Models of processing complex spoken words: the naïve, the passive, and the predictive

Suhail Matar (matar@nyu.edu)
Department of Psychology, New York University, 6 Washington Place
New York, NY 10003, USA

Alec Marantz (marantz@nyu.edu)
Department of Psychology & Department of Linguistics, 6 & 10 Washington Place
New York, NY 10003, USA
NYUAD Research Institute, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE

Abstract:

Speech processing involves segmenting a continuous input stream at various levels (e.g., sounds or words). But does the brain also segment spoken words into their meaningful subparts (called morphemes)? We gathered neurophysiological (MEG) data from participants as they heard complex words in Arabic, and compared the data against three different models of speech comprehension: a *naïve* model without morphological features, a *passive* model with morpheme onset information, and a *predictive* model with boundary anticipation and morpheme surprisal and entropy. The predictive model explains significantly more data variability in bilateral superior temporal cortex compared to the passive model, which in turn explains more variability than the naïve model in bilateral temporal and inferior frontal regions. We also test different predictive parsing strategies. Our results support speech comprehension models that segment the input into morphemes predictively, rather than passively wait for boundaries or full morpheme identification.

Keywords: speech; morphology; predictive processing; segmentation; magnetoencephalography (MEG)

Introduction A major challenge in speech processing is how the brain segments a continuous input stream, and it is typically studied at the word or sound level. But morphemes (the smallest meaningful language units; e.g., '*bake*' and '*-ing*' in the word '*baking*') comprise an intermediate level. Research has shown that the brain decomposes morphologically complex words into their parts in reading (see Leminen, Smolka, Duñabeitia & Pliatsikas, 2019 for a review). Does the brain segment speech at the morphological level? If so, what is the nature of this process?

Methods 27 participants listened to single words in Arabic while we recorded brain activity using magnetoencephalography (MEG). Words consisted of a verb stem morpheme and one of four direct object pronouns (Fig. 1a; e.g., 'qayyama-ni'='*(He) evaluated me.*'; hyphens represent morpheme boundaries). Verb stems were either long (all from the Arabic template '_a__a_a') or short (a shorter template with the same onset: '_a__a'). In Arabic, root consonants are substituted into the underscore slots to produce stems (e.g., root {**j**,**r**,**b**} produces '**j**a**rr**a**b**a'='*(he) tested*' in the long template; {**j**,**r**} produces '**j**a**rr**a'='*(he) dragged*' in the short template). We had two conditions (Fig. 1). Morphologically ambiguous stems were short or long: all long stems had corresponding shorter stems with the same onset (e.g., 'jarra'/'jarraba'), producing temporary ambiguity as they unfold. Morphologically unambiguous stems were all long (e.g., 'qayyama'), with no derivable shorter stems (i.e., 'qayya' is not a stem). Across conditions, stems became uniquely identifiable at the same stem uniqueness points (Fig. 1a; Balling & Baayen, 2012). Comprehension tasks targeting either stems or pronouns followed 25% of trials. Using source-localization, we estimated cortical activity in bilateral temporal and inferior frontal areas.

We contrast three hierarchically-nested models (Fig. 1b): (i) a morphologically *naïve* model that has acoustic, lexical and phonetic predictors, but is insensitive to morphological information or boundaries; (ii) a *passive* model, sensitive also to morpheme boundary/onset, and (iii) a *predictive* model, sensitive also to anticipatory segmentation, plus morphological surprisal and uncertainty (a function of corpus-calculated transition probabilities between morphemes). Specifically, we test three predictive parsing strategies: a *patient* parser waits until a morpheme is fully identified before anticipating a boundary; an *eager* parser begins the predictive process as soon as the input is congruent with any morpheme, even if not yet unique (e.g. '*jarra*', even if the actual stem will be '*jarraba*'), even at the cost of a later boundary revision (Fig. 1b); a *probabilistic* parser assigns weights to early and late boundary options depending on corpus frequency. We used a temporal response function framework (TRF; Brodbeck et al., 2021) to estimate typical responses to each predictor, and to measure the models' explanatory power.

Results & Interpretation Compared to a null model, all three models significantly explained activity in all regions ($p<0.0001$; corrected for multiple comparisons), which validates our models' predictive power. When comparing nested model pairs, the *passive* model with morpheme onset information explained significantly more activity than the *naïve* model in bilateral temporal and inferior frontal ROIs ($p=0.0001$; Fig. 2a). In turn, all *predictive* models explained more activity than the *passive* model in bilateral temporal cortex (*patient* & *probabilistic*: $p<0.0001$; *eager*: $p=0.0002$; Fig. 2b).

