

## SHALLOW MORPHOLOGICAL PROCESSING IN PSEUDOWORDS

Lexical innovation and encoding are key components in the process by which the lexicon grows and changes. Existing phonotactic and morphological patterns influence the encoding and adoption of new words in a feedback loop. In this study, we explore the role of partial and complete morphological decomposition in determining the acceptability of novel possible words (pseudowords). We show that, controlling for phonotactics, speakers prefer pseudowords with a morphological decomposition even if it is nonsensical. This result argues for the existence of shallow morphological parsing, and sheds light on the role of morphology in the evolution of the lexicon.

The influence of morphological decomposition in processing known words has been previously investigated via semantic priming. Transparently-derived words prime embedded words, but words without transparent derivation do not: ‘cleaner’ primes ‘clean’ but ‘tinsel’ does not prime ‘tin’ (Marslen-Wilson, et al, 1994; Rastle, et al, 2004). Such results suggest deep morphosemantic analysis of known words. But they shed little light on the processing of novel words, which have no established meanings. It is known that phonotactic cues to morphology influence acceptability ratings of pseudowords (Hay et al. 2004); here we focus on a different source of information, namely the presence of substrings matching established morphemes. 8400 pseudowords were generated using biphone and triphone statistical models. These were trained on monomorphemic words in CELEX, to limit the presence of phonotactic word boundary cues. Real words were excluded. Length and phonotactic probability were systematically varied. Using an online interface, 24 Likert-scale acceptability judgments per word from US English speakers were collected. There were 1440 participants in total. The pseudomorphology effect is tested for predictive power beyond the fit of a phonotactic predictor of acceptability using an LMER analysis.

The decomposability of the pseudowords was assessed using left/right substring matches in CELEX, and the Lancaster English stemmer from the Natural Language Toolkit (Bird, et al, 2009). Neither method uses syntactic or semantic analysis. The decompositions include novel compounds and complex words (e.g., ‘sellfilth’), concatenations of morphemes that lack a valid morphosyntactic analysis (e.g., ‘nullier’), and strings with a partial analysis, where the identification of a word or suffix leaves a meaningless residue.

Controlling for phonotactics, apparent suffixation and compounding both contribute to significantly increased acceptability. For compounding,  $X^2(2)=594.5$ ,  $p<0.001$ ; for suffixation,  $X^2(2)=264.2$ ,  $p<0.001$ . The effect is larger for complete decomposition than for partial decomposition: for pseudocompounds,  $\beta=0.30$ ,  $SE=0.016$ ,  $t=18.2$ ; for real compounds,  $\beta=0.89$ ,  $SE=0.044$ ,  $t=20.0$ ; for pseudosuffixation,  $\beta=0.16$ ,  $SE=0.016$ ,  $t=9.9$ ; for real suffixation,  $\beta=0.61$ ,  $SE=0.042$ ,  $t=14.5$ . Lexical neighborhood density is not a significant predictor.

These results imply the existence of shallow morphological parsing, in which potentially meaningful parts of words are identified without building a complete lexical representation. The parallel existence of shallow and deep processes has been suggested in other linguistic domains as an efficient, flexible strategy for perception in noisy and variable contexts (Sanford & Graesser, 2006). The enhanced acceptability of parseable pseudowords should give them an advantage in being added to the lexicon over phonotactically legal words of comparable length. The lesser advantage for words with a partial parse means that pseudowords containing ‘cran-morphemes’, though viable, should be less readily assimilated than fully decomposable pseudowords. A followup study that we also hope to present in July seeks to further validate these conjectures by eliciting explicit parses of pseudoword and real word stimuli.