Predicting the Dative Alternation

Imagine you are a child...
• 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

• 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:

\[Susan \text{ gave } [\text{toys}] [\text{to the children}]\]  
V NP PP
the prepositional dative structure:

Susan gave [toys] [to the children]

the double object structure:

Susan gave [the children] [toys]
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?
The problem is interesting for many reasons
The problem is interesting for many reasons

- the psychology of language: learning, production...
The problem is interesting for many reasons

- the psychology of language: learning, production...
- computer science: natural language generation systems
The problem is interesting for many reasons

- the psychology of language: learning, production...
- computer science: natural language generation systems
- education: second language acquisition
The problem is interesting for many reasons

- The psychology of language: learning, production...
- Computer science: natural language generation systems
- Education: second language acquisition
- English literature: quantitative studies of style
Traditional theoretical linguistics:

- Problem too difficult to tackle.  
  (We must idealize.)
Traditional theoretical linguistics:

- Problem too difficult to tackle. (We must idealize.)
- Problem uninteresting for us. ("Grammar is grammar and usage is usage.")
It’s not too difficult.

Using some tools which have been employed in other areas of our field (Baayen 2008)…

*we can correctly predict 94% of the actual choices of dative constructions in a corpus of natural spontaneous conversations* $^a$

$^a$the 3,000,000 word Switchboard collection of recorded telephone conversations
It *is* interesting for theoretical linguistics.

We learn that . . .

*using traditional methods of data collection in theoretical linguistics, we have underestimated the space of grammatical possibility*

*persistent questions about what kinds of data are valid to use for linguistic theory can be answered empirically*
Part I
A natural hypothesis:

predicting different dative structures from different meanings
Two ways of viewing the same giving event:\(^a\)

causing a change of state (possession)
causing a change of place (movement to a goal)

\[(\text{possession}) \Rightarrow V \ NP \ NP\]
\[\textit{Susan gave the children toys}\]

\[(\text{movement to goal}) \Rightarrow V \ NP \ [\text{to NP}]\]
\[\textit{Susan gave toys to the children}\]

\(^a\text{Pinker 1989; Gropen, Pinker, Hollander, Goldberg, and Wilson 1989}\)
Evidence from give idioms—a—
no movement to a goal, no PP dative:

That movie gave me the creeps.
*That movie gave the creeps to me.
The lighting here gives me a headache.
*The lighting here gives a headache to me.

---

*a* Oehrle 1976 and many linguists thereafter; recently in *Linguistic Inquiry* (2001: 261)
Try Google…

GIVE THE CREEPS TO
many examples like these:

This life-sized prop will **give the creeps to just about anyone**! Guess he wasn’t quite dead when we buried him!\(^a\)

...Stories like these must **give the creeps to people whose idea of heaven is a world without religion**. . . \(^b\)

\(^a\)http://www.frightshop.com/
\(^b\)enquirer.com/editions/2001/09/30/loc_lords_gym.html
Try Google…

GIVE A HEADACHE TO
many examples like these:

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 form gaze Sarah vertigo.a

Design? Well, unless you take pride in giving a headache to your visitors with a flashing background? no.b

---

ahttp://lair.echidnoyle.org/rpg/log/27.html
bhttp://members.tripod.com/~SailorMoonWorstOfWeb/archive/RunJan01.html
The Googled examples are not supposed to be possible.
Evidence from ‘verbs of continuous imparting of force’\textsuperscript{a}—focus on movement to goal, only PP dative:

\begin{quote}
I carried/pulled/pushed/schlepped/lifted/lowered/hauled the box to John.

\textit{*I carried/pulled/pushed/schlepped/lifted/lowered/hauled John the box.}
\end{quote}

\textsuperscript{a}Pinker (1989:110–111), Levin (1993: 46, 114), many other authors.
Try Google…
CARRIED HER A
PUSHED HIM THE
etc.
some examples like these:

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

As Player A pushed him the chips, all hell broke loose at the table.\(^b\)


\(^b\)www.cardplayer.com/?sec=afeature&art_id=165
Nothing like heart burn food. “I have the tums.” Nick joked. He **pulled himself a steaming piece of the pie.** “Thanks for being here.”

“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...”

