The relationship between lexical frequency, compositionality, and phonological reduction in English compounds

Forrest Davis[1] and Abigail C Cohn[1]
[1]Cornell University

This work investigates the effect of lexical frequency and compositionality on phonological reduction. A sample of compounds selected from the Buckeye Corpus (Pitt et al 2007) was analyzed to test the effects of lexical frequency and compositionality on the rime duration of compounds: do more opaque compounds (*cupboard*) exhibit more phonological reduction (shorter rimes) than transparent ones (*blueberry*)?

Degree of compositionality, a challenge in many areas of linguistics, psychology (Libben & Jarema 2006) and computer science (Reddy et al 2011), is determined for the present research on the basis of a survey conducted with 24 American English speakers, who were asked to rate the compounds, on a 7-point Likert scale (opaque to transparent). Lexical frequency was taken from the Corpus of Contemporary American English (Davies 2008), and degree of conventionalization was calculated from these data using Pointwise Mutual Information (Church & Hanks 1990) (PMI). See Table 1.

While phonological reduction in compounds can have numerous realizations, the present research focuses on the duration of the final rime of the second word. Crucially, when compared to the same rime in monosyllabic nouns, this duration offers a continuous measure that can be applied to all compounds, and also serve as a measurement reflective of loss of secondary stress. Given the potential effects of other factors, we also controlled for lexical frequency and conventionalization of the items.

Measurements were made of 21 different bisyllabic nominal compounds with VC(C) rimes to assess the extent to which compositionality is correlated with degree of phonological reduction. Two specific hypotheses were tested:

Hypothesis 1: The duration of the final rime of a compound is statistically different from the duration of the same rime in monosyllabic nouns.

Hypothesis 2: The degree of compositionality is predictive of the duration of the final rime of a compound.

Specifically, Hyp1 predicts that the ratio of the final rime to the rime in the monosyllabic noun will be positively correlated with the compositionality rating of the compound. Hyp2 predicts that the compositionality rating of a compound will be a statistically significant predictor of the duration of the final rime of a compound.

As seen in Table 2, Hyp1 was confirmed: a two-sample t-test revealed that compounds with lower compositionality ratings are significantly shorter than expected from the duration of the rimes of the corresponding individual elements. Indeed, transparent compounds such as *weekend* and *airport*, have a rime longer than expected, suggesting increased compositionality ratings are associated with further lengthening.

As seen in Table 3, compositionality is a significant predictor of degree of phonological reduction, as determined in a linear regression model: we observe a positive correlation, where the less compositional a compound is, the shorter its final rime is.

Taken together, the results demonstrate that conventionalization and compositionality are correlated with phonological reduction, while just the raw lexical frequency for the compound or its head are not significant predictors. Thus, lexical frequency and compositionality have distinct reflexes in phonological form, suggesting that there is a dynamic relationship between lexical representation and phonological realization.

Compound	rating	freq	MODfreq	HEADfreq	PMI	duration
homework	6.375	6068	196060	216060	4.1561	210.2434
airline	2.375	19170	119180	15920	6.1096	219.6274

Table 1: Compounds with compositionality rating, frequency of compound (freq), frequency of modifier (MODfreq), frequency of head (HEADfreq), PMI value, and mean duration (in ms).

compound	rating	p-value	tstat
software	1.875	0.027	-2.220
freshman	3.087	7.60e-06	-4.517
background	3.208	0.022	-2.340
airplane	4.167	0.002	-3.182
football	4.292	0.036	-2.116
airport	4.583	0.0006	3.475
weekend	6.625	0.008	2.659

Table 2: Compounds with significant (< 0.05) p-values from a two-sample t-test. Compositionality rating and t-statistic included.

Factor	Estimate	SE	tstat	p-value
rating	447.69	140.52	3.186	**0.002**
freq	0.002	0.002	1.3336	0.184
mod freq	0.004	0.001	2.795	**0.006**
head freq	-0.001	0.0005	-1.898	0.059
PMI	96.706	42.595	2.270	**0.024**
mono dur	17.0	5.514	3.085	**0.002**

Table 3: Main factors in linear regression for duration of final rime in compound. Freq is frequency of compound, mod freq is frequency of modifier of compound, head freq is frequency of head of compound, and mono dur is duration of rime in monosyllabic noun.