Applying quantitative methods to dialect Dutch verb clusters

Jeroen van Craenenbroeck

KU Leuven

May 8, 2014, Tartu, Estonia

Mapping Methods: Approaches to Language Studies
0 – Outline

1. Introduction & central research question
2. The data
3. A dialectometric analysis
4. Reversing the perspective
5. Summary and conclusions

Applying quantitative methods to dialect Dutch verb clusters
1 – Outline

1. Introduction & central research question
2. The data
3. A dialectometric analysis
4. Reversing the perspective
5. Summary and conclusions
1 – Introduction & central research question

- In Dutch (like in many head-final Germanic languages) verbs tend to cluster at the right periphery of the clause:

(1) Ik vind dat iedereen moet kunnen zwemmen.
    I find that everyone must can swim
    ‘I think everyone should be able to swim.’
in Dutch (like in many head-final Germanic languages) verbs tend to cluster at the right periphery of the clause:

(1)  Ik vind dat iedereen moet kunnen zwemmen.
    I find that everyone must can swim
    ‘I think everyone should be able to swim.’

such verbal clusters show a considerable degree of word order freedom within Dutch:

(2)  a.  Ik vind dat iedereen moet zwemmen kunnen.
    b.  Ik vind dat iedereen zwemmen moet kunnen.
    c.  Ik vind dat iedereen zwemmen kunnen moet.
    d.  *Ik vind dat iedereen kunnen zwemmen moet.
    e.  *Ik vind dat iedereen kunnen moet zwemmen.
1 – Introduction & central research question

- in Dutch (like in many head-final Germanic languages) verbs tend to cluster at the right periphery of the clause:

(1) Ik vind dat iedereen moet kunnen zwemmen. (123)
I find that everyone must can swim
‘I think everyone should be able to swim.’

- such verbal clusters show a considerable degree of word order freedom within Dutch:

(2) a. Ik vind dat iedereen moet zwemmen kunnen. (132)
b. Ik vind dat iedereen zwemmen moet kunnen. (312)
c. Ik vind dat iedereen zwemmen kunnen moet. (321)
d. *Ik vind dat iedereen kunnen zwemmen moet. (231)
e. *Ik vind dat iedereen kunnen moet zwemmen. (213)
1 – Introduction & central research question

- this wide range of word order variation poses a challenge for formal linguistic theories
this wide range of word order variation poses a challenge for formal linguistic theories.

E.g., under a uniform head-initial structure the 123-order follows naturally, but all other orders have to be derived:

\[
(3) \quad \begin{array}{c}
\text{VP1} \\
\text{V1} \\
\underline{\text{must}} \\
\text{V2} \\
\underline{\text{can}} \\
\text{VP3} \\
\text{V3} \\
\underline{\text{swim}}
\end{array}
\]
1 – Introduction & central research question

- there is an extensive theoretical literature on how (not) to derive particular cluster orders (see Wurmbrand (2005) for an overview)

Applying quantitative methods to dialect Dutch verb clusters
1 – Introduction & central research question

- there is an extensive theoretical literature on how (not) to derive particular cluster orders (see Wurmbrand (2005) for an overview)
- missing from the literature is a systematic study of the *correlations* between various cluster orders, e.g.
1 – Introduction & central research question

- There is an extensive theoretical literature on how (not) to derive particular cluster orders (see Wurmbrand (2005) for an overview).
- Missing from the literature is a systematic study of the correlations between various cluster orders, e.g.

(4) **Midsland Dutch**

a. *dat elkeen mot kanne zwemme.*
   that everyone must can swim
   ‘that everyone should be able to swim.’

b. dat elkeen mot zwemme kanne.

(123)  

(312)  

(213)  

(231)  

(321)  

(132)  

Applying quantitative methods to dialect Dutch verb clusters
there is an extensive theoretical literature on how (not) to derive particular cluster orders (see Wurmbrand (2005) for an overview)

missing from the literature is a systematic study of the **correlations** between various cluster orders, e.g.

(5) **Langelo Dutch**

a. dat iedereen moet kunnen zwemmen.
   that everyone must can swim
   ‘that everyone should be able to swim.’

b. *dat iedereen mot zwemmen kunnen.

c. dat iedereen zwemmen mot kunnen.

(✓312)

d. *dat iedereen zwemmen kunnen mot.

(✓321)

e. *dat iedereen kunnen zwemmen mot.

(✓231)

f. *dat iedereen kunnen mot zwemmen.

(✓213)
1 – Introduction & central research question

- in verb clusters of the type MODAL-MODAL-MAIN V, 231 and 213 are systematically excluded
1 – Introduction & central research question

- In verb clusters of the type MODAL-MODAL-MAIN V, 231 and 213 are systematically excluded.
- Of the remaining 15 combinations of cluster orders \( (2^4 - 1) \), 12 are attested.
1 – Introduction & central research question

- in verb clusters of the type MODAL-MODAL-MAIN V, 231 and 213 are systematically excluded

- of the remaining 15 combinations of cluster orders ($2^4-1$), 12 are attested:

<table>
<thead>
<tr>
<th>example dialect</th>
<th>123</th>
<th>132</th>
<th>321</th>
<th>312</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beetgum</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hippolytushoef</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>Warffum</td>
<td>✓</td>
<td>✓</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Oosterend</td>
<td>✓</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Schermerhorn</td>
<td>✓</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>Visvliet</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Kollum</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>Langelo</td>
<td>✓</td>
<td>*</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>Midsland</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>Lies</td>
<td>*</td>
<td>*</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>Bakkeveen</td>
<td>*</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Waskemeer</td>
<td>*</td>
<td>✓</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
1 – Introduction & central research question

- **goal of this talk**: to investigate the correlations between verb cluster orders from a quantitative perspective
goal of this talk: to investigate the correlations between verb cluster orders from a quantitative perspective
goal of this talk: to investigate the correlations between verb cluster orders from a quantitative perspective

Main research question:
Can a quantitative analysis of verb cluster orders shed new light on the theoretical analysis of this phenomenon?
1 – Introduction & central research question

- broader issues:

- the grammatical nature of microvariation: are there grammatical microparameters at play in the word order variation found in verb clusters and if so, how can we uncover them?
  - e.g. Barbiers (2005): grammar rules out 231 and 213 in MOD-MOD-V-cluster, but all other orders are freely available to all speakers; the choice between them is determined by sociolinguistic factors

- the interaction between quantitative-statistical and formal-theoretical approaches to language:
  - to what extent can a quantitative analysis of large datasets lead to new theoretical insights?
  - to what extent can theoretical analyses guide and inform quantitative analyses of language data?

