Subjective estimates of verb frame frequencies

Nick Huang & Serene Siow
National University of Singapore

The frequencies of verb frames play an important role in psycholinguistics. Much research assumes that speakers and learners can track abstract verb frames in their linguistic experience, as revealed indirectly through behavioral measures, e.g. reading times and acceptability judgments [1-5, etc.]. This study asks whether speakers can directly access and report frame frequencies, drawing inspiration from work in the lexical processing literature showing that speakers can do so for words [6, etc.]. We report initial results from a study on American English suggesting an affirmative conclusion: self-reported (subjective) estimates tend to be positively correlated with corpus-derived estimates (from [7]) of both absolute and relative frequency.

This finding has potential practical implications, as it points to a novel method of obtaining frame frequencies. Currently, frame frequencies are estimated via corpus analyses and asking participants to compose/complete sentences with these verbs [7,8]. Both methods have limitations: the first requires access to large corpora, which are unavailable for many languages and thus present challenges for comparative psycholinguistics. The second, on the other hand, risks undercounting longer frames, which are more effortful to write/produce [8]. These are limitations that can be avoided by surveying speakers for frequency estimates.

Frequency estimation experiment. The goal of this study is to validate subjective estimates of verb frame frequency against [7]'s corpus-derived estimates. We focus on three frames prominent in sentence processing research: transitive, passive, and finite declarative sentential complements ("SFIN"). We selected 100 verbs with higher frequency of SFIN as reported by [7] (0-74% of all frames, median=9%). These were randomly sorted into 4 lists of 25 verbs.

We created one "semantically bleached" sentence [9] for each verb-frame pair. In instructions, participants (30 per list) were told that each semantically-bleached sentence represents a set of structurally-similar sentences; examples were provided. Using sliders, participants provided percentages for each of the three frames and a fourth option of "all other frames," summing to 100 (Fig. 1). Attention checks were added to four verbs on the list, requiring participants to set the slider to 0. We only analyzed responses by 115 participants who passed at least half the attention checks.

Results. A. Absolute frequency. Fig. 2 plots correlations with [7]'s estimates for the three frames. We find significant correlations, although correlations are stronger for passive and SFIN frames than transitives ($r=.56$ and $.53$ vs $r=.23$, all $p<.05$).

B. Relative frequency. Here we ask whether participants can judge accurately that one frame is more frequent than another for a given verb. We calculated frequency differences by crossing all three frames with each other. Fig. 3 shows that these relative frequency estimates again are positively correlated with those from [7]'s data (all $p<.05$).

Discussion. Our results show that it is possible to collect subjective frame frequency estimates that correlate well with corpus-derived estimates. This in turn raises new questions for investigation. One important question is whether subjective estimates truly reflect judgments of frequency instead of, say, acceptability. This is a valid concern at least for transitives. Post-hoc analyses show that after excluding the 20% of verbs that are the least acceptable with this frame, the correlation between subjective and corpus-derived absolute frequency estimates becomes insignificant ($p=.60$). However, similar analyses suggest that this is less of a concern for SFIN ($r=.46$; $p<.01$); ongoing work will examine this issue for passives. Another open question for now is how much detail speakers can track: e.g. whether speakers distinguish between passives based on the choice of auxiliary (*Someone was/got scolded*), as well as how clear a correlation can be obtained for a larger variety of frames (e.g. various clausal complements).

Figure 1. Example of frequency estimation task (using the verb *emphasize* as illustration)

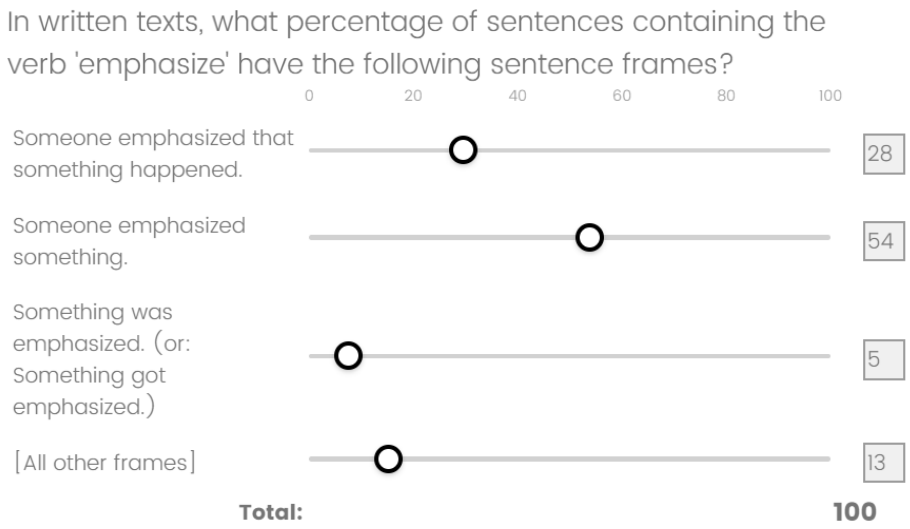


Figure 2. Pearson correlations of absolute frame frequency for 100 verbs (NB: unlike Gahl et al. we exclude from corpus frequencies the frames where the verb co-occurs with a particle, so we would treat e.g. *find* and *find out* separately. Our passives are intended to cover both Gahl et al.'s regular and adjectival passives.)

