

Structural frequency effects in comprehenders' noisy-channel inferences

Yingtong Liu (Harvard), Rachel Ryskin (MIT), Richard Futrell (UC Irvine), Edward Gibson (MIT)
y_liu@g.harvard.edu

Comprehenders must recover a speaker's intended meaning from linguistic input that may be corrupted by noise. Recent work suggests that they do so by combining prior knowledge $P(S_i)$ about what sentences are likely to be intended, with a noise model $P(S_i \rightarrow S_p)$ which encodes how an intended sentence S_i might be corrupted to a perceived sentence S_p in transmission, as in (1) (Gibson et al., 2013; Poppels & Levy, 2016; Ryskin et al., 2018). For example, given input such as "The table jumped onto the cat", comprehenders might infer that an error occurred between the speaker's intent and the ultimately-perceived sentence, such that the intended sentence was actually the more plausible "The cat jumped onto the table".

Previous studies have tested the noisy-channel framework by manipulating the prior probability of the *meaning* of the intended sentence (i.e., plausibility), but the theory also predicts that the *form* of the sentence matters. When presented with lower-frequency structures in the perceived sentence S_p , comprehenders should be more likely to infer that an error occurred and that the intended sentence S_i differs from S_p . In other words, low-frequency structures in the perceived sentence should bias the listener towards more plausible meanings/structures. This prediction follows from the noisy-channel framework because the prior probability $P(S_i)$ for an infrequent structure is low, so that other possible intended sentences with higher prior probability become more attractive for the comprehender. In a novel investigation of this hypothesis, we examined how comprehenders interpreted events of varying plausibility formed in word orders of varying frequencies in English and Mandarin.

Though it is standardly claimed that SVO is the dominant word order in English and Mandarin (Haegeman, 1994; Huang et al., 2008), there has not yet been a quantitative measure of the degree to which the two languages allow various word orders. **Expts 1-2** examined the acceptability of the six logically possible word orders for simple transitive events in English and Mandarin, as an approximate measure of the frequencies of these word orders in the two languages. We presented speakers with pictures depicting various events and the six logically possible word orders for each picture (2) ($n=30$ in each), along with fillers with clearly correct answers (e.g., "The boy is beside the couch"). A comma was inserted between two consecutive NPs to avoid ambiguity. Participants were asked to select all the acceptable descriptions among the six word orders in 8 test and 12 filler trials. Mixed effects logistic regressions with word order as predictor showed that SVO was more frequently chosen than OSV, the second preferred word order, in both languages, while other word orders are less common ($\beta_s > 8$, $z_s > 8$, $ps < 0.01$) (Figs. 1 & 2). Furthermore, the entropy in the distribution of allowable Mandarin word orders ($H=1.28$, 95% CI=[1.26, 1.3]) is larger than that in English ($H=0.95$, 95% CI=[0.9, 1]).

In Expts 3-4, English and Mandarin speakers ($n=97$ for English; $n=81$ for Mandarin) read implausible or plausible sentences in SVO or OSV word orders. Each sentence was paired with a comprehension question that distinguished literal and non-literal interpretations (3). There were 12 test and 24 filler trials. Mixed-effects logistic regression analyses with word order and plausibility as predictors revealed that OSV sentences were more likely to be interpreted non-literally compared to SVO sentences in both English and Mandarin ($\beta_s > 1.5$, $z_s > 6.5$, $ps < 0.01$; Figs. 3 & 4). Thus, as predicted, people are much more likely to interpret "The boy, the trash threw" as the more plausible "The boy threw the trash" than they are to interpret "The trash threw the boy" in the more plausible way (and similarly in Mandarin examples).

In sum, our results show structural frequency affects comprehenders' noisy-channel inference. This suggests that languages may tend to encode low probability meanings using more frequent forms or forms of higher redundancy (e.g., adding case markers in other languages), to facilitate efficient communication. Future production studies will investigate how message plausibility impacts the form speakers choose.

$$(1) P(S_i|S_p) \propto P(S_i)P(S_i \rightarrow S_p)$$

(2)



Please select ALL the acceptable English description(s) of the picture.

- | | |
|---------------------------|-------|
| The boy threw the trash. | (SVO) |
| The trash, the boy threw. | (OSV) |
| Threw the trash, the boy. | (VOS) |
| The boy, the trash threw. | (SOV) |
| The trash threw the boy. | (OVS) |
| Threw the boy, the trash. | (VSO) |

- (3) The trash threw the boy. (SVO_implausible) The boy threw the trash. (SVO_plausible)
 The boy, the trash threw. (OSV_implausible) The trash, the boy threw. (OSV_plausible)
Question: Did the boy throw something/someone?