---

[a]www.realityfanfiction.addr.com/storm3.html  
[b]www.angelfire.com/wa2/bozyby/hold1.html
Therefore, when he got to purgatory, Buddha lowered him the silver thread of a spider as his last chance for salvation.\textsuperscript{a}

\textsuperscript{a}www.inch.com/ fujimura/ImofGrmain.htm
The Googled examples are not supposed to be possible.
Valuable data or sporadic errors?
Valuable data or sporadic errors?

- they sound fine in their contexts of use
Valuable data or sporadic errors?

- they sound fine in their contexts of use
- their structures are principled
Compare:

*That movie gave the creeps to me.

...Stories like these must **give the creeps**
to people whose idea of heaven is a world without religion...^a

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

That movie gave me the creeps.

The longer phrase is placed at the end — the principle of end weight. (Behaghel 1910, Wasow 2002)

Idioms like *give the creeps* have a strong bias toward the double object construction, but the principle of end weight overrides it.
Compare:

??Karen hand-carried a man a form.

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

??He dragged a guest a can of beer.

‘Well...it started like this...’” Shinbo explained while Sumomo **dragged him a can of beer** and opened it for him, ...
Notice:

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

…Shinbo explained while Sumomo dragged him a can of beer and opened it for him, …

The referent of the first object in the double object construction is given in the immediately preceding discourse; the second object is new and indefinite.
In contrast, in the worse-sounding constructed examples—

**Karen hand-carried a man a form.**

**He dragged a guest a can of beer.**

—the referent of the first object in the double object construction is not given in the immediately preceding discourse, and in fact is new.
In other words, the referent of the first object is normally more *discourse accessible* than that of the second object, as well as *more definite, pronominal, and shorter* — the principle of “receiver-entity differentiation” in double object constructions (Collins 1985)
Collins’ (1985) data

NP NP

NP PP

count

Receiver Entity

count

Receiver Entity

given
accessible
new
The data are highly skewed: 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.
Collins’ (1985) data
(proportions of NPs in the two dative structures)
The dative structures tend to be chosen so that *given referents precede nongiven referents in linear order* (Halliday 1970, Thompson 1990).

Similarly, *pronouns precede nonpronouns, definites precede indefinites, and shorter precede longer.*

This is quantitative ‘*Harmonic Alignment’* of various scales with syntactic position (cf. Optimality Theory).
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.
Conclusions from Part I:

Linguistic intuitions of ungrammaticality are a poor guide to the space of grammatical possibility.

Usage data reveals generalizations which we are sometimes blind to.

English dative verbs have more syntactic flexibility than we thought.

We can’t predict the dative alternation from meaning alone.
Part II
Predicting the dative alternation from multiple variables
Corpus studies of English have found that various properties of the recipient and theme have a quantitative influence on dative syntax (Thompson 1990, Collins 1995, Snyder 2003, Gries 2003, ao):

- discourse accessibility
- relative length
- pronominality
- definiteness
- animacy

⇒

dative construction choice
Yet what really drives the English dative alternation remains unclear…

- correlated variables seem to support reductive theories
- pooled data may invalidate grammatical inference
- nominal factors may derive from verb sense semantics
- cross-corpus differences appear to undermine the relevance of corpus studies to grammatical theory
1. The Problem of Correlated Variables
What really drives the dative alternation remains unclear because of pervasive correlations in the data:

- short pronouns
  - definite
  - discourse-given
  - usually animate
  - often discourse-given

- animates
  - often definite
  - frequently referred to pronominally
  - usually have nicknames (short) …

Correlations tempt us into reductive theories that explain effects in terms of just one or two variables (e.g. Hawkins 1994, Snyder 2003)
A beautifully simple theory:

1. Givenness correlates with shorter, less complex expressions (less description needed to identify)

2. Shorter expressions occur earlier in order to facilitate parsing (more complex after less)

Apparent effects of givenness (and correlated properties like animacy) could reduce to the preference to process syntactically complex phrases later than simple ones (Hawkins 1994).
Question 1:
Are these effects of discourse accessibility, animacy, and the like the epiphenomena of syntactic complexity effects in parsing?
Use **logistic regression** to control simultaneously for multiple variables related to a binary response.\(^a\)

Use **large samples of richly annotated data**: 2360 dative observations from the three-million-word Switchboard collection of recorded telephone conversations.

