

The effect of dependency length minimization on acceptability

Motivation: Dependency Length Minimization (DLM; Temperley, 2007), which predicts a preference for shorter constituents to be closer to their syntactic heads, demonstrates how syntactic variation could potentially be shaped by processing efficiency (Hawkins, 2014). Liu (2020, 2021) conducted a corpus study investigating this principle across 34 languages, using the double PP construction as a test case (e.g. *we talked [PP1 about PPs] [PP2 with the reviewers]*). The results showed that although there is a crosslinguistic tendency for DLM, this preference is much weaker in head-final contexts (e.g. preverbal orderings in Hindi) as compared to head-initial ones (e.g. postverbal orderings in English).

Current study: We ask whether the tendency for shorter dependencies is reflected in acceptability judgments across languages with distinct typological features. In particular, we compare the double PP construction in English and Hindi. Our study extends prior work in two ways. First, most evidence for DLM comes from corpora (Futrell et al., 2015) and comprehension/production studies (Gibson, 1998); by contrast, acceptability judgment is a different empirical domain and it remains an open question if the preference for DLM holds in acceptability experiments for languages other than English (Bresnan et al., 2007). Second, the majority of studies on constituent orders have taken data from written texts/stimuli or post-processed transcriptions of spoken data, yet ordering preferences can differ between written and spontaneous spoken domains (Kramer, 2020, 2021), and the influence of DLM is weaker in naturalistic speech at least in English (Liu, 2019). This suggests the processing constraints of DLM could operate differently with natural spoken language stimuli.

Methods: We conducted a preregistered acceptability judgment study using 20 sets of audio stimuli (Sedarous & Namboodiripad, 2020) for English and 24 sets for Hindi. Items had animate subjects and intransitive head verbs immediately followed by two adjacent PP dependents (e.g. Table 1). Previous work of acceptability judgments manipulated dependency length via using syntactic alternations as stimuli, in which the order of constituents is switched (Bresnan, 2008). In comparison we included four conditions for each stimulus set: the baseline was the SHORT-SHORT condition where both PPs had a length of 3; the lengths of the PPs were then varied in the other three conditions via attaching a relative clause (RC; length 3) which modified the nominal head within the PP. In the English experiments, 128 English-speaking participants heard 5 items from each condition (plus 60 fillers of varying acceptability) and rated them on a 1-7 Likert scale. The ratings were transformed into by-subject z-scores, then subjected to Bayesian mixed-effect analyses (predicting z-scored ratings as a function of condition with long-long as the reference level; random effects for items and participants). The procedures were similar for the Hindi experiments, except that each of the total 73 Hind-speaking participants heard 6 items from each condition (plus 69 fillers of varying acceptability).

Results & Discussion: For English, we expected (1) the lowest mean ratings and correspondingly the lowest coefficient value in the mixed-effect models for LONG-LONG sentences; (2) the second lowest mean ratings and coefficient for LONG-SHORT sentences; (3) the potential rating differences between the SHORT-SHORT and SHORT-LONG conditions are less straightforward, since although the former has shorter dependencies, the latter abides by DLM. Results for English show LONG-LONG sentences do have the lowest mean ratings and coefficient value (mean z-score = -0.09; β = -0.09); SHORT-SHORT sentences are the most acceptable (mean z-score = 0.78; β = 0.88). However, we found no significant difference between the LONG-SHORT (mean z-score = 0.25; β = 0.35) and SHORT-LONG conditions (mean z-score = 0.22; β = 0.32). For Hindi, our analyses are comparatively more exploratory given lack of directly related findings except that Liu (2020) demonstrated no preference for DLM in written Hindi. Therefore we expected (1) also the lowest mean ratings and the lowest coefficient for LONG-LONG sentences; (2) comparable ratings and coefficients for LONG-SHORT and SHORT-LONG. Nevertheless, our results present no significant effect for any of the four conditions (LONG-LONG: mean z-score = 0.25; β = 0.22; SHORT-SHORT: mean z-score: 0.22; β = -0.02; LONG-SHORT: mean z-score: 0.19; β = -0.02; SHORT-LONG: mean z-score: 0.28; β = -0.02). These findings indicate that dependency length may not play a strong role in acceptability, at least with audio stimuli, as compared to how it predicts online processing behavior with written stimuli or patterns in corpora.

