Is knowledge of syntax probabilistic? Experiments with the English dative alternation*

Joan Bresnan
Stanford University
bresnan@stanford.edu

Is knowledge of syntax probabilistic? In recent work Bresnan, Cueni, Nikitina, and Baayen (2005) have shown that a conditional logistic regression model of the English dative alternation predicts 94% of the alternative dative syntax choices made by dialogue participants during telephone conversations. The model assigns a probability to each observation in previously unseen data as a function of multiple variables, including semantic class of the verb, discourse accessibility, constituent weight, animacy, definiteness, and pronominality of the arguments. The observations all contain alternating dative verbs, which are used in both prepositional dative constructions like *He brought the pony to my children* and double object constructions like *He brought my children the pony*. Thus, the model does not classify observations as grammatical or ungrammatical, but predicts the choice of alternative paraphrases which both have probability greater than zero.

Bresnan et al. (in press) show that the model generalizes beyond the contingencies of the particular collection of telephone conversations and predicts statistical differences in a very different written corpus of edited reportage. This result raises the question of whether such models could represent the implicit knowledge of English language users (which also generalizes across very different processing tasks). Could linguistic competence itself have a probabilistic, predictive capacity in the area of syntax? If so, language users could theoretically predict the dative syntax choices that dialogue participants make, as a function of the same variables. Where the model predicts a single outcome with high accuracy, subjects should be able to do so, and where the model predicts a single outcome with low accuracy (underdetermining dative syntax choices), subjects should also do so.

Figure 1 shows the model probabilities of a prepositional dative construction for a small random sample of one hundred observations of the alternating verbs from the

*Thanks to Jeff Elman and Steve Pinker for suggesting that I follow up my previous corpus studies with experiments along the lines of the present study, and to Marilyn Ford and Anette Rosenbach for helpful discussions of experimental design and interpretation. None of them can be blamed for the use I have made of their advice. Thanks also to Nick Romero for creative internet sampling and research assistance.

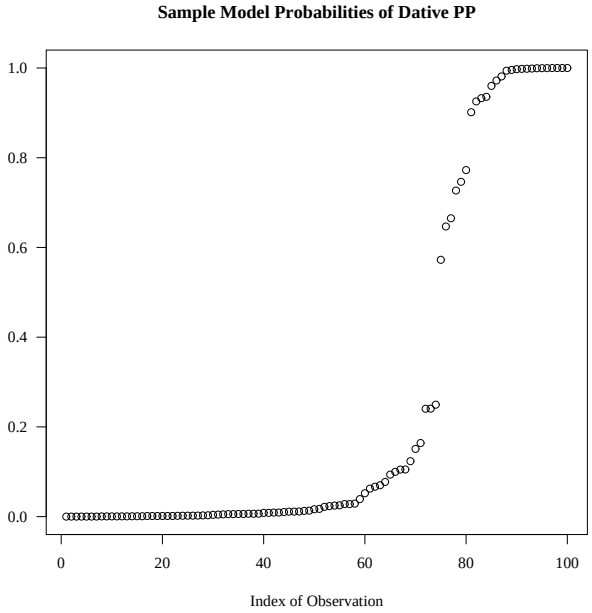

Figure 1: Sample probabilities from the corpus model of Bresnan et al.

corpus dataset. The data points at the lower and upper extremes of the *y* axis scale have probabilities of being a prepositional construction near 0 and 1, respectively. (Low probability of being a prepositional construction is equivalent to high probability of being a double object construction in this model.) The data points in the middle are cases where each alternative has substantial probability. The hypothesis is that language users' ratings of the naturalness of alternative dative constructions should correspond to the predicted model probabilities.

The hypothesis was investigated in Experiment 1, a forced-choice task eliciting scalar ratings of alternative dative constructions as continuations of natural dialogue contexts (cf. Rosenbach 2003). The contexts were included because discourse accessibility and animacy of recipient referents are significant predictors in the corpus study. The experimental materials were sampled from the dative corpus data in five equal probability bins defined by the corpus model, ranging from very low probability of being a prepositional dative to very high. For each sampled construction, an alternative paraphrase was constructed, and both were presented as choices in the original dialogue context, which was edited for readability by shortening and by removing disfluencies. Items were pseudo-randomized and construction choices were alternated to make up a questionnaire. The subjects were paid Stanford undergraduate students of both genders who reported being monolingual English speakers having not taken a course in syntax. Each subject received the same questionnaire, with items and construction choices in the same orders.

