"Boy and Girl" or "Girl and Boy": Dialect, Modality and Frequency Effects on the Production of Binomial Expressions

Ha Leem Son & Eunkyung Yi (Ewha Womans University)

Introduction The components of language lie on a continuum, ranging from fully productive to highly fixed expressions, and studying this entire span is essential to deepening our understanding of human language. Transparent *binomials* are a type of multi-word expression that exhibit both complete structural productivity and a high degree of fixedness [1-2]. For example, the binomial *boy and girl* simply refers to the set {boy, girl}, while its reversed order, *girl and boy*, is meaning-equivalent but considered less natural by native speakers. However, this intriguing phenomenon has been studied mostly in terms of linguistic and cognitive factors that may modulate the ordering of binomials. Informed by usage-based linguistics, the present study aims to conduct a comprehensive analysis of the use of 32 transparent binomials in naturally occurring language, focusing on factors likely to modulate their use, namely dialect (British & American), modality (written & spoken) and frequency (token & relative).

Methods A list of 32 fully transparent binomials was selected for analysis based on prior studies (e.g., [1]). These binomials were queried in both their canonical (*right and wrong*) and reversed (*wrong and right*) orders in a collection of naturally occurring spoken and written language, using the Corpus of Contemporary American English (COCA, 1 billion words, 1990-2019) and the British National Corpus (BNC, 100 million words, 1980-1993). Each resulting token (66,081 tokens in total) was coded for dialect (American/British), modality (spoken/written), order (canonical/reversed), and year (when available). Logistic regression analyses were conducted in two main areas: the first examined the effects of dialect and modality on the order of binomials, using data from both COCA and BNC. The second focused on the effects of modality and frequency, using COCA only.

Results First, when comparing the two corpora in their entirety, we found main effects of dialect ($b = 0.09$, $p < 0.001$) and modality ($b = 0.48$, $p < 0.001$) on the order of binomials and a marginally significant interaction between the two ($b = -0.33$, $p = 0.053$). The nature of the interaction was that the spoken modality was a stronger predictor in the BNC ($b = 0.48$, $p < 0.001$) than in COCA ($b = 0.15$, $p < 0.001$). However, when the time periods of the corpora were aligned to the last year of the BNC (~1993), the dialect effect disappeared. Next, when token (canonical) and relative (canonical/total) frequencies were included as predictors along with modality in the analysis of COCA, we observed main effects of modality ($b = 10.93$, $p < 0.001$) and canonical frequency ($b = -0.92$, $p < 0.001$), but no effect of relative frequency ($b = 0.005$, $p = 0.60$). Further investigation of their interactions revealed distinct patterns in spoken and written language (see Figure 1 for general distribution trends). Specifically, relative frequency was a significant predictor of binomial order in the spoken modality ($b = 0.09$, $p < 0.001$), while canonical frequency was significant in the written modality ($b = -1.04$, $p < 0.001$) with an observed negative coefficient. Post hoc analyses revealed that the negativity in the effect of canonical frequency was primarily driven by a few binomials with relatively high competition between their canonical and reversed orders.

Discussion Overall, the results demonstrated that binomials are more likely to occur in their canonical order in spoken than written language, while evidence for dialect effects remains inconclusive, necessitating further investigation with corpora that are more balanced in size and better aligned in time frame. In light of usage-based linguistics, the findings regarding frequency effects show that the higher proportion of canonical binomials relative to reversed ones increases the likelihood of their occurrences in spoken language, whereas their use in written language is more influenced by the absolute frequency of canonical binomials. These findings suggest that modality and frequency interact to shape the ordering of binomials in naturally occurring language, indicating the need for deeper research on these factors across diverse linguistic contexts**.**

References

[1] Morgan, E., & Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition, 157*, 384-402.

