Joan Bresnan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen

Predicting the dative alternation

Abstract

Theoretical linguists have traditionally relied on linguistic intuitions such as grammaticality judgments for their data. But the massive growth of computer-readable texts and recordings, the availability of cheaper, more powerful computers and software, and the development of new probabilistic models for language have now made the spontaneous use of language in natural settings a rich and easily accessible alternative source of data.

Surprisingly, many linguists believe that such 'usage data' are irrelevant to the theory of grammar. Four problems are repeatedly brought up in the critiques of usage data:

1. Correlated factors seeming to support reductive theories;
2. Pooled data invalidating grammatical inference;
3. Syntactic choices reducing to lexical biases; and
4. Cross-corpus differences undermining corpus studies.

Presenting a case study of work on the English dative alternation, we show first, that linguistic intuitions of grammaticality are deeply flawed and seriously underestimate the space of grammatical possibility, and second, that the four problems in the critique of usage data are empirical issues that can be resolved by using modern statistical theory and modeling strategies widely used in other fields.

The new models allow linguistic theory to solve more difficult problems than it has in the past, and to build convergent projects with psychology, computer science, and allied fields of cognitive science.

1 The problem

Imagine a child trying to convey the message that a person named Susan gave toys to some children.[1] Through an incremental process of formulating a sentence, the partial

---

[1] We are grateful to Jan Strunk and M. Catherine O'Connor for crucial help with data and presentation. We have also benefited from discussions with Jen Hay, Beth Levin, Richard Oehrle, and Fritz Newmeyer, although we could not make every improvement suggested in the present version. The graphics and analyses were prepared using R (R Development Core Team, 2004). The following description of the problem is adapted from Ferreira's (1996) description of a parallel, incremental model of language production by Bock (1982) and Levelt (1989).

expression *Susan gave___* has already been constructed. Two items from the message could now fill the position after the verb: *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 should be selected?

This is the problem addressed in this study. To establish our terminology, we provide (1). We restate the problem slightly more generally: How does an English speaker determine which of the alternative dative structures to choose to convey a given message about a giving event – the prepositional dative structure or the double object structure?

<div style="text-align:center">Terms used with the dative alternation</div>

Prepositional dative structure:	. . . *gave* [*toys*] [*to the children*]	V NP PP	
Double object structure:	. . . *gave* [*the children*] [*toys*]	V NP NP	
Dative PP:	. . . *gave* [*toys*] [***to the children***]	V NP **PP**	(1)
Dative NP:	. . . *gave* [***the children***] [*toys*]	V **NP** NP	
Theme:	. . . *gave* [***toys***] [*to the children*]	V **NP** PP	
	. . . *gave* [*the children*] [***toys***]	V NP **NP**	
Recipient:	. . . *gave* [*toys*] [***to the children***]	V NP **PP**	
	. . . *gave* [***the children***] [*toys*]	V **NP** NP	

The problem is interesting for many reasons. In the psychology of language, the problem bears on learning and production. In computer science, it is interesting for the development of natural language generation systems. In education, it relates to second language acquisition. Even in English literature, the problem may be of interest for quantitative studies of style and the determination of authorship.

However, in (traditional) theoretical linguistics problems of this kind have been considered too complex and difficult to tackle. The view is widespread that linguistic theory must make radical idealizations of its data. In fact, some would even consider our problem uninteresting for linguistic theory. The slogan "Grammar is grammar and usage is usage" – the title of the 2003 Presidential address to the Linguistic Society of America – seems to set problems like this outside the bounds of linguistic theory, at least for those who believe that the proper subject matter of linguistic theory is grammar itself, abstracted away from the choices of grammatical structures made by actual speakers.

In this essay we will demonstrate that the problem is *not* too difficult. Using some tools which have been employed in other areas of our field (see Baayen, 2004, and references), we can correctly predict 94% of the actual choices of dative constructions in a large corpus of natural spontaneous conversations, the three-million word Switchboard collection of recorded telephone conversations (Godfrey *et al.*, 1992).

We will show that the problem *is* interesting for theoretical linguistics, for several reasons. First, we learn that using traditional methods of data collection in theoretical linguistics, we have underestimated the space of grammatical possibility. Second, we find that persistent questions about what kinds of data are valid to use for linguistic theory can be answered empirically.

2 Predicting from different meanings

One very natural approach to the problem is the idea of predicting different dative structures from different meanings. Advanced by Green (1974) and Oehrle (1976), this idea was taken up in influential work on language learnability by Pinker and colleagues (Gropen *et al.*, 1989; Pinker, 1989, pp. 110–111). They argued that there are two ways of viewing the same giving event: as *causing a change of state (possession)* or as *causing a change of place (movement to a goal)*. They hypothesized that the different ways of conceptualizing the giving event are associated with different structures, the possession meaning with the double object structure and the movement meaning with the prepositional dative structure:

> *Meaning-to-structure mapping hypothesis*:
> causing a change of state (possession) $\Rightarrow$ V NP NP,
>   example: *Susan gave the children toys*;   (2)
> causing a change of place (movement to goal) $\Rightarrow$ V NP [to NP],
>   example: *Susan gave toys to the children*.

One kind of evidence provided for this hypothesis comes from *give* idioms, illustrated in (3) and (4):

> *That movie gave me the creeps.*   (3a)

> **That movie gave the creeps to me.*   (3b)

> *The lighting here gives me a headache.*   (4a)

> **The lighting here gives a headache to me.*   (4b)

These 'giving events' involve no movement to a goal. Giving someone the creeps is causing someone to have feelings of fear and revulsion. Giving someone a headache is causing someone to have a headache. In these examples neither 'the creeps' nor 'a headache' is transferred to a goal. Instead, they come into existence through a change of state in the possessor. In accordance with the meaning-to-structure mapping hypothesis (2), the examples are reported to allow only the double object structure.

Another kind of evidence comes from 'verbs of continuous imparting of force':

> *I carried/pulled/pushed/schlepped/lifted/lowered/hauled the box to John.*   (5a)

> **I carried/pulled/pushed/schlepped/lifted/lowered/hauled John the box.*   (5b)

These verbs describe particular manners of movement to a goal, while the movement all the way to the goal is caused by an agent's continuous imparting of force. Again in accordance with the meaning-to-structure mapping hypothesis (2), they are reported to allow only the PP dative [by Pinker (1989, pp. 110–111), Levin (1993, pp. 46, 114), and many other authors].

Yet when we search the World Wide Web for instances of *give* idioms used with the prepositional dative, we find many in current use. The examples in (6) and (7) are illustrative. As found in actual use, these examples are entirely well-formed and natural to many speakers.

Give the creeps to:

...Orson Welles, who as the radio character, "The Shadow", used to **give "the creeps" to countless child listeners**...
http://clps.k12.mi.us/platte/scifi/toppage21.htm (6a)

This story is designed to **give the creeps to people who hate spiders**, but is not true.
http://www.google.com/search?hl=en&ie=ISO-8859-1& (6b)
q=%22give+the+creeps+to%22&btnG=Google+Search
(cached)

This life-sized prop will **give the creeps to just about anyone**!
Guess he wasn't quite dead when we buried him! (6c)
http://www.frightshop.com/

Some of Andy's death screens are pretty nasty and the enemies are guaranteed to **give the creeps to the smaller set**. (6d)
http://www.ladydragon.com/a-heartofdarkness.html

...Stories like these must **give the creeps to people whose idea of heaven is a world without religion**... (6e)
http://enquirer.com/editions/2001/09/30/loc_lords_gym.html

Give a headache to:

She found it hard to look at the Sage's form for long. The spells that protected her identity also **gave a headache to anyone trying to determine even her size**, the constant bulging and rippling of her (7a)
form gaze Sarah vertigo.
http://lair.echidnoyle.org/rpg/log/27.html

From the heads, offal and the accumulation of fishy, slimy matter, a stench or smell is diffused over the ship that would **give a headache to the most athletic constitution**. (7b)
www.downhomer.com/Webmag/2000/0007/page36.html

Design? Well, unless you take pride in **giving a headache to your visitors** with a flashing background?no.
http://members.tripod.com/~SailorMoonWorstOfWeb/archive/ (7c)
RunJan01.html

Similarly, when we search the World Wide Web for instances of verbs of continuous imparting of force in the double object construction, we find some very natural sounding examples. See (8) for illustrative examples.

