# Is knowledge of syntax probabilistic?

## EXPERIMENTS WITH THE ENGLISH DATIVE ALTERNATION

JOAN BRESNAN

# The Problem

Imagine you are a child...

- your message: something happened – the giving of toys to some children by a person named Susan

- your message: something happened – the giving of toys to some children by a person named Susan

- you construct incrementally:

- your message: something happened – the giving of toys to some children by a person named Susan

- you construct incrementally: *Susan*

- your message: something happened – the giving of toys to some children by a person named Susan

- you construct incrementally: *Susan gave*

- your message: something happened – the giving of toys to some children by a person named Susan

- you construct incrementally: *Susan gave* __

- two items from message could fill postverbal position: *children* and *toys*

- your message: something happened – the giving of toys to some children by a person named Susan

- you construct incrementally: *Susan gave* __

- two items from message could fill postverbal position: *children* and *toys*

- if *toys* is inserted first, a prepositional dative structure is eventually built: *Susan gave toys to the children*

- your message: something happened – the giving of toys to some children by a person named Susan

- you construct incrementally: *Susan gave* __

- two items from message could fill postverbal position: *children* and *toys*

- if *toys* is inserted first, a prepositional dative structure is eventually built: *Susan gave toys to the children*

- if *children* is inserted first, a double object structure is eventually built: *Susan gave the children toys*

- your message: something happened – the giving of toys to some children by a person named Susan

- you construct incrementally: *Susan gave __*

- two items from message could fill postverbal position: *children* and *toys*

- if *toys* is inserted first, a prepositional dative structure is eventually built: *Susan gave toys to the children*

- if *children* is inserted first, a double object structure is eventually built: *Susan gave the children toys*

- which item to select?

the prepositional dative structure:

*Susan gave* [*toys*] [*to the children*]    V NP PP

the prepositional dative structure:

*Susan gave* [*toys*] [*to the children*]     V NP PP

the double object structure:

*Susan gave* [*the children*] [*toys*]     V NP NP

the "dative PP":

*Susan gave* [*toys*] [*to the children*]          V NP PP

the "dative PP":

*Susan gave* [*toys*] [*to the children*]     V NP PP

the "dative NP":

*Susan gave* [*the children*] [*toys*]     V NP NP

the "theme" NP

*Susan gave* [*toys*] [*to the children*]    V NP PP

*Susan gave* [*the children*] [*toys*]    V NP NP

the "recipient" NP

*Susan gave* [*toys*] [*to the children*]　　　　V NP PP

*Susan gave* [*the children*] [*toys*]　　　　V NP NP

How do we determine which structure to choose?

A psychological approach. . .

Assume a parallel, incremental model of sentence formulation (Bock 1982, Levelt 1989, Bock and Levelt 1994, ao)

implemented in an interactive activation network (V. Ferreira 1996)

- message: a person named Susan gave toys to some children

- incremental construction: *Susan gave* __

- two items from message could fill postverbal position: *children* and *toys*

- *Susan gave toys ...*

- *Susan gave the children ...*

- which item to select?

Select the item with highest activation at the time the postverbal position is to be filled.

## Activation is increased by

- *discourse accessibility* (Bock and Irwin 1980, Prat-Sala and Branigan 2000)

- *animacy* (Bock, Loebell, and Morey 1992 )

- *effects of prior processing* (Bock 1986; Pickering, Branigan, and McLean 2002)

- etc

V. Ferreira: alternating dative verbs (*give*) allow faster processing than nonalternating (*donate*) in a production task.

A linguistic approach...

## Bresnan, Cueni, Nikitina, and Baayen in press:

- collect a database of 2360 instances of dative constructions from a three-million word corpus of telephone conversations in English

- manually annotate the data for multiple variables

- fit a mixed-effect logistic regression model to the data and evaluate the model on randomly selected subsets of training and testing data

# Variables annotated include:[a]

verbal meaning

discourse accessibility

relative complexity ($\sim$length)

pronominality

definiteness

animacy

structural parallelism

---

[a]Thompson 1990; Hawkins 1994; Collins 1995; Lapata 1999; Arnold et al 2000; Snyder 2003; Wasow 2002; Gries 2003

The model predicts the choice of construction for *give* and other 37 other dative verbs in spoken English with 94% accuracy