Applying quantitative methods to dialect Dutch verb clusters
1 – Introduction & central research question

- broader issues:
  - the grammatical nature of microvariation:

Applying quantitative methods to dialect Dutch verb clusters
broader issues:

- the grammatical nature of microvariation:
  - are there grammatical microparameters at play in the word order variation found in verb clusters and if so, how can we uncover them?
broader issues:

- the grammatical nature of microvariation:
  - are there grammatical microparameters at play in the word order variation found in verb clusters and if so, how can we uncover them?
  - e.g. Barbiers (2005): grammar rules out 231 and 213 in MOD-MOD-V-cluster, but all other orders are freely available to all speakers; the choice between them is determined by sociolinguistic factors
1 – Introduction & central research question

**broader issues:**

1. the grammatical nature of microvariation:
   - are there grammatical microparameters at play in the word order variation found in verb clusters and if so, how can we uncover them?
   - e.g. Barbiers (2005): grammar rules out 231 and 213 in MOD-MOD-V-cluster, but all other orders are freely available to all speakers; the choice between them is determined by sociolinguistic factors

2. the interaction between quantitative-statistical and formal-theoretical approaches to language:

**Applying quantitative methods to dialect Dutch verb clusters**
broader issues:

1. the grammatical nature of microvariation:
   - are there grammatical microparameters at play in the word order variation found in verb clusters and if so, how can we uncover them?
   - e.g. Barbiers (2005): grammar rules out 231 and 213 in MOD-MOD-V-cluster, but all other orders are freely available to all speakers; the choice between them is determined by sociolinguistic factors

2. the interaction between quantitative-statistical and formal-theoretical approaches to language:
   - to what extent can a quantitative analysis of large datasets lead to new theoretical insights?
1 – Introduction & central research question

broader issues:

1. the grammatical nature of microvariation:
   - are there grammatical microparameters at play in the word order variation found in verb clusters and if so, how can we uncover them?
   - e.g. Barbiers (2005): grammar rules out 231 and 213 in MOD-MOD-V-cluster, but all other orders are freely available to all speakers; the choice between them is determined by sociolinguistic factors

2. the interaction between quantitative-statistical and formal-theoretical approaches to language:
   - to what extent can a quantitative analysis of large datasets lead to new theoretical insights?
   - to what extent can theoretical analyses guide and inform quantitative analyses of language data?
2 – Outline

1. Introduction & central research question
2. The data
3. A dialectometric analysis
4. Reversing the perspective
5. Summary and conclusions

Applying quantitative methods to dialect Dutch verb clusters
all data in this talk stem from the SAND-project:
2 – The data

- all data in this talk stem from the SAND-project:
2 – The data

- all data in this talk stem from the SAND-project:
  - universities of Amsterdam, Leiden, Antwerp, Brussels, Ghent, and the Meertensinstitute
2 – The data

- all data in this talk stem from the SAND-project:
  - universities of Amsterdam, Leiden, Antwerp, Brussels, Ghent, and the Meertensinstitute
  - four themes: left periphery of the clause, right periphery of the clause, pronominal reference, negation & quantification
2 – The data

- all data in this talk stem from the SAND-project:
  - universities of Amsterdam, Leiden, Antwerp, Brussels, Ghent, and the Meertensinstitute
  - four themes: left periphery of the clause, right periphery of the clause, pronominal reference, negation & quantification
  - 267 dialect locations in Belgium and the Netherlands
2 – The data

- all data in this talk stem from the SAND-project:
  - universities of Amsterdam, Leiden, Antwerp, Brussels, Ghent, and the Meertensinstitute
  - four themes: left periphery of the clause, right periphery of the clause, pronominal reference, negation & quantification
  - 267 dialect locations in Belgium and the Netherlands
  - has yielded two atlas volumes

Applying quantitative methods to dialect Dutch verb clusters
1.3.2.1  

Modal_{FINITE} • Modal_{INFINITIVE} • V_{INFINITIVE}

Ik vind dat iedereen moet kunnen zwemmen.
I think that everyone must.FIN can.INF swim.INF

‘I think that everyone should be able to swim.’

- $V_1$-$V_2$-$V_3$ (moet kunnen zwemmen) 242
- $V_1$-$V_3$-$V_2$ (moet zwemmen kunnen) 34
- $V_3$-$V_1$-$V_2$ (zwemmen moet kunnen) 83
- $V_3$-$V_2$-$V_1$ (zwemmen kunnen moet) 37
2 – The data

- the data analysis in this talk is based on the raw data from 13 SAND-maps:

- about two-verb clusters (3 ⇨ auxiliary-participle, 1 ⇨ modal-infinitive)
- about three-verb clusters (modal-modal-infinitive, modal-auxiliary-participle, auxiliary-auxiliary-infinitive, auxiliary-modal-infinitive)
- about particle placement inside the cluster
- about morphology of the past participle

for a total of 67 linguistic variables in 267 locations (=17889 data points)

Two caveats:
1. not all questions used the same methodology: translation tasks vs. direct judgment questions
2. not all questions were asked in all dialect locations (the data table contains 8% NAs)
2 – The data

- the data analysis in this talk is based on the raw data from 13 SAND-maps:
  - 4 about two-verb clusters (3×auxiliary-participle, 1×modal-infinitive)
- two caveats:
  1. not all questions used the same methodology: translation tasks vs. direct judgment questions
  2. not all questions were asked in all dialect locations (the data table contains 8% NAs)
2 – The data

- The data analysis in this talk is based on the raw data from 13 SAND-maps:
  - 4 about two-verb clusters (3× auxiliary-participle, 1× modal-infinitive)
  - 4 about three-verb clusters (modal-modal-infinitive, modal-auxiliary-participle, auxiliary-auxiliary-infinitive, auxiliary-modal-infinitive)

- About particle placement inside the cluster
- About morphology of the past participle

- For a total of 67 linguistic variables in 267 locations (=17889 data points)

Two caveats:
- Not all questions used the same methodology: translation tasks vs. direct judgment questions
- Not all questions were asked in all dialect locations (the data table contains 8% NAs)
2 – The data

- The data analysis in this talk is based on the raw data from 13 SAND-maps:
  - 4 about two-verb clusters (3×auxiliary-participle, 1×modal-infinitive)
  - 4 about three-verb clusters (modal-modal-infinitive, modal-auxiliary-participle, auxiliary-auxiliary-infinitive, auxiliary-modal-infinitive)
  - 3 about particle placement inside the cluster

- Two caveats:
  1. Not all questions used the same methodology: translation tasks vs. direct judgment questions
  2. Not all questions were asked in all dialect locations (the data table contains 8% NAs)
2 – The data

- The data analysis in this talk is based on the raw data from 13 SAND-maps:
  - 4 about two-verb clusters (3×auxiliary-participle, 1×modal-infinitive)
  - 4 about three-verb clusters (modal-modal-infinitive, modal-auxiliary-participle, auxiliary-auxiliary-infinitive, auxiliary-modal-infinitive)
  - 3 about particle placement inside the cluster
  - 2 about morphology of the past participle
2 – The data

- the data analysis in this talk is based on the raw data from 13 SAND-maps:
  - 4 about two-verb clusters (3 × auxiliary-participle, 1 × modal-infinitive)
  - 4 about three-verb clusters (modal-modal-infinitive, modal-auxiliary-participle, auxiliary-auxiliary-infinitive, auxiliary-modal-infinitive)
  - 3 about particle placement inside the cluster
  - 2 about morphology of the past participle

- for a total of 67 linguistic variables in 267 locations (=17889 data points)
2 – The data

- The data analysis in this talk is based on the raw data from 13 SAND-maps:
  - 4 about two-verb clusters (3×auxiliary-participle, 1×modal-infinitive)
  - 4 about three-verb clusters (modal-modal-infinitive, modal-auxiliary-participle, auxiliary-auxiliary-infinitive, auxiliary-modal-infinitive)
  - 3 about particle placement inside the cluster
  - 2 about morphology of the past participle

