeated as many times as needed to get fifty items for every one of the 6 conditions, resulting in 300 trials.

The items were fully randomized and equally divided into 6 blocks,

**Table 1**  
English transliterations of example stimuli, and equivalent examples in English with prefix *re-*. Materials (without hyphens) were presented in Greek orthography. English examples are presented for exposition purposes only.

Suffix	Grammatical	Ungrammatical	
		Cat Violations	ArgStr Violations
<i>-simos</i>	ananeo-simos (renewable)	varel-imos (barrel-able)	gela-simos (laugh-able)
<i>-tos</i>	sevas-tos (respectable)	lahano-tos (cabbage-able)	tremi-tos (tremble-able)
<i>-menos</i>	lavo-menos (wounded)	ahino-menos (urchin-able)	argi-menos (be late-able)
<i>Equivalent examples in English with re-</i>	re-form (re- requires a transitive verb)	re-knife (re- rejects a noun stem)	re-smile (re- rejects an intransitive, agentive stem)

each including a mixture of items from the different conditions. Each item was presented within a rectangular patch ( $\sim 5 \times 2$  in.) in the center of the screen for 60 ms with a 2 s inter-stimulus interval. Participants were instructed to focus on the pictures and avoid blinking to minimize MEG data artifacts. No active response to the stimulus was required.

### 2.1.2. The morphological processing experiment

The test items consisted of pseudowords with Cat or ArgStr violations and grammatical words, with one of the following three suffixes: *-tos*, *-simos*, and *-menos*, following Manouilidou (2007) and Manouilidou and Stockall (2014) who report clear differences between Cat violations and ArgStr violations formed with these specific suffixes.

Cat violation items were created by attaching the suffixes to noun stems instead of verb stems. For ArgStr violation items the suffixes were attached to intransitive verbs. Examples for each of the two violation types of stimuli, as well as of grammatical items are provided in Table 1.

Candidate items were drawn from the Greek experiment in Manouilidou and Stockall (2014) (see Manouilidou and Stockall (2014) for selection details), and supplemented with additional items, created using length and stem frequency as the main criteria. This resulted in an initial set of 346 items with 30–35 items in each of the 9 conditions. These items were then normed in order to exclude violation items that were judged as acceptable words, and grammatical words that were not reliably judged as acceptable. We used the Ixex Farm environment to administer an online task in which participants rated each item on a scale from 1 to 7, (1 = unacceptable). Eighty-three native Greek speakers completed the task online. Participants with a country of origin other than Greece were excluded from the analysis. All violation items with an average rating of  $\geq 5$  and all grammatical items with an average rating of  $\leq 4$  were excluded from the final set of stimuli. Nine items were excluded.

From the remaining 337 items, we selected the maximum number of items that would allow us to have an equal number of items in each of the nine conditions controlling for length and stem frequency, resulting in 25 items for each condition. The total number of test items was 150 ungrammatical items (25 items  $\times$  2 violation types  $\times$  3 suffixes) and 75 grammatical items (25 items  $\times$  3 suffixes). To have an equal number of grammatical and ungrammatical items, we added 75 grammatical Greek adjectives with suffixes different than the ones under investigation: *-ikos*, *-tikos*, *-inos*, and *-eros*.<sup>3</sup> The filler items were matched for length and stem frequency with the test items. In total, we had 300 items.

<sup>3</sup> Most of these items were taken from Manouilidou (2006). Items formed from the same lemma as a test item were removed and new items were added.

## 2.2. Methods

### 2.2.1. Participants

Twenty Greek native speakers (mean age = 32.3, 13 female) were recruited from the Greek community in Abu Dhabi to participate in this study. All participants were right-handed, with normal or corrected-to-normal vision. Two of the subjects were eliminated from the analysis due to excessive head movement during the test sessions.

### 2.2.2. Procedure

Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, [www.neurobs.com](http://www.neurobs.com)) was used as the presentation platform for both tasks. The stimuli were projected onto a screen that was located approximately 85 cm away from the participant.

The main task of the experiment was a visual lexical decision task, with simultaneous MEG data recording. Each trial began with a fixation cross (“+”) for 300 ms, followed by a single item that stayed on the screen until the participants gave a response, or for a maximum of 2 s. Participants were instructed to indicate whether the item was a real word in Greek by pressing one of two buttons with their left hand. In order to familiarize the participants with the task, we included a short practice with 8 items at the beginning of the session, none of which was included among the test items. For the practice, if the participant gave a wrong answer a red cross appeared. During the actual test session, no feedback was provided. The items were fully randomized and each participant received a unique randomization. The experiment consisted of three blocks and lasted around 20 min.

A 208-channel axial gradiometer whole-head MEG system (Kanazawa Institute of Technology, Kanazawa, Japan) was used to record the data continuously, at a sampling frequency of 1000 Hz. The data were filtered during acquisition between 0.1 Hz and 200 Hz. The head of every subject was digitized prior to entering the magnetically shielded room using a hand-held FastSCAN laser scanner (Polhemus, VT, USA). The head position during the experiment was determined using coils attached to predefined anatomical regions. The head scan and the coil measurements were then used for the co-registration process.