# Directions & magnitudes of effects in dative model
## (positive coefs $\Rightarrow$ V NP PP, negative $\Rightarrow$ V NP NP)

|  | Coefficient | Odds Ratio PP | 95% C.I. |
|---|---|---|---|
| nonpronominality of recipient | 1.73 | 5.67 | 3.25–9.89 |
| inanimacy of recipient | 1.53 | 5.62 | 2.08–10.29 |
| nongivenness of recipient | 1.45 | 4.28 | 2.42–7.59 |
| indefiniteness of recipient | 0.72 | 2.05 | 1.20–3.5 |
| plural number of theme | 0.72 | 2.06 | 1.37–3.11 |
| structural parallelism in dialogue | -1.13 | 0.32 | 0.23–0.46 |
| nongivenness of theme | -1.17 | 0.31 | 0.18–0.54 |
| length difference (log scale) | -1.16 | 0.31 | 0.25–0.4 |
| indefiniteness of theme | -1.74 | 0.18 | 0.11–0.28 |
| nonpronominality of theme | -2.17 | 0.11 | 0.07–0.19 |

Qualitative view of findings:

Harmonic alignment with syntactic position

discourse given ≻ not given

animate ≻ inanimate

definite ≻ indefinite

pronoun ≻ non-pronoun

less complex ≻ more complex

V NP NP

V NP PP

'Harmonic alignment' ∼ corpus frequency

Could these kinds of models represent language users' implicit knowledge of their language?

Does linguistic competence have a probabilistic, predictive capacity that weighs multiple information sources?

If a multivariable probabilistic model represents implicit knowledge of language, then language users could theoretically *predict what someone is going to say*, given a choice between two paraphrases in the same context.

*Can speakers assess the probability of construction choice as a function of the corpus model predictors?*
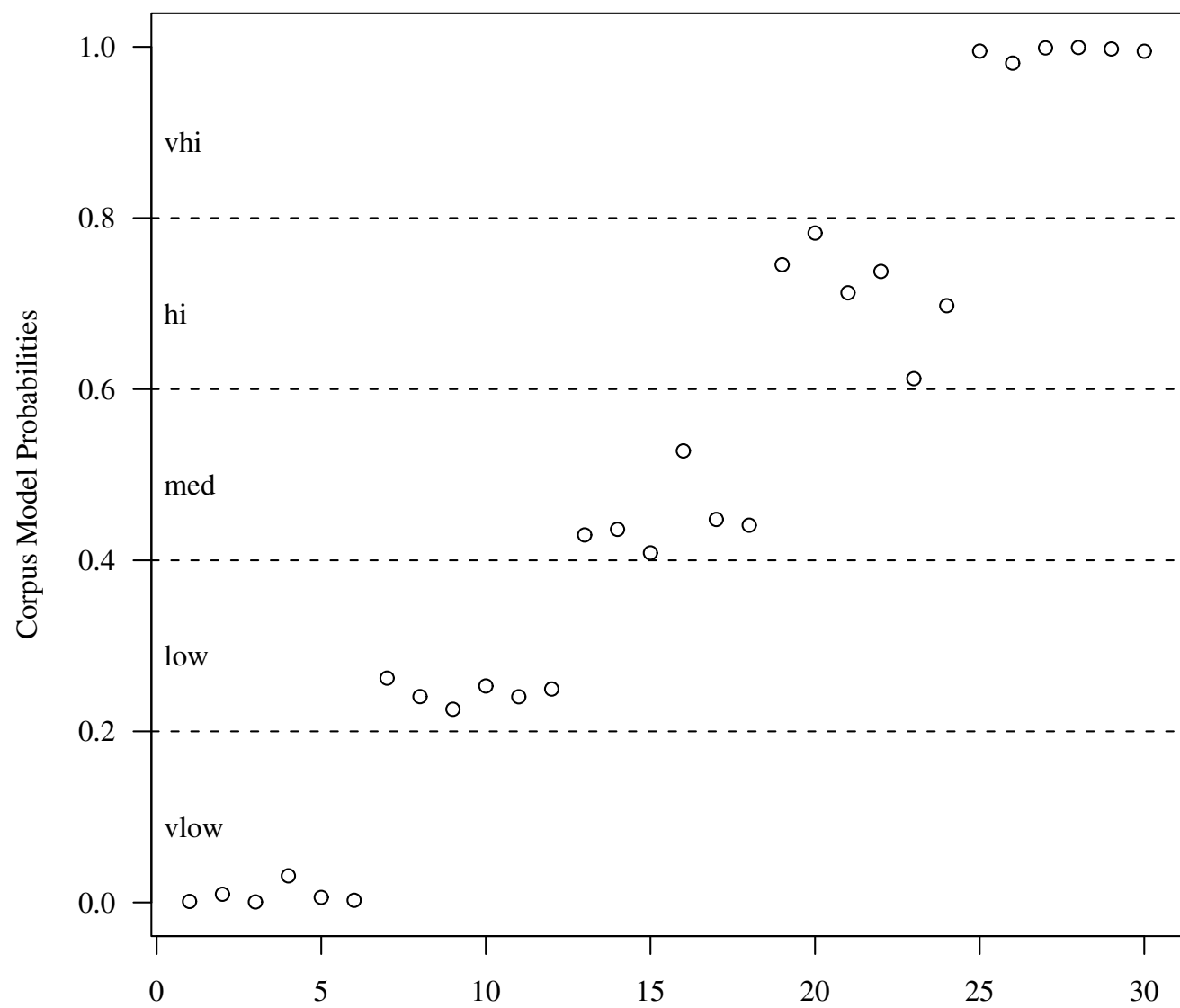
# Experiment 1