- For a total of 67 linguistic variables in 267 locations (=17889 data points)
- Two caveats:
2 – The data

- The data analysis in this talk is based on the raw data from 13 SAND-maps:
  - 4 about two-verb clusters (3×auxiliary-participle, 1×modal-infinitive)
  - 4 about three-verb clusters (modal-modal-infinitive, modal-auxiliary-participle, auxiliary-auxiliary-infinitive, auxiliary-modal-infinitive)
  - 3 about particle placement inside the cluster
  - 2 about morphology of the past participle

- For a total of 67 linguistic variables in 267 locations (=17889 data points)

- Two caveats:
  - Not all questions used the same methodology: translation tasks vs. direct judgment questions
2 – The data

- the data analysis in this talk is based on the raw data from 13 SAND-maps:
  - 4 about two-verb clusters (3 × auxiliary-participle, 1 × modal-infinitive)
  - 4 about three-verb clusters (modal-modal-infinitive, modal-auxiliary-participle, auxiliary-auxiliary-infinitive, auxiliary-modal-infinitive)
  - 3 about particle placement inside the cluster
  - 2 about morphology of the past participle

- for a total of 67 linguistic variables in 267 locations (=17889 data points)

- two caveats:
  1. not all questions used the same methodology: translation tasks vs. direct judgment questions
  2. not all questions were asked in all dialect locations (the data table contains 8% NAs)
3 – Outline

1. Introduction & central research question
2. The data
3. A dialectometric analysis
4. Reversing the perspective
5. Summary and conclusions
3 – A dialectometric analysis: Introduction

**note:** given that the SAND-data stems from questionnaires, a number of possible sociolinguistic variables has been controlled for (e.g. age and social status of speaker, register, choice of verb, etc.)

Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: Introduction

• **note:** given that the SAND-data stems from questionnaires, a number of possible sociolinguistic variables has been controlled for (e.g. age and social status of speaker, register, choice of verb, etc.)

• one variable that is explicitly—and deliberately—not kept constant is **location**
3 – A dialectometric analysis: Introduction

- **note:** given that the SAND-data stems from questionnaires, a number of possible sociolinguistic variables has been controlled for (e.g. age and social status of speaker, register, choice of verb, etc.)
- one variable that is explicitly—and deliberately—not kept constant is location
- so if Barbiers’s analysis is on the right track, we might expect location to be a deciding factor in accounting for the attested variation in verb cluster ordering
note: given that the SAND-data stems from questionnaires, a number of possible sociolinguistic variables has been controlled for (e.g. age and social status of speaker, register, choice of verb, etc.)

one variable that is explicitly—and deliberately—not kept constant is location

so if Barbiers’s analysis is on the right track, we might expect location to be a deciding factor in accounting for the attested variation in verb cluster ordering

dialectometry is a subdiscipline of linguistics that uses computational and quantitative techniques in dialectology (Nerbonne and Kretzschmar Jr., 2013)
3 – A dialectometric analysis: Introduction

- **note:** given that the SAND-data stems from questionnaires, a number of possible sociolinguistic variables has been controlled for (e.g. age and social status of speaker, register, choice of verb, etc.)
- one variable that is explicitly—and deliberately—not kept constant is **location**
- so if Barbiers’s analysis is on the right track, we might expect location to be a deciding factor in accounting for the attested variation in verb cluster ordering
- **dialectometry** is a subdiscipline of linguistics that uses computational and quantitative techniques in dialectology (Nerbonne and Kretzschmar Jr., 2013)
- it offers precisely the tools needed to trace the effect of location on the verb cluster data

Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: An MDS-analysis

- starting point: a $267 \times 67$ matrix with one row per location and one column per linguistic variable
## Applying quantitative methods to dialect Dutch verb clusters

<table>
<thead>
<tr>
<th>row.names</th>
<th>1_2_14a</th>
<th>2_1_14a</th>
<th>1_2_14b</th>
<th>2_1_14b</th>
<th>1_2_15a</th>
<th>2_1_15a</th>
<th>1_2_15b</th>
<th>2_1_15b</th>
<th>kunnen_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A001p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A001q</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A002p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A006p</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B001a</td>
<td>0</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B004p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B007p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B013b</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B035p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B041p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B046b</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>B052a</td>
<td>0</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>B062p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B085c</td>
<td>0</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>B127p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B128a</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C023p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C029p</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
starting point: a $267 \times 67$ matrix with one row per location and one column per linguistic variable

step 1: convert the table into a $267 \times 267$ (symmetric) distance matrix, whereby for each pair of locations a distance between them is calculated based on the linguistic features they share
Applying quantitative methods to dialect Dutch verb clusters

