Syntactic "Optionality" Reflects Performance rather than Competence

Spencer Caplan (University of Pennsylvania)

Introduction: Cognitive biases for so-called "communicative efficiency" have received popular attention (Jaeger 2011, Piantadosi et al. 2011, etc.) as instrumental in driving patterns of linguistic use and mental representation. On this view, probabilistic usage of syntactic alternations (as in 1) is representative of probabilistic grammatical knowledge (Bresnan, 2007; Bybee, 2006), and usage patterns reflect the "optimal" ordering under an information-theoretic framework (Jaeger, 2010).

Here we argue that such accounts need to be re-evaluated. An explicit psycholinguistic performance model, the "incremental" language production framework (Bock and Levelt, 2002), offers a better fit and causal explanation of empirical syntactic-optionality data compared with probabilistic or information-theoretic accounts of linguistic representation (Jaeger, 2010).

Previous Models: A prominent previous account, the "Uniform Information Density" hypothesis (UID) (Jaeger, 2010), proposes that syntactic optionality is driven by a speaker's implicit managing of computable information content to maximize communicative efficiency. On this model, conditional probability serves as a proxy for optimized information content. Previous corpus modeling of optional "that"-omission has supported UID (Jaeger, 2010). However, such evidence is potentially problematic. There are syntactic complications: not all cases of embedded complements are optional with respect to "that"-mentioning (Grimshaw, 2009). Sentential subjects, as well as factives with "it" as objects, require the presence of the complementizer (2-3).

(1) The coach knew (that) the players were tired
(2) The committee hated (it) **that**/*Ø the new students failed the exam
(3) **That**/*Ø the new students failed the exam annoyed the committee

Additionally, rates of "that"-omission show a great deal of variability by genre, ranging from 1% in formal writing to 85% in conversational speech (Biber, 1999). With so much variance attributable to sociolinguistic register, it is not clear what we might learn about the cognitive architecture of the production system.

Data: We propose the English verb-particle construction (e.g. 4-5) as a better case to evaluate theories of optionality.

(4) John picked up the book
(5) John picked the book up

We extracted a large database of verb-particle alternations from COCA, a balanced corpus of modern English (Davies, 2009). Verb-particles are a good test case to evaluate as they are subject to fewer syntactico-semantic confounds compared with "that"-omission and only show limited variation by register (less than 20% of variance). Our extracted data includes approximately 60,000 sentences including 500 verb-particle pairs. Such "big data" is important if we wish to test fine-grained predictions, particularly involving low-frequency predicates or rare interactions between conditions. Compare this to 3,000 instances in Bresnan (2007) and 7,000 in Jaeger (2010). From these data we can test the information-theoretic UID account (Jaeger, 2010) against the framework of Incremental Generation (IG) (Bock and Levelt, 2002). Under IG, the architecture of sentence generation requires several components: retrieve lemmas from memory, assign such elements their proper functional roles, as well as assign linear order in adherence with syntactic restrictions. If we assume such modules operate incrementally and in parallel, then variations in the order in which information is delivered from one component to the next can readily affect the linear order that elements appear in speech. So long as the system does not intentionally hold retrieved lemmas back in a buffer, any factors which speed up lexical access will also be proxies for spoken linear order (see Rayner, 1998 for a review of lexical access factors). While both UID and IG make convergent predictions regarding conditional probability, only IG predicts that the factors of frequency, definiteness, and constituent length, etc. should all predict linear order in optional constructions. This is because such factors empirically correlate with lexical retrieval times (Rayner, 1998) yet are orthogonal to "information density".

Results: Following Bresnan (2007) and Jaeger (2010), a multilevel logit model was used to evaluate predictions of an IG account of optionality compared with UID. The dependent variable was the binary outcome of linear order (particle-first rather than object-first) in verb-particle sentences. Evaluating over the entire database, we see a strong correlation between all IG-related factors and output order (Blue in Table). For example, as the frequency of the object increases, that speeds up lexical retrieval of that object. The negative coefficient indicates that this makes it more likely to be linearized before the particle. Since predictability (labeled in the table as "Info") is a convergent prediction of UID and IG, the dataset taken as a whole does not provide direct negative evidence against either account.

However, when limited to evaluation over even moderately long objects (at least four words), then the predictions of UID are not borne out while the other instantiations of IG remain significant (Green in Table). To whatever degree we can characterize the output of the language production system as "optimal" in information ordering, this is an emergent property of a mechanistic performance system (IG rather than a reflection of explicit optimization, cognitive biases, or probabilistic representation by speakers. Syntactic "optionality" reflects performance rather than competence.

	All particle cases (58,619 cases)				**Longer object particles (5,679 cases)**			
Factor	**Estimate**	**Std. Err**	**Z-Value**	**P**	**Estimate**	**Std. Err**	**Z-Value**	**P**
Freq(obj)	**-282.8**	**38.12**	**-7.42**	**~0**	**-502**	**77.9**	**-2.84**	**0.01**
Info(obj|verb)	**0.039**	**0.01**	**8.6**	**~0**	**0.02**	**0.03**	**0.69**	**0.49**
NP-length	**1.0**	**0.03**	**40.37**	**~0**	**0.63**	**0.12**	**5.18**	**~0**
Definite-Obj	**-0.59**	**0.02**	**-26.9**	**~0**	**-1.13**	**0.15**	**-7.33**	**~0**

Selected References:

Biber, D. (1999). A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse studies*, 1(2), 131-150.

Bock, K., & Levelt, W. (2002). Language production. *Psycholinguistics: Critical concepts in psychology*, 5, 405.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. *Cognitive foundations of interpretation*, 69-94.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711-733.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistics*, 14(2), 159-190.

Grimshaw, J. (2009). That's nothing: the grammar of complementizer omission. *Lecture presented at: CUNY Graduate Center and SUNY Stonybrook*.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, 61(1), 23-62.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526-3529.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372.