## 2.3. Data analysis

### 2.3.1. Behavioral data

Reaction Times (RTs) and yes/no responses were recorded for every trial. Coefficients were estimated with linear mixed-effects models (Baayen, Davidson, & Bates, 2008), using the lmer function of the lme4 package in R (Bates & Maechler, 2009) for the RT data, and the glmer function of the lme4 package in R to run logistic regressions for the acceptance rate data. The fixed effects were Condition Type and Suffix Type. By-subject slopes for all model terms, as well as by-subject and by-item intercepts were also fitted as random effects. Model comparisons were then performed by extracting one fixed term at a time.

### 2.3.2. MEG data processing

All MEG data, from both the localizer and the lexical decision task, were preprocessed similarly. First, the raw data were noise reduced using three gradiometer reference channels located away from the participant’s head, utilizing the Continuously Adjusted Least Squares Method (CALM; Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 2001) in the MEG160 software (Yokohawa Electric Corporation and Eagle Technology Corporation, Tokyo, Japan).

The noise-reduced data were further preprocessed and analyzed with MNE-Python (Gramfort et al., 2013, 2014) and Eelbrain (<https://pythonhosted.org/eelbrain>). Independent Component Analysis (ICA) was conducted to remove components related to specific noise patterns, while additional artifact rejection was performed through manual inspection of the data. The data were low-pass filtered at 40 Hz and epoched from  $-200$  ms to 600 ms, relative to the beginning of the

stimulus. Structural MRIs were reconstructed by scaling and orienting the Freesurfer average brain (CorTech Labs, La Jolla, CA and MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging, Charleston, MA) to each participant's head shape. A source space consisting of 5124 vertices, equally split into the two hemispheres, was generated on each reconstructed surface. Once we baseline corrected the data using a  $-200$  ms to 0 ms pre-stimulus interval, the BEM (boundary-element model) method was utilized on the activity at each of the vertices to compute the forward solution. Using the forward solution and the grand average of the data for all trials within each subject, the inverse solution was calculated.

Following Gwilliams et al. (2016), the inverse solution was calculated using signed fixed orientation for the source estimates. This means that the direction of the current normal to the cortex was defined, and the dipoles were projected perpendicular to the cortical surface, estimating activity from the magnitude of the current dipole normal to the cortex. The signed normed estimates were transformed into noise-normalized dynamic statistical parameter maps (dSPMs: Dale et al., 2000), using an SNR value of 3 for ANOVA analyses and a value of 2 for the regression analyses. Although polarity in MEG data is still far from being fully understood, opposite polarity activities reflect discrete response components (for further discussion, see Gwilliams et al., 2016).

**2.3.2.1. The Tarkiainen localizer.** The localizer was used to identify the Type I and Type II-Noise and -Letter responses. The statistical analysis used a combination of regression analysis and spatio-temporal permutation cluster tests. All procedures related to the cluster-based permutation tests were based on the method of Maris and Oostenveld (2007). The exact steps for this analysis are detailed in Gwilliams et al. (2016). For the purpose of the present experiment, only the Type II-Letter response was relevant.

**2.3.2.2. The morphological processing experiment.** The statistical analysis of the data from the lexical decision task varied depending on the question each analysis was aiming to answer, as well as on the region of interest. The spatio-temporal tests were run on anatomical ROIs, while the mixed effects models were run on fROIs. The distinction between the two analyses lies in the ROIs' size difference: anatomical ROIs are larger than fROIs, permitting a statistically valid search for spatial clusters of activity within the ROI, which the small fROIs do not.

To assess which linguistic variables correlate with brain activity in specific brain areas and time-windows we either ran a spatio-temporal analysis combining a regression test with permutation cluster tests, or a linear mixed-effects model analysis utilizing the lmer function of the lme4 package in R. The linear mixed-effects models included the neural activity averaged across both time and space as the dependent variable. Coefficients were estimated with fixed effects of length of the item, number of syllables, log stem frequency, log surface frequency, bigram frequency and stem:whole word TP. By-subject random slopes were fitted for all the lexical variables, as well as by-subject and by-item random intercepts. In order to test for an effect of Violation Type or Suffix Type, we ran 2 (Violation Type)  $\times$  3 (Suffix Type) AVOVAs, which were combined with permutation cluster tests. For all tests utilizing permutation cluster tests, the p-values were corrected for multiple comparisons over time and space as described by Maris and Oostenveld (2007).

### 3. Results

#### 3.1. Behavioral results

As shown in Table 2, the ArgStr violations yield the longest RTs ( $\bar{x}$  = 1227.86 ms), followed by the Cat violations ( $\bar{x}$  = 1146.26 ms), and then by the grammatical items ( $\bar{x}$  = 966.36 ms).

The fixed effects of the initial mixed effect model were Condition, Suffix and the interaction of Suffix with Condition. Condition was coded

**Table 2**

Mean RTs (ms)  $\times$  condition, within each suffix and across suffixes.