<table>
<thead>
<tr>
<th></th>
<th>A001p</th>
<th>A001q</th>
<th>A002p</th>
<th>A006p</th>
<th>B001a</th>
<th>B004p</th>
<th>B007p</th>
<th>B013b</th>
<th>B035p</th>
<th>B041p</th>
<th>B046b</th>
<th>B052a</th>
<th>B062p</th>
</tr>
</thead>
<tbody>
<tr>
<td>A001p</td>
<td>0.000</td>
<td>0.500</td>
<td>0.333</td>
<td>0.706</td>
<td>0.250</td>
<td>0.647</td>
<td>0.357</td>
<td>0.250</td>
<td>0.611</td>
<td>0.650</td>
<td>0.533</td>
<td>0.545</td>
<td>0.500</td>
</tr>
<tr>
<td>A001q</td>
<td>0.500</td>
<td>0.000</td>
<td>0.444</td>
<td>0.750</td>
<td>0.588</td>
<td>0.375</td>
<td>0.471</td>
<td>0.563</td>
<td>0.444</td>
<td>0.444</td>
<td>0.632</td>
<td>0.714</td>
<td>0.500</td>
</tr>
<tr>
<td>A002p</td>
<td>0.333</td>
<td>0.444</td>
<td>0.000</td>
<td>0.789</td>
<td>0.429</td>
<td>0.667</td>
<td>0.286</td>
<td>0.429</td>
<td>0.632</td>
<td>0.600</td>
<td>0.500</td>
<td>0.500</td>
<td>0.429</td>
</tr>
<tr>
<td>A006p</td>
<td>0.706</td>
<td>0.750</td>
<td>0.789</td>
<td>0.000</td>
<td>0.706</td>
<td>0.765</td>
<td>0.737</td>
<td>0.538</td>
<td>0.563</td>
<td>0.600</td>
<td>0.600</td>
<td>0.727</td>
<td>0.813</td>
</tr>
<tr>
<td>B001a</td>
<td>0.250</td>
<td>0.588</td>
<td>0.429</td>
<td>0.706</td>
<td>0.000</td>
<td>0.667</td>
<td>0.167</td>
<td>0.000</td>
<td>0.625</td>
<td>0.714</td>
<td>0.462</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>B004p</td>
<td>0.647</td>
<td>0.375</td>
<td>0.667</td>
<td>0.765</td>
<td>0.667</td>
<td>0.000</td>
<td>0.625</td>
<td>0.667</td>
<td>0.400</td>
<td>0.556</td>
<td>0.706</td>
<td>0.750</td>
<td>0.571</td>
</tr>
<tr>
<td>B007p</td>
<td>0.357</td>
<td>0.471</td>
<td>0.286</td>
<td>0.737</td>
<td>0.167</td>
<td>0.625</td>
<td>0.000</td>
<td>0.182</td>
<td>0.588</td>
<td>0.682</td>
<td>0.308</td>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>B013b</td>
<td>0.250</td>
<td>0.563</td>
<td>0.429</td>
<td>0.538</td>
<td>0.000</td>
<td>0.667</td>
<td>0.182</td>
<td>0.000</td>
<td>0.571</td>
<td>0.625</td>
<td>0.417</td>
<td>0.556</td>
<td>0.500</td>
</tr>
<tr>
<td>B035p</td>
<td>0.611</td>
<td>0.444</td>
<td>0.632</td>
<td>0.563</td>
<td>0.625</td>
<td>0.400</td>
<td>0.588</td>
<td>0.571</td>
<td>0.000</td>
<td>0.353</td>
<td>0.625</td>
<td>0.643</td>
<td>0.429</td>
</tr>
<tr>
<td>B041p</td>
<td>0.650</td>
<td>0.444</td>
<td>0.600</td>
<td>0.600</td>
<td>0.714</td>
<td>0.556</td>
<td>0.682</td>
<td>0.625</td>
<td>0.353</td>
<td>0.000</td>
<td>0.588</td>
<td>0.500</td>
<td>0.667</td>
</tr>
<tr>
<td>B046b</td>
<td>0.533</td>
<td>0.632</td>
<td>0.500</td>
<td>0.600</td>
<td>0.462</td>
<td>0.706</td>
<td>0.308</td>
<td>0.417</td>
<td>0.625</td>
<td>0.588</td>
<td>0.000</td>
<td>0.167</td>
<td>0.571</td>
</tr>
<tr>
<td>B052a</td>
<td>0.545</td>
<td>0.714</td>
<td>0.500</td>
<td>0.727</td>
<td>0.500</td>
<td>0.750</td>
<td>0.333</td>
<td>0.556</td>
<td>0.643</td>
<td>0.500</td>
<td>0.167</td>
<td>0.000</td>
<td>0.500</td>
</tr>
<tr>
<td>B062p</td>
<td>0.500</td>
<td>0.500</td>
<td>0.429</td>
<td>0.813</td>
<td>0.500</td>
<td>0.571</td>
<td>0.333</td>
<td>0.500</td>
<td>0.429</td>
<td>0.667</td>
<td>0.571</td>
<td>0.500</td>
<td>0.000</td>
</tr>
<tr>
<td>B085c</td>
<td>0.692</td>
<td>0.667</td>
<td>0.583</td>
<td>0.846</td>
<td>0.545</td>
<td>0.667</td>
<td>0.400</td>
<td>0.600</td>
<td>0.571</td>
<td>0.692</td>
<td>0.500</td>
<td>0.455</td>
<td>0.222</td>
</tr>
<tr>
<td>B127p</td>
<td>0.400</td>
<td>0.500</td>
<td>0.438</td>
<td>0.706</td>
<td>0.385</td>
<td>0.563</td>
<td>0.357</td>
<td>0.385</td>
<td>0.438</td>
<td>0.579</td>
<td>0.533</td>
<td>0.545</td>
<td>0.385</td>
</tr>
<tr>
<td>B128a</td>
<td>0.438</td>
<td>0.526</td>
<td>0.556</td>
<td>0.818</td>
<td>0.500</td>
<td>0.588</td>
<td>0.471</td>
<td>0.533</td>
<td>0.471</td>
<td>0.652</td>
<td>0.588</td>
<td>0.667</td>
<td>0.429</td>
</tr>
<tr>
<td>C023p</td>
<td>0.500</td>
<td>0.412</td>
<td>0.611</td>
<td>0.810</td>
<td>0.563</td>
<td>0.357</td>
<td>0.529</td>
<td>0.600</td>
<td>0.333</td>
<td>0.636</td>
<td>0.706</td>
<td>0.667</td>
<td>0.385</td>
</tr>
<tr>
<td>C029p</td>
<td>0.563</td>
<td>0.438</td>
<td>0.667</td>
<td>0.737</td>
<td>0.625</td>
<td>0.429</td>
<td>0.588</td>
<td>0.643</td>
<td>0.400</td>
<td>0.652</td>
<td>0.600</td>
<td>0.636</td>
<td>0.571</td>
</tr>
<tr>
<td>C041a</td>
<td>0.667</td>
<td>0.652</td>
<td>0.739</td>
<td>0.550</td>
<td>0.773</td>
<td>0.650</td>
<td>0.739</td>
<td>0.722</td>
<td>0.389</td>
<td>0.455</td>
<td>0.667</td>
<td>0.571</td>
<td>0.684</td>
</tr>
<tr>
<td>C108p</td>
<td>0.714</td>
<td>0.682</td>
<td>0.714</td>
<td>0.636</td>
<td>0.783</td>
<td>0.762</td>
<td>0.800</td>
<td>0.778</td>
<td>0.471</td>
<td>0.476</td>
<td>0.684</td>
<td>0.714</td>
<td>0.737</td>
</tr>
<tr>
<td>C123p</td>
<td>0.650</td>
<td>0.682</td>
<td>0.650</td>
<td>0.652</td>
<td>0.773</td>
<td>0.762</td>
<td>0.739</td>
<td>0.722</td>
<td>0.556</td>
<td>0.368</td>
<td>0.647</td>
<td>0.615</td>
<td>0.667</td>
</tr>
<tr>
<td>C146t</td>
<td>0.727</td>
<td>0.524</td>
<td>0.739</td>
<td>0.652</td>
<td>0.792</td>
<td>0.650</td>
<td>0.760</td>
<td>0.647</td>
<td>0.550</td>
<td>0.500</td>
<td>0.700</td>
<td>0.824</td>
<td>0.810</td>
</tr>
<tr>
<td>C148p</td>
<td>0.773</td>
<td>0.636</td>
<td>0.727</td>
<td>0.600</td>
<td>0.818</td>
<td>0.556</td>
<td>0.783</td>
<td>0.789</td>
<td>0.600</td>
<td>0.500</td>
<td>0.611</td>
<td>0.714</td>
<td>0.800</td>
</tr>
</tbody>
</table>
starting point: a $267 \times 67$ matrix with one row per location and one column per linguistic variable

step 1: convert the table into a $267 \times 267$ (symmetric) distance matrix, whereby for each pair of locations a distance between them is calculated based on the linguistic features they share

step 2: apply multidimensional scaling (MDS) to the distance matrix
3 – A dialectometric analysis: An MDS-analysis

- starting point: a $267 \times 67$ matrix with one row per location and one column per linguistic variable
- step 1: convert the table into a $267 \times 267$ (symmetric) distance matrix, whereby for each pair of locations a distance between them is calculated based on the linguistic features they share
- step 2: apply multidimensional scaling (MDS) to the distance matrix
- MDS is a technique for reducing a multidimensional distance matrix into a lower-dimensional (typically two or three, though see later) one that retains—as much as possible—the original distances
starting point: a $267 \times 67$ matrix with one row per location and one column per linguistic variable

step 1: convert the table into a $267 \times 267$ (symmetric) distance matrix, whereby for each pair of locations a distance between them is calculated based on the linguistic features they share

step 2: apply multidimensional scaling (MDS) to the distance matrix

MDS is a technique for reducing a multidimensional distance matrix into a lower-dimensional (typically two or three, though see later) one that retains—as much as possible—the original distances