\(^a\)Williams 1994; Arnold, Wasow, Losongco, and Ginstrom 2000; cf. Gries 2003
explanatory variables:

- discourse accessibility, definiteness, pronominality, animacy (Thompson 1990, Collins 1995)
- differential length in words of recipient and theme (Arnold et al. 2000, Wasow 2002, Szmrecsanyi 2004b)
- number, person (Aissen 1999, 2003; Haspelmath 2004; Bresnan and Nikitina 2009)
- concreteness of theme
plus 5 broad semantic classes of uses of verbs which participate in the dative alternation:

- abstract (abbreviated ‘a’): give it some thought
- transfer of possession (‘t’): give an armband, send
- future transfer of possession (‘f’): owe, promise
- prevention of possession (‘p’): cost, deny
- and communication (‘c’): tell, give me your name, said on a telephone
Model A:

Response ~

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)

The Logistic Regression Model

\[
\text{logit}[\text{Probability(} \text{Response} = 1)] = X\beta
\]

or

\[
\text{Probability(} \text{Response} = 1) = \frac{1}{1 + \exp(-X\beta)}
\]
## Classification Table for Model A

(1 = PP; cut value = 0.50)

<table>
<thead>
<tr>
<th>Predicted</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observed</th>
<th>0</th>
<th>1796</th>
<th>63</th>
<th>97%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1796</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>115</td>
<td>386</td>
<td></td>
<td>77%</td>
</tr>
</tbody>
</table>

Overall: 92%

% Correct from always guessing NP NP (=0): 79%
Model A plot of observed against predicted responses

![Model A plot of observed against predicted responses](image)
How well does the model generalize to new data?

Divide the data randomly 100 times into a training set of sufficient size for the model parameters ($n = 2000$) and a testing set ($n = 360$).

Fit the Model A parameters on each training set and score its predictions on the unseen testing set.

Mean overall score (average \% correct predictions on unseen data) = 92\%.
All of the model predictors except for number of recipient are significant.

All, $p < 0.001$ except person of recipient, number of theme, and concreteness of theme, $p < 0.05$.

—And most predictors contribute to the explanatory value of the model*

Increase in $-2 \log$ likelihood (decrease in model goodness-of-fit) for predictor removed from full model.

- **Pron Rec** ***
- **Pron Theme** ***
- **LogRThmDiff** ***
- **Def of Theme** ***
- **Animacy of Rec** ***
- **Previous** *
- **Num of Theme** *
- **Def of Rec**

increase in 2 log likelihood
(decrease in model goodness of fit)
Qualitative view of quantitative findings:

discourse given ▼ not given
animate ▼ inanimate
definite ▼ indefinite
pronoun ▼ non-pronoun
less complex ▼ more complex

V NP_{rec} NP_{thm}

V NP_{thm} PP_{rec}
The model formula:

\[
\text{Probability}\{\text{Response} = 1\} = \frac{1}{1 + e(-X\hat{\beta})}, \quad \text{where}
\]

\[
X\hat{\beta} = 
\]

0.95
-1.34\{c\} + 0.53\{f\} - 3.90\{p\} + 0.96\{t\}
+0.99\{\text{accessibility of recipient} = \text{nongiven}\}
-1.1\{\text{accessibility of theme} = \text{nongiven}\}
+1.2\{\text{pronominality of recipient} = \text{nonpronoun}\}
-1.2\{\text{pronominality of theme} = \text{nonpronoun}\}
+0.85\{\text{definiteness of recipient} = \text{indefinite}\}
-1.4\{\text{definiteness of theme} = \text{indefinite}\}
+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}\}
-1.1\{\text{parallelism} = 1\} - 1.2 \cdot \text{length difference (log scale)}

and \(\{c\} = 1\) if subject is in group \(c\), 0 otherwise; likewise for \(f, p, t\).
Positive coefficients favor PP dative, negative favor NP:

\[ +0.99 \{ \text{accessibility of recipient} = \text{nongiven} \} \]
\[ -1.1 \{ \text{accessibility of theme} = \text{nongiven} \} \]
\[ +1.2 \{ \text{pronominality of recipient} = \text{nonpronoun} \} \]
\[ -1.2 \{ \text{pronominality of theme} = \text{nonpronoun} \} \]
\[ +0.85 \{ \text{definiteness of recipient} = \text{indefinite} \} \]
\[ -1.4 \{ \text{definiteness of theme} = \text{indefinite} \} \]
\[ +2.5 \{ \text{animacy of recipient} = \text{inanimate} \} \]
\[ +0.48 \{ \text{person of rec} = \text{nonlocal} \} \]
\[ -1.2 \cdot \text{length difference (log scale)} < 0 \quad [\text{len(rec)} > \text{len(th)}] \]
\[ -1.2 \cdot \text{length difference (log scale)} > 0 \quad [\text{len(rec)} < \text{len(th)}] \]

