The geometry of linguistic variation

Koen Plevoets
Faculty of Applied Language Studies
University College Ghent
The geometry of linguistic variation

A lectometrical analysis of Belgian Dutch
The geometry of linguistic variation

- Lectometry: measurement of distances between language varieties or ‘lects’
- ‘Lect’: generalisation of dialect, sociolect, mesolect,…
- Convergence of
  - Dialectometry (Heeringa 2004)
  - Register analysis (Biber 1995)
- OUTCOME: geometrical representation
Overview

- Background: Belgian Dutch
- General research question
- Data
- Method
- Case studies
- Conclusion
Background

Belgian Dutch: Dutch spoken in Flanders, the Northern part of Belgium
Background

• 1585: fall of Antwerp
  • Southern provinces (i.e. Flanders) remained under foreign rule (Spanish, Austrian, French)
  • Northern provinces (i.e. The Netherlands) had a ‘Golden Century’ (17th C)
• The Netherlands developed Standard Dutch
Background

- 1830: independence of Belgium
- French is the overall dominant language
- 2nd half 19th C: industrialisation
  - Walloon cities of Charleroi, Mons, La Louvière, Liège, Verviers,...
  - Flanders remains agrarian and poor
- Reaction: Flemish Movement
Background

- Standard Dutch adopted from The Netherlands
  - 1873, 1878, 1883, 1898: ‘language laws’ grant official recognition of Dutch
  - 1st half 20th C: gradual social improvement
  - After WW II: large-scale standardisation efforts (mass-media)

- RESULT: diglossia
Background

~ Auer (2005)

Standard Dutch

Flemish dialects
Background

• Main sociolinguistic finding: *linguistic insecurity* (Deprez & Geerts 1977, Knops 1982, Geerts 1985,...)
  • DEF: tendency to depreciate items which one uses oneself
  • Hypercorrections
  • ‘Sunday suit’
Background

- 2nd half 20th C: economic development of Flanders
- Service and knowledge economy
- Prosperity and welfare
Background

- Upward mobility
- Formation of an elite
Background

~ Auer (2005): ‘diaglossia’
Background

‘Flemish Diamond’
Background

‘Flemish Diamond’
Background

~ Auer (2005)

Standard Dutch

Regiolects
Background

• ‘Brabantic expansion’: neighbouring regions adopt Brabantic forms (prestige)

• Attitude change: inhabitants of Brabantic area rate their vernacular forms higher than standard forms (linguistic confidence)
  (Jaspaert 1986, Impe 2010)

~ Arabic (exemplary case of diglossia)
Background
Background

CONCLUSION: disintegration of Belgian Dutch into a multitude of varieties
General research question

Depict the varieties of Belgian Dutch
Data

- **TSS**
  - Spoken Dutch Corpus (10M #)
  - Speech situations & social factors
  - 37 linguistic variables

- **COMURE**
  - Dutch Parallel Corpus (10M #)
  - Text types & source languages
  - 13 linguistic variables
Method

• Linguistic variable: alternation of linguistic variants

• E.g.
  • -ing vs. -in’
  • (TO BE) not vs. ain’t
  • ...

• ...
Method

<table>
<thead>
<tr>
<th></th>
<th>wvl</th>
<th>ovl</th>
<th>bra</th>
<th>lim</th>
</tr>
</thead>
<tbody>
<tr>
<td>deze</td>
<td>30</td>
<td>46</td>
<td>55</td>
<td>24</td>
</tr>
<tr>
<td>dezen</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>die</td>
<td>821</td>
<td>894</td>
<td>1222</td>
<td>611</td>
</tr>
<tr>
<td>dieje.n</td>
<td>27</td>
<td>61</td>
<td>345</td>
<td>89</td>
</tr>
<tr>
<td>diene.n</td>
<td>143</td>
<td>383</td>
<td>114</td>
<td>59</td>
</tr>
<tr>
<td>geen</td>
<td>215</td>
<td>239</td>
<td>353</td>
<td>167</td>
</tr>
<tr>
<td>gene.n</td>
<td>23</td>
<td>97</td>
<td>150</td>
<td>57</td>
</tr>
</tbody>
</table>

...
Method

- Correspondence Analysis: Chi-square as a distance metric (i.e. Mahalanobis distance)
  \[
  \sqrt{(bra - wvl)^T S^{-1} (bra - wvl)}
  \]
- Partitioning: Huyghens’ theorem
  \[
  X^2_{Total} = X^2_{Between} + X^2_{Within}
  \]
  \[
  X^2_{Within} = \sum \sum \sum \frac{(O - E_{sub})^2}{E}
  \]
Case study 1: TSS

• Spoken Dutch Corpus (10M #)
• Speech situations & social factors
• 37 linguistic variables
Case study 1: TSS

- a: Face-to-face conversations
- b: Interviews with teachers of Dutch
- c: Telephone dialogues (switchboard)
- d: Telephone dialogues (mini disc)
- f: Broadcast discussions/debates
- g: Non-broadcast discussions/debates
- h: Lessons recorded in classroom
Case study 1: TSS

- **i**: Live commentaries (sports)
- **j**: News reports/reportages
- **k**: News bulletins
- **l**: Commentaries/columns/reviews
- **m**: Ceremonious speeches/sermons
- **n**: Lectures/seminars
- **o**: Read texts
Case study 1: TSS

• **A**: Occupation in higher management or government
• **B**: Occupation requiring higher education
• **C**: Employed in the teaching or research staff of a university or a college
• **D**: Employed in an administrative office or a service organisation
• **E**: Occupation not requiring any level of specification
Case study 1: TSS

- **F**: Self-employed
- **G**: Politician
- **H**: Employed in the media (journalist, reporter) or artist
- **I**: Student, trainee
- **J**: Unemployed
Case study 1: TSS

- Informalisation (Wouters 2007): the growing welfare makes people more independent, which deregulates social norms

- Engel’s law
  - Ernst Engel (1821-1896)
  - As income increases, the expenditure on basic necessities (e.g. food) may increase in absolute numbers but decreases in relative numbers
  - As income increases, one proportionally invests more in luxury and/or leisurely activities
Engel's law
Engel's law

Expenditure on food vs. Income
Case study 2: COMURE

- Dutch Parallel Corpus (10M #)
- Text types & source languages
- 13 linguistic variables
Case study 2: COMURE

• Background:
  • Law of growing standardisation (Toury 1995) or Normalisation universal (Baker 1993)
  • Translations tend to be closer to the standard norms than non-translations

• Project:
  • Isabelle Delaere (Ph.D.)
  • Gert De Sutter (supervisor)
Conclusion

• (In)formality depends on context: formal situations ask for formal variants
• (In)formality depends on power: dominant individuals can afford themselves more leniency
Bibliography
Thank you!

koen.plevoets@hogent.be

isabelle.delaere@hogent.be
gert.desutter@hogent.be