Condition	-simos	-tos	-menos	across suffixes
ArgStr Viol.	1202.42	1226.04	1254.62	1227.86
Cat Viol.	1131.84	1117.62	1188.83	1146.26
Gram.	1020.66	943.4	988.22	966.36

into two orthogonal contrasts: Grammaticality, which compared grammatical items to the average of all the ungrammatical items, and Violation Type, which compared ArgStr violations to Cat violations. Neither removing the interaction term of suffix, or removing suffix completely significantly reduced model fit. This final model with Grammaticality ( $\beta$  =  $-1485.7$ , SE = 166.2,  $t$  =  $-8.94$ ) and Violation Type ( $\beta$  = 960.7, SE = 228.6,  $t$  = 4.2) as fixed variables was compared against the models containing only one of these two variables at a time. Statistically significant differences were found for both Grammaticality ( $\chi^2(11)$  = 197.04,  $p$  < .001) and Violation Type ( $\chi^2(11)$  = 36.97,  $p$  < .001).

In a *post-hoc* analysis of the RTs, we included the ratings collected in the acceptability judgment norming task we first ran to select materials for the study. The purpose of this analysis was to investigate to what extent the effect of Violation Type is still present once the variance due to these offline ratings is accounted for. If the ratings variance explains the Violation Type variance, it would suggest that the RTs reflect the graded judgements of wellformedness.

The fixed effects of the mixed effects model for this analysis were Violation Type, Rating and the interaction of those terms with Suffix Type. Grammaticality was not included as this analysis only included the pseudoword items. Following model comparison as described above, no statistically significant effect of suffix was found. When the model containing only Violation Type ( $\beta$  = 624.6, SE = 261.8,  $t$  = 2.39) and Rating ( $\beta$  = 452.4, SE = 160.9,  $t$  = 2.81) was compared against the models containing only one of these two variables at a time, statistically significant differences were found both for Violation Type ( $\chi^2(11)$  = 23.59,  $p$  < .001) and Rating ( $\chi^2(11)$  = 21.19,  $p$  < .001), suggesting that the two effects are independent from each other, and reflecting an additional effect of wellformedness in the observed RTs.

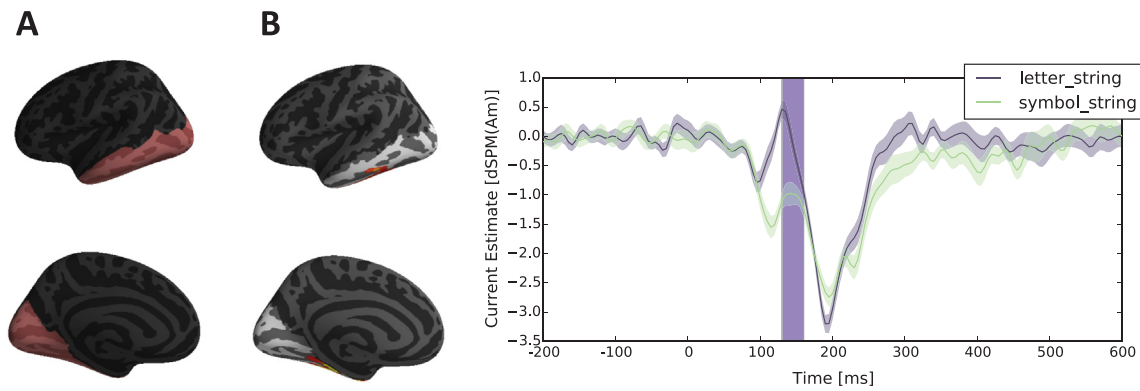
Acceptance rates were calculated both for the two violation conditions and the grammatical items for each of the three suffixes separately (Table 3). The average acceptance rate across the different suffixes was 33.3% for the ArgStr violation items, 10.67% for the Cat violation items, and 92% for the grammatical items.

As in the RT analysis, the fixed effects of the initial mixed effect model were Condition, coded into the orthogonal contrasts of Grammaticality ( $\beta$  =  $-0.96$ , SE = 0.28,  $z$  =  $-3.36$ ) and Violation Type ( $\beta$  = 1.97, SE = 0.25,  $z$  = 7.73), Suffix Type (with *-menos* as baseline: for *-simos*:  $\beta$  =  $-0.64$ , SE = 0.26,  $z$  =  $-2.41$ ; for *-tos*:  $\beta$  =  $-0.17$ , SE = 0.26,  $z$  =  $-0.67$ ) and the interaction of Suffix Type with Grammaticality (with *-menos* as baseline: for *-simos*:  $\beta$  =  $-0.11$ , SE = 0.61,  $z$  =  $-0.19$ ; for *-tos*:  $\beta$  = 2.21, SE = 0.67,  $z$  = 3.30) and with Violation (with *-menos* as baseline: for *-simos*:  $\beta$  =  $-1.19$ , SE = 0.44,  $z$  =  $-2.70$ ; for *-tos*:  $\beta$  = 1.05, SE = 0.39,  $z$  = 2.63). Removing the interaction term yielded a significant difference, suggesting a significant interaction of Suffix Type with Grammaticality and Violation type. After correcting for multiple comparisons, only the difference between ArgStr violations and Cat violations for *-tos*

**Table 3**

Mean Acceptance Rates  $\times$  condition, within each suffix and across suffixes.

