Brain & Language 147 (2015) 1–13

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Non-linear processing of a linear speech stream: The influence of morphological structure on the recognition of spoken Arabic words

L. Gwilliams [a,*], A. Marantz [a,b,c,1]

[a] NYUAD Institute, New York University Abu Dhabi, United Arab Emirates
[b] Department of Psychology, New York University, United States
[c] Department of Linguistics, New York University, United States

ARTICLE INFO

Article history:
Received 2 December 2014
Accepted 20 April 2015

Keywords:
MEG
Morphology
Spoken word recognition
Roots
Decomposition
Prediction
Surprisal
Obligatory Contour Principle
Arabic

ABSTRACT

Although the significance of morphological structure is established in visual word processing, its role in auditory processing remains unclear. Using magnetoencephalography we probe the significance of the root morpheme for spoken Arabic words with two experimental manipulations. First we compare a model of auditory processing that calculates probable lexical outcomes based on whole-word competitors, versus a model that only considers the root as relevant to lexical identification. Second, we assess violations to the root-specific Obligatory Contour Principle (OCP), which disallows root-initial consonant gemination. Our results show root prediction to significantly correlate with neural activity in superior temporal regions, independent of predictions based on whole-word competitors. Furthermore, words that violated the OCP constraint were significantly easier to dismiss as valid words than probability-matched counterparts. The findings suggest that lexical auditory processing is dependent upon morphological structure, and that the root forms a principal unit through which spoken words are recognised.

1. Introduction

1.1. Routes to word recognition

In modelling the structure of the mental lexicon, one of the most prevalent questions is the role that morphology plays in the organisation, production and comprehension of words. Historically the debate has been between "decompositional" and "whole word" theories of word recognition, with evidence over the past decade supporting a morphologically sensitive, decompositional approach in the visual modality of lexical processing. Behavioural masked priming studies for example, which have somewhat dominated the field of enquiry, have found consistent evidence for the decomposition of words with regular suffixation and pseudo-suffixation (e.g., teacher-TEACH; corner-CORN; Rastle, Davis, & New, 2004; see Rastle & Davis, 2008 for a review). Corresponding results have also been established in the neurophysiological literature, supporting decomposition of regularly derived (e.g., Solomyak & Marantz, 2010) irregularly derived (e.g., Stockall & Marantz, 2006) and pseudo-suffixed forms (e.g., Lewis, Solomyak, & Marantz, 2011; Whiting, Shtyrov, & Marslen-Wilson, 2014). This body of research indicates that comprehending a visual word entails decomposition into constituent morphemes, which are linked to abstract representations in the lexicon for processing.

The influence of word-internal structure in spoken word recognition has been explored to a much lesser extent, and contention remains regarding the role of morphology in auditory processing. Methodologies for exploring the decomposition of complex words into morphemes during spoken word recognition include cross-modal priming, whereby an individual is presented with a masked visual word and asked to make a lexical decision on an auditorily presented target. Evidence from this paradigm appears to coincide with evidence from the visual domain of processing, whereby the root of a regularly derived complex word (e.g., government-GOVERN; Kielar & Joanisse, 2010; Marslen-Wilson, Tyler, Waksler, & Older, 1994) or suffixed non-word (e.g., rapidifier-RAPID; Meunier & Longtin, 2007) is primed for recognition. Responses to morphological violations such as the incorrect use of verbal inflection have also been evidenced to elicit specific ERP response components, independent from semantic or syntactic lexical errors (Friederici, Pfeifer, & Hahne, 1993). Furthermore, compound words that consist of two free stems (e.g., teacup) also

---

[*] Corresponding author at: New York University Abu Dhabi, Saadiyat Island, PO Box 129188, Abu Dhabi, United Arab Emirates.

E-mail addresses: laura.gwilliams@nyu.edu (L. Gwilliams), marantz@nyu.edu (A. Marantz).

[1] Address: New York University, 10 Washington Place, New York, NY 10003, United States.

http://dx.doi.org/10.1016/j.bandl.2015.04.006
0093-934X/© 2015 The Authors. Published by Elsevier Inc.

appear to be decomposed into their constituents and incrementally integrated (Koester, Holle, & Gunter, 2009), aided by prosodic information (Koester, 2014).

Two main theories of spoken word recognition can currently be recognised. A "continuous", non-decompositional approach supports a strictly linear and morphologically insensitive method of auditory processing: The Shortlist B model as proposed by Norris and McQueen (2008) posits that auditory word recognition is based on the probability distribution of acoustic signals over time, whereby the likelihood of each incoming phoneme is predicted based upon all prior phoneme(s) that have been processed, regardless of word-internal structure. This theory is considered a full listing model as it assumes a lexicon that is structured in terms of whole word units rather than morphological constituents, in accordance with Butterworth (1983) and Janssen, Bi, and Caramazza (2008). A recurring and prevalent objection to such a theory, however, is the necessary redundancy that would be caused by holding separate entries in the lexicon for morphologically related words such as "cover", "uncover" and "covering", for example (Wurm, 1997); although some suggest that using storage size as a measure of efficiency is misguided given the capacity of the human brain (Bybee, 1988; Sandra, 1994). In addition, from a linguist's perspective, full listing models are not obviously compatible with the results of linguistic morphology (see Marantz, 2013).

