Deriving the Structure of Variation from the Structure of Non-Variation in the English Dative

Max Bane
bane@uchicago.edu

Department of Linguistics
University of Chicago

February 20, 2010
28th West Coast Conference on Formal Linguistics
University of Southern California
Introduction

- Nondeterministic variability in the choice of form is rife in language usage.
Introduction

- 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></th>
<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>
<td>1,024</td>
</tr>
<tr>
<td>% deleted</td>
<td></td>
<td>62%</td>
<td>45%</td>
<td>37%</td>
</tr>
<tr>
<td>Tejano English (San Antonio)</td>
<td>n</td>
<td>1,738</td>
<td>947</td>
<td>564</td>
</tr>
<tr>
<td>% deleted</td>
<td></td>
<td>62%</td>
<td>25%</td>
<td>46%</td>
</tr>
<tr>
<td>AAVE (Washington, DC)</td>
<td>n</td>
<td>143</td>
<td>202</td>
<td>37</td>
</tr>
<tr>
<td>% deleted</td>
<td></td>
<td>76%</td>
<td>29%</td>
<td>73%</td>
</tr>
</tbody>
</table>
Examples of Variable Alternations

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

<table>
<thead>
<tr>
<th></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...
    b. ... I thocht you was a scuba diver.
    c. ... we were lyin at anchor.
    d. ... we was tired...
This Talk

(3) 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:

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.


- Syntactic, semantic, prosodic, etymological properties proposed to explain which verbs allow which variants with which arguments.
This Talk

- 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.
This Talk

- 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.
This Talk

- 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].
(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.

  - Not found in today’s corpus.

- Crucially, 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

- Prepositional (69.8%)
- Double Object (26.9%)
- NP Shift (3.31%)

N = 1601
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 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.
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 ≠ PW or whose stem = PW ≠ 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)
The Model Outputs

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 \text{VERB THEME GOAL} linearizations.

WRAP-XP
- An XP must be contained in a prosodic phrase.
- Penalizes \text{VERB(GOAL)}, but not \text{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

(38) The tableau for the input ‘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>*P-PHRASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. → (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>
<td>***</td>
</tr>
<tr>
<td>b. → (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>
<td>***</td>
</tr>
<tr>
<td>c. (give)(the book)(my sister)</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>d. (give)(my sister)(the book)</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>e. → (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>
<td>**</td>
</tr>
<tr>
<td>f. (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>
<td>*</td>
</tr>
<tr>
<td>g. (give the book)(my sister)</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>h. → (give my sister)(the book)</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
</tbody>
</table>

(Anttila et al to appear)
(39) The tableau 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>

(Anttila et al to appear)
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 (Kiparsky 1994, Anttila 1997):
Predicting usage frequency with this model

► The basic idea (Kiparsky 1994, Anttila 1997):
  ► The number of constraint rankings that yield a given input-output pairing ("ranking volume") affects how likely that pairing is to be observed.
  ► Cf. “variation from many grammars” (Kroch 1989).
Predicting usage frequency with this model

The basic idea (Kiparsky 1994, Anttila 1997):

- The number of constraint rankings that yield a given input-output pairing ("ranking volume") affects how likely that pairing is to be observed.
  - Cf. “variation from many grammars” (Kroch 1989).
- 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 $k$. 

Results...

- 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”`

- Frequencies are **conditional** on input.
  - Only **mutually independent** frequencies included.
Results...

- In the following plots, each point corresponds to a single input-output mapping, e.g.:
  - `give(her, it) \mapsto \text{“give it to her”}`
  - `donate(my sister, the book) \mapsto \text{“donate my sister the book”}`
  - `give(my little sister, it) \mapsto \text{“give to my little sister it”}`

- Frequencies are conditional on input.
  - Only mutually independent frequencies included.

- Different phonological phrasings collapsed, because we can only observe linearizations in the textual corpus.
  - R-volumes of different phrasings for same linearization added together.
Results…

- 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”`.

- Frequencies are **conditional** on input.
  - Only **mutually independent** frequencies included.

- Different phonological phrasings **collapsed**, because we can only observe linearizations in the textual corpus.
  - R-volumes of different phrasings for same linearization **added together**.

- Grammar sampling predicts **simple linear relation**.
Unrestricted: Counting among all 10! rankings

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

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

![Graph showing the relationship between number of rankings and corpus frequency, with a weak correlation indicated by R^2 = 0.20.](image)

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, near-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, near-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.
Empirical Adequacy

- Unspecified grammar: Not a very great linear relationship, accuracy below baseline.
- 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.
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.
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.
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.

- The solution found depends on what partial order we start with.
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

Good linear trend, $R^2 = 0.98$, accuracy = 80.6%.
Where does an underspecified grammar come from?

- Brute-force search of partial orders establishes how well usage frequencies can be predicted in principle.
Where does an underspecified grammar come from?