Condition	-simos	-tos	-menos	across suffixes
ArgStr Viol.	38%	30%	33%	33.67%
Cat Viol.	6%	16%	10%	10.67%
Gram.	90%	91%	95%	92.00%



**Fig. 1.** The Tarkiainen localizer analysis. (A) Location of spatial vertices used in the analysis, as defined by Gwilliams et al. (2016). (B) Location of the most significant cluster of the Type II – Letter response and its average activation.

remained as a significant contrast of interaction ( $p < .001$ ). Extracting Suffix Type, Grammaticality and Violation Type one at a time also yielded significant differences, suggesting statistical main effects for all three predictors (Suffix Type:  $\chi^2(55) = 85.71$ ,  $p < .001$ ; Grammaticality  $\chi^2(55) = 232.53$ ,  $p < .001$ ; Violation Type:  $\chi^2(55) = 114.75$ ,  $p < .001$ ).

### 3.2. MEG results

#### 3.2.1. The Tarkiainen localizer

The Type II–Letter response of the localizer was characterized by a positive correlation between the brain activity in the left anterior occipital-temporal region and the String Type. Our analysis revealed two significant clusters of positive polarity, one from 130 to 160 ms ( $p = .007$ , number of vertices = 55) and one from 130 to 180 ms ( $p = .044$ , number of vertices = 24). The most significant cluster was chosen for the fROI analyses (Fig. 1).

Type I and Type II–Noise responses were also identified and found to have spatial and temporal profiles consistent with those identified in Gwilliams et al. (2016) for English, thus confirming the general reliability of the localizer.

#### 3.2.2. The morphological processing experiment

In the very early window from 100 to 200 ms, during which the early, obligatory decomposition process is expected to take place, we used linear mixed-effects models to investigate the effect of a number of lexical variables: item length, number of syllables, log stem frequency, log surface frequency, bigram frequency and stem:whole word TP. The left fusiform gyrus was used as the anatomical region of interest, and following Gwilliams et al. (2016), the most significant of the clusters we identified in the Type II–Letter analysis of the Tarkiainen localizer was used as an fROI to more accurately identify the M170 response.

The activity in left fusiform gyrus was positively correlated with the TP from stem to whole word ( $p = .006$ ) for the grammatical test items. The activity in the localizer-based fROI was modulated both by stem:whole word TP ( $p = .046$ ) for the grammatical test items, and the length ( $p = .037$ ) for all the test items. None of the other lexical variables showed a significant effect in the activity of either of the two brain areas. Comparisons of these models with models containing the suffix type as a fixed variable did not show a significant difference, suggesting no significant differences between suffixes.

Examination of the 200–300 ms time-window using a spatio-temporal  $2 \times 3$  ANOVA in the temporal lobe<sup>4</sup> (Fig. 2A), where the effects of syntactic licensing are expected to manifest, revealed an effect of Violation Type. As shown in Fig. 2B, Cat violations evoked more activity

than the ArgStr violation items for all the three suffixes, an effect that persisted across the entire time-window we analyzed. The analysis revealed a significant cluster ( $p = .015$ ) spreading across the ventral and posterior portions of the region we analyzed (Fig. 2C).

From 300 to 500 ms, we performed another spatio-temporal  $2 \times 3$  ANOVA between Suffix Type and Violation Type in the OF (Fig. 3A) to investigate the semantic composition processing stage. The results revealed a significant effect of Violation Type ( $p = .024$ ) from 425 to 500 ms (Fig. 3B–C), with ArgStr violations evoking greater activity than Cat violations. Suffix Type also had a significant effect ( $p < .001$ ) on evoked OF activity throughout the entire time-window (Fig. 3D–E). Both *-tos* and *-simos* items showed positive activity, with *-tos* evoking more positive activity than *-simos*, whereas *-menos* items evoked more negative activity. There was no interaction between Violation Type and Suffix Type.

