The effect of register on syntactic alternations

An exploratory study of regional variation in the English dative alternation

Melanie Röthlisberger (University of Zurich)
The English dative alternation

Mary gave \[ \text{John} \] \[ \text{the apple} \]

recipient  theme

Mary gave \[ \text{the apple} \] to \[ \text{John} \]

theme  recipient
The English dative alternation

Mary gave [him] [the apple]  
\[ditransitive dative\]

Mary gave [the apple] to [him]  
\[prepositional dative\]
Why register?

• **regional variation** in the probabilistic constraints that influence grammatical variation, i.e. the choice between two variants (e.g. Bresnan & Hay 2008, Röthlisberger et al. 2017)

• Text type as important in other syntactic alternations (e.g. Grafmiller 2014: genitive alternation)

→ **Inter-register variation**: Different effects of register on the choice of dative alternation

→ **Intra-register variation**: Different effects of language-internal constraints depending on register
Research objectives

1. **Inter-register variation**: What is the influence of register/text type on the choice of dative variant in the different varieties?

2. **Intra-register variation**: To what extent do language-internal factors differ in their effect on syntactic variation in different registers?

→ regional variation vs. register variation
Data & Methodology
Corpus data
Corpus data I

- International Corpus of English (ICE) series
  - 60% spoken (transcriptions), 40% written texts = 1m words per subcorpus
  - 500 texts, 2,000 words per text
  - 12 different registers, same corpus structure

<table>
<thead>
<tr>
<th>SPOKEN 300</th>
<th>WRITTEN 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialogues  180 Private 100 Face-to-face conversations 90 s1a Phonecalls 10 s1b</td>
<td></td>
</tr>
<tr>
<td>Public 80 Classroom lessons 20 Broadcast Discussions 20 s1b</td>
<td></td>
</tr>
<tr>
<td>Broadcast Interviews 10 Parliamentary Debates 10</td>
<td></td>
</tr>
<tr>
<td>Legal cross-examinations 10 Business Transactions 10</td>
<td></td>
</tr>
<tr>
<td>Monologues 120 Unscripted 70 Spontaneous commentaries 20 s2a Unscripted Speeches 30 Demonstrations 10</td>
<td></td>
</tr>
<tr>
<td>Legal Presentations 10</td>
<td></td>
</tr>
<tr>
<td>Scripted 50 Broadcast News 20 s2b Broadcast Talks 20 Non-broadcast Talks 10</td>
<td></td>
</tr>
<tr>
<td>200 Non-printed 50 Student Writing 20 Student Essays 10 w1a Exam Scripts 10</td>
<td></td>
</tr>
<tr>
<td>Letters 30 Social Letters 15 w1b Business Letters 15</td>
<td></td>
</tr>
<tr>
<td>Printed 150 Academic writing 40 Humanities 10 w2a Social Sciences 10</td>
<td></td>
</tr>
<tr>
<td>Natural Sciences 10 Technology 10</td>
<td></td>
</tr>
<tr>
<td>Popular Writing 40 Humanities 10 w2b Social Sciences 10</td>
<td></td>
</tr>
<tr>
<td>Natural Sciences 10 Technology 10</td>
<td></td>
</tr>
<tr>
<td>Reportage 20 Press news reports 20 w2c</td>
<td></td>
</tr>
<tr>
<td>Instructional writing 20 Administrative Writing 10 w2d Skills/Hobbies 10</td>
<td></td>
</tr>
<tr>
<td>Persuasive writing 10 Press editorials 10 w2e</td>
<td></td>
</tr>
<tr>
<td>Creative writing 20 Novels &amp; short stories 20 w2f</td>
<td></td>
</tr>
</tbody>
</table>
Corpus data II

- Corpus of Global web-based English (GloWbE) (Davies and Fuchs 2015)
  - general websites and blogs
  - data sampled in 2012-13 – https://corpus.byu.edu/glowbe/
  - same varieties sampled as represented in ICE
  - 500,000 words per variety sampled from random subsets
Data extraction and annotation

(e.g. Bresnan et al. 2007)