MDS makes the data accessible to visual inspection and exploration (Borg and Groenen, 2005)
Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: An MDS-analysis

- **note:** the data is not randomly distributed → this suggests that the distribution of verb cluster orderings across the Dutch-speaking area is not random
3 – A dialectometric analysis: An MDS-analysis

- **note:** the data is not randomly distributed → this suggests that the distribution of verb cluster orderings across the Dutch-speaking area is not random

- we can discern roughly three regions
Applying quantitative methods to dialect Dutch verb clusters
A dialectometric analysis: An MDS-analysis

- **note:** the data is not randomly distributed → this suggests that the distribution of verb cluster orderings across the Dutch-speaking area is not random
- we can discern roughly three regions
- recall: each point represents a location → this means we can plot these (three groups of) points against an actual map
Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: An MDS-analysis

- **note:** the data is not randomly distributed → this suggests that the distribution of verb cluster orderings across the Dutch-speaking area is not random

- we can discern roughly three regions

- recall: each point represents a location → this means we can plot these (three groups of) points against an actual map

- the three regions uncovered through MDS correspond to three homogenous geographical regions → this seems to suggest that geography is indeed largely responsible for the variation found in verb clusters (and hence that Barbiers’s hypothesis might be on the right track)
there are three reasons to think that the geography-based analysis is an oversimplification:

1. The blue (Netherlandic) region is linguistically quite diversified.
2. Breaking up the data per province disrupts the clean MDS-picture.
3. An analysis in terms of scree plot suggests the data is at least four-dimensional.
there are three reasons to think that the geography-based analysis is an oversimplification:

1. the blue (Netherlandic) region is linguistically quite diversified
there are three reasons to think that the geography-based analysis is an oversimplification:

1. the blue (Netherlandic) region is linguistically quite diversified
2. breaking up the data per province disrupts the clean MDS-picture

Applying quantitative methods to dialect Dutch verb clusters
there are three reasons to think that the geography-based analysis is an oversimplification:

1. the blue (Netherlandic) region is linguistically quite diversified
2. breaking up the data per province disrupts the clean MDS-picture
3. an analysis in terms of scree plot suggests the data is at least four-dimensional
there are three reasons to think that the geography-based analysis is an oversimplification:

1. the blue (Netherlandic) region is linguistically quite diversified
2. breaking up the data per province disrupts the clean MDS-picture
3. an analysis in terms of scree plot suggests the data is at least four-dimensional

linguistic profile of the three regions
there are three reasons to think that the geography-based analysis is an oversimplification:

1. the blue (Netherlandic) region is linguistically quite diversified
2. breaking up the data per province disrupts the clean MDS-picture
3. an analysis in terms of scree plot suggests the data is at least four-dimensional

linguistic profile of the three regions

when we map which constructions occur in which region, the green (Frisian) and the red (Belgian) region have a clear linguistic profile, while the blue (Netherlandic) region presents a very messy picture
Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):

Applying quantitative methods to dialect Dutch verb clusters
Applying quantitative methods to dialect Dutch verb clusters
Frisia (green):

- Consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
- Overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
- Some 123 and 132 in MOD-MOD-V
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)

Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V

- Flanders (red):
Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V

- Flanders (red):
  - consistently 21 in AUX-V-clusters and 12 in MOD-V-clusters

Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V

- Flanders (red):
  - consistently 21 in AUX-V-clusters and 12 in MOD-V-clusters
  - consistently 123 in MOD-MOD-V- and AUX-MOD-V-clusters
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V

- Flanders (red):
  - consistently 21 in AUX-V-clusters and 12 in MOD-V-clusters
  - consistently 123 in MOD-MOD-V- and AUX-MOD-V-clusters
  - predominantly 132 in MOD-AUX-V, with some 312
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V

- Flanders (red):
  - consistently 21 in AUX-V-clusters and 12 in MOD-V-clusters
  - consistently 123 in MOD-MOD-V- and AUX-MOD-V-clusters
  - predominantly 132 in MOD-AUX-V, with some 312
  - predominantly 231 in AUX-AUX-V, with some 123
Frisia (green):
- consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
- overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
- some 123 and 132 in MOD-MOD-V

Flanders (red):
- consistently 21 in AUX-V-clusters and 12 in MOD-V-clusters
- consistently 123 in MOD-MOD-V- and AUX-MOD-V-clusters
- predominantly 132 in MOD-AUX-V, with some 312
- predominantly 231 in AUX-AUX-V, with some 123

Netherlands (blue):
Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: A closer look at the data

- Frisia (green):
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V

- Flanders (red):
  - consistently 21 in AUX-V-clusters and 12 in MOD-V-clusters
  - consistently 123 in MOD-MOD-V- and AUX-MOD-V-clusters
  - predominantly 132 in MOD-AUX-V, with some 312
  - predominantly 231 in AUX-AUX-V, with some 123

- Netherlands (blue):
  - predominantly 123 in AUX-AUX-V and AUX-MOD-V
3 – A dialectometric analysis: A closer look at the data

- **Frisia (green):**
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V

- **Flanders (red):**
  - consistently 21 in AUX-V-clusters and 12 in MOD-V-clusters
  - consistently 123 in MOD-MOD-V- and AUX-MOD-V-clusters
  - predominantly 132 in MOD-AUX-V, with some 312
  - predominantly 231 in AUX-AUX-V, with some 123

- **Netherlands (blue):**
  - predominantly 123 in AUX-AUX-V and AUX-MOD-V
  - no clear preference in two-verb clusters
3 – A dialectometric analysis: A closer look at the data

- **Frisia (green):**
  - consistently 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V

- **Flanders (red):**
  - consistently 21 in AUX-V-clusters and 12 in MOD-V-clusters
  - consistently 123 in MOD-MOD-V- and AUX-MOD-V-clusters
  - predominantly 132 in MOD-AUX-V, with some 312
  - predominantly 231 in AUX-AUX-V, with some 123

- **Netherlands (blue):**
  - predominantly 123 in AUX-AUX-V and AUX-MOD-V
  - no clear preference in two-verb clusters
  - no clear preference in MOD-MOD-V or in MOD-AUX-V
**Conclusion:** the three-way MDS-split glosses over significant linguistic complexity in the verb cluster data.
3 – A dialectometric analysis: A closer look at the data

- **conclusion**: the three-way MDS-split glosses over significant linguistic complexity in the verb cluster data
- this intuition is further confirmed if we split up the data by province
3 – A dialectometric analysis: A closer look at the data

- **Conclusion:** The three-way MDS-split glosses over significant linguistic complexity in the verb cluster data.

- This intuition is further confirmed if we split up the data by province.

- **Data per province**
3 – A dialectometric analysis: A closer look at the data

- **Conclusion:** The three-way MDS-split glosses over significant linguistic complexity in the verb cluster data.
- This intuition is further confirmed if we split up the data by province.

