A Combinatorial Model of Variation in the English Dative Alternation

Max Bane
bane@uchicago.edu

Department of Linguistics
University of Chicago

February 7, 2010
Berkeley Linguistics Society 36
Nondeterministic variability in the choice of form is rife in language usage.
Nondeterministic variability in the choice of form is rife in language usage.

- Within-speaker: Across a single individual’s utterances.
- Between-speaker: Across the utterances of different individuals.
Examples of Variable Alternations

- **Phonological:**
  - Variable final -t/-d deletion in English dialects:

<table>
<thead>
<tr>
<th>(Coetzee 2004)</th>
<th>Pre-C</th>
<th>Pre-V</th>
<th>Pre-pause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicano English (Los Angeles)</td>
<td>n</td>
<td>3,693</td>
<td>1,574</td>
</tr>
<tr>
<td></td>
<td>% deleted</td>
<td>62%</td>
<td>45%</td>
</tr>
<tr>
<td>Tejano English (San Antonio)</td>
<td>n</td>
<td>1,738</td>
<td>947</td>
</tr>
<tr>
<td></td>
<td>% deleted</td>
<td>62%</td>
<td>25%</td>
</tr>
<tr>
<td>AAVE (Washington, DC)</td>
<td>n</td>
<td>143</td>
<td>202</td>
</tr>
<tr>
<td></td>
<td>% deleted</td>
<td>76%</td>
<td>29%</td>
</tr>
</tbody>
</table>
Examples of Variable Alternations

- **Phonological:**
  - Variable final -t/-d deletion in English dialects:

<table>
<thead>
<tr>
<th>Dialect</th>
<th>Pre-C</th>
<th>Pre-V</th>
<th>Pre-pause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicano English (Los Angeles)</td>
<td>3,693</td>
<td>1,574</td>
<td>1,024</td>
</tr>
<tr>
<td>% deleted</td>
<td>62%</td>
<td>45%</td>
<td>37%</td>
</tr>
<tr>
<td>Tejano English (San Antonio)</td>
<td>1,738</td>
<td>947</td>
<td>564</td>
</tr>
<tr>
<td>% deleted</td>
<td>62%</td>
<td>25%</td>
<td>46%</td>
</tr>
<tr>
<td>AAVE (Washington, DC)</td>
<td>143</td>
<td>202</td>
<td>37</td>
</tr>
<tr>
<td>% deleted</td>
<td>76%</td>
<td>29%</td>
<td>73%</td>
</tr>
</tbody>
</table>

- **Morphosyntactic:**
  - *was/were* in Buckie Scottish English (Smith 2000, Adger 2006, Adger & Smith 2005):

  (2) a. ... I thocht you were a diver... (7:262.41)
  b. ... I thocht you was a scuba diver. (7:259.21)
  c. ... we were lyin at anchor. (g:875.32)
  d. ... we was tired... (b:254.15)
The dative alternation in English:

a. I gave the dusty old book to my sister.
b. I gave my sister the dusty old book.
c. I gave to my sister the dusty old book.
The dative alternation in English:

- I gave the dusty old book to my sister.
- I gave my sister the dusty old book.
- I gave to my sister the dusty old book.


Syntactic, semantic, prosodic, etymological properties proposed to explain which verbs allow which variants with which arguments.
I will describe Anttila’s (2008, to appear) OT, prosodic model of the dative alternation. Will show how the model can be used to predict the variants’ frequency of usage in conversational corpora.
I will describe Anttila’s (2008, to appear) OT, prosodic model of the dative alternation.

Will show how the model can be used to predict the variants’ **frequency of usage** in conversational corpora.

- By supposing that the ranking of constraints (i.e. the grammar) is **underspecified** or **partial**.
- The number of possible ways to resolve the underspecification (i.e. number of consistent **total orders**) predicts corpus frequency.
- ⇒ Speakers resolve underspecification by **randomly sampling** total grammars consistent with partial.
I will describe Anttila’s (2008, to appear) OT, prosodic model of the dative alternation.

Will show how the model can be used to predict the variants’ frequency of usage in conversational corpora.

