DIALECTAL DITRANSITIVE PATTERNS IN BRITISH ENGLISH

LANGUAGE-INTERNAL VS. LANGUAGE-EXTERNAL CONSTRAINTS

Johanna Gerwin (Kiel) & Melanie Röthlisberger (Zürich)
(1) SETTING THE SCENE

DITRANSITIVES

- She gave a book to him. (PREP)
- She gave him a book. (DOC)
- She bought a book for ihm. (PREP)
- She bought him a book. (DOC)

DIALECTAL VARIANTS

- She gave it him. (altDOC)
- She gave it to him. (PREP)
- She gave him it. (DOC)

• He had such a go at me for showing it my mum (BNCreg)
• If I find out I 'll give it you back. (BNCreg)
• I 'll fetch it you down anyway, next time (BNCreg)
(1) SETTING THE SCENE

ORIGINAL STUDY:
DITRANSITIVES IN BRITISH ENGLISH DIALECTS (GERWIN 2014)

• Ditransitive patterns as a Labovian sociolinguistic variable:
• How is pattern choice determined by language-external factors such as ‘region/origin of the speaker’, ‘time’, and ‘spoken vs written mode’?

<table>
<thead>
<tr>
<th></th>
<th>1960/70s</th>
<th>1990s</th>
</tr>
</thead>
<tbody>
<tr>
<td>spoken data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2,962</td>
<td>5,147</td>
</tr>
<tr>
<td>written data</td>
<td>872</td>
<td>917</td>
</tr>
</tbody>
</table>

⇒ total: 9,899 tokens
(1) SETTING THE SCENE

ORIGINAL STUDY:
DITRANSITIVES IN BRITISH ENGLISH DIALECTS (GERWIN 2014)

• regional preferences, especially with pronominal themes (it/them)
(1) SETTING THE SCENE

ORIGINAL STUDY:
DITRANSLITES IN BRITISH ENGLISH DIALECTS (GERWIN 2014)

- regional preferences, especially with pronominal themes (it/them)
- slight diachronic effect between FRED and BNCweb data
• regional preferences, especially with pronominal themes (it/them)
(1) To which extent is pattern choice determined by the sociolinguistic factors ‘region/origin of the speaker’ and ‘time’ when well-established language-internal predictors are modelled simultaneously?

(2) Do language-internal predictors differ in their importance on dative choice across different British regions?
(2) DATA

FREIBURG ENGLISH DIALECT CORPUS (FRED)

- 2.5 million words
- 1970-1989
- 431 speakers (mainly NORMs born between 1890 & 1919)

BRITISH NATIONAL CORPUS ONLINE EDITION (BNCWEB)

- spoken-demographic part: ca. 10 million words
- regionally tagged data: 3.3 million words
- Early 1990s
- 124 speakers (from age 15)

Language-external predictor ‘time’
→ real-time difference of about 20 years
→ apparent-time difference of about 2-3 generations of speakers
(2) DATA

BRITISH REGIONS

- Southeast (SE)
- Southwest (SW)
- Midlands (MID)
- North (N)
- Wales (WAL)
- Scotland (SCOT)

Language-external predictor ‘region/origin of speaker’ → 6 British regions
(2) DATA

DATA SELECTION

- only two post-verbal objects
- prototypical declarative sentences
- 14 alternating verbs

→ 7070 ditransitives
→ 2661 in FRED
→ 4449 in BNC
(2) DATA

Language-internal predictor ‘verb semantics’
→ 3 different senses

• TRANSFER:
  All the time they was bringin' them in their living  
  (FRED: KEN 011)

• COMMUNICATION:
  It tells you all about it in the book and also here  
  (BNCreg: KDL 207)

• ABSTRACT/IDIOMATIC:
  And you put your foot behind the stone, and give it a good (e 'cracking sound') crack  
  (FRED: PEE 001)
(2) DATA

Language-internal predictor ‘recipient animacy’
→ 2-way distinction: animate – inanimate

Language-internal predictor ‘syntactic weight’
→ number of typed characters
→ including spaces in multi-word objects
→ excluding prepositions to and for

Language-internal predictors ‘recipient pronominality’ & ‘theme pronominality’
→ pronoun v. full noun phrase
Language-internal predictor ‘particle’
→ intervening, following, none at all

EXAMPLES:

• Particle intervening between two objects

  *All the time they was bringing ‘em in their living*  
  *(FRED: KEN 011)*

  *Now look you’ve got to give that ticket back to Mr […]*  
  *(BNCreg: KCT 14186)*

• Particle following two objects

  *My mum used to give me a penny back.*  
  *(FRED: SFK 032)*
(3) METHODOLOGY

CONDITIONAL RANDOM FORESTS (CRFS)

- Party package
- trial-and-error
- ntree ctrees on random subsample of the data
- bagging over testset to predict out-of-bag-sample
- permutation scheme

FOR THIS STUDY:
- 1 random forest on full dataset
- 6 separate conditional random forests for each region
(3) METHODOLOGY

COMPARING THE RANKING OF 6 CRFS

1. fit one forest per region
2. define the ranking of predictors (e.g. 1 = most important, 2 = second important)
3. calculate spearman’s rank correlation coefficient as measure of similarity
4. transform similarity to distance: $1 - \text{similarity coefficient}$
5. use Multidimensional scaling analysis to visualize distance between regions (smacof package for non-metric MDS solution)
(4) RESULTS

CRF ON FULL DATASET

RESEARCH QUESTION (1):
To which extent is pattern choice determined by the sociolinguistic factors ‘region/origin of the speaker‘ and ‘time‘ when well-established language-internal predictors are modelled simultaneously?

