Toward Lexical Access Mechanisms for Regular and Irregular Forms
Constantine Lignos (University of Pennsylvania), Kyle Gorman (Oregon Health & Science University)

Background: Experimental investigations of processing of complex words can lead to better understanding of how these words are represented in the mind, a subject of debate both from theoretical (e.g., Jackendoff 1975; Halle and Marantz 1993) and experimental (e.g., Baayen et al. 1997; Taft 2004) perspectives. We present evidence from frequency effects that both regularly and irregularly inflected forms are accessed as decomposed forms through serial access (rank-based) mechanisms.

Proposal: We use the English Lexicon Project (ELP; Balota et al. 2007), the largest set of lexical decision data for English words, to explore the nature of lexical access using reaction times (RT). We examine the observed frequency effects and determine they are best explained by the rank hypothesis (Murray and Forster 2004). We then extend this finding to processing of irregular verbs, finding that increasing frequency of a class of irregular verbs facilitates lexical access beyond mere word frequency, providing support for models of irregular form storage in which classes of exceptions are listed in order of use (Yang 2005).

Procedure: We analyzed correct-response trials of regularly inflected (6,681 items, 201,758 trials) and irregular past tense (58 items, 1,856 trials) English words. Maximal mixed effects linear regression models of log RT were fitted with random intercepts for subjects and fixed effects of trial number, education level, gender, and item-based predictors: squared orthographic length, orthographic neighborhood size, and number of syllables. To compare rank and frequency, the residuals of these models were examined using linear models for additional effects of frequency and rank predictors for whole words (e.g., playing) and bases (e.g., all words containing play) beyond a baseline model.

Rank provides a mechanism for frequency effects: Figure 1 shows the effect of frequency across a wide range of items after other predictors are accounted for. We find that the simple model of accessing roots/bases of lexical items as a frequency ordered list (Murray and Forster 2004) may best explain this, as highest-frequency items have the lowest rank (order) in the list and can be accessed first. Figure 2 shows the effect of base rank on RT, with the near coincidence of the linear model and LOESS fits demonstrating that RT is proportional to rank, confirmed by the finding that expressing rank as degree two polynomial does not improve fit (F(1) = 3.20, p = 0.073). This represents the first large-scale evidence that lexical access of base forms by serial access best explains frequency effects observed in morphologically complex words.

Irregulars are decomposed and accessed through classes: With these decomposition and rank effects in evidence, we extend our analysis of regular inflections to irregular past tense forms. Elsewhere Condition-based models (e.g., SPE, Yang 2005) model exceptions to regular processes as a list of exception classes (e.g., rime → /aʊt/ for think, catch, etc.) that must be traversed before the default rule, with the order of classes determined by their frequency. We computed the frequency and rank of each class, finding significant effects for both word frequency ($\chi^2_{LR}(1) = 26.4, p = 2.83 \times 10^{-7}$) and irregular class rank ($\chi^2_{LR}(1) = 7.27, p = 0.007$), with the lowest rank (most frequent) class facilitating lexical decision 22ms compared with the highest rank class. This demonstrates that in addition to the known effects of frequency of irregulars (Clahsen et al. 2004), the frequency of an irregular verb’s class also facilitates access, providing evidence for their storage as decomposed forms.

Conclusion: This study presents the most comprehensive effort to date to characterize frequency effects for inflected forms using specific lexical access mechanisms. We conclude that the simplest explanation for frequency effects in lexical access is the serial rank hypothesis and demonstrate that this hypothesis predicts RT data for both regular and irregularly inflected English words. We are currently extending this work to provide an explicit algorithm for lexical access of these forms and compare its predictions against models such as Minimum Generalization (Albright 2002) on irregular forms.

1Residualization was performed among correlated predictors ($r > .20$) to eliminate multicollinearity. All frequencies were computed using US English SUBTLEX norms.
Figure 1: LOESS fit line of RT for word frequency.

Figure 2: LOESS (red) and linear model (blue) fit lines of RT for base (root) rank.

References