- **Data per province**

  When we divvy up the scatterplot according to the 18 provinces of the Netherlands and Flanders, the picture of three homogeneous verb cluster regions disappears.
MDS-representation of 13 SAND-maps about verb clusters (split up by province)

Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: A closer look at the data

- note:
3 – A dialectometric analysis: A closer look at the data

- note:
  - geographically contiguous provinces are not necessarily MDS-contiguous
3 – A dialectometric analysis: A closer look at the data

**note:**
- geographically contiguous provinces are not necessarily MDS-contiguous
  - e.g. North Brabant vs. Antwerp

Applying quantitative methods to dialect Dutch verb clusters
note:
- geographically contiguous provinces are not necessarily MDS-contiguous
  - e.g. North Brabant vs. Antwerp
- several provinces are ‘spread out’ along certain axes
3 – A dialectometric analysis: A closer look at the data

- note:
  - geographically contiguous provinces are not necessarily MDS-contiguous
    ▶ e.g. North Brabant vs. Antwerp
  - several provinces are ‘spread out’ along certain axes
    ▶ e.g. Drenthe, North Holland, Groningen, possibly French Flanders
3 – A dialectometric analysis: A closer look at the data

● note:
  ● geographically contiguous provinces are not necessarily MDS-contiguous
    ▶ e.g. North Brabant vs. Antwerp
  ● several provinces are ‘spread out’ along certain axes
    ▶ e.g. Drenthe, North Holland, Groningen, possibly French Flanders

● this suggests that there are dimensions of variation other than purely geographical ones
3 – A dialectometric analysis: A closer look at the data

- **note:**
  - geographically contiguous provinces are not necessarily MDS-contiguous
    - e.g. North Brabant vs. Antwerp
  - several provinces are ‘spread out’ along certain axes
    - e.g. Drenthe, North Holland, Groningen, possibly French Flanders

- this suggests that there are dimensions of variation other than purely geographical ones

- **scree plot**
3 – A dialectometric analysis: A closer look at the data

- **note:**
  - geographically contiguous provinces are not necessarily MDS-contiguous
    - e.g. North Brabant vs. Antwerp
  - several provinces are ‘spread out’ along certain axes
    - e.g. Drenthe, North Holland, Groningen, possibly French Flanders

- this suggests that there are dimensions of variation other than purely geographical ones

- **scree plot**
  - a scree plot maps the Stress of the MDS-representation against its dimensionality
A dialectometric analysis: A closer look at the data

- **note:**
  - geographically contiguous provinces are not necessarily MDS-contiguous
    - e.g. North Brabant vs. Antwerp
  - several provinces are ‘spread out’ along certain axes
    - e.g. Drenthe, North Holland, Groningen, possibly French Flanders

- this suggests that there are dimensions of variation other than purely geographical ones

### scree plot

- a scree plot maps the Stress of the MDS-representation against its dimensionality
- Stress is a measure of ‘badness of fit’: how poorly does the MDS-analysis represent the actual distance matrix?
note:
- geographically contiguous provinces are not necessarily MDS-contiguous
  - e.g. North Brabant vs. Antwerp
- several provinces are ‘spread out’ along certain axes
  - e.g. Drenthe, North Holland, Groningen, possibly French Flanders
this suggests that there are dimensions of variation other than purely geographical ones

scree plot

- a scree plot maps the Stress of the MDS-representation against its dimensionality
- Stress is a measure of ‘badness of fit’: how poorly does the MDS-analysis represent the actual distance matrix?
- the goal is to strike a balance between reducing Stress and keeping the number of derived dimensions low
Applying quantitative methods to dialect Dutch verb clusters
Applying quantitative methods to dialect Dutch verb clusters
3 – A dialectometric analysis: A closer look at the data

- the scree plot suggests that the data is at least four- but possibly even ten-dimensional
the scree plot suggests that the data is at least four- but possibly even ten-dimensional

this makes an interpretation purely in terms of geography unlikely; it suggests that there are other sources to the variation in verb cluster ordering found in Dutch
while a dialectometric MDS-based analysis of the verb cluster data suggests that there is a geographical—and hence possibly non-grammatical—dimension to the data, a closer inspection of the facts revealed that there are additional sources of variation at work.
4 – Outline

1. Introduction & central research question
2. The data
3. A dialectometric analysis
4. Reversing the perspective
5. Summary and conclusions
4 – Reversing the perspective: A reverse MDS-analysis

- a traditional dialectometric MDS-analysis allows us to approach the main research question only indirectly, by comparing different dialect locations
4 – Reversing the perspective: A reverse MDS-analysis

- a traditional dialectometric MDS-analysis allows us to approach the main research question only indirectly, by comparing different dialect locations
- so let’s reverse the perspective: let’s directly map the differences between linguistic constructions (i.e. verb cluster orders) based on their geographical spread
Reversing the perspective: A reverse MDS-analysis

- A traditional dialectometric MDS-analysis allows us to approach the main research question only indirectly, by comparing different dialect locations.

- So let’s reverse the perspective: let’s directly map the differences between linguistic constructions (i.e. verb cluster orders) based on their geographical spread.

- Starting point: a $31 \times 267$ matrix with one row per verb cluster order and one column per location.

Applying quantitative methods to dialect Dutch verb clusters
Applying quantitative methods to dialect Dutch verb clusters

<table>
<thead>
<tr>
<th>row.names</th>
<th>A001p</th>
<th>A001q</th>
<th>A002p</th>
<th>A006p</th>
<th>B001a</th>
<th>B004p</th>
<th>B007p</th>
<th>B013b</th>
<th>B035p</th>
<th>B...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_2_14a</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2_1_14a</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1_2_14b</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2_1_14b</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1_2_15a</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2_1_15a</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1_2_15b</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2_1_15b</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>kunnen_zwemmen_moet</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>moet_kunnen_zwemmen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>moet_zwemmen_kunnen</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>zwemmen_kunnen_moet</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>zwemmen_moet_kunnen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>123_moet_hebben_gemaakt</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>132_moet_gemaakt_hebben</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>312_gemaakt_moet_hebben</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>321_gemaakt_hebben_moet</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>v1v2v3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>1</td>
</tr>
</tbody>
</table>
a traditional dialectometric MDS-analysis allows us to approach the main research question only indirectly, by comparing different dialect locations

so let’s reverse the perspective: let’s directly map the differences between linguistic constructions (i.e. verb cluster orders) based on their geographical spread

starting point: a $31 \times 267$ matrix with one row per verb cluster order and one column per location

step 1: convert the table into a $31 \times 31$ (symmetric) distance matrix, whereby for each pair of verb cluster orders a distance between them is calculated based on their geographical spread
Applying quantitative methods to dialect Dutch verb clusters

