Numbers in the Grammar?
Further Evidence from Experiments with English Relativizer and Complementizer Optionality

Robin Melnick
San José State University

T. Florian Jaeger
University of Rochester

Tom Wasow
Stanford University
How gradient is “gradient grammar”?

• “Gradient grammar”…Grammaticality is not (in all cases) categorical (Bresnan & Hay 2006)

• Experiments with English dative (Bresnan 2007) show that the same factors that influence production also influence judgment, thus arguably part of competence

• But dative is only “somewhat gradient” so similar results for a more fully gradient phenomena might offer further evidence for “numbers in the grammar”

• The present work offers such evidence by working with English “that”-optionality
Agenda

- Background
- Methodology
- Experiments
Probabilistic syntactic variation

- **Paraphrase denial**
  - “A difference in syntactic form always spells a difference in meaning” (Bolinger)

- **Paraphrase accommodation**
  - Semantically equivalent options in the grammar (e.g., Adger & Smith in MP)

- **Paraphrase explanation**
  - Many factors influence both production and judgment
Some “alternations”
(with multivariate studies)

- Dative
  - Bresnan et al.
- Genitive
  - O’Connor et al.
- Particle
  - Gries; Wasow & Arnold
- Topicalization
  - Snider & Zaenen
- Dislocation
  - Snider & Zaenen
- Lexical
  - Gries; Zwicky
- Agreement
  - Tagliamonte; Riordan; Melnick
- Reduction
  - Wasow, Jaeger & Orr; Jaeger
Factors

- Animacy
- Definiteness
- Discourse accessibility
- Length
- Pronominality
- Person
- Number
- Structural parallelism
- Semantic class
- etc.
Factors example: The genitive alternation

The landlord's children

- Statistically speaking, *animate* modifiers tend to be expressed in the S-Genitive alternant.
- But they may be expressed in the Of-Genitive alternant as well, of course.

The children of the landlord
Factors example: The genitive alternation

- Statistically speaking, *discourse-old* or *highly accessible* modifiers tend to be expressed in the S-Genitive alternant.
- But they may be expressed in the Of-Genitive alternant as well.
Factors example: The genitive alternation

The plans of my brother in Florida

- Statistically speaking, heavy modifiers tend to be expressed in the Of-Genitive alternant.
- But they may be expressed in the S-Genitive alternant as well.

My brother in Florida's plans
Factor interaction

- How can we tell when these three dimensions are contributing independently to an alternation?

**His** penchant for betting on the ponies

Humans are often topical

Topics are often expressed as pronouns

Pronouns are light
Example: Dative shift (Bresnan et al. 2007)

Positive coefficients favor PP dative, negative favor NP:

• +0.99 \{accessibility of recipient = nongiven\}
• −1.1 \{accessibility of theme = nongiven\}
• +1.2 \{pronominality of recipient = nonpronoun\}
• −1.2 \{pronominality of theme = nonpronoun\}
• +0.85 \{definiteness of recipient = indefinite\}
• −1.4 \{definiteness of theme = indefinite\}
• +2.5 \{animacy of recipient = inanimate\}
• +0.48 \{person of recipient = nonlocal\}
• −0.03 \{number of recipient = plural\}
• +0.5 \{number of theme = plural\}
• −0.46 \{concreteness of theme = nonconcrete\}
• −1.1 \{parallelism = 1\} − 1.2 * length difference (log scale)
Issues for logistic regression

- Data sparsity
  → Eliminate variables; use coarse granularity
- Normal distribution
  → Log, power, or other transform, plus centering
- Overfitting
  → Sample size / # of parameters
- Collinearity
  → Combine variables
- Clumping
  → Bootstrapping; mixed models
Dative model

(Bresnan et al. 2007)
Dative experiment: Stimuli bins

(Bresnan 2007)
Dative experiment: Results

(Bresnan 2007)
OK, back to Relativizer/Complementizer Optionality

- Probabilistic Reduction Hypothesis (Jaeger 2006, 2009)
  - Both REL and COMPL reduction
  - Prior REL work: Fox & Thompson; Wasow, Jaeger & Orr
- **Predictability** is most significant factor

\[
\text{We } v [\text{hope}] (\text{that})_{\text{COMPL}} \text{ you enjoy } \\
\text{NP } [\text{the talk}]_i (\text{that})_{\text{REL}} \text{ we give } \underline{\text{____}}_i \text{ today}
\]
Factors in CC model (Jaeger 2006, 2009)