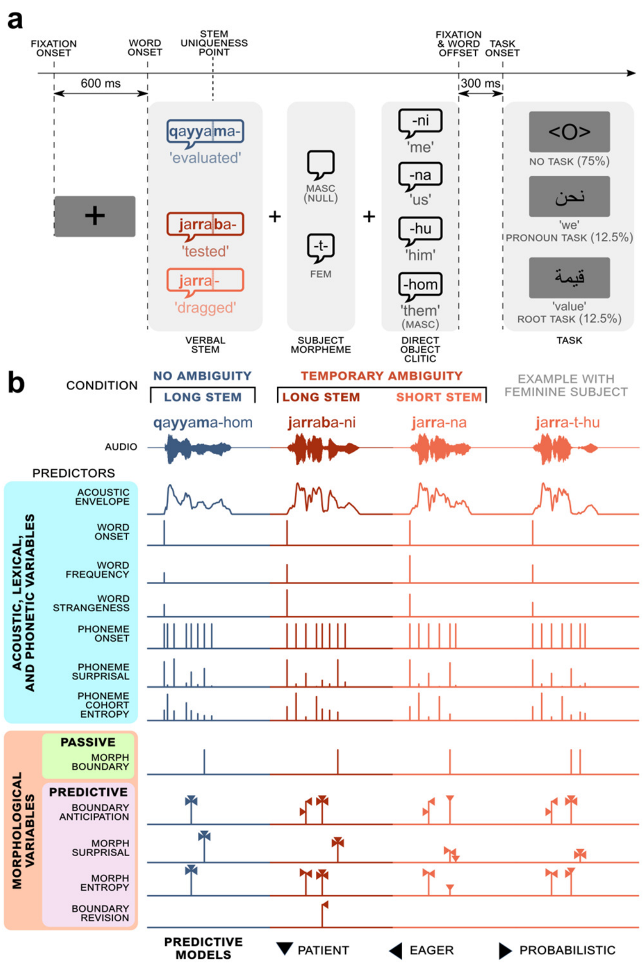

Figure 1: a) Experimental design & trial structure. Words had a stem morpheme, a subject morpheme (null if masculine subject), and a direct object morpheme. Stems had identical uniqueness points (vertical bars), after which stem were disambiguated from all possible stems. Stems had either no ambiguity (blue; always long-stem), or temporary ambiguity (red=long, orange=short stem); b) Example predictor values. We compare three hierarchically-nested models: the *naïve* model only had acoustic, lexical, and phonetic features (cyan); the *passive* model had also a morpheme boundary feature (cyan+green); the *predictive* models had also predictive morphological features (cyan+orange). We considered three predictive parsing strategies: patient (▼), eager (◀), and probabilistic (▶)

Our results provide evidence for morphological segmentation during speech processing, and support models where the brain (i) predictively segments words to morphemes, rather than passively wait for uniqueness points or boundaries, and (ii) uses morphological information to predictively process speech (Ettinger, Linzen & Marantz, 2014). This challenges some cohort-based models for speech comprehension, in which words are undecomposable units (e.g., Norris & McQueen, 2008).

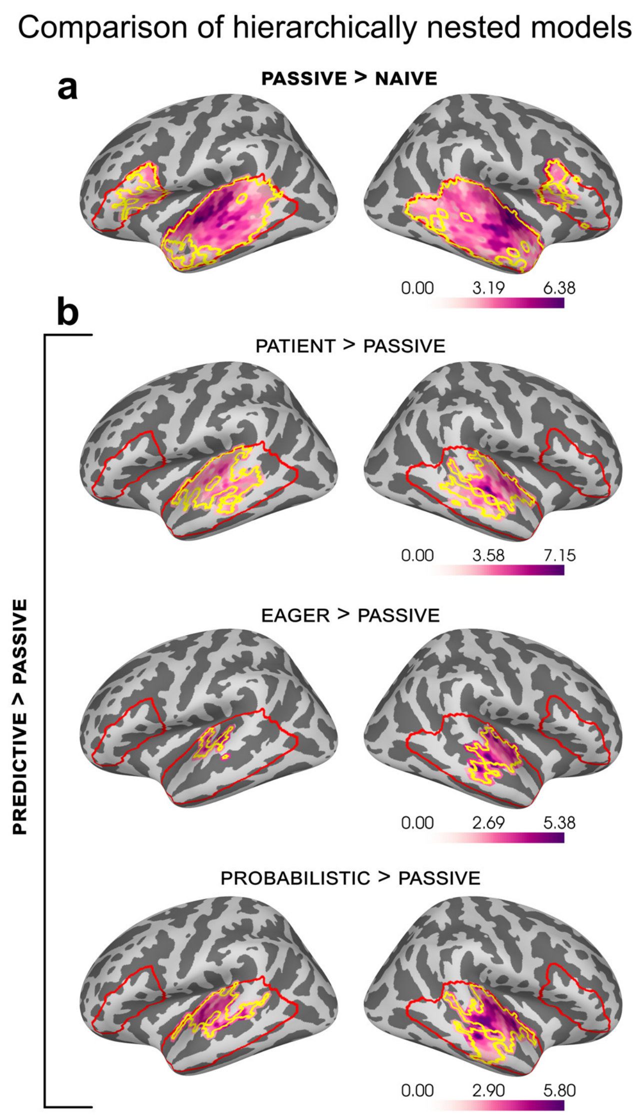

Figure 2: Comparing predictive power in a) passive vs. naïve models; b) predictive models (all three strategies) vs. passive model. Each panel shows t-value map (colorbar) from a one-tailed related t-test between models. Red borders outline test ROIs. Yellow borders show clusters of greater explanatory power.

Acknowledgments

This research was supported by the NYUAD Research Institute under Grant G1001.

References

Balling, L. W., & Baayen, R. H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*, *125*(1), 80–106.

Brodbeck, C., Das, P., Kulasingham, J. P., Bhattasali, S., Gaston, P., Resnik, P., & Simon, J. Z. (2021). Eelbrain: A Python toolkit for time-continuous analysis with temporal response functions. *bioRxiv*.

Ettinger, A., Linzen., T., & Marantz, A., (2014). The role of morphology in phoneme prediction: Evidence from MEG. *Brain & Language*, *129*, 14-23.

Leminen, A., Smolka., A., Duñabeitia, J. A., & Pliatsikas, C. (2019). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex*, *116*, 4-44.

Norris, D., & McQueen, J., (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357-395.