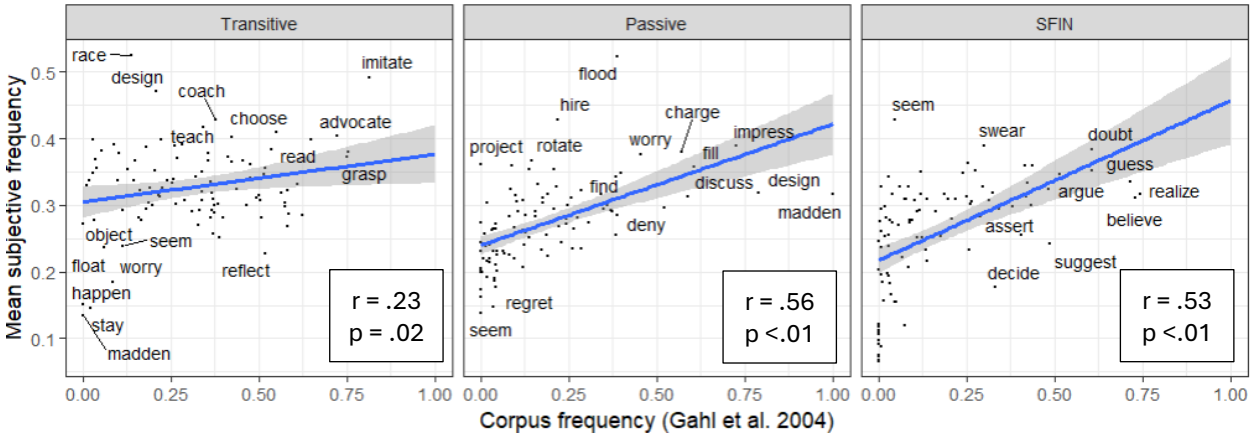
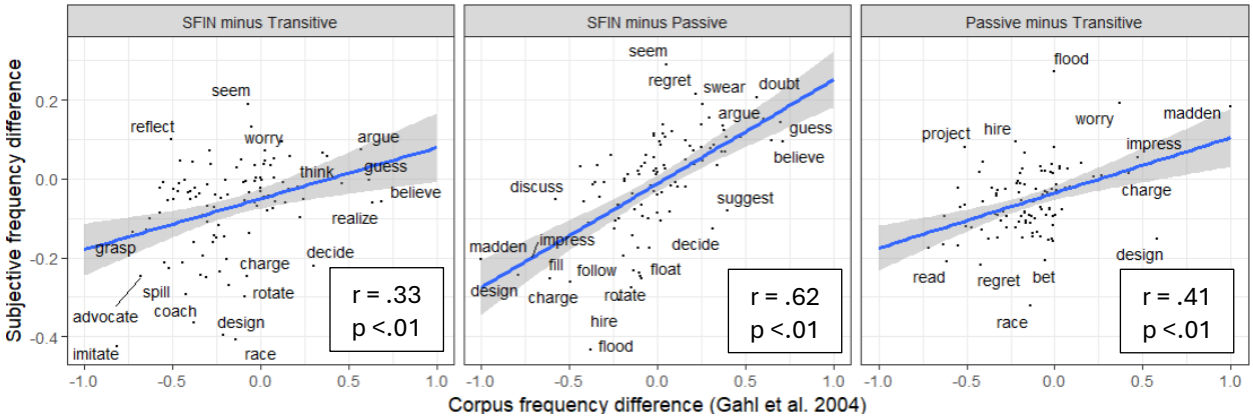


Figure 3. Pearson correlations of relative frame frequency for 100 verbs



Selected references. [1] Trueswell et al. 1994. *JML*. [2] Garnsey et al. 1997. *JML*. [3] Brooks & Tomasello. 1999. *Dev. Psych.* [4] Liu et al. 2022. *Cognition*. [5] Schuler et al. 2016. *CogSci Proceedings*. [6] Balota et al. 2001 *Memory & Cognition.* [7] Gahl et al. 2004. *BRMIC*. [8] Roland et al. 2007. *JML*. [9] White & Rawlins. 2020. *Glossa*.