Verbs of continuous imparting of force:

Karen spoke with Gretchen about the procedure for registering a complaint, and **hand-carried her a form**, but Gretchen never completed it. 2 June 1999, Nampa Controversy Summary – Idaho Library Association, http://www.idaholibraries.org/nampa.controversy.summary.htm	(8a)
As Player A **pushed him the chips**, all hell broke loose at the table. http://www.cardplayer.com/?sec=afeature&art_id=165	(8b)
Nothing like heart burn food. "I have the tums". Nick joked. He **pulled himself a steaming piece of the pie**. "Thanks for being here". http://www.realityfanfiction.addr.com/storm3.html	(8c)
"Well. . . it started like this. . .". Shinbo explained while Sumomo **dragged him a can of beer** and opened it for him, "We were having dinner together and. . ." http://www.angelfire.com/wa2/bozyby/hold1.html	(8d)
Therefore, when he got to purgatory, Buddha **lowered him the silver thread of a spider** as his last chance for salvation. http://www.inch.com/~fujimura/ImofGrmain.htm	(8e)

Note that *pushed him the chips* comes from a website for poker players; *the chips* are tokens representing money amounts and transfer of possession takes place by moving chips from one player to another across the poker table. Note also that *Sumomo dragged him a can of beer* describes an action by a tiny robot servant.

None of these examples from the World Wide Web is supposed to be grammatically possible, though sporadic counterexamples of the types given were already noticed by Gropen *et al.* (1989). Are these valuable data then, or simply sporadic errors?

In pioneering work, Green (1971) already showed that intuitions of grammaticality for various types of idiomatic dative constructions can be overridden by factors such as the pronominality of arguments. Reasons for considering the present examples to be authentically representative of grammatical possibility are first, that they sound fine as found in actual use, and second, that their structures are principled. We have already commented on the naturalness of the WWW examples. Concerning the principles underlying their structure, first compare (9)–(10):

*That movie gave the creeps to me.	(9a)
. . . Stories like these must **give the creeps to people whose idea of heaven is a world without religion**. . .	(9b)

??Stories like these must **give people whose idea of heaven is a world without religion the creeps**. . . (10a)

That movie gave me the creeps. (10b)

In both of the natural sounding examples, (9b) and (10b), the longer phrase is placed at the end by *the principle of end weight* (Behagel, 1909; Wasow, 2002). The unnatural sounding constructed examples (9a) and (10a) violate the principle of end weight.

We infer that idioms like *give the creeps* have a strong bias toward the double object construction, but the principle of end weight overrides it.

Next compare the (a) and (b) examples of (11) and (12):

??Karen hand-carried a man a form. (11a)

Karen spoke with Gretchen about the procedure for registering a complaint, and **hand-carried her a form**, but Gretchen never completed it. (11b)

??He dragged a guest a can of beer. (12a)

"Well. . . it started like this. . . ". Shinbo explained while Sumomo **dragged him a can of beer** and opened it for him, . . . (12b)

Notice that in the authentic examples, the referent of the first object in the double object construction is given in the immediately preceding discourse or even the same sentence, and is definite and pronominal, in contrast to the second object:

Karen spoke with *Gretchen* about the procedure for registering a complaint, and hand-carried *her* a form, but Gretchen never completed it. (13)

. . .*Shinbo* explained while Sumomo dragged *him* a can of beer and opened it for him, . . . (14)

In contrast, in the worse-sounding constructed examples the referent of the first object in the double object construction is not given in the immediately preceding discourse, and in fact is new, as well as nonpronominal and indefinite:

??Karen hand-carried **a man** a form.
??He dragged **a guest** a can of beer. (15)

In other words, the referent of the first object in the authentic examples is more *discourse accessible* than that of the second object, as well as *more definite*, *pronominal*, and *shorter* – exemplifying a principle observed by Collins (1995).

In a corpus study of Australian English dative constructions Collins noted that double object constructions are polarized on scales of discourse accessibility, definiteness, pronominality, and length in words, with the 'Receiver' (recipient) having the more
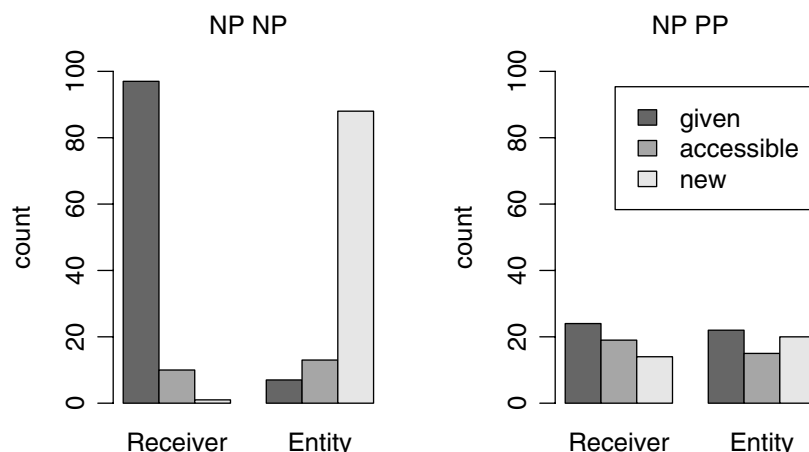
Figure 1. Plots of Collins' (1995) tabular data showing distribution of discourse accessibility in double object and prepositional dative structures.

prominent (topic-like) properties on these scales than the 'Entity' (theme). Collins (1995, p. 47) referred to the alignment of these properties with specific syntactic positions in double object structures as 'Receiver/Entity Differentiation'. To illustrate the phenomenon, Collins' tabular data for discourse accessibility in double object (NP NP) and prepositional dative (NP PP) structures are graphed in Figure 1.

The data are highly skewed in that most 'receivers' (recipients) are given and most 'entities' (themes) are nongiven. If we consider the *proportional distribution* of discourse accessibility across double object and prepositional dative structures, a familiar pattern emerges: 0.8 of given recipients and 0.76 of given themes occur in immediately postverbal position; 0.76 of nongiven themes and 0.75 of nongiven recipients occur in final position (in the second NP position or the dative PP position, respectively).

As illustrated in Figure 2, the dative structures tend to be chosen so that given referents precede nongiven referents in linear order (Halliday, 1970; Thompson, 1990). In exactly the same way, pronouns precede nonpronouns, definites precede indefinites, and shorter precede longer. In the parlance of Optimality Theory these phenomena are cases of "Harmonic Alignment" of various scales with syntactic position (Aissen, 2003, 1999).

We infer that the verbs of continuous imparting of force have a strong lexical bias toward the prepositional dative, but the principle of receiver-entity differentiation (harmonic alignment) can override it.