- By supposing that the ranking of constraints (i.e. the grammar) is underspecified or partial.
- The number of possible ways to resolve the underspecification (i.e. number of consistent total orders) predicts corpus frequency.
- ⇒ Speakers resolve underspecification by randomly sampling total grammars consistent with partial.

The Possible Variant Constructions

- English ditransitive verbs like *give* permit several syntactic arrangements of their arguments:

  (5) The Double Object Construction (D):
  a. A legitimate charity will give [you] [proof that your donation is tax deductible].
  b. Today I told [Sally] [the truth].

  (6) The Prepositional Construction (P):
  a. They’re willing to allocate [time and money] [to our cause].
  b. Please convey [my condolences] [to his family].
The Possible Variant Constructions

(7) The Heavy NP Shift Construction (H):
   a. I am going to reveal [to you] [everything I’ve learned in this business].
   b. At exactly midnight, radio [to the police] [your precise location and status].
The Possible Variant Constructions

(9) The Heavy NP Shift Construction (H):
   a. I am going to reveal [to you] [everything I’ve learned in this business].
   b. At exactly midnight, radio [to the police] [your precise location and status].

- Prepositional construction is always available, but different kinds of verbs and arguments may or may not allow double-object and NP shift.

(10) a. * Donate [my sister] [the book].  (No D)
     b. * I am going to reveal [to you] [it].  (No H)
The Possible Variant Constructions

- One logical possibility is never acceptable (except in some dialects, apparently):

  (11) * Give the book my sister.

- Furthermore, variation is possible for many combinations of verbs and arguments:
  - Please convey [my most heartfelt condolences] [to his family].
  - Please convey [to his family] [my most heartfelt condolences].
Corpus

  - Gotten by searching for 16 alternating/nonalternating verbs (6 1-foot, 10 2-foot) and examining first 100 results with relevant ditransitive context.
Corpus

  - Gotten by searching for 16 alternating/nonalternating verbs (6 1-foot, 10 2-foot) and examining first 100 results with relevant ditransitive context.

![Overall Corpus Frequencies](image)

N = 1601

- Prepositional (69.8%)
- Double Object (26.9%)
- NP Shift (3.31%)
Anttila’s (2008, to appear) OT Model

- Treats the alternation as a result of syntax/prosody interactions.
Anttila's (2008, to appear) OT Model

- Treats the alternation as a result of syntax/prosody interactions.
- Input Forms: specification of three quantities:
  - The number of feet in the verb
  - The number of lexical stresses in the goal
  - The number of lexical stresses in the theme
Anttila’s (2008, to appear) OT Model

- Treats the alternation as a result of syntax/prosody interactions.
- Input Forms: specification of three quantities:
  - The number of feet in the verb
  - The number of lexical stresses in the goal
  - The number of lexical stresses in the theme
- For convenience:
  - “give” refers to any verb whose stem = prosodic word = 1 foot, “donate” to any whose stem ≠ PW or whose stem = PW ≠ 1 foot.
  - “her” = stressless goal, “my sister” = 1-stress goal, “my little sister” = 2-stress goal.
Anttila’s (2008, to appear) OT Model

- Treats the alternation as a result of syntax/prosody interactions.
- Input Forms: specification of three quantities:
  - The number of feet in the verb
  - The number of lexical stresses in the goal
  - The number of lexical stresses in the theme
- For convenience:
  - “give” refers to any verb whose stem = prosodic word = 1 foot, “donate” to any whose whose stem $\neq$ PW or whose stem = PW $\neq$ 1 foot.
  - “her” = stressless goal, “my sister” = 1-stress goal, “my little sister” = 2-stress goal.
- Written as “give(her, the book)”, “donate(my little sister, it)”, etc.
Foot Structure of Verbs

Follows Grimshaw’s (2005:115) footing analysis:

1. stem = PW = 1 foot (give-type)
2. stem = PW ≠ 1 foot (donate-type)
3. stem ≠ PW (donate-type)
Introduction

The Dative Alternation

The OT Model

Variation from Sampling

Summary

donate(her, it)

N = 37

P (100%)

D (0%)

H (0%)

P (48.5%)

D (38.5%)

H (13%)