This is harmonic alignment with syntactic position
Answer to Question 1:

The Harmonic Alignment effects on syntactic choice cannot be reduced to one single predictor. In particular, the syntactic complexity in parsing hypothesis does not explain the influence of given-ness (and animacy, etc.) on the choice of dative syntax.
2. The problem of pooling different speakers’ data
Question 2
A persistent question about corpus studies of grammar . . .

in Newmeyer’s (2003: 696) words:

“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.”
This is an empirical question:

*What the speakers share in their choices of dative syntax might outweigh their differences.*
The Switchboard Corpus is annotated for speaker identity.

424 total speakers ⇒ total of 2360 instances of dative constructions

- 228 speakers ⇒ 4 – 7 each
- 106 speakers ⇒ 8 – 12 each
- 42 speakers ⇒ 13 – 19 each
- 11 speakers ⇒ 20+ each

The data are extremely unbalanced.
Speaker identity is a source of unknown dependencies in the data.

The effect of these unknown dependencies on the reliability of the estimates can be estimated from the observed data using modern statistical techniques:

When data dependencies fall into many small clusters (each speaker defines a ‘cluster’), assume a ‘working independence model’ (our Model A) and revise the covariance estimates using bootstrap sampling with replacement of entire clusters.

---

aEfron and Tibshirani (1986, 1993); Feng, McLerran, Grizzle (1996); Harrell (2001)
in other words…

Create multiple copies of the data by resampling from the speakers. The same speakers’ data can randomly occur many times in each copy.

Repeatedly re-fit the model to these copies of the data and use 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.*
Result: the model coefficients are the same; the confidence intervals of the odds ratios\(^a\) are wider, reflecting the reduction of independent observations in our data caused by the presence of clusters of speaker dependencies.

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%–50%, at chance.

We want the confidence intervals to stay nicely away from 1!

\(^a\)—the intervals in which you can be confident that the chance of error stays below threshold (\(<5\%)\)
## Model A

Relative magnitudes of significant effects with corrected error estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficient</th>
<th>Odds Ratio PP</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>inanimacy of recipient</td>
<td>2.54</td>
<td>12.67</td>
<td>5.56–28.87</td>
</tr>
<tr>
<td>nonpronominality of recipient</td>
<td>1.17</td>
<td>3.22</td>
<td>1.70–6.09</td>
</tr>
<tr>
<td>nongivenness of recipient</td>
<td>0.99</td>
<td>2.69</td>
<td>1.37–5.3</td>
</tr>
<tr>
<td>transfer semantic class</td>
<td>0.96</td>
<td>2.61</td>
<td>1.44–4.69</td>
</tr>
<tr>
<td>indefiniteness of recipient</td>
<td>0.85</td>
<td>2.35</td>
<td>1.25–4.43</td>
</tr>
<tr>
<td>plural number of theme</td>
<td>0.50</td>
<td>1.65</td>
<td>1.05–2.59</td>
</tr>
<tr>
<td>person of recipient</td>
<td>0.48</td>
<td>1.62</td>
<td>1.06–2.46</td>
</tr>
<tr>
<td>nongivenness of theme</td>
<td>-1.05</td>
<td>0.35</td>
<td>0.19–0.63</td>
</tr>
<tr>
<td>structural parallelism in dialogue</td>
<td>-1.13</td>
<td>0.32</td>
<td>0.22–0.47</td>
</tr>
<tr>
<td>nonpronomininality of theme</td>
<td>-1.18</td>
<td>0.31</td>
<td>0.19–0.50</td>
</tr>
<tr>
<td>length difference (log scale)</td>
<td>-1.21</td>
<td>0.3</td>
<td>0.22–0.4</td>
</tr>
<tr>
<td>communication semantic class</td>
<td>-1.34</td>
<td>0.26</td>
<td>0.13–0.55</td>
</tr>
<tr>
<td>indefiniteness of theme</td>
<td>-1.37</td>
<td>0.25</td>
<td>0.15–0.44</td>
</tr>
</tbody>
</table>
Answer to Question 2:
The influence of discourse accessibility, animacy, and the like on dative syntax remain significant when differences in speaker identity are taken into account.