We draw the following conclusions from this discussion of evidence for predicting the dative alternation from meaning. First, linguistic intuitions of ungrammaticality are a poor guide to the space of grammatical possibility. Second, usage data reveals generalizations which we are sometimes blind to. Third, English dative verbs have more syntactic flexibility than we thought, occurring more freely in alternative constructions. And fourth, we cannot predict the dative alternation from meaning alone.
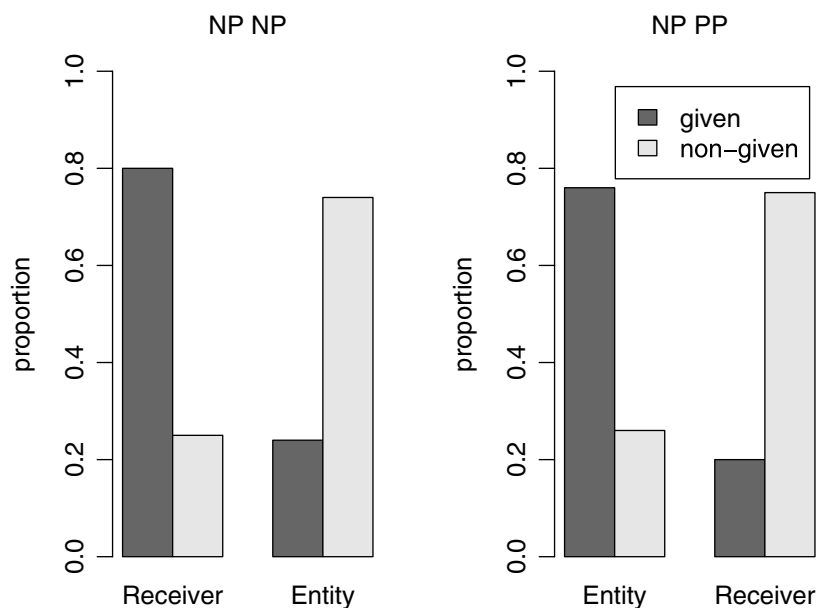
Figure 2. Collins' (1995) data: proportions of NPs in the two dative structures.

3 Predicting from multiple variables

Corpus studies of English have found that various properties of the recipient and theme have a quantitative influence on dative syntax, including discourse accessibility, relative length, pronominality, definiteness, and animacy (Thompson, 1990; Collins, 1995; Snyder, 2003; Gries, 2003, a.o.). Yet what really drives the English dative alternation remains unclear because of four problems inherent to the use of corpus data. First, pervasive correlations in the data seem to support reductive theories which explain the phenomena in terms of a single variable. Second, pooled data from multiple speakers may invalidate grammatical inference if our subject matter is the internalized grammar of the individual. Third, properties such as animacy and discourse accessibility, which characterize the referents of noun phrases, might derive from verb sense semantics. And fourth, cross-corpus differences appear to undermine the relevance of corpus studies to grammatical theory. In the next sections we explain each problem in turn and show how it can be addressed in our data.

3.1 The problem of correlated variables

What really drives the dative alternation remains unclear because of pervasive correlations in the data. Personal pronouns are short, definite, have referents which are discourse-given, and are usually animate. The referents of animate nominals are often discourse-given and definite, frequently referred to pronominally, and usually have nicknames short in length.

Such correlations tempt researchers into reductive theories that explain effects in terms of just one or two variables (e.g., Hawkins, 1994; Snyder, 2003). One such beautifully simple and appealing theory is that of Hawkins (1994). Discourse givenness correlates with shorter, less complex expressions, because less description is needed to identify a referent. Shorter expressions occur earlier in sentences in order to facilitate parsing, more complex processes being deferred after less complex. Therefore, *apparent effects of givenness* (*and correlated properties like animacy*) *could reduce to the preference to process syntactically complex phrases later than simple ones.*

Question 1. Are these effects of discourse accessibility, animacy, and the like the epiphenomena of syntactic complexity effects in parsing?

To answer this question, we used *logistic regression* to control simultaneously for multiple variables related to a binary response [Williams (1994), Arnold *et al.* (2000); Gries (2003) uses discriminant analysis, which makes assumptions about the data distributions that are seldom satisfied]. We also used *larger samples of richly annotated data*: 2360 dative observations from the three-million-word Switchboard collection of recorded telephone conversations. (Previous researchers of the English dative alternation have used very small samples, which permit fewer variables to be investigated simultaneously.)

All 2360 instances of dative constructions used by speakers in the full Switchboard collection of recorded and transcribed telephone conversations were annotated for the explanatory variables. (See the Appendix for information on the sample.)

Our model makes use of fourteen explanatory variables which were considered likely to influence the choice of alternative dative structures. Both discourse accessibility and relative length of recipient and theme were found to be significant by Arnold *et al.* (2000) in a previous logistic regression study. Because the measures of syntactic complexity or 'weight' are highly correlated (Arnold *et al.*, 2000; Wasow, 2002; Szmrecsanyi, 2004a), we used the difference in number of graphemic words between the theme and recipient to measure their relative weight, taking a sign-preserving log transform of the absolute value of the difference to reduce the effect of outliers. The factors of animacy, definiteness, and pronominality were already mentioned. Animacy and definiteness were coded using the coding practices of Garretson *et al.* (2004), and discourse accessibility was coded using Michaelis and Hartwell (forthcoming), which is based on Prince (1981) and Gundel *et al.* (1993). Pronominality was defined to distinguish phrases headed by pronouns (personal, demonstrative, and indefinite) from those headed by nonpronouns such as nouns and gerunds. In addition to these predictors, structural parallelism – the existence of the same kind of structure in the same dialogue – seemed likely to affect the choice of dative structures in our corpus data, based on experimental studies (see Szmrecsanyi, 2004b, 2005, for a review of a wide range of evidence).

From cross-linguistic evidence, number (singular/plural) and person could also have an influence (Aissen, 1999, 2003; Haspelmath, 2004) and a bivariate pilot study of the Switchboard corpus showed person to have an influence parallel to grammaticalized person effects on alternative dative structures (Bresnan and Nikitina, 2003).

Five broad semantic classes of uses of verbs which participate in the dative alternation were also considered: abstract (abbreviated 'a') as in *give it some thought*; transfer of possession ('t') as in *give an armband*, *send*; future transfer of possession ('f'), exemplified by *owe*, *promise*; prevention of possession ('p'), exemplified by *cost*, *deny*; and communication ('c') as in *tell*, *give me your name*, said on a telephone.

To have sufficient data for simultaneous comparisons, we eliminated variables having a very sparse value (person of theme, animacy of theme) and simplified the accessibility and animacy variables to binary values: accessibility was simplified to given and not given, animacy to animate (= human + animal) and inanimate (not human or animal). The factor of 'concreteness of theme' was added to compensate for our simplification of animacy into a binary category.

The specification of this model ('Model A') is shown in (16). The "Response" is the choice of the dative NP or dative PP, respectively denoted by 0 and 1.

> **Model A**: Response modeled as depending on
> semantic class + accessibility of recipient + accessibility of theme + pronominality of recipient + pronominality of theme + definiteness of recipient + definiteness of theme + animacy of recipient + person of recipient + number of recipient + number of theme + concreteness of theme + structural parallelism in dialogue + length difference (log scale). (16)

The mathematical structure of a logistic regression model is shown in (17).[2]

> **The Logistic Regression Model**:
>
> $$\text{logit}\big[\text{Probability}(\text{Response} = 1)\big] = \mathbf{X}\boldsymbol{\beta} \quad \text{or} \tag{17}$$
>
> $$\text{Probability}(\text{Response} = 1) = \frac{1}{1 + e^{-\mathbf{X}\boldsymbol{\beta}}}.$$

Because logistic regression directly models probability without assuming a particular distribution of data, it is robust for skewed and otherwise nonnormally distributed data.

After fitting Model A to the data, we evaluated the fit. As Table 1 shows, the model correctly classifies 92% of the data overall.

The accuracy measure in Table 1 counts any probability >0.5 as correct for the dative PP response, and so does not distinguish model probabilities near chance from those near 1. Another measure of fit compares the grouped predicted probabilities to the observed proportions of dative PPs. A perfect fit would correspond to the straight line shown in Figure 3. Because the data points are close to the straight line, this plot indicates a very good fit when proportions are considered.

---

[2] The function *logit* is the natural log of the odds of the Response 1 (here a dative PP realization) over 0 (dative NP realization). $\mathbf{X}$ is the $n \times p$ model matrix and $\boldsymbol{\beta}$ is the $p \times 1$ vector of coefficients, where n is the number of observations and p is the number of model parameters. Maximum likelihood estimation is used to estimate the coefficients $\boldsymbol{\beta}$.