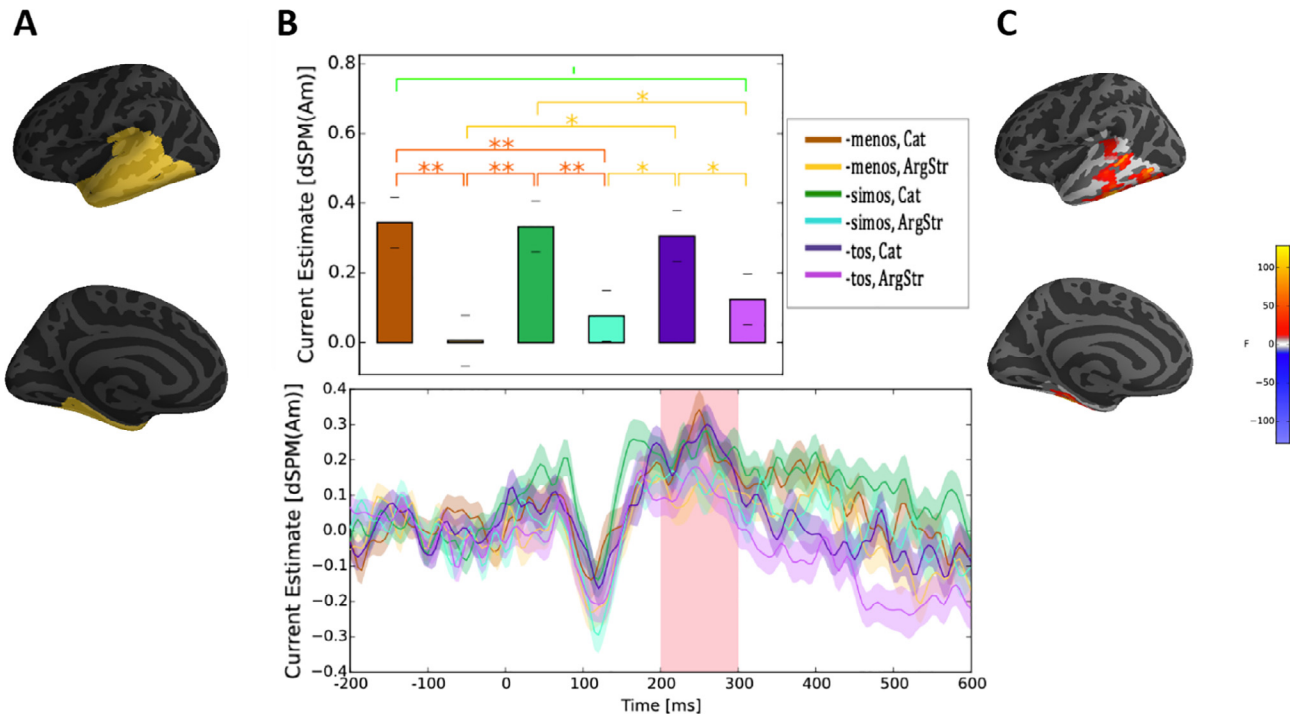
From 200 to 500 ms we looked for effects related to lexeme lookup in the temporal lobe, by investigating the effect of stem frequency in the temporal lobe for all the grammatical items. The spatio-temporal regression revealed a significant cluster in more ventral parts of the temporal lobe ( $p < .001$ ). The cluster extended in the entirety of the time window we analyzed, suggesting a sustained effect of stem frequency in this region. Based on the predictions stemming from the Full Decomposition model, we also looked for effects of surface frequency. A spatio-temporal regression on surface frequency in the temporal lobe from 350 to 500 ms revealed a cluster from 400 to 465 ms, which however failed to reach significance after correcting for multiple comparisons. No effects of surface frequency were found in earlier time-windows.

## 4. Discussion

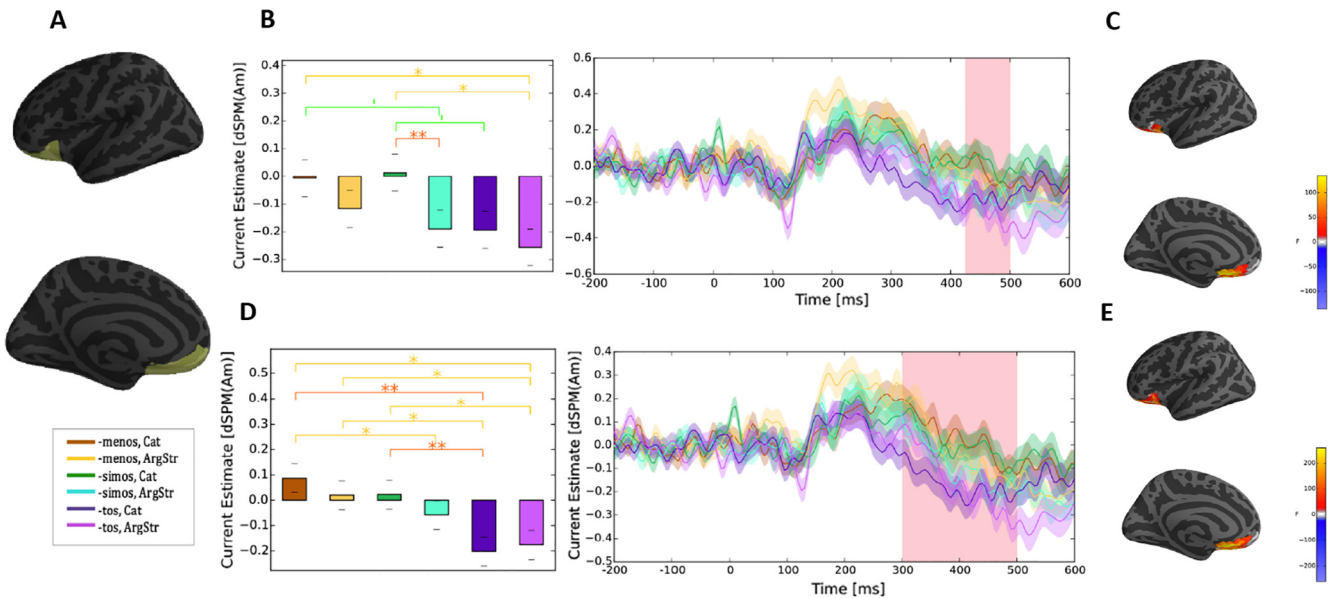
The main goal of this experiment was to test the Full Decomposition model by investigating both decomposition and recomposition of the same items as well as outlining the neurobiological profile of the different responses associated with complex morphological processing. Within a single experiment, we tried to find evidence for early, form-based decomposition, as well as to test both the prediction that stem frequency effects should precede whole word frequency effects and the prediction that licensing and composition are distinct processing stages. The analysis of both the behavioral and the MEG data provide support for this model, corroborating the idea that the constituent morphemes of morphologically complex items are processed separately.