Subjects were asked to rate the naturalness of the alternatives in the given context by distributing 100 rating points over the two alternatives in accordance with their own intuitions. Any pair of scores summing to 100 was permitted, including 0–100, 63–27, 50–50, etc. The following is a sample item:

```
----------------------------
Speaker:

About twenty-five, twenty-six years ago, my brother-in-law
showed up in my front yard pulling a trailer. And in this
trailer he had a pony, which I didn't know he was bringing.
And so over the weekend I had to go out and find some wood
and put up some kind of a structure to house that pony,

 (1) because he brought the pony to my children.

 (2) because he brought my children the pony.
----------------------------
```

Plots of the data show that subjects' scores of the naturalness of the alternative syntactic paraphrases appear to correlate with the corpus model probabilities. In Figure 2 each panel shows a single subject's mean scores for the items in each corpus probability bin. All of the subjects' mean ratings of items from the lowest probability bin are below their mean ratings of items from the highest probability bin. The ratings of items from the middle bins show the most variation within and across subjects, as expected from the original corpus model probabilities (Figure 1).

The results were analyzed using a linear mixed effects regression model (Pinheiro and Bates 2000, Baayen 2004), which fit the scores using adjustments for both subject and verb sense as random effects and adjustments for fixed effects conditioned on the random effects. All of the fixed effects are significant (pronoun theme, given theme, definite recipient, given recipient, and parallel PP dative in dialogue, $p < 0.0001$; animate recipient, $p < 0.001$; length differential of recipient and theme on a log scale, $p < 0.01$). Two insignificant effects—the order of items and the order of construction choice—were eliminated from the model because their coefficients were less than their standard errors.

The fit of the model ($R^2 = 0.61$) is displayed in Figure 3, a trellis graph with nonparametric smoothers to facilitate visualization of the data (Cleveland 1979). Each panel of the trellis plot is a scatterplot of the data from a single subject, showing all thirty scores (represented on the y axis) plotted against the fitted model values (represented on the x axis). A roughly linear relation appears in each panel.

These results show that subjects' scores of the naturalness of the alternative syntactic paraphrases correlate very well with the corpus model probabilities and can be

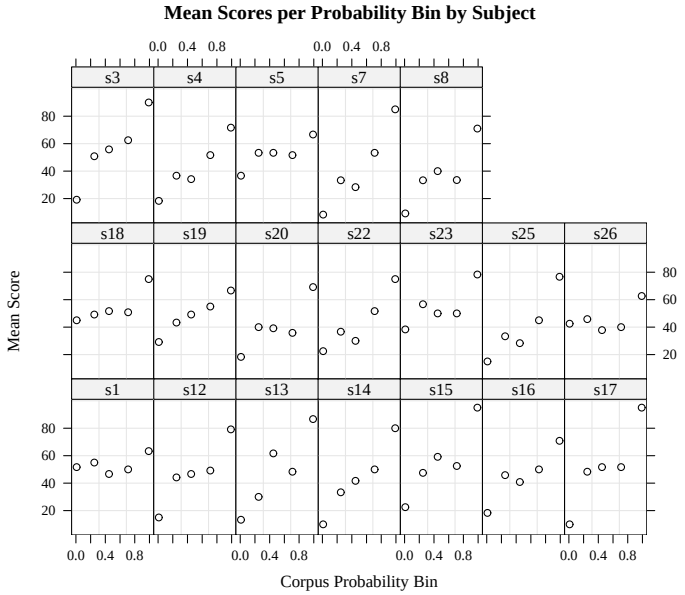

Figure 2: Experiment 1 Scores by Probability Bin

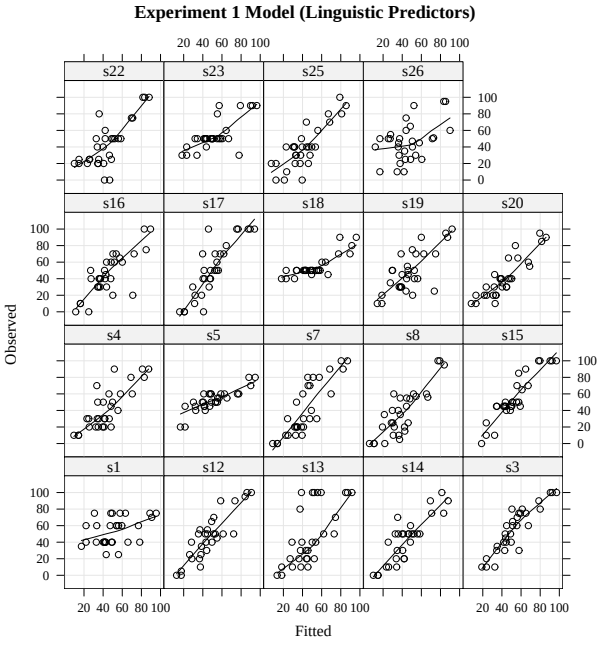

Figure 3: Fit of linear mixed effects model to Experiment 1 scores

substantially explained as a function of the same linguistic predictors as the original corpus model. In fact, the subjects' choices, which were made according to their own intuitions, reliably tend to pick out the same choices made by the original dialogue participants in the corpus transcriptions. They were in effect predicting with better than chance accuracy what the speaker would say.