Table 1. Model A accuracy. Classification table for Model A (1 = PP; cut value = 0.50)

		Predicted		% correct
		0	1	
Observed	0	1796	63	97
	1	115	386	77
			Overall	92

Note. "% correct" from always guessing NP NP (=0): 79%.

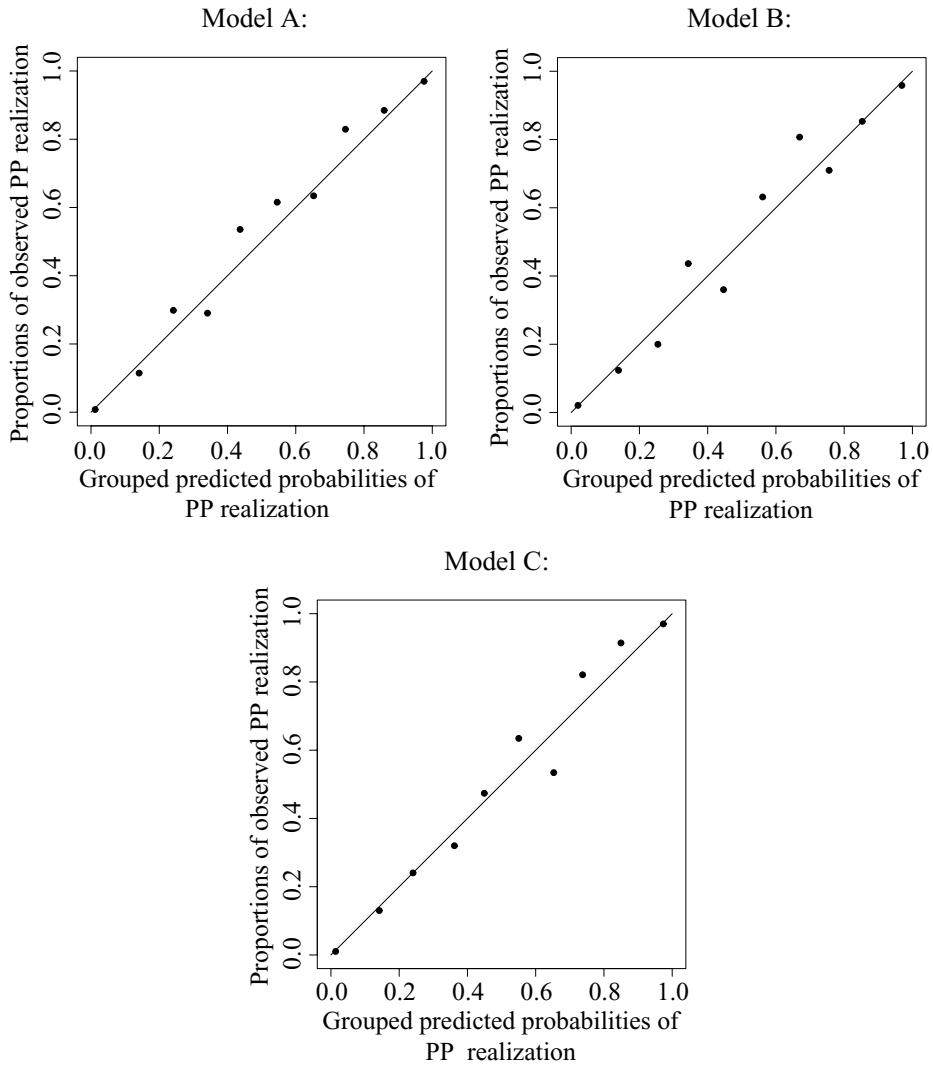Model A:

Model B:

Model C:

Figure 3. Model plots of observed against estimated responses.

To determine how well the model generalizes to unseen data, we divided the data randomly 100 times into a training set of sufficient size for estimating the model parameters ($n = 2000$) and a testing set ($n = 360$), fit the Model A parameters on each training set, and scored its predictions on the unseen testing set. The mean overall score (average % correct predictions on unseen data) is 92%, which is excellent, showing that the model is not overfitting the data.

All of the model predictors except for number of recipient are significant, all $p < 0.001$ except for person of recipient, number of theme, and concreteness of theme, for which $p < 0.05$. As illustrated qualitatively in (18), what Model A shows us is that even when we control simultaneously for all fourteen explanatory variables, the harmonic alignment effects noted by previous researchers are real. The bolded properties (discourse givenness, animacy, definiteness, pronominality, and relative length of recipient and theme) are independently aligned with the immediately postverbal position in both the double object and the prepositional dative structures. Similarly, the unbolded properties (nongiven, nonpronoun, etc.) are aligned with the final position, whether second object or dative PP.

Interpretation of Model A
Harmonic alignment with syntactic position:
(a) **discourse given** $\succ$ nongiven
(b) **pronoun** $\succ$ nonpronoun
(c) **animate** $\succ$ inanimate   (18)
(d) **definite** $\succ$ indefinite
(e) **recipient shorter than theme** $\succ$ recipient longer than theme
$$\text{V } \mathbf{NP} \text{ NP}$$
$$\text{V } \mathbf{NP} \text{ PP}$$

The model formula is shown in Figure 4. The signs of the coefficients show the directions of the effects. Positive coefficients favor the prepositional dative structure, and negative favor the double object structure.

In Figure 4 we see, for example, that in the first line labeled (a) the predictor {accessibility of recipient = nongiven} has a positive coefficient of $+0.99$, thus favoring the NP PP structure, as in *give the toys to a child*. In this structure the recipient (*a child*) occurs in the dative PP position at the right end following the postverbal NP *the toys*. Looking at line (a) of the diagram of qualitative harmonic alignment in (18), we see that the nongiven property is not bolded, and aligns (for recipients) with the unbolded final phrase in V **NP** PP. In contrast, when {accessibility of recipient = given}, this predictor has the value 0, which is negative compared to $+0.99$. Compared to a nongiven recipient, a given recipient therefore has a greater tendency to favor the double object structure V NP NP, as in *give the children toys*, placing the recipient (*the children*) in the first NP position preceding the second NP (*the toys*). And again looking at line (a) the diagram of qualitative harmonic alignment in (18), we see that the given property is bolded, and aligns (for recipients) with the first NP in V **NP** NP.

For an example of a predictor with a negative coefficient in Figure 4, consider the predictor in the second line labeled (a): {accessibility of theme = nongiven}. This pre-

$$\text{Probability}\{\text{Response} = 1\} = \frac{1}{1+e^{-\mathbf{X}\boldsymbol{\beta}}},$$
where

$$
\begin{aligned}
X\hat{\beta} = \ &0.95 \\
&- 1.34\{c\} + 0.53\{f\} - 3.90\{p\} + 0.96\{t\} \\
\text{(a)} \quad &+ 0.99\{\text{accessibility of recipient} = \text{nongiven}\} \\
\text{(a)} \quad &- 1.1\{\text{accessibility of theme} = \text{nongiven}\} \\
\text{(b)} \quad &+ 1.2\{\text{pronominality of recipient} = \text{nonpronoun}\} \\
\text{(b)} \quad &- 1.2\{\text{pronominality of theme} = \text{nonpronoun}\} \\
\text{(c)} \quad &+ 0.85\{\text{definiteness of recipient} = \text{indefinite}\} \\
\text{(c)} \quad &- 1.4\{\text{definiteness of theme} = \text{indefinite}\} \\
\text{(d)} \quad &+ 2.5\{\text{animacy of recipient} = \text{inanimate}\} \\
&+ 0.48\{\text{person of recipient} = \text{nonlocal}\} \\
&- 0.03\{\text{number of recipient} = \text{plural}\} \\
&+ 0.5\{\text{number of theme} = \text{plural}\} \\
&- 0.46\{\text{concreteness of theme} = \text{nonconcrete}\} \\
\text{(e)} \quad &- 1.1\{\text{parallelism} = 1\} - 1.2 \cdot \text{length difference (log scale)}
\end{aligned}
$$

and {c} = 1 if subject is in group c, 0 otherwise (and likewise for other categories).