The "dis-continuous", decompositional group of models holds a contrastive view. These theories support a morphologically structured lexicon and therefore a morphologically centred mechanism of auditory processing. From this perspective morphologically complex words are decomposed during word recognition, production and storage, and representations are formed on the basis of morphological constituents rather than whole words. By implication, a dis-continuous model of auditory processing would work on the basis of *morphological* recognition rather than whole word recognition. Consequently, each subsequent phoneme in the input is compared to possible morphemes and morphological continuations.

Experimental work has considered the uniqueness point (UP) to be an important factor in adjudicating between these two routes of auditory word recognition. The classic definition of UP refers to the point at which the word deviates from all onset-aligned words apart from inflectionally suffixed words and compounds, and has been shown to be an important determiner of lexical decision reaction-time (Marslen-Wilson & Welsh, 1978). This measure of UP assumes a continuous model of auditory word recognition, in agreement with Shortlist B, as it posits that the multimorphemic status of a cohort competitor formed through the affixation of derivational morphemes, whether or not these morphemes are productive in a language, is irrelevant to lexical recognition, with derived and un-derived forms treated equivalently. Recently, morphologically sensitive measures of UP have also been defined and positively assessed as predictors of lexical processing. For example, Balling and Baayen (2012) define the *complex uniqueness point* (CUP) as the point at which a suffixed word becomes uniquely distinguishable from all words that share the same stem, therefore considering derived morphological continuations as (morphological) competitors during recognition. Wurm (1997) focuses on the importance of prefixes to spoken word recognition and formulates the *conditional root uniqueness point* (CRUP) as the uniqueness point of the root given a particular prefix. Both the CUP and the CRUP were found to contribute significant predictive value to models of auditory (Wurm, 1997) and visual (Balling & Baayen, 2012) lexical decision tasks, in addition to the classic UP measure. Both authors therefore suggest that a combination of full-form processing and decomposition are involved in word recognition. Although these calculations do not constitute a processing model in their own right, they indicate that morphological structure is relevant to word recognition and motivate the formulation of a morphologically sensitive model of lexical processing.

1.2. Neuroimaging research of phoneme processing and prediction

Neurophysiological investigations into spoken word recognition suggest that the superior temporal gyrus (STG) is responsible for both low- and high-order processing of speech (Obleser & Eisner, 2009; Scott, Blank, Rosen, & Wise, 2000; Scott & Johnsrude, 2003). A recent study (Mesgarani, Cheung, Johnson, & Chang, 2014) investigated the role of the STG in processing acoustic information such as phonetic features, in order to establish how phoneme distinction arises during processing. Participants listened to 500 sentences of natural speech samples across a range of 400 native English speakers, and neurophysiological responses were recorded at the onset of each phoneme using direct inter-cranial recordings from the cortical surface of the STG. Distinct neural responses were found for phonemes differing on certain feature dimensions, such as manner of articulation for consonants (e.g., plosive vs. fricative), or the place and manner of articulation of vowels (e.g., low-back, high-front or glide); consistent responses were established across phonemes with shared features, regardless of the physical difference in acoustic realisation as a consequence of speaker differences. The neural populations recorded were found to be sensitive to phonetic features within the time-window of 150–200 ms post-phoneme onset-suggesting that the STG is responsible for low-level (but "abstract" categorical) processing of speech during this time course of activation.

Later in the time-course, the STG has also been associated with high-level processes such as the encoding of phonological prediction based on lexical knowledge. Gagnepain, Henson, and Davis (2012) conducted a study that compared responses of learned novel words (e.g., *formubo*) as compared to existing similar words (e.g., *formula*) and baseline words to which the participants had no prior exposure (e.g., *formuty*). The learned novel word "formubo" served to delay the UP of "formula" until the final consonant, thus modifying the possible phonemes that could be predicted at "formu" and allowing for an assessment of segment prediction at the following phoneme. The authors used magnetoencephalography (MEG) to measure neurophysiological responses to experimental items pre- and post-UP (e.g., before and after the "l" in "formula") in order to assess how the trained novel words modified phoneme prediction. When comparing learnt and existing items, sensor-space analysis of the root mean square (RMS) of left-temporal MEG gradiometers found a reliable temporal cluster 280–350 ms after the onset of the UP; more activity was elicited for the novel over the existing words, suggesting that activity negatively correlated with the predictability of the divergent phoneme. No differences were observed pre-UP, also in accordance to theories of segmental prediction, as all information prior to the divergent phoneme supports both the existing and learnt lexical items. Source reconstruction of these neural responses localised the effect to the STG. In a model proposed by the authors they suggest that the STG is responsible for establishing a set of co-activated lexical candidates given the sensory input, in order to form competing hypotheses about which phonemes will be heard next. With each additional speech segment, any competitors that become incongruent with the input are eliminated, and the remaining cohort receive increased activation as likely lexical targets. If the materialised phoneme sequence does not match the expectations formed by possible outcomes, the resultant activity reflects an "error prediction signal". This model therefore places competitors for word recognition at the forefront of segmental prediction, and the STG as the focal location for encoding responses related to segment prediction.