The effects of website reading experience on constructional preferences
Michelle Perdomo, Duane Watson (Vanderbilt University)

According to constraint-based and surprisal theories, language users build statistical models to efficiently process new linguistic input [1-3]. This predicts that individual differences in language experience will lead to individualized language preferences; however, this prediction has not been directly tested. We assess individual variation in language experience and exposure by analyzing website texts provided by participants to gauge constructional preferences for phrasal verbs (PVs) and datives in English, constructions that are susceptible to verb biases.

Participants (n=178) were given an adapted version of a reading habits questionnaire. They estimated the number of hours spent reading different categories of print and website texts and listed their top 5 to 10 favorite websites. Participants then completed a two-alternative forced choice task in which they selected one of two construction alternations for English PVs and datives (see Table 1). The task included transitive PVs in English that alternate between the continuous construction and the discontinuous construction. Idiomatic phrasal verbs, i.e., phrasal verbs whose meanings are more opaque, frequently occur in the continuous construction, while compositional phrasal verbs, i.e., verbs whose meanings are more transparent, occur in the discontinuous construction [4]. English datives allow for a verb to take a double object (DO) frame or a prepositional (PP) frame. For datives, verbs that express possession occur more frequently in the DO construction, and datives expressing movement occur more frequently in the PP construction [5].

Usage-based approaches posit that more reading hours lead to statistical models that mirror real-world constructional distributions, which would lead readers with higher reading exposure to prefer PV and dative constructions based on compositionality and manner. As a result, participants with larger amounts of text exposure should exhibit a stronger preference for idiomatic PVs in the continuous construction, compositional PVs in the discontinuous construction, possession datives in the DO construction, and movement datives in the PP construction. These theories also predict that readers will reflect the constructional biases mirrored in the text content they consume. We scraped 405 websites provided by participants and extracted linguistic features using a suite of NLP tools [6-8]. We extracted features that govern dative and phrasal verb constructional preferences, such as lemma-construction properties, word frequency, phrasal complexity, etc. A PCA was used to group the resulting 73 selected variables to describe website text. We extracted four principal components explaining 69% of the variance: (RC1) spoken-like text, (RC2) bigram and trigram frequency and range, (RC3) construction typicality and lexical diversity, and (RC4) semanticity. As a whole, we find evidence for and against usage-based accounts.

Inconsistent with usage-based accounts, for datives, participants who read *fewer* hours had constructional preferences that matched real world distributions, while participants who read *more* hours had no constructional preferences ($\beta=0.19$, SE=0.07, p<0.01). No effect of hours was found for PV constructions. Consistent with usage-based accounts, those who read websites with higher RC3 values had stronger preferences for the dative PP frame ($\beta=0.25$, SE=0.10, p<0.01) and the PV continuous frame ($\beta=-0.34$, SE=0.11, p<0.01)—the dative PP and the continuous PV frames are more frequent in writing. These findings suggest that constructional preferences are linked to individualized reading experience, including constructional frequency.

Table 1. Example trials for phrasal verbs and datives and their alternating constructions.

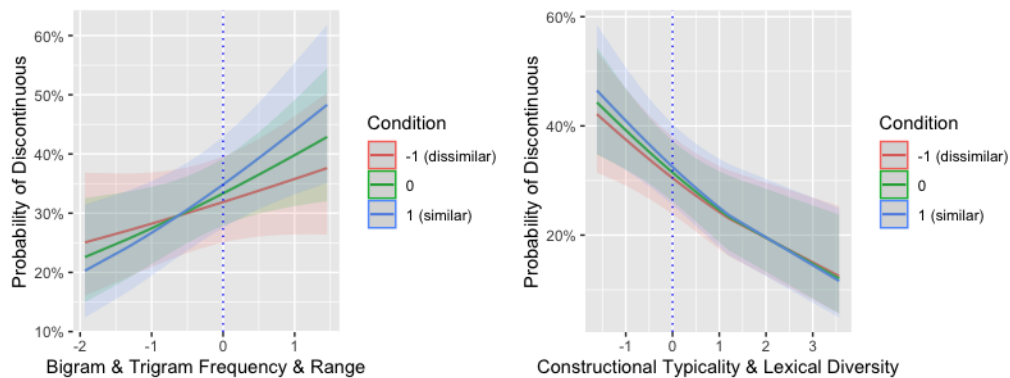| Sentence Type	Alternation	Example Sentence
Phrasal Verbs	Continuous	*She set off the alarm.*
	Discontinuous	*She set the alarm off.*
Datives	Double Object	*The teacher gave the student the exam.*
	Prepositional	*The teacher gave the exam to the student.*

Figure 1. The probability of a discontinuous construction PV response based on (a) RC2 and (b) RC3.
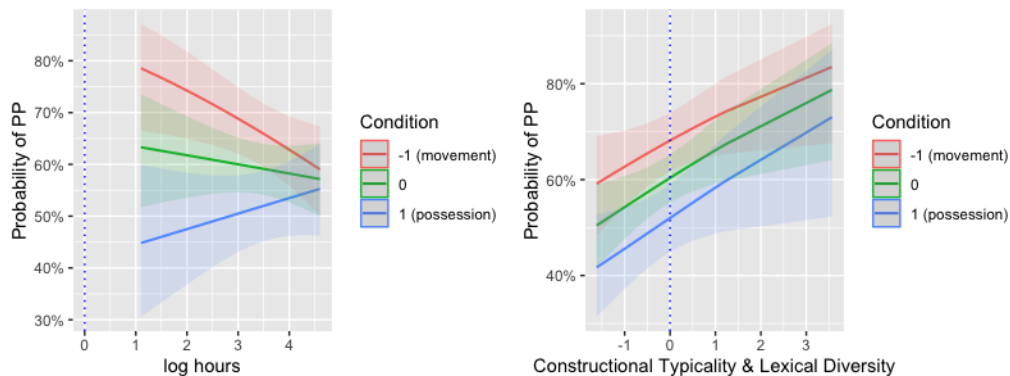

Figure 2. The probability of a prepositional dative response based on (a) log hours spent reading per week and (b) RC3.

References: [1] MacWhinney & Bates, 1989. Functionalism and the Competition Model [2] MacDonald, Pearlmutter, & Seidenberg, 1994. Perspectives on sentence processing [3] Levy, 2008. *Cognition* [4] Gries, 1999. *Cognitive Linguistics* [5] Krifka, 2003. *Korean Journal of English Language and Linguistics.* [6] Crossley, Kyle, & Dascalu, 2019. *Behavioral Research Methods* [7] Kyle, 2016. Measuring syntactic development in *L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Doctoral Dissertation) [8] Kyle, Crossley, & Berger, 2018. *Behavior Research Methods*