Figure 4. The model A formula.

dictor has the negative coefficient −1.1, which favors the NP NP structure *give the children toys*. That structure places the theme *toys* at the end. Looking at line (a) of the diagram of qualitative harmonic alignment in (18), we see that the nongiven property aligns (for themes) with the unbolded final phrase in V **NP** NP. In contrast, when {accessibility of theme = given}, this predictor has the value 0, which is positive compared to −1.1. This predictor therefore has a greater tendency to favor the prepositional dative structure, as in *give the toys to a child*. That structure places the theme (*the toys*) in immediate postverbal position preceding the PP (*to a child*). Looking again at line (a) of the harmonic alignment diagram in (18), we see that the bolded given property aligns with the bolded position (for themes) shown for V **NP** PP.

Now observe that all of the recipient predictors in the lines labeled (a)–(d) of the model formula have positive coefficients, while all of the theme predictors in these lines have negative coefficients. Note further that the values of each of these predictors are drawn from the right hand sides of the correspondingly labeled lines in (18); these are the lower ends of the scales that undergo harmonic alignment with the indicated syntactic positions. Thus by the same reasoning just illustrated for the accessibility predictors, we see that the Model A formula is showing quantitatively the pattern of harmonic alignment that is shown qualitatively in (18) for these factors.

As for the interval variable in Model A – the length difference predictor in line (e) – it, too, reveals the quantitative pattern of harmonic alignment. The length

difference coefficient -1.2 is negative, so when the length difference value itself is negative, the predictor will be positive, favoring the NP PP structure. By definition the length difference value is negative when the recipient is longer than the theme. This corresponds to the right-hand side of line (e) in (18), which harmonically aligns (for the long recipient) with the unbolded PP in V **NP** PP. On the other hand, when the length difference value is positive, the predictor will be negative, favoring the NP NP structure which places the recipient in the first postverbal NP position. By definition, a positive length difference value means that the theme is longer than the recipient, which corresponds to the left hand side of line (e) in (18). The shorter recipient harmonically aligns with the bolded first NP position in V **NP** NP. *In short, the model formula shows quantitatively the harmonic alignment pattern shown qualitatively in* (18).

Model A gives us our answer to Question 1. The effects of discourse accessibility, animacy, definiteness, pronominality, and syntactic weight on dative construction choice are not reducible to syntactic complexity in parsing.

3.2 *The problem of pooling different speakers' data*

One question is persistently asked about corpus studies of grammar.

Question 2. In the words of Newmeyer (2003, p. 696):

> The Switchboard Corpus explicitly encompasses conversations from a wide variety of speech communities. But how could usage facts from a speech community to which one does not belong have any relevance whatsoever to the nature of one's grammar? There is no way that one can draw conclusions about the grammar of an individual from usage facts about communities, particularly communities from which the individual receives no speech input.

The speakers of course differ from each other in many ways. However, what they share in determining their choices of dative syntax might outweigh their differences. This is therefore an empirical question.

The Switchboard Corpus is annotated for speaker identity. The 2360 instances of dative constructions in our sample were produced by 424 total speakers. As indicated in (19), the data are extremely unbalanced.

$$
\begin{aligned}
&228 \text{ speakers produced 4 to 7 instances each,} \\
&106 \text{ speakers produced 8 to 12 instances each,} \\
&42 \text{ speakers produced 13 to 19 instances each,} \\
&11 \text{ speakers produced 20+ instances each.}
\end{aligned}
\tag{19}
$$

We approached this problem as follows. Speaker identity is a source of unknown dependencies in the data. The effects of these unknown dependencies on the reliability of the estimates can be themselves estimated from the observed data using modern statistical techniques (Efron and Tibshirani, 1986, 1993; Feng *et al.*, 1996; Harrell, 2001). When data dependencies fall into many small clusters (each speaker defines a 'cluster'), we can assume a 'working independence model' (our Model A) and revise the covariance estimates using *bootstrap sampling with replacement of entire clusters*.

In other words, we can create multiple copies of the data by resampling *from the speakers*. The same speakers' data can randomly occur many times in each copy. We repeatedly re-fit the model to these copies of the data and used the average regression coefficients of the re-fits to correct the original estimates for intra-speaker correlations. *If the differences among speakers are large, they will outweigh the common responses and the findings of Model A will no longer be significant.*

The table in (20) shows the results of applying this procedure using Model A as our working independence model. The coefficients shown are identical to those in the Model A formula (ignoring rounding). Widened confidence intervals (abbreviated 'C.I.' in the table) reflect the reduction of independent observations in our data caused by the presence of clusters of speaker dependencies. The confidence intervals show the ranges of the odds ratios for which there is a similar chance of error ($<5\%$). An odds ratio of 1 means that the odds of a dative PP and a dative NP are the same, so the outcome is 50% and 50%. If a confidence interval contains 1, then we simply have no reason to say that a predictor leads to a higher (or lower) probability for the prepositional dative structure. Notice that in our table the confidence intervals nicely stay away from 1.[3]

Model A: Relative magnitudes of significant effects with corrected error estimates

	Coefficient	Odds ratio PP	95% C.I.
Inanimacy of recipient	2.54	12.67	5.56–28.87
Nonpronominality of recipient	1.17	3.22	1.70–6.09
Nongivenness of recipient	0.99	2.69	1.37–5.3
Transfer semantic class	0.96	2.61	1.44–4.69
Indefiniteness of recipient	0.85	2.35	1.25–4.43
Plural number of theme	0.50	1.65	1.05–2.59
Person of recipient	0.48	1.62	1.06–2.46
Nongivenness of theme	-1.05	0.35	0.19–0.63
Structural parallelism in dialogue	-1.13	0.32	0.22–0.47
Nonpronominality of theme	-1.18	0.31	0.19–0.50
Length difference (log scale)	-1.21	0.3	0.22–0.4
Communication semantic class	-1.34	0.26	0.13–0.55
Indefiniteness of theme	-1.37	0.25	0.15–0.44

(20)

We can therefore infer from our table that an inanimate recipient is over twelve times more likely to be expressed in a dative PP structure than an animate recipient (though the confidence interval here is wide, indicating much variability in the data);

---

[3]In analyzing the relative magnitudes of predictors, we chose to include in our models all variables that were considered, including those which turned out to have an insignificant effect on the response in our data. Eliminating such variables often biases the results by inflating the apparent magnitudes of the effects of other variables (Harrell, 2001). The question of whether the magnitudes of discourse effects on syntax are consequential compared to structural parallelism is raised by previous work on the agentless passive (Weiner and Labov, 1983), and our results show that they are. Nevertheless, only the significant predictors are displayed in our tables of relative magnitudes of effects. It is for this reason that 'concreteness of theme', for example, is omitted from (20) and (26), although it is included in the model formulae in Figures 4 and 5.

a nonpronominal recipient is over three times more likely to be in a dative PP than a pronominal recipient; and a nongiven recipient, almost three times more likely to be in a dative PP than a given recipient.

Thus we have an answer to Question 2. The influence of discourse accessibility, animacy, and the like on dative syntax remains significant when differences in speaker identity are taken into account. *What the speakers share in the choice of dative syntax outweighs their differences.*

3.3 The problem of lexical biases

What really drives the dative alternation *still* remains unclear. We have assumed that NPs can be drawn out of the database and examined independently for their properties of discourse accessibility, animacy, pronominality, and the like. But these observations of NP properties are not independent. Just as the pooling of data from different speakers introduced unknown dependencies among the observations (contrary to the fundamental mathematical assumptions of the models), so does the pooling of NP observations from different verbs.

