

Mixed-category effects on syntactic alternation

Many important new insights have emerged in recent years from the nascent study of the role of gradient grammatical constraints in patterns of spontaneous language use. To name a few: hard constraints in one language may emerge as soft constraints in another (Bresnan et al 2001); syntactic choices in usage cannot be reduced to lexical biases (Bresnan et al 2007; Jaeger 2006); and the soft constraints that can be inferred from spontaneous usage patterns are reflected in elicited acceptability judgments (Bresnan 2006). These findings only touch upon the range of questions raised by the gradient perspective on grammar. Here we use a little-studied alternation – *needs to be done* [the ToBE form] ~ *needs doing* [the -ING form] – to investigate how the mixed-category status of an alternant affects usage patterns. This alternation is examined because the gerund in the -ING alternant is a mixed verbal/nominal category (Malouf 2000), while the past participle in the ToBE alternant has no nominal characteristics. We find that speaker choice in this alternation is driven by gradient, not categorical, constraints, and that speakers are aware of the items' mixed-category status when choosing an alternant.

First, we demonstrate that previous claims about the semantic non-equivalence of the alternants are incorrect. We consider four proposals that semantic factors drive the alternation: one of our own, and three attributed to Lynne Murphy and discussed in Murray et al 1996. Murphy proposed that -ING requires the subject to already exist (cf. (1)) and to benefit from the action (2), and ToBE requires that the subject's possessor be the agent of the needed action (3). We proposed that achievement and state verbs (Vendler 1957) require ToBE (cf. (4)) because these verbs do not have a progressive form. However, we find counter-examples to all these proposals in the British National Corpus (BNC):

- (1) ...I believe that [the project] *needs **building***.
- (2) The recipe given here is a quick-to-make fondant that *does not need **cooking***.
- (3) ...their implicit optimism *needs to be **tempered*** by the realities of this kind of economic growth...
- (4) A continual drop in the pond level [...] can be attributed to a leak, which will *need **finding*** and sealing.

As categorical constraints do not account for speaker choice in this alternation, we employ a quantitative-constraint framework: we sampled and hand-annotated 1004 examples of the alternation from the BNC, using retrospective sampling to include equal numbers of each alternant. We then fit a mixed-effects logistic regression model to this data, treating verb (i.e. *build* in (1)) as a random effect. Treating the verb as a random effect accounts for the possibility that different verbs have idiosyncratic preferences for one alternant over the other. We find significant effects of verb length (in syllables) and frequency, quantity of post-verbal material (favoring ToBE), concreteness and definiteness of the subject, and use of the construction in a subject-extracted relative clause (favoring -ING).

We focus on the theoretical relevance of the effect of post-verbal material by considering environment prototypicality. The -ING form is a mixed verbal/nominal category, while the ToBE form lacks nominal characteristics. Thus we hypothesize that in a more prototypically nominal environment, the partially nominal -ING form will be more likely to occur. We use structural bias from the post-verbal phrase XP to estimate this prototypicality effect; structural bias is the ratio of probability of seeing an XP post-modifier in an NP to the probability seeing an XP post-modifier in a VP. Structural bias is tightly correlated with post-verbal material quantity ($\rho=-0.91$) and, when introduced into the logistic regression model, decreases the effect of post-verbal material quantity in the model to marginal significance, suggesting that even after accounting for a variety of other factors, structural bias affects speaker choice. We conclude that speaker choice in this alternation is driven by gradient factors and that speakers are aware of mixed-category status when choosing an alternant.

References

- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and Harald Baayen. 2007. "Predicting the Dative Alternation". In *Cognitive Foundations of Interpretation*, ed. by G. Boume, I. Kraemer, and J. Zwarts. Amsterdam: Royal Netherlands Academy of Science, pp. 69-94.
- Bresnan, Joan, Shipra Dingare, and Christopher D. Manning. 2001. "Soft Constraints Mirror Hard Constraints: Voice and Person in English and Lummi". *Proceedings of the LFG '01 Conference*. CSLI Publications.
- Bresnan, Joan. 2006. "Is syntactic knowledge probabilistic? Experiments with the English dative alternation." Invited paper to appear in the proceedings of the International Conference on Linguistic Evidence, Tuebingen, 2-4 February 2006, *Roots: Linguistics in search of its evidential base..* Series: Studies in Generative Grammar, edited by Sam Featherston and Wolfgang Sternefeld. Berlin: Mouton de Gruyter.
- Jaeger, T.F. 2006. *Redundancy and Syntactic Reduction in Spontaneous Speech*. Ph.D. thesis, Stanford University.
- Malouf, Robert. 2000. *Mixed Categories in the Hierarchical Lexicon*. CSLI Publications, Stanford.
- Murray, Thomas, Timothy Frazer, and Beth Lee Simon (1996) Need + Past Participle in American English. *American Speech* 71(3):255-271.
- Vendler, Zeno. 1957. Verbs and Times. *The Philosophical Review*. 66(2):143-160.