The dative corpus model

- defines a probability distribution over types of dative constructions

- as a function of givenness, pronominality, verb meaning in context, and other predictors.

**Sample Model Probabilities of Dative PP**

Index of Observation
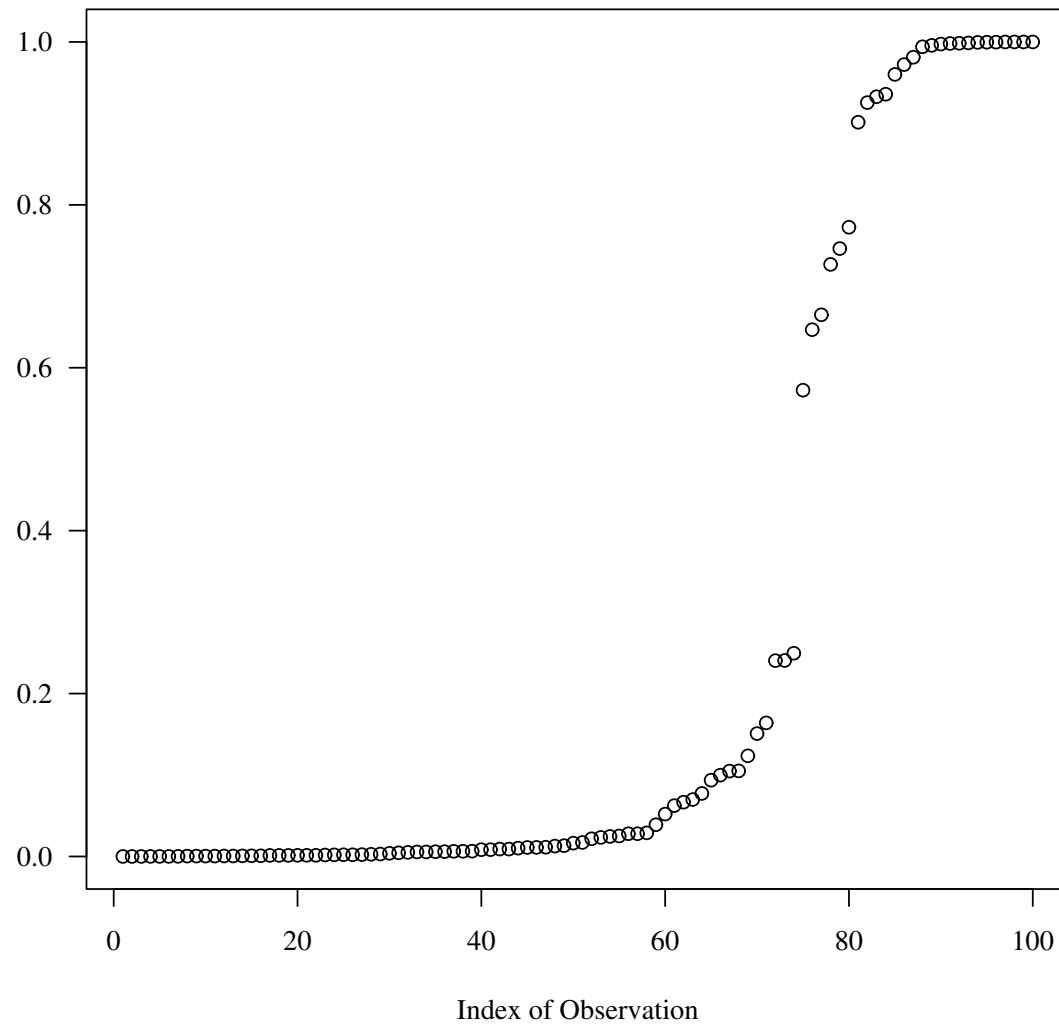
Where the model predicts high or low probabilities, subjects should also do so, and where the model predicts middle-range probabilities (underdetermining dative syntax choices), subjects should do so as well.