P (98.4%)

D (0.407%)

H (1.22%)

P (100%)

D (0%)

H (0%)

P (100%)

D (0%)

H (0%)

donate(my little sister, the book)

N = 246

P (94.4%)

D (5.71%)

H (0.93%)

give(her, it)

N = 2

P (100%)

D (0%)

H (0%)

P (5.64%)

D (94.4%)

H (0%)

give(my little sister, the book)

N = 134

P (47%)

D (53%)

H (0%)

give(my sister, the old book)

N = 149

P (100%)

D (0%)

H (0%)
Output forms: 4 linearizations $\times$ 2 prosodic phrasings.

(VERB) (GOAL) (THEME)  (VERB GOAL) (THEME)
(VERB) (THEME) (to GOAL)  (VERB THEME) (to GOAL)
(VERB) (to GOAL) (THEME)  (VERB to GOAL) (THEME)
(VERB) (THEME) (GOAL)  (VERB THEME) (GOAL)
The Constraint Set
(Anttila 2008, Anttila et al to appear)

- 10 constraints on what output form may be chosen for a given input form.

SYNTAX
- The goal NP must form an XP with its head.
- Penalizes VERB THEME GOAL linearizations.

WRAP-XP
- An XP must be contained in a prosodic phrase.
- Penalizes (VERB)(GOAL), but not (VERB)(to GOAL).
*(X)*

- Penalizes prosodic phrases without any lexical stresses.

*CLASH*

- Avoid multiple lexical stresses in a prosodic phrase.
- Penalizes each lexical stress beyond the first in a single phrase.

*TERNARY*

- No ternary prosodic phrases.
- Penalizes a two-foot verb sharing a phrase with anything else.

*TO*

- Penalizes prepositions.
STRESS-TO-STRESS

- Word stress and sentence stress (i.e., final) should coincide.
- Violations equal to the number of stresses in non-final prosodic phrases.

*P-PHRASE

- Avoid prosodic phrases.
- Violations equal to the number phrases.

FOCUS(GOAL)

- Penalizes non-final goal.

FOCUS(THEME)

- Penalizes non-final theme.
Example Tableaux

The tableau for the input ‘\textit{give(my sister, the book)}’. Four possible winners (a, b, e, h).

<table>
<thead>
<tr>
<th></th>
<th>SYNTAX</th>
<th>WRAP-XP</th>
<th>(x)</th>
<th>*TERNARY</th>
<th>*CLASH</th>
<th>FOCUS(Theme)</th>
<th>FOCUS(Goal)</th>
<th>STRESS-TO-STRESS</th>
<th>to</th>
<th>*-PHRASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>( \rightarrow ) (give)(the book)(to my sister)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>b.</td>
<td>( \rightarrow ) (give)(to my sister)(the book)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>c.</td>
<td>(give)(the book)(my sister)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>d.</td>
<td>(give)(my sister)(the book)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>e.</td>
<td>( \rightarrow ) (give the book)(to my sister)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>f.</td>
<td>(give to my sister)(the book)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>g.</td>
<td>(give the book)(my sister)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>h.</td>
<td>( \rightarrow ) (give my sister)(the book)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
</tbody>
</table>
The tableaux for the input ‘give(my sister, it)’. One possible winner (e).

<table>
<thead>
<tr>
<th></th>
<th>SYNTAX</th>
<th>WRAP-XP</th>
<th>*(x)</th>
<th>*TERNARY</th>
<th>*CLASH</th>
<th>FOCUS(Theme)</th>
<th>FOCUS(Goal)</th>
<th>STRESS-TO-STRESS</th>
<th>to</th>
<th>PHRASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>(give)(it)(to my sister)</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>b</td>
<td>(give)(to my sister)(it)</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>c</td>
<td>(give)(it)(my sister)</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>d</td>
<td>(give)(my sister)(it)</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>e</td>
<td>(give it)(to my sister)</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>f</td>
<td>(give to my sister)(it)</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>g</td>
<td>(give it)(my sister)</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>h</td>
<td>(give my sister)(it)</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>
Many rankings of the constraints give the same “language”
  (set of input-output pairings)
10! = 3,628,800 possible constraint rankings.
217 distinct possible languages.
Predicting usage frequency with this model

- The basic idea:
Predicting usage frequency with this model

The basic idea:

- The **number of constraint rankings** that yield a given input-output pairing ("ranking volume") affects how likely that pairing is to be observed.