The behavioral results of the lexical decision task replicated previous findings by Manouilidou and Stockall (2014). The processing of ArgStr violations seems to be more challenging than that of Cat violations. Participants took more time to decide on the grammaticality of a pseudoword like ‘gelasimos’ (*laugh-able*) and they also accepted such words as grammatical more often than pseudowords like ‘varelimos’

<sup>4</sup> The temporal lobe region used in this analysis consisted of the BA areas 20, 21, 22, 37, 38, 41 and 42..



**Fig. 2.** Comparison of brain activity between the two violation types for each of the three suffixes in the temporal lobe, from 200 to 300 ms. (A) Location of spatial vertices used in the analysis. (B) Averaged activity for each condition as an effect of Violation Type in the statistically significant spatio-temporal cluster and plot of the average brain activation for each condition in the entire epoch, from 0 to 600 ms. (C) Location of the spatio-temporal cluster, with every source that was part of the cluster at any point in time being color-coded with the sum *F* statistic.



**Fig. 3.** Comparison of brain activity between the two violation types for each of the three suffixes in the orbitofrontal lobe, from 300 to 500 ms. (A) Location of spatial vertices used in the analysis. (B) Averaged activity for each condition as an effect of Violation Type in the statistically significant spatio-temporal cluster and plot of the average brain activation for each condition in the entire epoch, from 0 to 600 ms. (C) Location of the Violation Type-cluster. (D) Averaged activity for each condition as an effect of Suffix Type in the statistically significant spatio-temporal cluster and plot of the average brain activation for each condition in the entire epoch, from 0 to 600 ms. (E) Location of the Suffix Type-cluster, with every source that was part of the cluster at any point in time being color-coded with the sum *F* statistic.

(*barrel-able*). The *post-hoc* analyses on the RTs of the two violation conditions, which found that the offline graded acceptability ratings accounted for significant RT variability, confirmed that RT was sensitive to whole word wellformedness.

The MEG results suggest a coherent account of visual word processing, comprising four major steps, which have not been summarized

together within a single experiment before. Our results favor the Full Decomposition hypothesis (Taft & Forster, 1975), wherein the processing of a morphologically complex item starts with its decomposition into stem and affix. Transition Probability from stem to whole word showed a significant effect 100–200 ms after the onset of the stimulus in the left fusiform gyrus, an effect not found for surface frequency.



Similar results have been previously reported for English (Gwilliams et al., 2016; Lewis et al., 2011). This discrepancy between the two lexical variables supports the hypothesis that complex words are parsed into their constituent morphemes before they are parsed as whole words.

Theories supporting the decomposition of morphologically complex words into stem and affix also argue for a stage at which the stem is looked up (Taft & Forster, 1975; Taft, 2004). This stage of lexeme lookup has previously been associated with the M350 response, originating in left MTG/STG. In the current data, stem frequency showed a sustained effect from 200 to 500 ms PSO in the temporal lobe for the grammatical items, interpreted as reflecting this stage. By contrast, surface frequency had a marginal effect starting 200 ms later than the stem frequency effect. Schreuder & Baayen (1995), in articulating the multi-stage model we explicitly tested here, are clear: “we assume that frequency effects are autonomous, in the sense that any word leaves a memory trace, irrespective of its morphological properties. Thus, we do not exclude that, due to repeated exposure, fully regular complex words (including inflections) may develop their own lexical representations” (p. 133). What is clear from many previous studies, and confirmed by the current study, is that frequency effects for morphological constituents are robust and early, while frequency effects for whole words occur later, if they are found at all.

Studies on syntactic and morphological processing suggest that an initial syntactic structure is assigned at ~200 ms after the item has been presented. King et al. (2018) found that brain activity in the ATL was modulated by N/V entropy around 200 ms PSO, and interpreted this effect as evidence for a step during which the syntactic category of the stem is identified. In our study, the syntactic category of the stem in ArgStr violation items matches the target category (verb), whereas in Cat violation items it does not. Based on the previously reported neurocognitive findings, we predicted a difference in temporal lobe activity between the two conditions, which would index the *licensing* stage, proposed to be based on the “appropriateness of morpheme combinations” (Schreuder & Baayen, 1995, p. 149). The current results confirm this prediction: increased activity evoked by Cat violation items, as compared to ArgStr violation items, was found in the posterior and ventral portions of the temporal lobe, from 200 to 300 ms (consistent with Dikker et al., 2009; Flick et al., 2018).