Thirty instances of dative constructions were randomly drawn from the centers of five probability bins of the dative corpus model distribution. (Potentially ambiguous items were replaced.)

The contexts of the sampled instances were retrieved from the full Switchboard corpus transcriptions and edited for readability by removing disfluencies and backchannelings.

An alternative to each target construction was constructed, the order of passages was randomized, and the order of target constructions alternated. A questionnaire was created containing the thirty passages.

## Sample passage:

_____

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.
_____

19 subjects from Stanford summer term undergraduates were recruited and paid.

The subjects were instructed to rate the relative naturalness of the alternatives in the given context passage, according to their own intuitions, on a scale of 0 to 100; the scores of the alternatives must sum to 100.

Mean Scores by Probability Bin
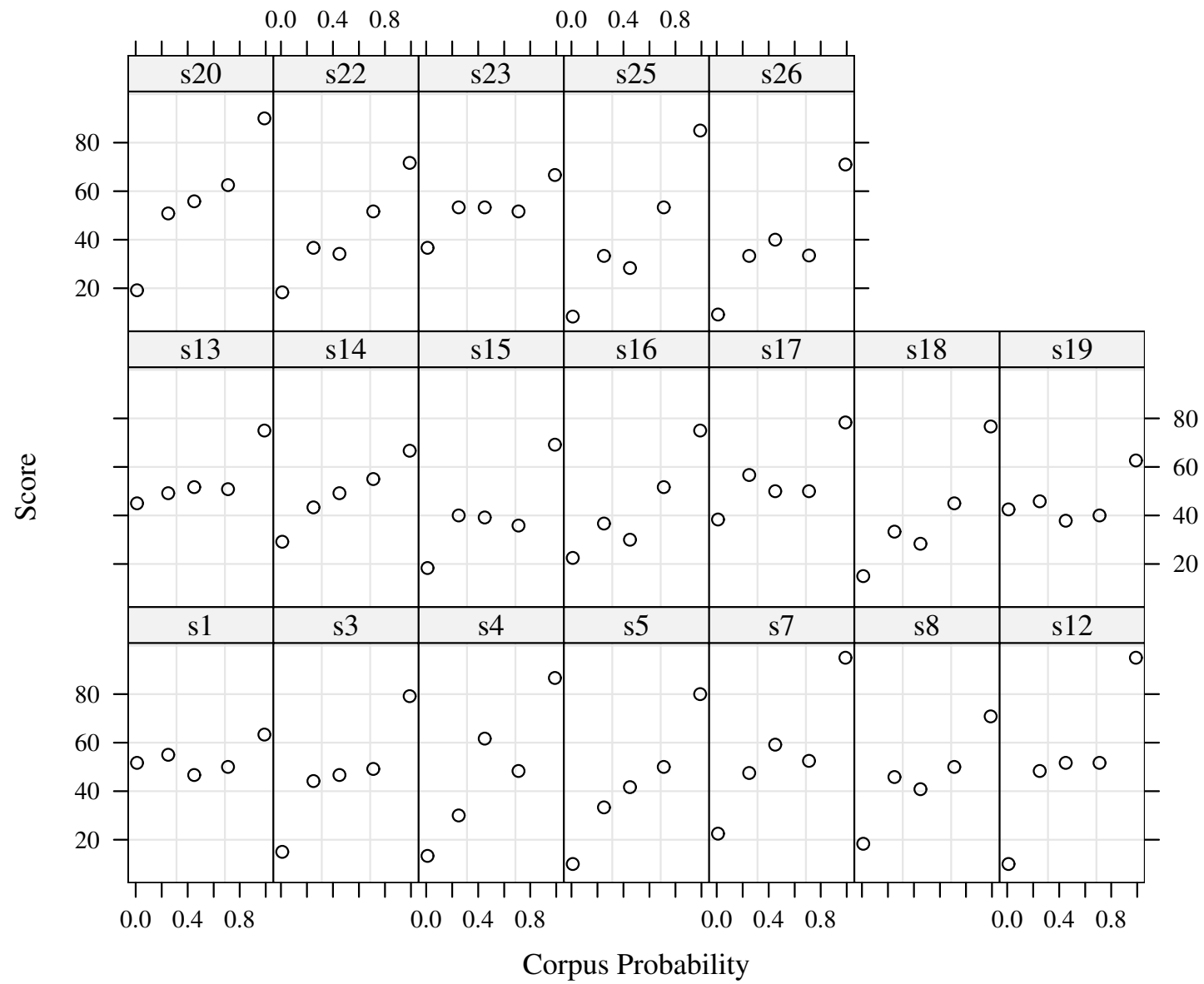
Every subject rated the PP alternatives from the vlow bin below those of the vhi bin.