<table>
<thead>
<tr>
<th>Factor name</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>L(PRECEDING)</td>
<td>Log length of material preceding CC*</td>
<td>cont.</td>
</tr>
<tr>
<td>NON-ADJACENCY</td>
<td>Interveners between embedding verb and CC*</td>
<td></td>
</tr>
<tr>
<td>L(CC*)</td>
<td>Log length of CC* in words</td>
<td>cont.</td>
</tr>
<tr>
<td>SUBJECT FORM</td>
<td>Accessibility coding of CC* subject</td>
<td></td>
</tr>
<tr>
<td>SUBJECT ANIMACY</td>
<td>Animacy of CC* subject</td>
<td></td>
</tr>
<tr>
<td>SUBJECT IDENTITY</td>
<td>Are the matrix and CC* subject identical?</td>
<td></td>
</tr>
<tr>
<td>SUBJECT AMBIGUOUS</td>
<td>CC* subject could create temporary ambiguity?</td>
<td></td>
</tr>
<tr>
<td>P(PRECEDING WORD)</td>
<td>Log frequency of word preceding CC*</td>
<td>cont.</td>
</tr>
<tr>
<td>P(INITIAL WORD)</td>
<td>Log frequency of word after complementizer site</td>
<td>cont.</td>
</tr>
<tr>
<td>MATRIX SUBJECT FORM</td>
<td>Accessibility coding of CC* subject</td>
<td>cat(3)</td>
</tr>
<tr>
<td>MATRIX NEGATION</td>
<td>Is the matrix clause negated?</td>
<td>cat(2)</td>
</tr>
<tr>
<td>MATRIX EMBEDDED</td>
<td>Is the matrix clause itself embedded?</td>
<td>cat(2)</td>
</tr>
<tr>
<td>MATRIX FINITE</td>
<td>Is the embedding verb finite?</td>
<td>cat(2)</td>
</tr>
<tr>
<td>MATRIX PRESENT TENSE</td>
<td>Is the embedding verb in present tense?</td>
<td>cat(2)</td>
</tr>
<tr>
<td>PRECEDING DISFLUENCY</td>
<td>Disfluencies between head and CC*</td>
<td>count</td>
</tr>
<tr>
<td>PRECEDING PAUSE</td>
<td>Pause immediately preceding CC*?</td>
<td>cat(2)</td>
</tr>
<tr>
<td>INITIAL DISFLUENCY</td>
<td>Disfluencies in CC* before/in subject</td>
<td>count</td>
</tr>
<tr>
<td>LATER DISFLUENCY</td>
<td>Disfluencies in remainder of CC*</td>
<td>count</td>
</tr>
<tr>
<td>LOG SPEECH RATE</td>
<td>Log speech rate around CC* start</td>
<td>cont.</td>
</tr>
<tr>
<td>SQ LOG SPEECH RATE</td>
<td>Squared log speech rate around RC* start</td>
<td>cont.</td>
</tr>
<tr>
<td>WITHIN-Speaker Persist.</td>
<td>Preceding within-speaker prime (if any)</td>
<td>cat(3)</td>
</tr>
<tr>
<td>ACROSS-Speaker Persist.</td>
<td>Preceding across-speaker prime (if any)</td>
<td>cat(3)</td>
</tr>
<tr>
<td>WORD FORM OCP</td>
<td>First word following complementizer site <em>that</em>?</td>
<td>cat(2)</td>
</tr>
<tr>
<td>OCP PRECEDING PHON</td>
<td>Is the preceding phon [+fricative] or [+dental]?</td>
<td>cat(2)</td>
</tr>
<tr>
<td>OCP FOLLOWING PHON</td>
<td>Is the following phon [+plosive] or [+alveolar]?</td>
<td>cat(2)</td>
</tr>
<tr>
<td>SPEAKER GENDER</td>
<td>Gender of speaker</td>
<td>cat(2)</td>
</tr>
</tbody>
</table>

Total control parameters in full model | 47
Importance of factors in CC model

![Importance of factors in CC model](image)
CC model

Sample Model Probabilities for “that” Inclusion (1) vs. Omission (0)
### Verb Bias

<table>
<thead>
<tr>
<th>Verb Lemma</th>
<th>Percent of database</th>
<th>that-bias in database</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>52%</td>
<td>11%</td>
</tr>
<tr>
<td>guess</td>
<td>14%</td>
<td>1%</td>
</tr>
<tr>
<td>know</td>
<td>8%</td>
<td>32%</td>
</tr>
<tr>
<td>say</td>
<td>8%</td>
<td>27%</td>
</tr>
<tr>
<td>remaining 29 verbs</td>
<td>17%</td>
<td>47%</td>
</tr>
</tbody>
</table>
## Factors in Rel model (Jaeger 2006)