Table 1. Sample Stimuli in the English Experiment; RCs are underlined; the NPs within the PPs were all definite.

| Condition | Sample Stimulus |
|-------------|---|
| short-short | The researcher looked [PP1 through the lens] [PP2 at the sky]. |
| long-long | The researcher looked [PP1 through the lens <u>that was adjusted</u>] [PP2 at the sky <u>that was darkening</u>]. |
| short-long | The researcher looked [PP1 through the lens] [PP2 at the sky <u>that was darkening</u>]. |
| long-short | The researcher looked [PP1 through the lens <u>that was adjusted</u>] [PP2 at the sky]. |

Figure 1. Density plot of the z-scored acceptability ratings in the four conditions; the dashed line represents the mean z-scored rating.

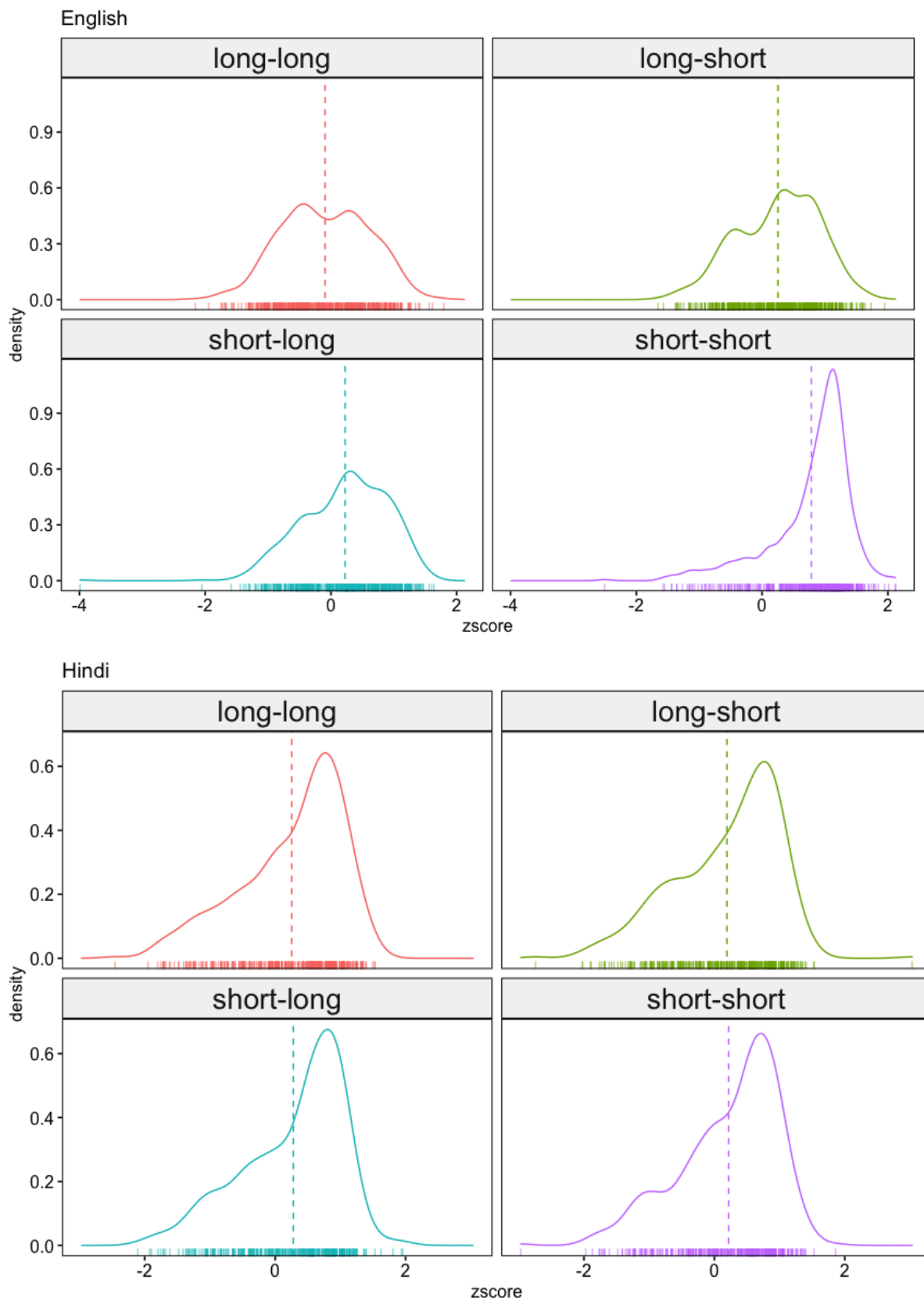


Figure 2. Coefficients of the four conditions predicting acceptability ratings; 95% confidence intervals were derived from the posterior of each parameter in the Bayesian mixed-effect model.