What the speakers share in the choice of dative syntax outweighs their differences.
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 NPs come from different verbs and different senses of the same verb!
The properties of recipients and themes depend on the verbs which describe the transfer events they are participating in. For example:

*bring* is nearly three times more likely to have a given recipient than *take*

*take* is over seven times more likely to have a nongiven recipient than *bring*.

(The goal of bringing is usually located near the speaker, the goal of taking is usually located away from the speaker)
The properties of the NP arguments are conditional not only on the verb they occur with, but also on the specific sense of the verb used.

For example, *give* has a larger than average proportion of inanimate recipients:

Um, but still, it *gives it some variety.*

but I’m going to *give it thumbs down.*

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* is having *only* animate recipients in our dataset (because we normally communicate with people).
Another example: the recipients of paying money are more likely to be animate and given than the ‘recipients’ of paying attention.

<table>
<thead>
<tr>
<th></th>
<th>animate</th>
<th>inanimate</th>
<th>given</th>
<th>nongiven</th>
</tr>
</thead>
<tbody>
<tr>
<td>pay (transfer)</td>
<td>83</td>
<td>1</td>
<td>61</td>
<td>23</td>
</tr>
<tr>
<td>pay (abstract)</td>
<td>17</td>
<td>40</td>
<td>31</td>
<td>26</td>
</tr>
</tbody>
</table>

(We are more likely to pay money to recipients that we already know, who are also likely as money-users to be animate. We are more likely to pay attention to less expected, nongiven things.)
Question 3:
Do the apparent effects of givenness and animacy on the choice of dative syntax hold, when they are conditioned on specific verb senses?
38 verbs $\times$ 5 semantic classes

Examples:

give.t = transfer: give you an armband

give.c = communication: give me this cock and bull story . . .

give.a = abstract: give that a lot of thought

pay.t = transfer: pay somebody good money

pay.a = abstract: pay attention to cats

55 verb senses in use in dataset
Use a **multilevel model** to condition the binary response on the verb sense:

**Model B: Response ~**

**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

**A Generalized Linear Model with a Single Random Intercept**

\[
\text{logit}[\Pr(Y_{ij} = y_{ij}|u_i)] = X_{ij}\beta + u_i
\]

The conditional probability of a response given a cluster \(i\) is systematically linked to a linear combination of fixed cross-cluster explanatory variables \(X_{ij}\) and a randomly varying normally distributed cluster effect.
% Classification Table for Model B

(1 = PP; cut value = 0.50)

<table>
<thead>
<tr>
<th>Predicted</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>68</td>
</tr>
<tr>
<td>Overall:</td>
<td></td>
</tr>
</tbody>
</table>
Model B plot of observed against predicted responses

Proportions of observed PP realization

Grouped predicted probabilities of PP realization
How well does Model B generalize to new data?

Divide the data randomly 100 times into a training set of sufficient size for the model parameters \((n = 2000)\) and a testing set \((n = 360)\).

Fit the Model B parameters on each training set and score its predictions on the unseen testing set.

Mean overall score (average \% correct predictions on unseen data) = 94\%. Very good!
The model formula showing harmonic alignment:

\[
\text{Probability}\{\text{Response} = 1\} = \frac{1}{1 + e^{(-X\hat{\beta} + u)}}, \quad \text{where}
\]

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

and \(\hat{u} \sim N(0, 2.27)\).
Relative magnitudes of significant effects in Model B