The last step outlined in Schreuder and Baayen (1995) is *composition*. Pykkänen, Olivieri, and Smart (2009), who investigated the effects associated with the violation of semantic constraints, identified a response peaking at ~300 ms in the prefrontal cortex. In the current experiment, a comparison between the two types of violations from 300 to 500 ms shows increased activity for ArgStr violation items compared to Cat violation items in the OF. This pattern, which is opposite to the one found in the temporal lobe in the earlier time-window, further supports the argument by Manouilidou and Stockall (2014) that the processing of Cat features temporally precedes the processing of ArgStr features. If we look at the broader picture, these results corroborate the idea that syntactic *licensing* and semantic *composition* occur at two distinct stages, with the former preceding the latter. Equally importantly, the current findings support the idea that distinct processing stages are associated with distinct spatio-temporal profiles. Comprehension studies on various linguistic levels, from single word to sentence level, have extensively characterized the OF as a *semantic* area of the brain (Fruchter & Marantz, 2015; Pykkänen, Olivieri & Smart, 2009; Pykkänen, Bemis & Elorrieta, 2014). The current findings, which are based on processing responses to pseudowords, and thus cannot possibly be due to any effect of whole word storage or access, provide additional support for a critical role of the OF in semantic composition. Further research focused on this area, with a wider range of languages, exemplifying a wider range of different morphological structure building restrictions is thus motivated.

Finally, as part of the analysis of the data for the current experiment, we included a localizer task to identify a better-defined region of

interest in the fusiform gyrus, where morphological decomposition occurs in the very early window, from 100 to 200 ms. Importantly, the analysis of the localizer data replicated previous findings by Gwilliams et al. (2016). Both the temporal and the spatial features of the clusters we identified were very similar to those of the clusters found for English. The identification of comparable brain responses in two distinct languages, with two distinct orthographies, cross-validates the involvement of the specific brain areas during the early stages of visual word processing, that is the role of the VWFA in early morphological decomposition. Of equal importance is the fact that these results confirmed the efficacy of using this test as a localizer across languages.

## 5. Conclusion

In summary, the present study provides evidence supporting the Full Decomposition model for the visual processing of morphologically complex items in Greek, while outlining the spatio-temporal profile of the neurocognitive responses associated with it. Unlike any of the previous studies conducted so far, this is the first attempt to provide support for all the steps involved both during the decomposition as well as the recomposition of such items. Evidence for the early, form-based decomposition stage was based on the early effects of TP in the fusiform gyrus. Recomposition, which involves three distinct stages, was supported by a number of other results. The effects of stem frequency in the temporal lobe, which preceded the effects of surface frequency, supported *lexeme lookup*. The increased response to Cat violations compared to ArgStr violations in the temporal lobe is suggested to show an effect of syntactic *licensing*, whereas the reverse effect found later in the OF is argued to be an effect of semantic *composition*. Since the Cat and ArgStr violation responses were evoked by pseudowords, formed by differentially violating the restrictions on deverbal suffixation in Greek, these results clearly argue for models of complex word processing which do not depend on access to stored, whole word representations.

Importantly, this is the first attempt to understand the online, neurocognitive processing of morphologically complex items in Greek. The fact that the current results pattern consistently with the findings of previous studies in languages like English and Dutch provides support for the crosslinguistic validity of the Full Decomposition model. However, further investigation is needed to better understand the neurocognitive profile of the distinct steps involved in complex morphological processing and the idiosyncratic differences that might exist between various suffixes.

## Statement of significance

Our results uncover the neural bases of morphological decomposition and composition in visual word recognition. Connecting to previous neurobiological research, we confirm the role of the visual word form area in early morphological processing. We relate the syntactic licensing of stem-suffix combinations to the ventral visual processing stream, somewhat independent from lexical access for the stem. We further elucidate the role of orbito-frontal regions in semantic composition. Overall, we connect the cognitive “full decomposition” theory of complex word recognition to brain regions already implicated in language processing, allowing for a tighter connection between neurobiological and cognitive neuro-scientific studies of language.

## Acknowledgments

We would like to thank Laura Gwilliams and Stephen Politzer-Ahles for their invaluable assistance with the data analysis.

## Funding

This work was supported by the NYU Abu Dhabi Research Institute, under Grant G1001.