Probabilistic knowledge of syntax is also indicated by certain mismatches between grammaticality judgments and language use. Dative verbs reported by linguists to be ungrammatical in the double object construction are nevertheless found in actual usage on the World Wide Web (Bresnan and Nikitina 2003, Bresnan et al. in press). Although we lack specific frequency measures for all of the relevant verbs, we know that differing alternation classes of dative verbs correspond to differing frequencies of use in internet samples (Lapata 1999), and that different argument types are more likely to occur in different syntactic positions following dative verbs (Thompson 1990, Collins 1995, Bresnan et al. in press). In particular, double object constructions in which a pronoun precedes a lexical NP are far more frequent than those in which two lexical NPs occur, and it is in the more frequent contexts that reportedly nonalternating dative verbs can most readily be found in actual use. For example, *whisper* is reported to be ungrammatical in the double object construction, but Google queries yield *whisper me the answer*, along with *whisper the password to the fat lady*. The reportedly ungramatical examples constructed by linguists tend to utilize the far less frequent positionings of argument types, like *whisper the fat lady the answer*.

Experiment 2 used the same task as Experiment 1 to investigate how subjects rate the reportedly ungrammatical dative constructions when they occur in the most frequent context types. Six alternating and eight reportedly nonalternating verbs were sampled from the internet, half belonging to communication semantic classes ('manner' type – *whisper, mutter, mumble, yell* vs. 'instrument' type – *phone, text, IM*) and half to motional transfer semantic classes ('continuous' type –*carry, push, drag, lower* vs. 'instantaneous' type – *flip, throw, toss*) (Pinker 1989). All of the verbs were sampled in the construction types found to be most frequent in corpus studies—the double object construction with pronoun recipient preceding lexical NP theme or the dative construction with a lexical NP as prepositional object. The same method of creating a questionnaire was used as in Experiment 1. The following is a sample item for the reportedly nonalternating verb *push*.

```
-----------------------------------
Money in the pot is dead money. It does not belong
to anyone until the hand is over

 (1) and the dealer pushes someone the pot.

 (2) and the dealer pushes the pot to someone.
-----------------------------------
```

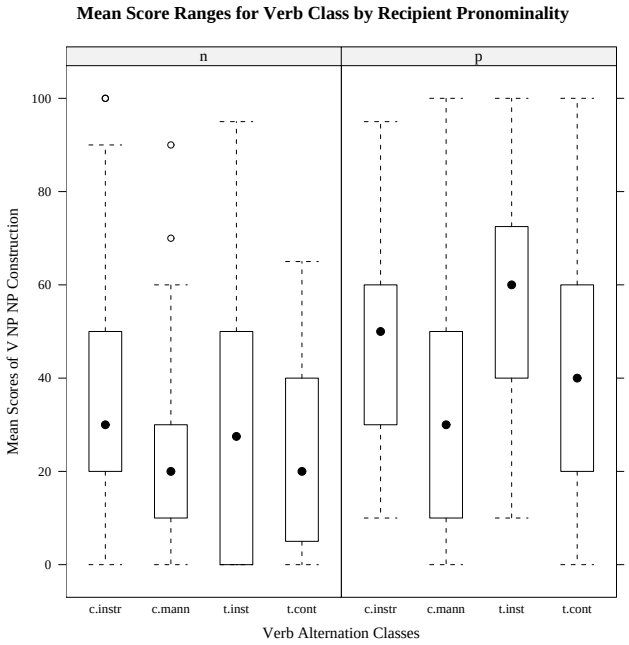

Figure 4: Experiment 2 Scores

The subjects were twenty paid Stanford undergraduates of both genders who reported that they were monolingual and had not taken a syntax course. They were given the same forced-choice scalar rating task as in Experiment 1.

A plot of the data in Figure 4 shows that the ranges of mean verb class scores of the double object constructions across subjects differ by both semantic type and pronominality of recipient. The black dots designate the middles of the ranges of mean scores in each category, the boxes are the interquartile ranges, and circled points falling outside of the dashed lines are outliers. Panel 'n' on the left represents the less frequent type in which both objects are lexical NPs; panel 'p' on the right represents the very frequent type in which a pronoun object precedes a lexical NP object. The columns represent the verb classes: in each panel, the first and third classes are alternating (instrument of communication 'c.instr' and transfer by instantaneous motion 't.inst', respectively), while the second and fourth are reportedly nonalternating (manner of communication 'c.mann' and transfer by continuous motion 't.cont', respectively).