<table>
<thead>
<tr>
<th>Factor name</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>L(PRECEDING)</td>
<td>Log length of material preceding modified NP</td>
<td>cont.</td>
</tr>
<tr>
<td>L(ADJACENCY)</td>
<td>Log length of material between head and RC*</td>
<td>cont.</td>
</tr>
<tr>
<td>L(RC*)</td>
<td>Log length of RC* in words</td>
<td>cont.</td>
</tr>
<tr>
<td>GAP-EMBEDDING</td>
<td>Gap part of a clausal complement in RC*?</td>
<td>cat(2)</td>
</tr>
<tr>
<td>GAP-EMBEDDING × L(RC*)</td>
<td></td>
<td>cont.(2)</td>
</tr>
<tr>
<td><strong>SUBJECT FORM</strong></td>
<td>Accessibility coding of RC* subject</td>
<td>cat(3)</td>
</tr>
<tr>
<td><strong>SUBJECT ANIMACY</strong></td>
<td>Animacy of RC* subject</td>
<td>cat(3)</td>
</tr>
<tr>
<td><strong>P(PRECEDING WORD)</strong></td>
<td>Log frequency of word preceding RC*</td>
<td>cont.</td>
</tr>
<tr>
<td><strong>P(INITIAL WORD)</strong></td>
<td>Log frequency of word after relativizer site</td>
<td>cont.</td>
</tr>
<tr>
<td><strong>MATRIX NEGATION</strong></td>
<td>Is the matrix clause negated?</td>
<td>cat(2)</td>
</tr>
<tr>
<td><strong>MATRIX VERB</strong></td>
<td>Type of matrix verb</td>
<td>cat(3)</td>
</tr>
<tr>
<td><strong>PRECEDING DISFLUENCY</strong></td>
<td>Disfluencies between head and RC*</td>
<td>count</td>
</tr>
<tr>
<td><strong>PRECEDING PAUSE</strong></td>
<td>Pause immediately preceding RC*?</td>
<td>cat(2)</td>
</tr>
<tr>
<td><strong>INITIAL DISFLUENCY</strong></td>
<td>Disfluencies in RC* before/in subject</td>
<td>count</td>
</tr>
<tr>
<td><strong>LATER DISFLUENCY</strong></td>
<td>Disfluencies in remainder of RC*</td>
<td>count</td>
</tr>
<tr>
<td><strong>LOG SPEECH RATE</strong></td>
<td>Log speech rate around RC* start</td>
<td>cont.</td>
</tr>
<tr>
<td><strong>SQ. LOG SPEECH RATE</strong></td>
<td>Squared log speech rate around RC* start</td>
<td>cont.</td>
</tr>
<tr>
<td><strong>WITHIN-SPEAKER PERSIST.</strong></td>
<td>Preceding within-speaker prime (if any)</td>
<td>cat(3)</td>
</tr>
<tr>
<td><strong>ACROSS-SPEAKER PERSIST.</strong></td>
<td>Preceding across-speaker prime (if any)</td>
<td>cat(3)</td>
</tr>
<tr>
<td><strong>OCP PRECEDING PHON</strong></td>
<td>Preceding phon [+fricative] or [+dental]?</td>
<td>cat(2)</td>
</tr>
<tr>
<td><strong>SPEAKER GENDER</strong></td>
<td>Gender of speaker</td>
<td>cat(2)</td>
</tr>
<tr>
<td><strong>Total control parameters in full model</strong></td>
<td></td>
<td>28</td>
</tr>
</tbody>
</table>
Rel model

Sample Model Probabilities for “that” Inclusion (1) vs. Omission (0)
8 Trials

1. Complementizer vs. Relativizer
2. 100-point system (Bresnan 2006) vs. Timed, forced choice
3. “Live” vs. Mechanical Turk

$2^3 = 8$
Relativizer experiments: Stimuli bins
Results: relativizer, 100-point, live

Spearman’s $\rho = 0.638$

$p < 0.001$
Results: relativizer, *forced-choice*, live

Spearman’s $\rho = 0.710$

$p < 0.0001$
Results: relativizer, forced-choice, Turk

Spearman’s ρ = 0.407

p < 0.025
Results: relativizer, 100-point, Turk

Spearman’s $\rho = 0.539$

$p < 0.01$
CC model

Sample Model Probabilities for “that” Inclusion (1) vs. Omission (0)
Results: compl, 100-point, Turk

Spearman’s $\rho = 0.600$

$p < 0.001$
Results: compl, forced, Turk

Spearman’s $\rho = 0.639$

$p < 0.0001$
Complementizer, Turk – token-by-token comparison
Conclusions

Content

• “Raised the bar” with highly gradient REL and COMP optionality phenomenon

• Just as with the more bifurcated dative data, factors that influence production here again influence judgment

• The strong effect of predictability (\%s) offers greater evidence for “numbers in the grammar”

Methodology

• Timed, forced choice scheme provides better results in “live” experiments

• Some validation for Mechanical Turk method, with significant results (if slightly less so)