## References

- Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & 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*, 669–672. <http://dx.doi.org/10.1109/77.919433>.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.
- Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using Eigen and R package version 0.999375-32 [Software]. Available from < <http://CRAN.R-project.org/package=lme4> > .
- Bemis, D. K., & Pyllkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience*, *31*(8), 2801–2814.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience*, *17*(1), 353–362.
- Burani, C., Dovetto, F. M., Spuntarelli, A., & Thornton, A. M. (1999). Morphological access and naming: The semantic interpretability of new root–suffix combinations. *Brain and Language*, *68*(1–2), 333–339.
- Butterworth, B. (Ed.). (1983). *Language production*. New York: Academic Press.
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area. *Brain*, *123*(2), 291–307.
- Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, *26*(1), 55–67.
- Dikker, S., Rabagliati, H., & Pyllkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition*, *110*(3), 293–321.
- Flick, G., Oseki, Y., Kaczmarek, A., Al Kaabi, M., Marantz, A., & Pyllkkänen, L. (2018). Building words and phrases in the left temporal lobe (submitted for publication).
- Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*(3), 259–281.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.
- Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*(5), 262–268.
- 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.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–460.
- Gwilliams, L., Lewis, G. A., & Marantz, A. (2016). Functional characterisation of letter-specific responses in time, space and current polarity using magnetoencephalography. *NeuroImage*, *132*, 320–333.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423.
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205.
- Hahne, A., & Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research*, *11*(2), 199–212.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.
- Holcomb, P. J., & Grainger, J. (2007). Exploring the temporal dynamics of visual word recognition in the masked repetition priming paradigm using event-related potentials. *Brain Research*, *1180*, 39–58.
- Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101–144.
- King, J., Linzen, T., & Marantz, A. (2018). Syntactic Categories as Lexical Features or Syntactic Heads: An MEG Approach. *Linguistic Inquiry* (in press).
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933.
- 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.
- Lehtonen, M., Monahan, P. J., & Poeppel, D. (2011). Evidence for early morphological decomposition: Combining masked priming with magnetoencephalography. *Journal of Cognitive Neuroscience*, *23*(11), 3366–3379.
- Lehtonen, M., Vorobyev, V. A., Hugdahl, K., Tuokkola, T., & Laine, M. (2006). Neural correlates of morphological decomposition in a morphologically rich language: An fMRI study. *Brain and Language*, *98*(2), 182–193.
- Lewis, G., Solomyak, O., & Marantz, A. (2011). The neural basis of obligatory decomposition of suffixed words. *Brain and Language*, *118*(3), 118–127.
- Linzen, T., Marantz, A., & Pyllkkänen, L. (2013). Syntactic context effects in visual word recognition: An MEG study. *The Mental Lexicon*, *8*(2), 117–139.
- Manouilidou, C., & Stockall, L. (2014). Teasing apart syntactic category vs. argument structure information in deverbal word formation: A comparative psycholinguistic study. *Italian Journal of Linguistics*, *26*(2), 71–98.
- Manouilidou, C. (2006). On the processing of thematic features in deverbal nominals (Doctoral dissertation, University of Ottawa).
- Manouilidou, C. (2007). Thematic constraints in deverbal word formation: Psycholinguistic evidence from pseudo-words. In Proceedings of the 7th International Conference on Greek Linguistics.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164*(1), 177–190.
- Marslen-Wilson, W. D., & Tyler, L. K. (2007). Morphology, language and the brain: The decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 823–836.
- Monahan, P. J., Fiorentino, R., & Poeppel, D. (2008). Masked repetition priming using magnetoencephalography. *Brain and Language*, *106*(1), 65–71.
- 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.
- Moscato del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*(1), 1–18.
- Pyllkkänen, L., Bemis, D. K., & Elorrieta, E. B. (2014). Building phrases in language production: An MEG study of simple composition. *Cognition*, *133*(2), 371–384.
- Pyllkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, *7*(5), 187–189.
- Pyllkkänen, L., Martin, A. E., McElree, B., & Smart, A. (2009). The anterior midline field: Coercion or decision making? *Brain and Language*, *108*(3), 184–190.
- Pyllkkänen, L., Oliveri, B., & Smart, A. J. (2009). Semantics vs. world knowledge in prefrontal cortex. *Language and Cognitive Processes*, *24*(9), 1313–1334.
- Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7–8), 942–971.
- Royle, P., Drury, J. E., Bourguignon, N., & Steinhauer, K. (2010). Morphology and word recognition: an ERP approach. In Proceedings of the 2010 annual conference of the Canadian Linguistic Association. Concordia University: Montreal.
- Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing*. Hillsdale, NJ: Lawrence Erlbaum.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379.
- Solomyak, O., & Marantz, A. (2009). Lexical access in early stages of visual word processing: A single-trial correlational MEG study of heteronym recognition. *Brain and Language*, *108*(3), 191–196.
- Solomyak, O., & Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, *22*(9), 2042–2057.
- Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon*, *1*(1), 85–123.
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A*, *57*(4), 745–765.
- Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 638–647.
- Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, *122*(11), 2119–2132.
- Tyler, L. K., Stamatakis, E. A., Post, B., Randall, B., & Marslen-Wilson, W. (2005). Temporal and frontal systems in speech comprehension: An fMRI study of past tense processing. *Neuropsychologia*, *43*(13), 1963–1974.
- Vannest, J., Newport, E. L., Newman, A. J., & Bavelier, D. (2011). Interplay between morphology and frequency in lexical access: The case of the base frequency effect. *Brain Research*, *1373*, 144–159.
- Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* *30*, 31–56.
- Zweig, E., & Pyllkkänen, L. (2009). A visual M170 effect of morphological complexity. *Language and Cognitive Processes*, *24*(3), 412–439.