In Figure 4 the four black dots in the panel on the right are higher than those on the left. This means that more subjects, on average, rated double object constructions of the more frequent argument type (pronoun precedes lexical NP) better than the less frequent type (two lexical NP objects), regardless of verb class. Further, in both panels the black dots in the first and third columns are higher than those in the second and fourth columns. Thus controlling for pronominality of recipient, the alternating classes are rated higher than the reportedly nonalternating classes by more of the subjects.

	Estimate	Std. Error	DF	t value	Pr($>\|t\|$)	
semantic class = c.instr	26.27	5.51	594	4.77	2.375e-06	***
semantic class = c.mann	14.49	4.65	594	3.12	0.001917	**
semantic class = t.inst	31.24	5.84	594	5.34	1.296e-07	***
semantic class = t.cont	21.30	4.93	594	4.32	1.804e-05	***
recipient = pronoun	13.88	3.78	594	3.67	0.000265	***
item order	0.43	0.16	594	2.74	0.006294	**

Significance codes: 0 '***' 0.001 '**'

Table 1: Model Coefficients for Experiment 2

Strikingly, the black dots in the second and fourth columns in the right panel are higher than the those in the first and third columns of the left panel. This means that more subjects rated the reportedly ungrammatical verb classes better in the frequent context than rate the grammatical verb classes in the infrequent context. (The latter are supposed to be fully grammatical by definition as alternating verbs.) In other words, the contextual information and relative frequency of argument types overrides and reverses linguists' reported classifications of relative grammaticality.

To analyze the significance of the results, we fit a linear mixed effects regression model in which both verb and subject are random effects and the fixed effects are pronominality of recipient, semantic class, and item order. We also modeled an interaction between the random effect of verb and pronominality of recipient. Finally, the fixed effect intercept was removed to allow the semantic class coefficients to be interpreted as mean rating values over the population of subjects rather than as contrasts to a reference semantic class value.

Table 1 shows that all of the fixed effects are significant (as is the random interaction between verb and pronominal recipient). (Construction order was not significant and was dropped from the final model because its coefficient was less than its standard error.) Note that the model estimates adjust for the effect of item order.

The model provides us with quantitative evidence for the significance of the relations observed in the plot in Figure 4. In this model a higher coefficient corresponds to a higher mean score for the double object construction across subjects. Thus, having a pronominal recipient raises the mean score of the double object construction for all subjects by nearly 14 points, compared to having a lexical NP recipient. The coefficients of the reportedly non-alternating classes (c.mann and t.cont) are lower than those of the alternating classes (c.instr and t.inst), meaning that they are rated as less natural in the double object construction. Notice, however, that when we add the pronominal recipient coefficient to the nonalternating semantic class coefficients, the resulting values exceed or approximate (within the standard error) the alternating class coefficients with two lexical NP objects. Thus the Figure 4 observations are significant when

taking subject, verb, verb-pronoun interactions, and item order into account.

In sum, the present study provides evidence that syntactic competence has a probabilistic and predictive capacity that generalizes across very different processing tasks including speaking, writing, and reading-based judgments. It further suggests that grammaticality judgments based on subtle contrasts between examples may actually reflect implicit knowledge of syntactic probabilities.

References

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. Mental Lexicon Working Papers, Edmonton, **1**:1–45.

Bresnan, J., A. Cueni, T. Nikitina, and R. H. Baayen (in press). Predicting the dative alternation. In Proceedings of the Royal Netherlands Academy of Science Colloquium on Foundations of Interpretation, October 28-28, 2004.

Bresnan, J. and T. Nikitina (2003). On the gradience of the dative alternation. On-line, Stanford University: http://www-lfg.stanford.edu/bresnan/download.html.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association, **74**:829–836.

Collins, P. (1995). The indirect object construction in English: an informational approach. Linguistics, **33**:35–49.

Lapata, M. (1999). Acquiring lexical generalizations from corpora: A case study for diathesis alternations. In Proceedings of the 37th Meeting of the North American Chapter of the Association for Computational Linguistics, pp. 397–404. College Park, Maryland.

Pinheiro, J. C. and D. M. Bates (2000). Mixed-Effects Models in S and S-PLUS. Springer, New York.

Pinker, S. (1989). Learnability and Cognition: The Acquisition of Argument Structure. The MIT Press, Cambridge, Mass.

Rosenbach, A. (2003). Aspects of iconicity and economy in the choice between the *s*-genitive and the *of*-genitive in english. In G. Rohdenburg and B. Mondorf, eds., Determinants of Grammatical Variation in English (Topics in English Linguistics/[TiEL], pp. 379–411. de Gruyter.

Thompson, S. A. (1990). Information flow and dative shift in English discourse. In J. A. Edmondson, C. Feagin, and P. Mühlhausler, eds., Development and Diversity: Language Variation across Time and Space, pp. 239–253. Summer Institue of Linguistics and University of Texas at Arlington, Dallas.