The intermediate bins vary more across subjects, as expected from the dative corpus model probabilities, since these bins are where there is more variation in actual usage.

(The questionnaires of subjects who had taken a syntax course, as well as bilinguals and non-native speakers of English, were discarded.)

What explains the apparent positive correlations between subjects' ratings and corpus model probabilities?

Are the ratings a function of the same kinds of linguistic predictors used in the original dative corpus model or they the result of opportunistic strategies or heuristics?

We fit a mixed-effect linear regression model (Pinheiro and Bates 2000, Baayen 2004) to the data:

**fixed effects:** same as in Bresnan et al. model: givenness, pronominality, animacy, verbal semantics in context, etc.

**random effects:**

- an adjustment for each subject (representing that subject's individual bias toward PP datives

- an adjustment for each verb sense in its context (e.g. *give an armband* vs. *give your name*)

Model $R^2 = 0.61$

All fixed effects significant, $p < 0.0001$, except for length differential of theme and recipient ($p < 0.05$)

Insignificant effects eliminated from final model: order of items, order of constructions, verb lemma frequency (CELEX)

# Model Coefficients showing Harmonic Alignment

| | Estimate | S.E. | DF | t val | Pr(>\|t\|) | |
|---|---|---|---|---|---|---|
| (Intercept) | 73.19 | 12.93 | 560 | 5.66 | 2.422e-08 | *** |
| pron theme | 16.91 | 3.20 | 560 | 5.29 | 1.777e-07 | *** |
| indef theme | -12.48 | 2.59 | 560 | -4.81 | 1.928e-06 | *** |
| ngiv theme | -14.77 | 2.46 | 560 | -6.01 | 3.272e-09 | *** |
| pron rec | -22.47 | 5.47 | 560 | -4.11 | 4.595e-05 | *** |
| indef rec | 14.13 | 4.44 | 560 | 3.19 | 0.001526 | ** |
| ngiv rec | -9.00 | 5.31 | 560 | -1.69 | 0.091024 | . |
| inanim rec* | -29.48 | 6.93 | 560 | -4.25 | 2.493e-05 | *** |
| paral pp | 16.70 | 4.01 | 560 | 4.17 | 3.585e-05 | *** |
| diff len (log) | -4.77 | 2.34 | 560 | -2.04 | 0.041980 | * |

*Animacy: only 2 exx, abstract sense: *give something to the country, pay attention to that*