[2] Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech, 63*(1), 95-122.
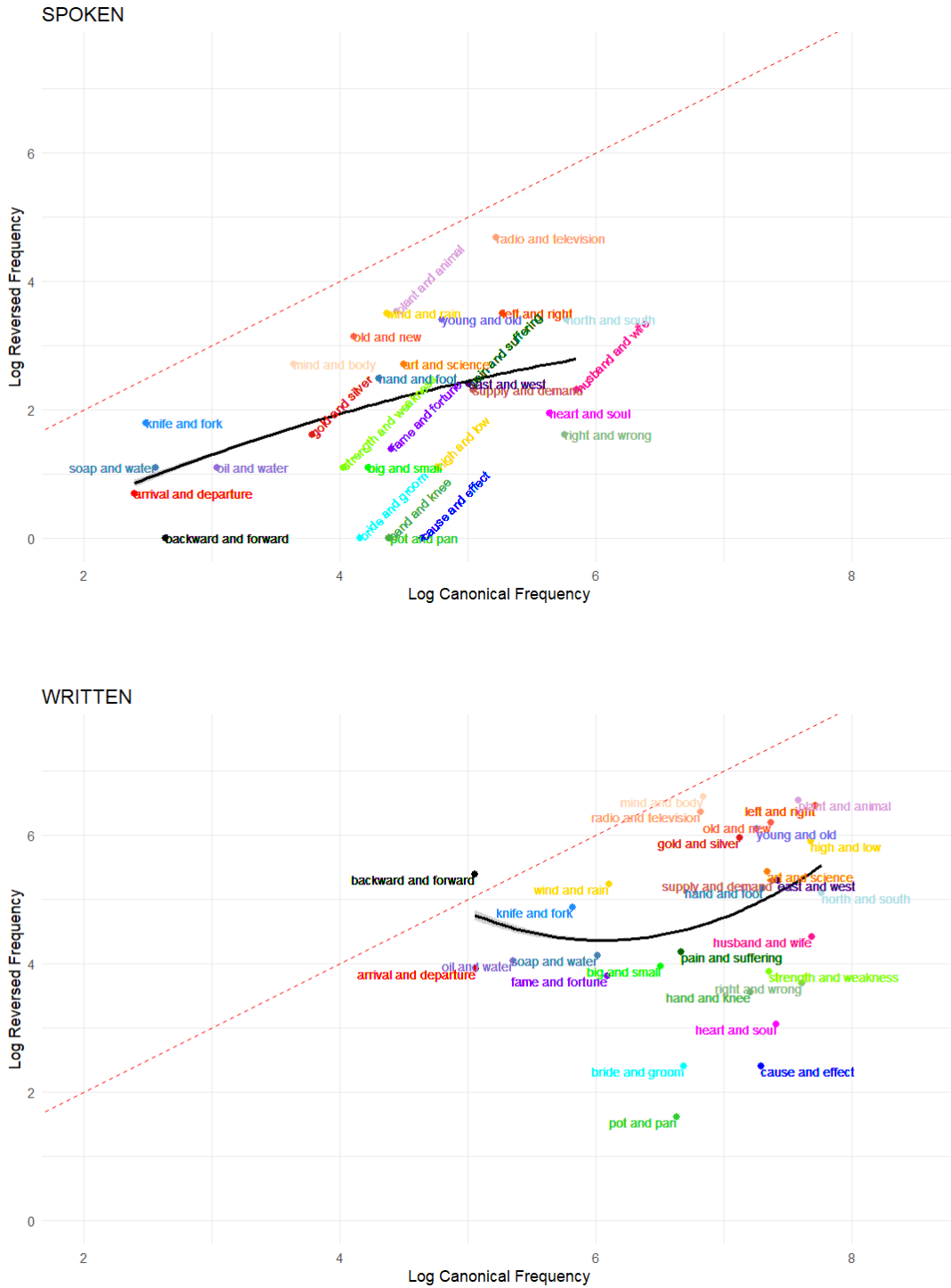
Figure 1. Distribution of relative and absolute frequencies of binomials used in the analyses