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Odds Ratio PP</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonpronominality of recipient</td>
<td>1.73</td>
<td>5.67</td>
<td>3.25–9.89</td>
</tr>
<tr>
<td>inanimacy of recipient</td>
<td>1.53</td>
<td>5.62</td>
<td>2.08–10.29</td>
</tr>
<tr>
<td>nongivenness of recipient</td>
<td>1.45</td>
<td>4.28</td>
<td>2.42–7.59</td>
</tr>
<tr>
<td>indefiniteness of recipient</td>
<td>0.72</td>
<td>2.05</td>
<td>1.20–3.5</td>
</tr>
<tr>
<td>plural number of theme</td>
<td>0.72</td>
<td>2.06</td>
<td>1.37–3.11</td>
</tr>
<tr>
<td>structural parallelism in dialogue</td>
<td>-1.13</td>
<td>0.32</td>
<td>0.23–0.46</td>
</tr>
<tr>
<td>nongivenness of theme</td>
<td>-1.17</td>
<td>0.31</td>
<td>0.18–0.54</td>
</tr>
<tr>
<td>length difference (log scale)</td>
<td>-1.16</td>
<td>0.31</td>
<td>0.25–0.4</td>
</tr>
<tr>
<td>indefiniteness of theme</td>
<td>-1.74</td>
<td>0.18</td>
<td>0.11–0.28</td>
</tr>
<tr>
<td>nonpronominality of theme</td>
<td>-2.17</td>
<td>0.11</td>
<td>0.07–0.19</td>
</tr>
</tbody>
</table>
Answer to Question 3:
The influence of givenness, animacy and the other variables on the choice of dative syntax remain significant when they are conditioned on specific verb senses.
4. The problem of cross-corpus differences
Question 4:

*Does it make sense to relate frequencies of usage to grammar? (Keller and Asudeh 2002: 240)*

*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.*
In fact it is true that
the frequencies of double-object constructions in the Switchboard collection of recordings of telephone conversations ≠
frequencies in the Treebank Wall Street Journal collection of news and financial reportage

V NP NP’s = 79% of total Switchboard datives
(n = 2360)

V NP NP’s = 62% of total Wall Street Journal datives (n = 905)
Fit the same model to the combined data from two different corpora and compare fits to the components.

**Model C: Response ~**

**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

**random effect:** verb sense

Model C = Model B minus three factors (person, number, and parallelism) not marked in our Wall Street Journal dative dataset

Model C data has 110 different verb senses
# Model C Classification Table

(1 = PP; cut value = 0.50)

<table>
<thead>
<tr>
<th>Predicted:</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>96%</td>
</tr>
<tr>
<td>1</td>
<td>86%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observed:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2320 96</td>
</tr>
<tr>
<td>1</td>
<td>119 730</td>
</tr>
</tbody>
</table>

Overall: 93%
Model C plot of observed against predicted responses

Proportions of observed PP realization

Grouped predicted probabilities of PP realization
How well does Model C generalize to new data?

Divide the data randomly 100 times into a training set of sufficient size for the model parameters ($n = 2000$) and a testing set ($n = 1265$) and score its predictions on the unseen testing set.

Mean overall score (average % correct predictions on unseen data) = 92%.
## Model C on component corpora

<table>
<thead>
<tr>
<th>% NP NP’s</th>
<th>Switchboard</th>
<th>Wall Street Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>predicted</td>
<td>79%</td>
<td>63%</td>
</tr>
<tr>
<td>actual</td>
<td>79%</td>
<td>62%</td>
</tr>
</tbody>
</table>

*How can this be?*
Inputs vary. For example:

Wall Street Journal recipients: nouns outnumber pronouns 5 to 1
Switchboard recipients: pronouns outnumber nouns almost 4 to 1

The pressure 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:

**Model C on component corpora**

<table>
<thead>
<tr>
<th>% NP NP’s (non-pronouns)</th>
<th>Switchboard</th>
<th>Wall St. Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>predicted</td>
<td>49%</td>
<td>58%</td>
</tr>
<tr>
<td>actual</td>
<td>49%</td>
<td>55%</td>
</tr>
</tbody>
</table>

*Again, how can this be?*
Again, inputs vary. For example, among non-pronoun complements to dative verbs:

- **Wall Street Journal** median length differential (log scale) = 1.1
- **Switchboard** median length differential (log scale) = 0.69

The pressure 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.*
Answer to Question 4:

Some striking differences between different corpora can be explained as the response of the same model to quantitatively different inputs.

The statistical 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.
There is 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.

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 differences will be revealed in modeling the data...

Conclusions from Part II:

The kinds of questions that have been raised about usage data are *empirical questions*.

- correlated factors seeming to support reductive theories
- pooled data invalidating grammatical inference
- nominal factors deriving from verb sense semantics
- cross-corpus differences undermining corpus grammar
Answers can be found by using modern statistical theory and modeling strategies widely used in other fields (biology, medicine).