Scores as a Function of Model Linguistic Predictors
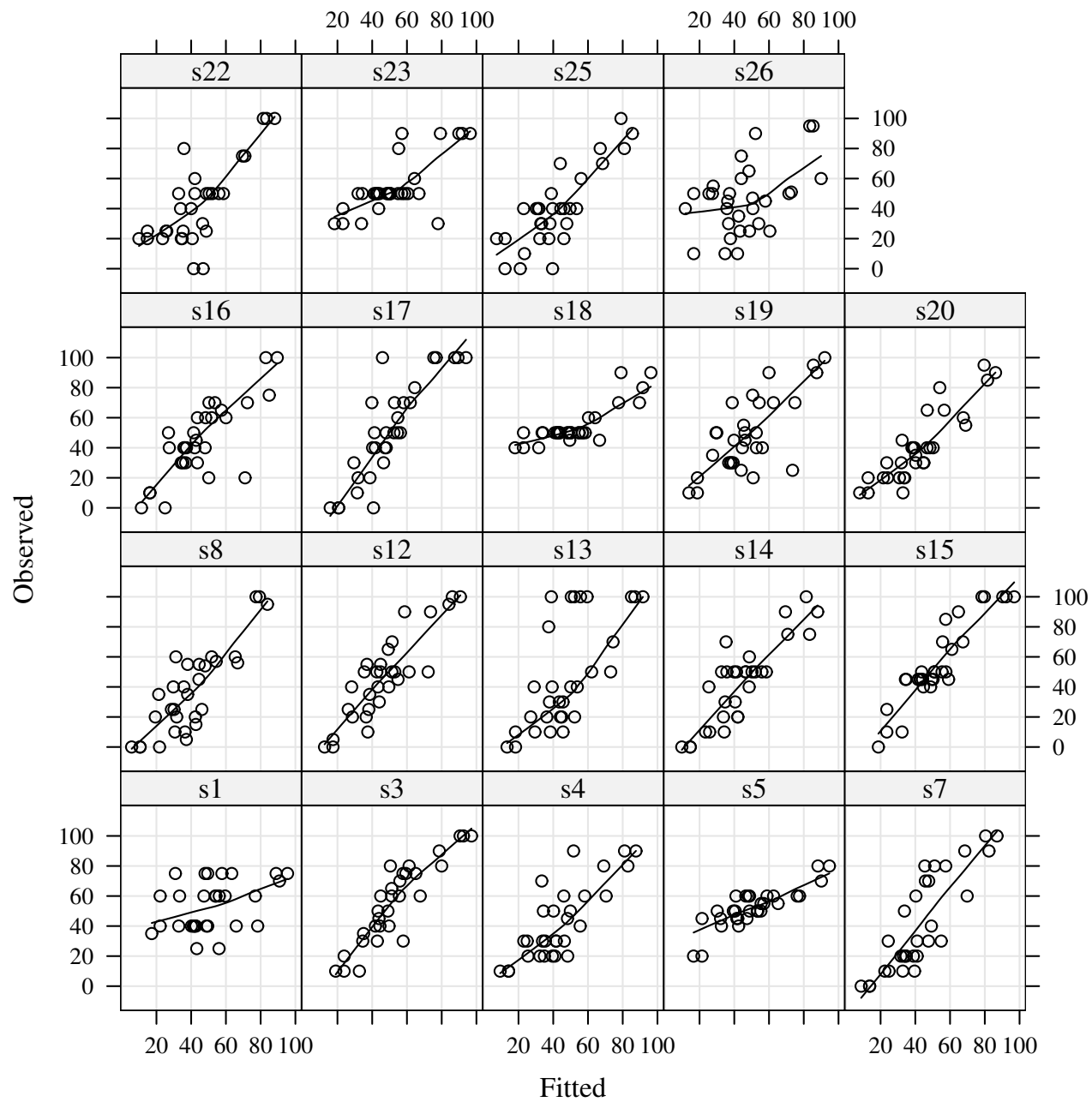
Interestingly, we can also compare each subject's ratings with the actual choices by the speakers in the original conversations. Baseline = 0.57.

Proportions of Subjects' Ratings
Favoring Actual Corpus Choices

0.63  0.83  0.80  0.70

0.80  0.80  0.67  0.77

0.73  0.83  0.80  0.77

0.80  0.77  0.77  0.73

0.73  0.87  0.67

Subjects' intuitions of syntactic probabilities are reliably more accurate than chance

(t = 13.4243, df = 18, p-value = 8.13e-11).

If linguistic competence has a probabilistic, predictive capacity that weighs multiple information sources, as Experiment 1 suggests, this could explain some puzzling mismatches between actual usage and generalizations based on grammaticality judgments.

What linguists report–

Verbs of continuous imparting of force impossible with double objects:

*I carried/pulled/pushed/schlepped/lifted/ lowered/hauled John the box.*

What is found in use (Bresnan and Nikitina 2003):

> *Karen spoke with Gretchen about the procedure for registering a complaint, and **hand-carried her a form**, but Gretchen never completed it.*

> *As Player A **pushed him the chips**, all hell broke loose at the table.*

What linguists report–

Manner-of-speaking verbs impossible with double objects:

*Susan whispered/yelled/mumbled/barked/ muttered Rachel the news.

What is found in use (Bresnan and Nikitina 2003):

*Shooting the Urasian a surprised look, she* **muttered him a hurried apology** *as well before skirting down the hall.*

*"Hi baby." Wade says as he stretches. You just* **mumble him an answer**. *You were comfy on that soft leather couch. Besides …*

What explains these mismatches?

We lack statistics for the specific examples, but we know:

Different alternation classes of dative verbs correspond to different frequencies of use in internet samples (Lapata 1999).