<table>
<thead>
<tr>
<th>1_2_14a</th>
<th>2_1_14a</th>
<th>1_2_14b</th>
<th>2_1_14b</th>
<th>1_2_15a</th>
<th>2_1_15a</th>
<th>1_2_15b</th>
<th>2_1_15b</th>
<th>k1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0.755905511</td>
<td>0.580357142</td>
<td>0.876190476</td>
<td>0.916666666</td>
<td>0.738775510</td>
<td>0.702020202</td>
<td>0.75</td>
<td>0.000</td>
</tr>
<tr>
<td>0.755905511</td>
<td>0.000</td>
<td>0.648760330</td>
<td>0.280851063</td>
<td>0.971074380</td>
<td>0.040816326</td>
<td>0.231092436</td>
<td>0.5512820510</td>
<td>0.000</td>
</tr>
<tr>
<td>0.580357142</td>
<td>0.648760330</td>
<td>0.000</td>
<td>0.925490196</td>
<td>0.934065934</td>
<td>0.661354581</td>
<td>0.600985221</td>
<td>0.727272727</td>
<td>0.000</td>
</tr>
<tr>
<td>0.876190476</td>
<td>0.280851063</td>
<td>0.925490196</td>
<td>0.000</td>
<td>0.989071038</td>
<td>0.276422764</td>
<td>0.440677966</td>
<td>0.647342995</td>
<td>0.000</td>
</tr>
<tr>
<td>0.916666666</td>
<td>0.971074380</td>
<td>0.934065934</td>
<td>0.989071038</td>
<td>0.000</td>
<td>0.992307692</td>
<td>0.959595959</td>
<td>0.974137931</td>
<td>0.000</td>
</tr>
<tr>
<td>0.738775510</td>
<td>0.040816326</td>
<td>0.661354581</td>
<td>0.276422764</td>
<td>0.992307692</td>
<td>0.000</td>
<td>0.219512195</td>
<td>0.543568464</td>
<td>0.000</td>
</tr>
<tr>
<td>0.702020202</td>
<td>0.600985221</td>
<td>0.440677966</td>
<td>0.959595959</td>
<td>0.219512195</td>
<td>0.000</td>
<td>0.744</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.551282051</td>
<td>0.727272727</td>
<td>0.647342995</td>
<td>0.974137931</td>
<td>0.543568464</td>
<td>0.744</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

kunnen_zwe: 0.985915492 | 0.995798319 | 0.989130434 | 1.000 | 1.000 | 0.995918367 | 1.000 | 0.999099099 |
moet_kunne: 0.691304347 | 0.105263157 | 0.603448275 | 0.356275303 | 0.962025316 | 0.087301587 | 0.149779735 | 0.607438016 |
moet_zwem: 0.892473118 | 0.870292887 | 0.891891891 | 0.885869565 | 1.000 | 0.869387755 | 0.918660287 | 0.756521739 |
zwemmen_k: 0.950980392 | 0.848739495 | 0.932203389 | 0.850828729 | 1.000 | 0.865306122 | 0.959090909 | 0.684684684 |
zwemmen_r: 0.654867256 | 0.681818181 | 0.645161290 | 0.803827751 | 0.940476190 | 0.689516129 | 0.695238095 | 0.563909774 |
123_moet_h: 0.518518518 | 0.673387096 | 0.532786885 | 0.831111111 | 0.910112359 | 0.674509803 | 0.643192488 | 0.639455782 |
132_moet_g: 0.795698924 | 0.389344262 | 0.728643216 | 0.454128440 | 0.975757575 | 0.373015873 | 0.314285714 | 0.837719298 |
312_gemaak: 0.644808743 | 0.318548387 | 0.578125 | 0.530864197 | 0.961748633 | 0.303149606 | 0.355263157 | 0.562814070 |
321_gemaak: 0.955357142 | 0.808333333 | 0.896 | 0.812834224 | 1.000 | 0.823293172 | 0.920704845 | 0.619469026 |
v1v2v3: 0.608433734 | 0.381147540 | 0.554913294 | 0.593220338 | 0.944099378 | 0.388888888 | 0.438053097 | 0.519553072 |
4 – Reversing the perspective: A reverse MDS-analysis

- A traditional dialectometric MDS-analysis allows us to approach the main research question only indirectly, by comparing different dialect locations.
- So let’s reverse the perspective: let’s directly map the differences between linguistic constructions (i.e. verb cluster orders) based on their geographical spread.
- Starting point: a $31 \times 267$ matrix with one row per verb cluster order and one column per location.
- Step 1: convert the table into a $267 \times 367$ (symmetric) distance matrix, whereby for each pair of locations a distance between them is calculated based on the linguistic features they share.
- Step 2: apply MDS to this distance matrix.
Reverse MDS-analysis of 31 verb cluster orders
Note: each point now represents a particular cluster order and closeness of points indicates how alike two verb cluster orders are based on their geographical spread.
4 – Reversing the perspective: A reverse MDS-analysis

- **note**: each point now represents a particular cluster order and closeness of points indicates how alike two verb cluster orders are based on their geographical spread.
- if this likeness is the result of grammatical microparameters, then verb cluster orders that are ‘closeby’ should be the result of the same parameter.
Reverse MDS-analysis of 31 verb cluster orderings

Applying quantitative methods to dialect Dutch verb clusters
Applying quantitative methods to dialect Dutch verb clusters
advantage of this approach: we can use the reverse MDS-analysis as a touchstone for theoretical analyses of verb cluster orders

Barbiers (2005) derives verb cluster orders as follows:

- base order is uniformly head-initial
- derives 12 and 123
- movement is VP-intraposition
- derives 21 and 231 (movement of VP2), 312 and 132 (movement of VP3) and fails to derive 213 (because VP2 contains VP3)
- VP-intraposition can pied-pipe other material
- derives 321 (movement of VP3 to specVP1 via specVP2 and with pied-piping of VP2)
- VP intraposition is triggered by feature checking: modal and aspectual auxiliaries enter into a(n eventive) feature checking relation with the main verb, while perfective auxiliaries enter into a perfective checking relationship with their immediately dominated verb
- rules out 231 in the case of MOD-MOD/AUX-V-clusters (there is no checking relation between the two auxiliaries and hence no movement of VP2 to specVP1 is allowed) and 312 in the case of AUX-AUX/MOD-V-clusters

Applying quantitative methods to dialect Dutch verb clusters
4 – Reversing the perspective: A reverse MDS-analysis

- **advantage** of this approach: we can use the reverse MDS-analysis as a touchstone for theoretical analyses of verb cluster orders
- e.g. Barbiers (2005) derives verb cluster orders as follows:
4 – Reversing the perspective: A reverse MDS-analysis

- **advantage** of this approach: we can use the reverse MDS-analysis as a touchstone for theoretical analyses of verb cluster orders
- e.g. Barbiers (2005) derives verb cluster orders as follows:
  - base order is uniformly head-initial → derives 12 and 123
4 – Reversing the perspective: A reverse MDS-analysis

- **advantage** of this approach: we can use the reverse MDS-analysis as a touchstone for theoretical analyses of verb cluster orders
- e.g. Barbiers (2005) derives verb cluster orders as follows:
  - base order is uniformly head-initial → derives 12 and 123
  - movement is VP-intraposition → derives 21 and 231 (movement of VP2), 312 and 132 (movement of VP3) and fails to derive 213 (because VP2 contains VP3)
4 – Reversing the perspective: A reverse MDS-analysis

- **advantage** of this approach: we can use the reverse MDS-analysis as a touchstone for theoretical analyses of verb cluster orders

- e.g. Barbiers (2005) derives verb cluster orders as follows:
  - base order is uniformly head-initial → derives 12 and 123
  - movement is VP-intraposition → derives 21 and 231 (movement of VP2), 312 and 132 (movement of VP3) and fails to derive 213 (because VP2 contains VP3)
  - VP-intraposition can pied-pipe other material → derives 321 (movement of VP3 to specVP1 via specVP2 and with pied-piping of VP2)
advantage of this approach: we can use the reverse MDS-analysis as a touchstone for theoretical analyses of verb cluster orders

e.g. Barbiers (2005) derives verb cluster orders as follows:
- base order is uniformly head-initial → derives 12 and 123
- movement is VP-intraposition → derives 21 and 231 (movement of VP2), 312 and 132 (movement of VP3) and fails to derive 213 (because VP2 contains VP3)
- VP-intraposition can pied-pipe other material → derives 321 (movement of VP3 to specVP1 via specVP2 and with pied-piping of VP2)
- VP intraposition is triggered by feature checking: modal and aspectual auxiliaries enter into a(n eventive) feature checking relation with the main verb, while perfective auxiliaries enter into a perfective checking relationship with their immediately dominated verb → rules out 231 in the case of MOD-MOD/AUX-V-clusters (there is no checking relation between the two auxiliaries and hence no movement of VP2 to specVP1 is allowed) and 312 in the case of AUX-AUX/MOD-V-clusters