It is easy to see that the properties of recipients and themes depend on the verbs which describe the transfer events they are participating in. In our dataset, for example, the verb *bring* is nearly three times more likely to have a given recipient than the verb *take*, while *take* is over seven times more likely to have a nongiven recipient than *bring* – even though the two verbs belong to the same broad semantic class of transfer of possession. (This may be because of the differing points of view implied by the two verbs: the goal of bringing is usually located near the speaker, while the goal of taking is usually located away from the speaker.) There are thirty-eight different verbs in our dataset.

The properties of the NP arguments are conditional not only on the verb they occur with, but also on the specific *use* of the verb. For example, the verb *give* has a larger than average proportion of inanimate recipients, because of the many abstract uses it allows (Bresnan and Nikitina, 2003):

> Example inanimate recipients with *give*:
>> Um, but still, it gives **it** some variety.
>> But I'm going to give **it** thumbs down.   (21)
>> You know, give **it** a great deal of thought,
>> and you can add hamburger if you want to give **it** a little more body.

But the communicative sense of *give*, as in *give me your name*, is like the verb *tell* in having only animate recipients in our dataset, because we normally communicate with people.

Another example of the dependence of recipient and theme properties on the particular *use* of the verb being used comes from *pay*. The recipients of paying in the

transfer sense (such as paying money) are far more likely to be animate and given than the 'recipients' of paying in the abstract sense (paying attention or heed), as shown in (22).

Distribution of animacy and discourse accessibility for two senses of pay					
	Animate	Inanimate	Given	Nongiven	
Pay (transfer)	83	1	61	23	(22)
Pay (abstract)	17	40	31	26	

Presumably we are more likely to pay money to recipients that we already know, who are also likely as money-users to be animate. Similarly, we are probably more likely to pay attention to less expected, nongiven things.

These facts motivate our third question.

Question 3. Do the apparent effects of givenness and animacy on the choice of dative syntax still hold when they are conditioned on specific verbs and verb uses?

In the case of speaker dependencies, we used the technique of bootstrap sampling with replacement of entire clusters to estimate the reliability lost by violating the assumption of independence of observations. That technique works well for many small clusters, but in the case of the verb-use dependencies, there are several extremely large clusters: the abstract uses of the verb *give* alone make up one third of the entire dataset, and the transfer uses of *give*, one sixth! Fortunately, an alternative approach – variously called "multilevel regression", "mixed effects modeling", and "conditional regression" – is available which allows us to build the clusters into the model as an additional layer (Pinheiro and Bates, 2000).

To define our clusters, we crossed the thirty-eight individual verbs participating in the dative alternation in our dataset with the five broad semantic classes we had annotated. There are fifty-five such theoretical 'verb senses' in use in our dataset of dative structures from the full Switchboard corpus. Examples are given in (23).

Example verb senses:
give.t = *give* in transfer sense: *give you an armband*
give.c = *give* in communication sense: *give me this cock and bull story . . .*
give.a = *give* in abstract sense: *give that a lot of thought*   (23)
pay.t = *pay* in transfer sense: *pay somebody good money*
pay.a = *pay* in abstract sense: *pay attention to cats*.

We used a *multilevel logistic regression model* to condition the binary response on the verb sense. The model specification is shown in (24).[4]

> **Model B**: Response modeled as depending on
> **fixed effects**:   semantic class + accessibility of recipient + accessibility of theme + pronominality of recipient + pronominality of theme + definiteness of recipient + definiteness of theme + animacy of recipient + person of recipient + number of recipient + number of theme + concreteness of theme + structural parallelism in dialogue + length difference (log scale) -1,
> **random effect**:   verb sense.

(24)

The mathematical structure of the model is shown in (25):[5]

A generalized linear model with a single random intercept

$$\mathrm{logit}\big[\mathrm{Pr}(\mathbf{Y}_{ij} = y_{ij}|u_i)\big] = \mathbf{X}_{ij}\boldsymbol{\beta} + u_i.$$

(25)

In Model B the conditional probability of a response given a group i is systematically linked to a linear combination of fixed cross-group explanatory variables $\mathbf{X}_{ij}$ and a randomly varying normally distributed group effect.

After fitting Model B to the data, we evaluated the fit. As Table 2 shows, the model correctly classifies 95% of the data overall. The estimated probabilities of Model B shown in Figure 3 also show an excellent fit.

How well does Model B generalize to unseen data? As before, we divided the data randomly 100 times into a training set of sufficient size for estimating the model parameters ($n = 2000$) and a testing set ($n = 360$), fit the Model B parameters on each training set, and scored its predictions on the unseen testing set. The mean overall score (average % correct predictions on unseen data) is 94%, which is very good, showing only a slight overfitting.

Table 2. Model B accuracy. Classification table for Model B (1 = PP; cut value = 0.50)

		Predicted		% correct
		0	1	
Observed	0	1809	50	97
	1	68	433	86
			Overall	95

---

[4]The model intercept was defined to be zero by subtracting 1. This allows the different verb sense groups to be contrasted with each other rather than with an arbitrary baseline.

[5]There are i groups of data (one for each verb sense), each group having j observations, so that the total of observations $n = i \times j$. $\mathbf{X}_{ij}$ is the $n \times p$ model matrix, where p is the number of parameters and $\boldsymbol{\beta}$ is the $p \times 1$ vector of coefficients. The random variable u_i is normally distributed, $u_i \sim N(0, \sigma)$, so a single parameter σ can be estimated for the set of verb senses. We used the glmmPQL algorithm described by Venables and Ripley (2002) to estimate the model parameters.

The model formula is shown in Figure 5. Notice that the directions of the effects again reveal the same pattern of harmonic alignment already seen in (18) and in the Model A formula in Figure 4: the positive coefficients for accessibility, pronominality, animacy, and definiteness favor the PP dative, the negative favor the NP dative, and the negative coefficient for the length differential is exactly the same as in Model A.

Finally, the relative magnitudes of the effects are shown in (26). This table shows that an inanimate recipient is over five times more likely to be expressed in a dative PP structure than an animate recipient. Animacy remains a major significant effect, along with pronominality and givenness. A nonpronominal recipient is also over five times more likely to be in a dative PP than a pronominal recipient; and a nongiven recipient, over four times more likely to be in a dative PP than a given recipient.

Model B: Relative magnitudes of significant effects.

	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

(26)

Model B thus gives us an answer to Question 3. The influence of givenness, animacy, pronominality and the other variables on the choice of dative syntax remain independently significant when they are conditioned on specific verb senses.

3.4 The problem of cross-corpus differences

We now take up the fourth problem for corpus studies of grammar.[6]

Question 4. Does it make sense to relate frequencies of usage to grammar?

> After all, unlike the grammaticality of a linguistic form, which is an idealization over usage, the actual frequency of usage of a form is a function of both grammatical structure and extra-grammatical factors such as memory limitations, processing load, and the context.

The data we have examined so far come from the Switchboard corpus, which reflects the on-line processing of spontaneous speech. How could our probabilistic generalizations hold of a very different corpus consisting of edited, written reportage not

---

[6]This problem is raised by Keller and Asudeh (2002, p. 240) as part of their critique of stochastic optimality theory, but the problem applies more generally to probabilistic theories of grammar based on usage data. See Boersma (2004) for a response.

$$\text{Probability}\{\text{Response} = 1\} = \frac{1}{1 + e^{-\mathbf{X}\boldsymbol{\beta} + u}},$$
where

$$
\begin{aligned}
X\hat{\beta} = {} & 1.5\{a\} + 0.58\{c\} + 0.96\{f\} - 3.28\{p\} + 2.7\{t\} \\
& + 1.5\{\text{accessibility of recipient} = \text{nongiven}\} \\
& - 1.2\{\text{accessibility of theme} = \text{nongiven}\} \\
& + 1.7\{\text{pronominality of recipient} = \text{nonpronoun}\} \\
& - 2.2\{\text{pronominality of theme} = \text{nonpronoun}\} \\
& + 0.7\{\text{definiteness of recipient} = \text{indefinite}\} \\
& - 1.7\{\text{definiteness of theme} = \text{indefinite}\} \\
& + 1.5\{\text{animacy of recipient} = \text{inanimate}\} \\
& + 0.4\{\text{person of recipient} = \text{nonlocal}\} \\
& - 0.2\{\text{number of recipient} = \text{plural}\} \\
& + 0.7\{\text{number of theme} = \text{plural}\} \\
& + 0.35\{\text{concreteness of theme} = \text{nonconcrete}\} \\
& - 1.1\{\text{parallelism} = 1\} - 1.2 \cdot \text{length difference (log scale)}
\end{aligned}
$$

and $\hat{u} \sim N(0, 2.27)$.