Different argument types are more frequent in certain complement positions of dative verbs (Thompson 1990, Collins 1995, Bresnan et al)

In particular —

$$V [\ldots \text{Pronoun} \ldots] \text{ NP}$$

*far more frequent in spoken English than*

$$V [\ldots \text{Noun} \ldots] \text{ NP}$$

(1530 vs. 178 in Switchboard corpus)

In the reportedly ungrammatical examples, linguists tend to use the less frequent positionings of argument types

# Experiment 2

14 verbs in 4 semantic classes were sampled from the internet together with the immediate syntactic and discourse contexts they occured in.

# Verbs used in Experiment 2

|  Communication |  |  Transfer |  |
| --- | --- | --- | --- |
| Alternating | Non-Alternating | Alternating | Non-Alternating |
| 'a_cm' | 'n_cm' | 'a_tr' | 'n_tr' |
| *phone* | *whisper* | *flip* | *carry* |
| *text* | *mutter* | *throw* | *push* |
| *IM* | *mumble* | *toss* | *drag* |
|  | *yell* |  | *lower* |

Each verb was sampled in the two most frequent argument type configurations:

```
V [...Pronoun...]  NP
```

and

```
V NP to [...Noun...]
```

(The data also included two instances of *someone* sampled in the prepositional dative construction and one instance of *someone* sampled in the double object construction.)

Using the same method as in Experiment 1, a natural discourse passage with alternative syntactic continuations was constructed for each item, and a questionnaire was created with the 28 passages (each of 14 verbs collected in two different naturally occuring constructions — V Pron NP and V NP to NP).

Examples –

*whisper me the price* $\Rightarrow$ *whisper the price to me*

*whisper the password to the fat lady* $\Rightarrow$
*whisper the fat lady the password*

*toss the ball to Worthy* $\Rightarrow$ *toss Worthy the ball*

*toss me the socks* $\Rightarrow$ *toss the socks to me*

# Syntactic contexts for each verb

V [...Pronoun...] NP      (sampled)

V NP to [...Pronoun...]      (constructed)

V NP to [...Noun...]      (sampled)

V [...Noun...] NP      (constructed)

## Sample item:

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

   (1) and the dealer pushes the pot to someone.

   (2) and the dealer pushes someone the pot.

20 subjects from Stanford summer term undergraduates were recruited and paid. (Subjects who had taken a syntax course were excluded, as well as bilinguals and non-native speakers of English.)

Subjects were given the same forced-choice scalar scoring task as in Experiment 1: to rate the naturalness of the examples in their context in accordance with their own intuitions.

Mean score ranges of V NP NP as a function of verb class and NP type
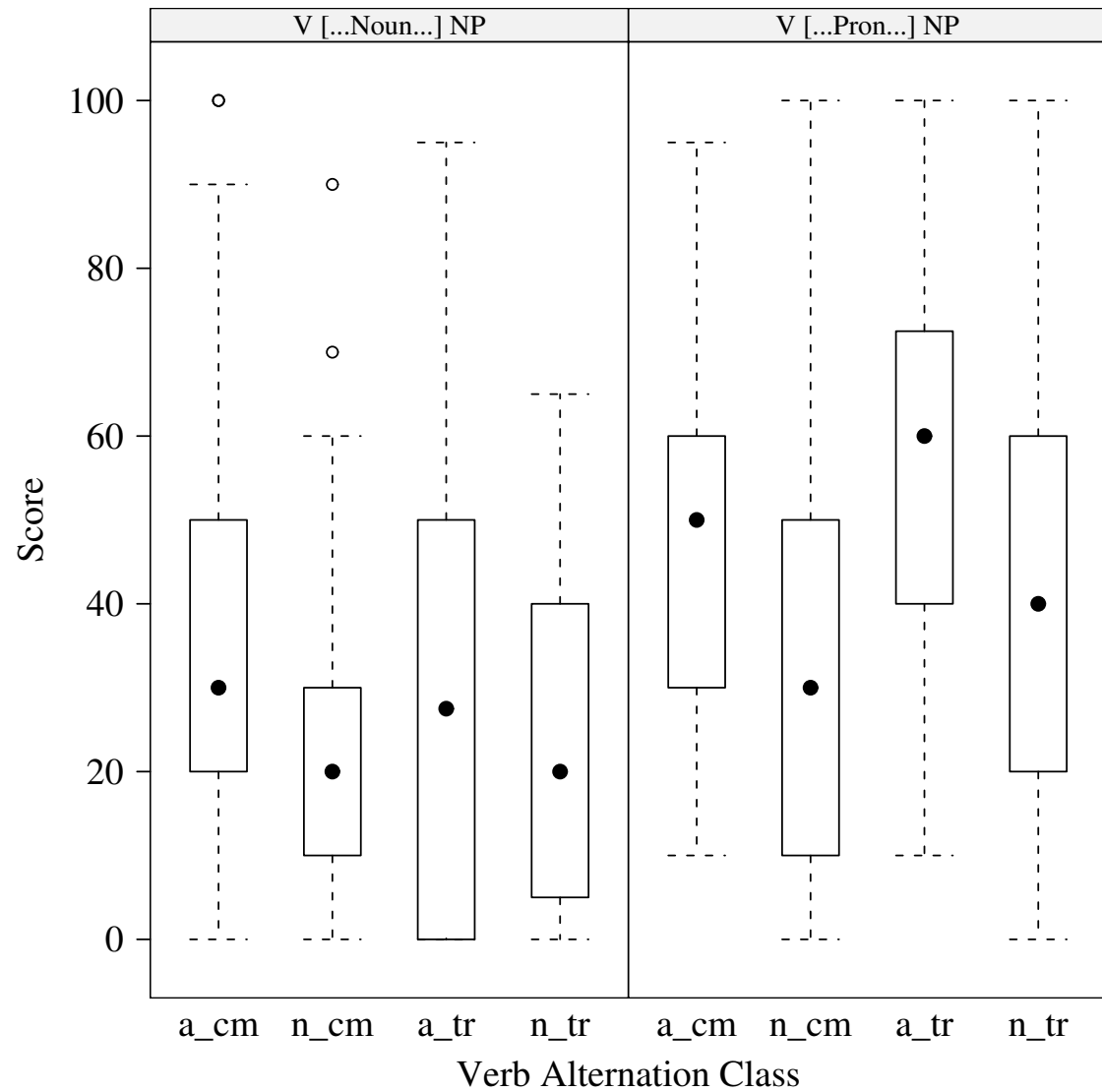
Strikingly, the reportedly ungrammatical verb classes are rated as highly or higher in the frequent context than the grammatical verb classes in the infrequent context. (The latter are supposed to be fully grammatical by definition as alternating verbs.)