• retrieval of dative variants using verb list and perl script
• restrict to choice context (incl. pronouns)
• code for numerous (language-internal) factors: length (weight ratio), complexity, pronominality, givenness, definiteness, person, animacy, concreteness of theme, verb sense
• code for language-external factors: Register and Mode (spoken vs written)

$N = 13,171$
Register coding

Koch & Oesterreicher (1985)
Mixed-effects logistic regression

1. Inter-register variation (regional variation in register effects):
   - fixed effects: 5 most important language-internal predictors (acc. to random forest)
   - interaction REGISTER * VARIETY
   - random effects: lexical idiosyncrasies, file ID

2. Intra-register variation (internal variation):
   - fit one model per variety
   - interaction of REGISTER with 5 most important language-internal predictors
   - random effects: lexical idiosyncrasies, file ID
Results
RQ 1: Regional variation in register effects

Model formula:
Dative variant ~ (1 | Verb) + (1 | ThemeHeadFilter) + (1 | RecHeadFilter) + (1 | FileID)
+ ThemeHeadFreq + ThemeComplexity + RecPron + ThemePron + WeightRatio
+ Register + Variety + Variety:Register

• numeric predictors are standardized by two standard deviations (ThemeHeadFreq + WeightRatio)
• WeightRatio is log-scaled (cf. Bresnan et al. 2007)
• Register and Variety are coded with sum coding instead of treatment coding (comparison NOT versus one reference level but versus the grand mean of all other levels)
Regional variation in register effects: Register * Variety interaction

Evaluation

C-statistic: 0.98
Accuracy: 93.3%

no indication of overdispersion / multicollinearity
Two corpora - same register?

What is the effect of adding GloWbE data to the written informal register?

with GloWbE:

without GloWbE:
RQ 2: Register-internal variation (intra-register variation)

Model formula:

Dative variant ~ (1|Verb) + (1|ThemeHead) + (1|RecHead) + (1|FileID)  
  + Register * (WeightRatio + RecPron + ThemeComplexity + ThemePron  
  + ThemeHeadFreq)

• One model per variety (9 models in total)
• numeric predictors are standardized by two standard deviations (ThemeHeadFreq + WeightRatio)
• binary predictors are transformed into numbers and centered around the mean
• WeightRatio is log-scaled (cf. Bresnan et al. 2007)
• Register is again coded with sum coding instead of treatment coding
Register-internal variation

- No statistically significant interactions in: BrE, HKE, IrE, JamE, NZE, SinE
- Theme complexity & theme head frequency significantly different by Register in CanE
- Recipient pronominality significantly different by Register in IndE
In sum

1. **Inter-register variation:** What is the influence of register/text type on the choice of dative variant in the different varieties?

Register differs in its effect across different varieties of English. Overall,
- formal registers are more likely to use the prepositional dative (with the exception of HKE)
- In PhiE, SinE the differences between the registers are negligible → homogeneous registers
- In HKE, BrE, IrE and JamE, register effects are diverging
- other varieties lay somewhere in between

The addition of GloWbE data results in slightly different results, regarding the effect of spoken informal vs written informal data:
**GloWbE data might be more different from the written informal spectrum than assumed.**
In sum

2. **Intra-register variation**: To what extent do language-internal factors differ in their effect on syntactic variation in different registers?

- Language-internal predictors differ in their effect size, notably recipient pronominality and theme complexity
- Apart from CanE and IndE, all varieties are very homogeneous regarding intra-register variation
to conclude…
Take-home messages

→ registers display variety-specific patterns (at least wrt the dative alternation) (Neumann, this conference)

→ comparison between registers of ICE corpora requires reconsiderations

→ probabilistic grammars are fairly stable between varieties (we already know that) but are even more homogeneous among registers (sic!)
Open questions

• Change the coding of register to a more fine-grained coding (e.g. text type)
• Investigate competition between formality and mode (are they as similarly distinct as the terminology implies?)
• …
Thank you!

melanie.roethlisberger@es.uzh.ch
@m_rothlisberger
slides available at: www.melanie-roethlisberger.ch
References