- Brute-force search of partial orders establishes how well usage frequencies can be predicted in principle.
- Not a theory of how to infer an underspecified grammar.
  - Overfitting!
  - Need to show how speakers might infer an underspecified grammar from the variable data presented to them.
  - Hopefully predict frequencies nearly as well as the best grammar above.
The Structure of Variation from the Structure of Non-Variation

- Variation is problematic for grammatical inference.
  - Multiple observed outputs per input necessarily imply grammatical contradictions.
- Task cannot be to acquire a (single) total constraint order.
  - Need the partial bits of a grammar that are consistent with generalizations in the data.
The Structure of Variation from the Structure of Non-Variation

- Variation is problematic for grammatical inference.
  - Multiple observed outputs per input necessarily imply grammatical contradictions.
- Task cannot be to acquire a (single) total constraint order.
  - Need the partial bits of a grammar that are consistent with generalizations in the data.
- A simple hypothesis:
  - Just ignore the variable data.
  - Resulting underpsecified grammar is whatever is implied by the categorical data.
  - Structure of variation follows from that.
Non-variable mappings

- The blogspot corpus exhibits four non-variable mappings:
  - donate(her, it) → Prepositional
  - donate(my sister, it) → Prepositional
  - give(her, it) → Prepositional
  - give(my sister, it) → Prepositional
- 114 tokens out of 1,600.
Partial Orders vs ERC Sets

- Partial orders are a simple way of describing sets of rankings (partial grammars).
- **Elementary Ranking Conditions** (ERCs; Prince 2002) are a more expressive way.
Partial Orders vs ERC Sets

- Partial orders are a simple way of describing sets of rankings (partial grammars).
- Elementary Ranking Conditions (ERCs; Prince 2002) are a more expressive way.
- Statements in a three-valued logic.
  - Describe exactly the conditions under which a given candidate could win over the other candidates in its tableaux.
Partial Orders vs ERC Sets

- Partial orders are a simple way of describing sets of rankings (partial grammars).

- **Elementary Ranking Conditions** (ERCs; Prince 2002) are a more expressive way.

- Statements in a three-valued logic.
  - Describe *exactly* the conditions under which a given candidate could win over the other candidates in its tableaux.

- \( \Rightarrow \) What a learner can infer about the grammar from observing a winning output.
Partial grammar implied by non-variable mappings

- The invariant mappings imply an incomplete grammar, an ERC-set restriction on constraint rankings.
Partial grammar implied by non-variable mappings

- The invariant mappings imply an incomplete grammar, an ERC-set restriction on constraint rankings.
- But we have homophony
  - Even among the non-variable observations, we can’t distinguish the 2-phrase candidates from the 3-phrase candidates.
The invariant mappings imply an incomplete grammar, an ERC-set restriction on constraint rankings.

But we have homophony

Even among the non-variable observations, we can’t distinguish the 2-phrase candidates from the 3-phrase candidates.

Three ways to handle this:

- Assume they are all 2-phrases.
- Assume they are all 3-phrases (where not harmonically bounded).
- Use ERC-set disjunction to construct a “weaker” partial grammar that permits either.
Assuming all 2-phrases

Partial grammar implied by non-variable mappings

77,472 possible rankings.
Assuming all 2-phrases
Counting rankings consistent with invariant data

\[ R^2 = 0.8674, \quad p \leq 6.3 \times 10^{-21} \]

\[ R^2 = 0.86, \quad \text{Accuracy 79.1}\%. \quad \text{The categorical data contain the seed of the variable frequencies.} \]
Assuming all 3-phrases

Partial grammar implied by non-variable mappings

26,424 possible rankings.
Assuming all 3-phrases
Counting rankings consistent with invariant data

\[ R^2 = 0.86, \text{ Accuracy 69.8\%}. \]
Taking disjunction of 2-, 3-phrases

Partial grammar implied by non-variable mappings

108,864 possible rankings.
Taking disjunction of 2-, 3-phrases
Counting rankings consistent with invariant data

\[ R^2 = 0.8892, \quad p \leq 6.67e-15 \]

\[ R^2 = 0.89, \quad \text{Accuracy } 80.6\%. \]
Argued for the empirical adequacy of a grammar-sampling model of variation in the English dative.

Summary

- Argued for the empirical adequacy of a grammar-sampling model of variation in the English dative.
- Can in principle explain at least 98% of usage-frequency variance and achieve 80.6% predictive accuracy
  - When the data are grouped according to a prosodic parameterization.
Summary

- Argued for the empirical adequacy of a grammar-sampling model of variation in the English dative.
- Can in principle explain at least 98% of usage-frequency variance and achieve 80.6% predictive accuracy
  - When the data are grouped according to a prosodic parameterization.
- Offered a simple, preliminary method by which similar performance (90% variance, 80.6% accuracy) can follow from grammatical inference.
  - Much of the structure of variation can be derived from the structure of non-variation.
A Connection

- Adger’s (2006) “variation from the combinatorics of underspecification”
- Constructs a Minimalist model of variable was/were agreement in Buckie Scottish English (Smith 2000).
- Featural agreement requirements are underspecified.
- Tries to relate variants’ frequencies to how many ways the underspecification could be resolved.
A Connection

- Adger’s (2006) “variation from the combinatorics of underspecification”
- Constructs a Minimalist model of variable was/were agreement in Buckie Scottish English (Smith 2000).
- Featural agreement requirements are underspecified.
- Tries to relate variants’ frequencies to how many ways the underspecification could be resolved.
- The grammar-sampling idea is framework-neutral.
Questions

- What do significant deviations from linear relationship between ranking-volume and corpus frequency mean?
  - Variables and factors this model is not sensitive to; sample is from many speakers.
  - Lexical idiosyncracies? (mixed effects model)
Questions

- What do significant deviations from linear relationship between ranking-volume and corpus frequency mean?
  - Variables and factors this model is not sensitive to; sample is from many speakers.
  - Lexical idiosyncracies? (mixed effects model)
- Relation to other models of the alternation?
  - Bresnan et al’s (2007, et seq) logistic regressions on various predictors (arguments’ pronominality, animacy, givenness, etc.).
  - Achieve token accuracies \( \sim 90\% \) (different corpus, prepositional vs double-object).
  - What does it mean when very different models make comparable predictions?
Thanks

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