To assess significance, we fit a mixed-effect linear regression model to the data:

**fixed effects:** semantic class, pronominality of recipient, and item order

**random effects:**

- an adjustment for each subject

- an adjustment for each verb

- an interaction between verb and pronominality of recipient (representing possible effects of the specific Verb + Pronoun or V + NP)

Construction order and verb lemma frequency were not significant and were dropped from the final model because their coefficients were less than their standard errors.

To measure the influence of the specific context on the choice of syntactic construction, all of the items were annotated for discourse givenness of recipient and theme and the presence of a parallel construction—double object or prepositional dative—in the preceding context.

All of these factors were tested in the model and found to be insignificant for this dataset, with coefficients less than the standard errors, and were dropped from the final model.

All remaining fixed effects are significant: semantic class and pronominality of recipient, $p < 0.0001$, item order $p < 0.01$.

The model shows that the relations visible in the plotted data are significant, even after taking into account the effects of experimental subject, verb, verb-pronoun interactions, and item order.

In sum,

language users' ability to weigh multiple conflicting constraints not only enables them to reliably make predictive and probabilistic syntactic judgments (Experiment 1), it can reliably override and reverse reported classifications of relative grammaticality (Experiment 2).

The second point was already recognized by Green (1971).

Textbook examples of verbs that occur only in the double object construction, not prepositional dative:

*Ted denied Kim the opportunity to march.*
*\*Ted denied the opportunity to march to Kim.*

*The brass refused Tony the promotion.*
*\*The brass refused the promotion to Tony.*

Compare (Green):

*Ted gave Joey permission to march, but **he denied it to Kim**.*

*The brass gave Martin permission to sit, but **they denied it to Tony**.*

# A concluding reflection

Chomsky (1957): language users can make binary classifications of grammatical and ungrammatical sentences of their language, including those they have never encountered –

(a) *Colorless green ideas sleep furiously.*
(b) **Furiously sleep ideas green colorless.*

(a), (b) are equally improbable, but only (a) is grammatical—it is nonsensical, but syntactically well-formed.

Hence, statistical models are irrelevant to linguistic theories of grammaticality.

Nearly fifty years later—

Fernando Pereira (2000):

(a) and (b) are *not* equally improbable. A simple statistical model trained on newspaper text can in fact distinguish between Chomsky's (a) and (b), separating their probability by five orders of magnitude.

In restricting our attention to categorical models and highly idealized linguistic evidence for the past decades, we linguists have been seriously underestimating the capacities of language users and mislocating the boundaries of grammatical possibility.