Predicting usage frequency with this model

- The basic idea:
  - The number of constraint rankings that yield a given input-output pairing ("ranking volume") affects how likely that pairing is to be observed.
  - Possible rankings may be restricted by an underspecified grammar. E.g.:
    - A partial ordering of constraints. Statements of the form $X \gg Y; \ Z \gg W;$ etc., but with some constraints left unranked relative to each other.
    - A collection of Elementary Ranking Conditions (ERCs; Prince 2002). Restrictions implied by observed optima.
Counting Rankings

Need to be able to count the number of rankings (subject to some partial grammar) that generate a given input-output pair.

Riggle (2008) describes a feasible way of doing this

Without simply enumerating all $k!$ rankings, which is infeasible even for small $k$. 
In the following plots, each point corresponds to a single input-output mapping, e.g.:

- `give(her, it) → “give it to her”`
- `donate(my sister, the book) → “donate my sister the book”;`
- `give(my little sister, it) → “give to my little sister it”`.

Different phonological phrasings collapsed, because we can only observe linearizations in the textual corpus.

- R-volumes of different phrasings for same linearization added together.
Unrestricted: Counting among all 10! rankings

If we suppose a **fully unspecified** grammar, so that all 10! rankings are possible choices:
If we suppose a **fully unspecified** grammar, so that all 10! rankings are possible choices:

![Graph showing significant but weak correlation](image)

- **Corpus Frequency**
- **Number of Rankings**

Significant but weak correlation $R^2 = 0.20$. Many rankings choose categorically bad (0-frequency) mappings. **Not** a very good linear trend.
A note about regressions

- Regressing corpus frequency against number of consistent rankings.
- Must use **weighted** least-squares regressions.
- Each frequency point weighted by the inverse of its squared standard binomial error.
  - Corrects for unequal variance/certainty of each datapoint.
  - Verified by standard diagnostics (not significantly non-normal residuals, linear Q-Q line).
A note about regressions

- Regressing corpus frequency against number of consistent rankings.
- Must use **weighted** least-squares regressions.
- Each frequency point weighted by the inverse of its squared standard binomial error.
  - Corrects for unequal variance/certainty of each datapoint.
  - Verified by standard diagnostics (not significantly non-normal residuals, linear Q-Q line).
- Alternative approach would be multinomial logistic regression.
Assessing performance

- A problem with $R^2$ is that it only describes percentage of variance explained when data are grouped according to how the model distinguishes inputs (number of feet in verb, stresses in arguments).
Assessing performance

- A problem with $R^2$ is that it only describes percentage of variance explained when data are grouped according to how the model distinguishes inputs (number of feet in verb, stresses in arguments).
- An alternative: “accuracy”.
  - For each input, the regression assigns a probability to each corresponding output.
  - Suppose that the output with highest probability were always chosen (“plurality wins”).
  - Accuracy: how often, across corpus tokens, would the plurality-wins prediction match the observed output?
Assessing performance

- A problem with $R^2$ is that it only describes percentage of variance explained when data are grouped according to how the model distinguishes inputs (number of feet in verb, stresses in arguments).
- An alternative: “accuracy”.
  - For each input, the regression assigns a probability to each corresponding output.
  - Suppose that the output with highest probability were always chosen (“plurality wins”).
  - Accuracy: how often, across corpus tokens, would the plurality-wins prediction match the observed output?
- Baseline 69.8% accuracy if always guess prepositional output.
- Theoretical maximum accuracy is 81.2%, if predicted probabilities equal empirical frequencies.
- 48.2% accuracy with fully unspecified grammar.
Anttila proposes that three of the constraints, SYNTAX, WRAP-XP, *(X), are undominated (though unranked relative to each other), to rule out most categorically bad constructions.
Anttila proposes that three of the constraints, SYNTAX, WRAP-XP, *(X), are undominated (though unranked relative to each other), to rule out most categorically bad constructions.