Figure 5. The model B formula.

subject to memory limitations, processing load, or the speaker-hearer context? In fact it is true that the frequencies of double-object constructions in the Switchboard collection of recordings of telephone conversations differ substantially from the frequencies in the Treebank *Wall Street Journal* collection of news and financial reportage, as shown in (27).

V NP NP's = 79% of total Switchboard datives ($n = 2360$), (27)
V NP NP's = 62% of total Wall Street Journal datives ($n = 905$).

On the face of it, such facts seem problematic for our usage-based studies.

In order to answer Question 4, we fit the same model to the combined data from two different corpora and compare the component fits. (On the specification, see n. 4.)

Model C: Response modeled as depending on
fixed effects: semantic class + accessibility of recipient + accessibility of theme + pronominality of recipient + pronominality of theme + definiteness of recipient + definiteness of theme + animacy of recipient + concreteness of theme + length difference (log scale) -1, (28)
random effect: verb sense.

Table 3. Model C accuracy. Model C classification table (1 = PP; cut value = 0.50)

		Predicted		% correct
		0	1	
Observed	0	2320	96	96
	1	119	730	86
			Overall	93

Model C is Model B after removing three factors (person, number, and parallelism) not marked in our *Wall Street Journal* dative dataset. There are 110 verb senses in the combined corpora for Model C.

After fitting Model C to the data, we evaluated the fit. As Table 3 shows, the model correctly classifies 93% of the data overall. The estimated probabilities of Model C shown in Figure 3 also show a very good fit.

To determine how well Model C generalizes to unseen data, we again divided the data randomly 100 times into a training set of sufficient size for estimating the model parameters ($n = 2000$) and a testing set ($n = 1265$) and score its predictions on the unseen testing set. The mean overall score (average % correct predictions on unseen data) is 92%, showing only a slight overfitting.

Model C fits the combined data very well, and interestingly, it captures the substantial difference in frequencies of double object constructions in data from the component corpora, as shown in (29).

Model C on data from component corpora

% NP NP's	Switchboard	*Wall Street Journal*
Predicted	79%	63%
Actual	79%	62%

(29)

How is this possible?

The answer is that the inputs to the model vary. For example, in the *Wall Street Journal* dataset, recipient nouns outnumber pronouns 5 to 1. In the Switchboard dataset recipient pronouns outnumber nouns almost 4 to 1. Thus the tendency for pronominal recipients to appear in the NP object position is about the same across the two corpora. *There are more double object constructions in the Switchboard corpus in part because there are simply more recipient pronouns.*

Setting pronouns aside, the proportion of dative NP NP constructions is higher in the *Wall Street Journal* data than in the Switchboard data, and Model C captures this difference between the corpora as well, as shown in (30).

Model C on nonpronominal data from component corpora

% NP NP's (nonpronouns)	Switchboard	*Wall Street Journal*
Predicted	49%	58%
Actual	49%	55%

(30)

Again, how is this possible?

Again, the answer is that inputs vary. For example, among nonpronoun complements to dative verbs, median length differential (log scale) in the Treebank *Wall Street Journal* is 1.1, but the median length differential (log scale) in the Switchboard corpus is 0.69. The tendency for longer themes to appear at the end, favoring the V NP NP construction, is about the same in both of the two corpora. *There are more double object constructions in the Wall Street Journal corpus when we set pronouns aside in part because there are simply longer theme noun phrases.*

Our answer to Question 4 is therefore that some striking differences between different corpora can be explained as the response of the same model to quantitatively different inputs. The probabilistic structure embedded in the model has generality and captures significant structural properties of language beyond the contingencies of a particular corpus.

But is there really no difference between the two corpora with respect to how strong the predictors are? We investigated this question by adding to Model C an additional factor "modality" whose value is 's' for the Switchboard data and 'w' for the *Wall Street Journal* data and then developing further models to study all interactions with modality. We found a small but significant higher probability of using the V NP PP structure in the *Wall Street Journal* data, but there is no indication whatsoever that the other parameters of the model are different for data from the two corpora. The simplest model, which treats modality as a simple main effect, is also the most accurate, as shown in (31).

Comparison of models incorporating modality

Model	Proportion correct	Parameters	
All interactions with modality	0.935069	28	
Stepwise model selection	0.935069	26	(31)
A simple main effects model with modality	0.9353752	15	
A model with hand selected sign interactions	0.9353752	21	

We conclude that the model for spoken English transfers beautifully to written, *except that in written English*, *there is a slightly higher probability of using the prepositional dative structure.* (Of course, it is always possible that in other registers and corpora and other regional varieties of English, further changes are required, but for the present data, there is only the simple main effect of modality to be added to Model C.)

To summarize, we have examined four problems inherent to the use of corpus data in linguistic theory – the problem of correlated factors seeming to support reductive theories, the problem of pooled data invalidating grammatical inference, the problem of nominal factors possibly deriving from verb sense semantics, and the problem of cross-corpus differences. We have shown how answers can be found by using modern statistical theory and modeling strategies used in other areas of our field and widely used in other fields such as biology and education.

Along with formal syntactic and semantic properties, the properties of animacy and discourse accessibility have an irreducible effect on dative syntax across written and spoken modalities, across verb senses, and across speakers.

4 Concluding remarks

We have found that linguistic data are more probabilistic than has been widely recognized in theoretical linguistics. We have examined a body of ecologically valid data – spontaneous language use in natural settings – using statistical techniques for analyzing multiple variables. And we have constructed a model that can predict the choice of dative structures with 94% accuracy, and can resolve persistent questions about usage data.

Our results are corroborated by findings from research on language production in controlled laboratory settings. On the effects of discourse accessibility on syntactic choice in datives and passives, see Bock and Irwin (1980) and Prat-Sala and Branigan (2000) and references. On the effects of animacy, see Bock *et al.* (1992) and references.

Recall where we began this essay, with (traditional) theoretical linguistics regarding the problem of predicting the dative alternation as too difficult to tackle and as outside the proper subject matter for linguistic theory. We note that field linguists and typologists have found that these factors have become part of the categorical grammatical structure of dative syntax in a number of languages (see, for example, Hawkinson and Hyman, 1974; Morolong and Hyman, 1977; Polinsky, 1994, 1996, 1997; Evans, 1997; Haspelmath, 2003, 2004). But even if this were not so, we suggest that by tackling problems of this kind, theoretical linguistics has an opportunity to build collaborative research with psychology, computer science, and allied fields and thereby deepen our understanding of the cognitive foundations of interpretation.

Appendix: Data sample

The sample was selected by taking all verbs that appear in either the prepositional dative structure or the double object structure in the one-million word Treebank Switchboard corpus (Marcus *et al.*, 1993), less benefactives (*buy*, *cook for*), those verbs for which there are three or more alternative constructions (*ask people a question*, *ask a question of/to people*; *provide John* (*with*) *a book*, *provide a book to John*),[7] and non-alternating verbs (those for which there were not at least five instances in each dative structure in that portion of the Internet indexed by Google). The thirty-eight resulting verbs were then used to search for dative structures in the full three-million word Switchboard corpus. Instances were excluded which lacked two overt objects, had a passivized object as subject, occurred in highly fixed expressions (*to tell you the truth*, *I'll tell you what*), *were concealed questions* (*I'll tell you another plant that is purply*), or had unambiguously spatial goals (*take my dog to Saint Louis*).