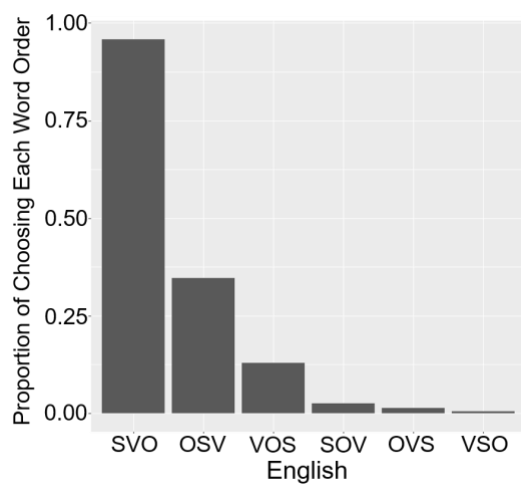


Fig.1. Proportion of choosing the six logically possible word orders in English.

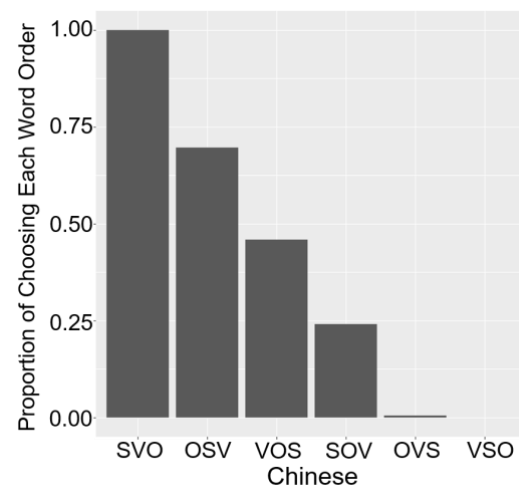


Fig.2. Proportion of choosing the six logically possible word orders in Mandarin.

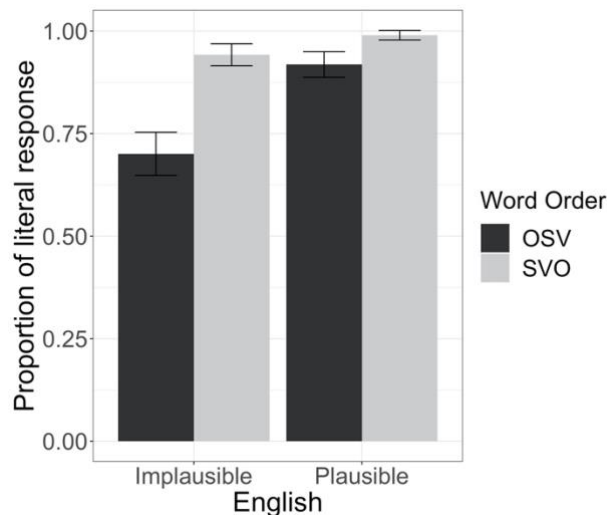


Fig.3. Proportion of choosing literal interpretation by word order and plausibility in English (95% CI).

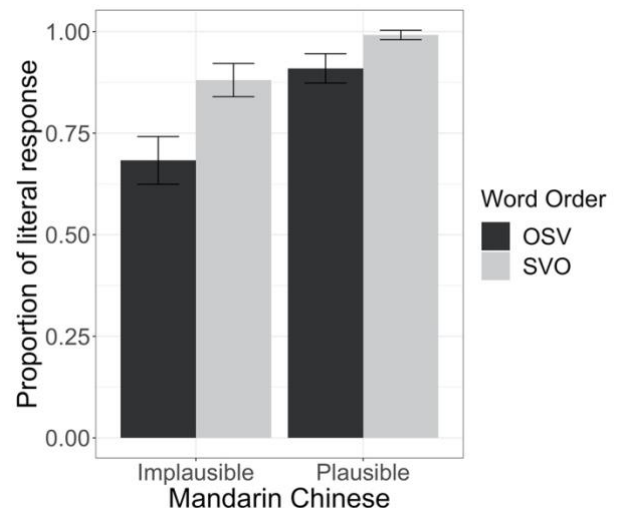


Fig.4. Proportion of choosing literal interpretation by word order and plausibility in Mandarin (95% CI).

References: [1] Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *PNAS*. [2] Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*. [3] Poppels T, & Levy R (2016). Structure-sensitive noise inference: Comprehenders expect exchange errors. *CogSci*.