Corresponds to this partial order:
Anttila proposes that three of the constraints, SYNTAX, WRAP-XP, *(X), are undominated (though unranked relative to each other), to rule out most categorically bad constructions.

Corresponds to this partial order:

- If we only count among the rankings consistent with this partial grammar...
Restricted: Counting with 3 undominated constraints

Now counting among the $3! \times 7!$ rankings consistent with the partial order. $R^2 = 0.697$, Accuracy = 68.0%.
Still not a very great linear relationship.

How well in principle can the grammar-sampling model predict the structure of variation?

Free parameter: the partial grammar.
- What is the partial grammar that best predicts the usage frequencies?
- Best $R^2$, best token accuracy.
A heuristic procedure for finding the underspecified grammar (partial order) within which the correlation between number of rankings and corpus frequency is maximal.
Finding a highly predictive partial grammar

- A heuristic procedure for finding the underspecified grammar (partial order) within which the correlation between number of rankings and corpus frequency is maximal.

- A local hill-climbing search
  1. Begin with some partial order $P$.  
  2. Consider all neighboring partial orders gotten by adding or removing one binary statement ($X \gg Y$) to/from $P$. 
  3. Set $P$ equal to the best such neighbor (highest $R^2$).
  4. Repeat.  
     - … until no improvements on $P$ can be found.
A heuristic procedure for finding the underspecified grammar (partial order) within which the correlation between number of rankings and corpus frequency is maximal.

A local hill-climbing search

1. Begin with some partial order $P$.
2. Consider all neighboring partial orders gotten by adding or removing one binary statement ($X \gg Y$) to/from $P$.
3. Set $P$ equal to the best such neighbor (highest $R^2$).
4. Repeat.
   - ... until no improvements on $P$ can be found.

The solution found depends on what partial order we start with.
Starting with an empty partial order

If we start from no ranking statements, the best partial order found for maximizing \( R^2 \) is:

\[
\begin{align*}
\text{focusGoal} & \quad \text{syntax, wrap-xp} \\
\text{noTo} & \quad \text{noClash}
\end{align*}
\]

(consistent with 90,720 total orders)
Regression

Linear regression under $\hat{P}_{\varnothing R^2}$

- $R^2 = 0.9099$
- $p \leq 1.26e-24$

Searching for best $R^2$ gives: $R^2 = 0.91$, accuracy 80.6%.
Most predictive partial order

- Best partial order found by the search procedure after 100 restarts from random initial partial orders:
Sampling within the best partial order of 100 random restarts

Weighted linear regression under best partial order of 100 random restarts

Good linear trend, $R^2 = 0.98$, accuracy = 80.6%.
Argued for the empirical adequacy of:
Summary

- Argued for the empirical adequacy of:
  - An essentially **prosodic** analysis of the dative alternation
Argued for the empirical adequacy of:

- An essentially prosodic analysis of the dative alternation
- Relating variable usage to the combinatorics of an underspecified grammar.
Argued for the empirical adequacy of:

- An essentially prosodic analysis of the dative alternation
- Relating variable usage to the combinatorics of an underspecified grammar.
- A model in which speakers randomly sample the ways of resolving an underspecified grammar.
  - In contrast to models where speakers learn and apply real-valued statistical parameters/probabilities directly:
    - Stochastic OT (Boersma and Hayes 2001), MaxEnt (Goldwater and Johnson 2003, Hayes and Wilson 2008), Variable Rules (Labov 1969)
Where does an underspecified grammar come from?

Brute-force search of partial orders establishes how well usage frequencies can be predicted in principle.
Brute-force search of partial orders establishes how well usage frequencies can be predicted in principle.

Not a theory of how speakers acquire an underspecified grammar.

Forthcoming work (Bane to appear) shows how speakers might infer an underspecified grammar from the variable data presented to them.

Predicts frequencies nearly as well as the best grammars here.
Many thanks to Matthew Adams, Arto Anttila, Karlos Arregi, Jason Riggle, and Morgan Sonderegger for discussion and assistance!

Slides to be posted: http://clml.uchicago.edu/~max
References I


References III