---

[7]Heavy NP shift examples (V PP NP) were not considered, because of their extreme rarity. For the verbs we examined, there were only four instances in the full Switchboard corpus, compared to over five hundred instances in the unshifted order (V NP PP).

References

Aissen, J. 1999. Markedness and subject choice in Optimality Theory. Natural Language and Linguistic Theory **17**: 673–711.

Aissen, J. 2003. Differential object marking: Iconicity vs. economy. Natural Language and Linguistic Theory **21**: 435–483.

Arnold, J., T. Wasow, A. Losongco and R. Ginstrom. 2000. Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. Language **76**: 28–55.

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

Behagel, O. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. Indogermanische Forschungen, 25/110.

Bock, J.K. 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. Psychological Review **89**: 1–47.

Bock, J.K. and D.E. Irwin. 1980. Syntactic effects of information availability in sentence production. J. Verbal Learning and Verbal Behavior **19**: 467–484.

Bock, J.K., H. Loebell and R. Morey. 1992. From conceptual roles to structural relations: Bridging the syntactic cleft. Psychological Review **99**: 150–171.

Boersma, P. 2004. A Stochastic OT account of paralinguistic tasks such as grammaticality and prototypicality judgments. On-line, Rutgers Optimality Archive, http://roa.rutgers.edu/.

Bresnan, J. and T. Nikitina. 2003. The gradience of the dative alternation. On-line at http://lfg-stanford.edu/bresnan/download.html.

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

Efron, B. and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science **1**: 54–77.

Efron, B. and R. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Evans, N. 1997. Role or cast? Noun incorporation and complex predicates in Mayali. In A. Alsina, J. Bresnan and P. Sells (eds), *Complex Predicates*, pp. 397–430. CSLI, Stanford.

Feng, Z., D. McLerran and J. Grizzle. 1996. A comparison of statistical methods for clustered data analysis with Gaussian error. Statistics in Medicine **15**: 1793–1906.

Ferreira, V.S. 1996. Is it better to give than to donate? Syntactic flexibility in language production. J. Memory and Language **35**: 724–755.

Garretson, G., M.C. O'Connor, B. Skarabela and M. Hogan. 2004. Coding practices used in the project Optimal Typology of Determiner Phrases. On-line at http://npcorpus.bu.edu/documentation/index.html.

Godfrey, J., E. Holliman and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of ICASSP-92*, San Francisco, pp. 517–520.

Green, G. 1971. Some implications of an interaction among constraints. In *Papers from the Seventh Regional Meeting, Chicago Linguistic Society*, pp. 85–100. Chicago Linguistic Society, The University of Chicago, Chicago.

Green, G. 1974. *Semantics and Syntactic Regularity*. Indiana University Press, Bloomington.

Gries, S.Th. 2003. Towards a corpus-based identification of prototypical instances of constructions. Annual Review of Cognitive Linguistics **1**: 1–28.

Gropen, J., S. Pinker, M. Hollander, R. Goldberg and R. Wilson. 1989. The learnability and acquisition of the dative alternation. Language **65**: 203–257.

Gundel, J.K., N. Hedberg and R. Zacharsky. 1993. Referring expressions in discourse. Language **69**: 274–307.

Halliday, M.A.K. 1970. Language structure and language function. In J. Lyons (ed.), *New Horizons in Linguistics*, pp. 140–165. Penguin Books, Harmondsworth.

Harrell, F.E., Jr. 2001. *Regression Modeling Strategies, with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, New York.

Haspelmath, M. 2003. Ditransitive constructions in the world's languages: Alignment types, alignment splits, and inverse patterns. Berkeley Handout, March 2003. University of California, Berkeley.

Haspelmath, M. 2004. Explaining the ditransitive person-role constraint: A usage-based approach. Constructions 2/2004. On-line, http://www.digijournals.de/constructions/articles/35.

Hawkins, J. 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press, Cambridge.

Hawkinson, A. and L. Hyman. 1974. Hierarchies of natural topic in Shona. Studies in African Linguistics **5**: 147–170.

Keller, F. and A. Asudeh. 2002. Probabilistic learning algorithms and Optimality Theory. Linguistic Inquiry **33**: 225–244.

Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago.

Marcus, M., B. Santorini and M.A. Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. Computational Linguistics **19**: 313–330.

Michaelis, L.A. and S.F. Hartwell. Forthcoming. Lexical subjects and the conflation strategy. In N. Hedberg and R. Zacharski (eds), *Topics in the Grammar–Pragmatics Interface: Papers in Honor of Jeanette K. Gundel*. Benjamins, Amsterdam.

Morolong, M. and L. Hyman. 1977. Animacy, objects, and clitics in Sesotho. Studies in African Linguistics **8**: 199–218.

Newmeyer, F. 2003. Grammar is grammar and usage is usage. Language **79**: 682–707.

Oehrle, R.T. 1976. *The grammar of the English dative alternation*. PhD dissertation, Cambridge, MA, MIT Department of Linguistics and Philosophy.

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*. MIT Press, Cambridge, MA.

Polinsky, M. 1994. Double objects in causatives: towards a study of coding conflict. Studies in Language **19**: 129–221.

Polinsky, M. 1996. The double object construction in spoken Eastern Armenian. In *NSL. 8. Linguistic Studies in the Non-Slavic Languages of the Commonwealth of Independent States and the Baltic Republics*, pp. 307–335. Chicago Linguistic Society, The University of Chicago, Chicago.

Polinsky, M. 1997. Dominance in precedence: SO/OS languages. In *CLS 33. Papers from the Panels on Linguistic Ideologies in Contact, Universal Grammar, Parameters and Typology, the Perception of Speech and other Acoustic Signals*, pp. 253–269. Chicago Linguistic Society, The University of Chicago, Chicago.

Prat-Sala, M. and H.P. Branigan. 2000. Discourse constraints on syntactic processing in language production: A cross-linguistic study of English and Spanish. J. Memory and Language **42**: 168–182.

Prince, E. 1981. Toward a new taxonomy for given-new information. In P. Cole (ed.), *Radical Pragmatics*, pp. 223–255. Academic Press, New York.

R Development Core Team. 2004. *R: A language and environment for statistical computing*. On-line at http://www.R-project.org.

Snyder, K. 2003. *The relationship between form and function in ditransitive constructions*, PhD dissertation. University of Pennsylvania, Department of Linguistics, Philadelphia.

Szmrecsanyi, B. 2004a. On operationalizing syntactic complexity. In G. Purnelle, C. Fairon and A. Dister (eds), *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-la-Neuve, March 10–12, 2004*, pp. 1032–1039. Presses Universitaires de Louvain, Louvain-la-Neuve.

Szmrecsanyi, B. 2004b. *Persistence phenomena in the grammar of spoken English*, PhD dissertation, Albert-Ludwigs-Universität Freiburg Philology Faculty, Freiburg.

Szmrecsanyi, B. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. Corpus Linguistics and Linguistics Theory **1**: 113–149.

Thompson, S. 1990. Information flow and dative shift in English discourse. In J.A. Edmondson, F. Crawford and P. Mühlhäusler (eds), *Development and Diversity, Language Variation Across Space and Time*, pp. 239–253. Summer Institute of Linguistics, Dallas, TX.

Venables, W.N. and B.D. Ripley. 2002. *Modern Applied Statistics with S-Plus*. Springer, New York.

Wasow, T. 2002. *Postverbal Behavior*. CSLI, Stanford.

Weiner, E.J. and W. Labov. 1983. Constraints on the agentless passive. J. Linguistics **19**: 29–58.

Williams, R.S. 1994. A statistical analysis of English double object alternation. Issues in Applied Linguistics **5**: 37–58.