Applying quantitative methods to dialect Dutch verb clusters
4 – Reversing the perspective: A reverse MDS-analysis

- from this theoretical account we can distill the following microparameters:
from this theoretical account we can distill the following microparameters:

- [±base-generation]: can the order be base-generated?
4 – Reversing the perspective: A reverse MDS-analysis

- from this theoretical account we can distill the following microparameters:
  - [±base-generation]: can the order be base-generated?
  - [±movement]: can the order be derived via movement?
from this theoretical account we can distill the following microparameters:

- [±base-generation]: can the order be base-generated?
- [±movement]: can the order be derived via movement?
- [±pied-piping]: does the derivation involve pied-piping?
from this theoretical account we can distill the following microparameters:

- \( \pm \text{base-generation} \): can the order be base-generated?
- \( \pm \text{movement} \): can the order be derived via movement?
- \( \pm \text{pied-piping} \): does the derivation involve pied-piping?
- \( \pm \text{feature-checking violation} \): does the order involve a feature checking violation?
from this theoretical account we can distill the following microparameters:

- [±base-generation]: can the order be base-generated?
- [±movement]: can the order be derived via movement?
- [±pied-piping]: does the derivation involve pied-piping?
- [±feature-checking violation]: does the order involve a feature checking violation?

and these microparameters we can map against the output of the reverse MDS-analysis
Applying quantitative methods to dialect Dutch verb clusters
from this theoretical account we can distill the following microparameters:

- [±base-generation]: can the order be base-generated?
- [±movement]: can the order be derived via movement?
- [±pied-piping]: does the derivation involve pied-piping?
- [±feature-checking violation]: does the order involve a feature checking violation?

and these microparameters we can map against the output of the reverse MDS-analysis

preliminary conclusion: the four microparameters from Barbiers (2005) do point to verb cluster orders that pattern together in the MDS-analysis, but they don’t yet account for the full pattern of variation
4 – Reversing the perspective: Dimensionality & parameters

- just like with the ‘regular’ MDS-analysis we can use a scree plot to determine the true dimensionality of the verb cluster data
Applying quantitative methods to dialect Dutch verb clusters

Scree plot for the reverse MDS-analysis of 31 verb cluster orders
just like with the ‘regular’ MDS-analysis we can use a scree plot to determine the true dimensionality of the verb cluster data

the plot suggests that our data is four-dimensional → this suggests that four microparameters should suffice to account for the variation in verb cluster ordering
just like with the ‘regular’ MDS-analysis we can use a scree plot to determine the true dimensionality of the verb cluster data.

The plot suggests that our data is four-dimensional → this suggests that four microparameters should suffice to account for the variation in verb cluster ordering.

We can plot the output of a four-dimensional MDS-analysis in a $4 \times 4$ scatter plot matrix.
Applying quantitative methods to dialect Dutch verb clusters
just like with the ‘regular’ MDS-analysis we can use a scree plot to determine the true dimensionality of the verb cluster data

the plot suggests that our data is four-dimensional → this suggests that four microparameters should suffice to account for the variation in verb cluster ordering

we can plot the output of a four-dimensional MDS-analysis in a $4 \times 4$ scatter plot matrix

just like with two-dimensional reverse MDS, we can color code the plotted cluster orders based on linguistic variables in order to determine the grammatical meaning of the four coordinates
just like with the ‘regular’ MDS-analysis we can use a scree plot to determine the true dimensionality of the verb cluster data

the plot suggests that our data is four-dimensional → this suggests that four microparameters should suffice to account for the variation in verb cluster ordering

we can plot the output of a four-dimensional MDS-analysis in a 4×4 scatter plot matrix

just like with two-dimensional reverse MDS, we can color code the plotted cluster orders based on linguistic variables in order to determine the grammatical meaning of the four coordinates

e.g. Barbiers’s [±base-generation]-parameter doesn’t seem to correspond to any of the four dimensions
Applying quantitative methods to dialect Dutch verb clusters
4 – Reversing the perspective: Conclusion

- using MDS to map the differences between verb cluster orders in terms of their geographical spread rather than the other way around offers more insight into the theoretical analysis of this phenomenon.
4 – Reversing the perspective: Conclusion

- using MDS to map the differences between verb cluster orders in terms of their geographical spread rather than the other way around offers more insight into the theoretical analysis of this phenomenon.
- not only can such MDS-graphs be used to test existing linguistic theories.
4 – Reversing the perspective: Conclusion

- using MDS to map the differences between verb cluster orders in terms of their geographical spread rather than the other way around offers more insight into the theoretical analysis of this phenomenon
- not only can such MDS-graphs be used to test existing linguistic theories
- they can also help us detect the relevant microparameters at play in the data
5 – Outline

1. Introduction & central research question
2. The data
3. A dialectometric analysis
4. Reversing the perspective
5. Summary and conclusions

Applying quantitative methods to dialect Dutch verb clusters
Main research question: Can a quantitative analysis of verb cluster orderings shed new light on the theoretical analysis of this phenomenon?
5 – Summary and conclusions

- **Main research question:** Can a quantitative analysis of verb cluster orderings shed new light on the theoretical analysis of this phenomenon?
- **Answer:** Yes, it allows us to directly test existing analyses to determine to what extent they predict the observed variation, and it offers a way to probe for the relevant factors (i.e. microparameters) at play in the data.
Main research question: Can a quantitative analysis of verb cluster orderings shed new light on the theoretical analysis of this phenomenon?

Answer: Yes, it allows us to directly test existing analyses to determine to what extent they predict the observed variation, and it offers a way to probe for the relevant factors (i.e. microparameters) at play in the data.

broader issues:
Main research question: Can a quantitative analysis of verb cluster orderings shed new light on the theoretical analysis of this phenomenon?

Answer: Yes, it allows us to directly test existing analyses to determine to what extent they predict the observed variation, and it offers a way to probe for the relevant factors (i.e. microparameters) at play in the data.

Broader issues:

- **the grammatical nature of microvariation**: there are clear indications that the distribution of verb cluster orders is not purely sociolinguistic, but that there are linguistic factors at work.
Main research question: Can a quantitative analysis of verb cluster orderings shed new light on the theoretical analysis of this phenomenon?

Answer: Yes, it allows us to directly test existing analyses to determine to what extent they predict the observed variation, and it offers a way to probe for the relevant factors (i.e. microparameters) at play in the data.

broader issues:

- **the grammatical nature of microvariation:** there are clear indications that the distribution of verb cluster orders is not purely sociolinguistic, but that there are linguistic factors at work

- **quantitative-statistical vs. formal-theoretical linguistics:** the interaction between the two approaches can be mutually beneficial: the latter can inform the former, and the former can be used to test predictions of the latter