- ntree = 2000
- mtry = 3
- 93.2 % accuracy
(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS

RESEARCH QUESTION (2):

Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region
(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS

RESEARCH QUESTION (2):
Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region
2. Determine rank of each predictor per region

<table>
<thead>
<tr>
<th>Factor</th>
<th>Midlands</th>
<th>North</th>
<th>Scotland</th>
<th>Southeast</th>
<th>Southwest</th>
<th>Wales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme pronominality</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Weight ratio</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Recipient pronominality</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Verb semantics</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Recipient animacy</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Corpus</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Particle</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
**COMPARING FACTOR IMPORTANCE ACROSS REGIONS**

**RESEARCH QUESTION (2):**

Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region
2. Determine rank of each predictor per region
3. Calculate Spearman's rank correlation coefficient between all regions (15 comparisons)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Midlands</th>
<th>North</th>
<th>Scotland</th>
<th>Southeast</th>
<th>Southwest</th>
<th>Wales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme pronounality</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Weight ratio</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Recipient pronounality</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Verb semantics</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Recipient animacy</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Corpus</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Particle</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
RESEARCH QUESTION (2):

Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region
2. Determine rank of each predictor per region
3. Calculate Spearman’s rank correlation coefficient between all regions (15 comparisons)
4. Take inverse for distance (1 – coefficient)
(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS

RESEARCH QUESTION (2):
Do language-internal predictors differ in their importance on dative choice across different British regions?

1. Fit random forest per region
2. Determine rank of each predictor per region
3. Calculate Spearman’s rank correlation coefficient between all regions (15 comparisons)
4. Take inverse for distance (1 – coefficient)
5. Fit MDS: Reduce number of comparisons to 2

<table>
<thead>
<tr>
<th></th>
<th>Midlands</th>
<th>North</th>
<th>Scotland</th>
<th>Southeast</th>
<th>Southwest</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>0.036</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td>0.714</td>
<td>0.857</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>0.179</td>
<td>0.321</td>
<td>0.607</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southwest</td>
<td>0.071</td>
<td>0.143</td>
<td>0.786</td>
<td>0.107</td>
<td></td>
</tr>
<tr>
<td>Wales</td>
<td>0.107</td>
<td>0.179</td>
<td>0.679</td>
<td>0.071</td>
<td>0.071</td>
</tr>
</tbody>
</table>
(4) RESULTS

COMPARING FACTOR IMPORTANCE ACROSS REGIONS

![Graph showing factor importance across regions with regions labeled as Midlands, North, Scotland, Southeast, Southwest, and Wales. The graph plots factors on a two-dimensional scale, with labeled regions positioned accordingly.]
RESEARCH QUESTION (1)
To which extent is pattern choice determined by the sociolinguistic factors ‘region/origin of the speaker’ and ‘time’ when well-established language-internal predictors are modelled simultaneously?

- Region = third most important factor
- Time (‘corpus’) = least important factor

RESEARCH QUESTION (2)
Do language-internal predictors differ in their importance on dative choice across different British regions?

- Yes.
- Differences between regions in the way language-internal predictors impact dative choice
- Close correlation with geographic distance between regions

SUMMARY
- Region = third most important factor
- Time (‘corpus’) = least important factor
- Yes.
- Differences between regions in the way language-internal predictors impact dative choice
- Close correlation with geographic distance between regions
(5) SUMMARY

- **Importance of theme pronominality**
  due to higher frequency of altDOC datives?

- **Scotland behaves differently from the other regions**
  influence of Danelaw, language contact?

- **Fundamental Dialectology Principle (Nerbonne & Kleiweg 2007: 154)**
  importance of geographic proximity to predict linguistic similarity

- **On the methodological level**
  Important to include other possible (dialectal) variants in the analysis to contribute to an integrated approach to syntactic variability in non-standard varieties
THANK YOU!

JOHANNA GERWIN
gerwin@anglistik.uni-kiel.de

MELANIE RÖTHLISBERGER
melanie.roethlisberger@es.uzh.ch
@m_rothlisberger

More information in our upcoming paper!
Proportional distributions by region

Proportional distributions by corpus
APPENDIX

REGIONAL DISTRIBUTION OF DITRANS. PATTERNS IN FRED

... AND IN THE BNC_{REG}
APPENDIX

Proportional distributions by recipient pronominality

- **Recipient pronominality**
  - Dative variant
    - $altDOC$ vs $canDOC$ vs $canPREP$

Proportional distributions by theme pronominality

- **Theme pronominality**
  - Dative variant
    - $altDOC$ vs $canDOC$ vs $canPREP$
APPENDIX

Proportional distributions by recipient animacy

Proportional distributions by verb semantics
Proportional distributions by weight ratio