Reverse dialectometry

Geography as a probe into linguistic theory

Jeroen van Craenenbroeck

KU Leuven/CRISSP

Maps and Grammar

September 17–18, 2014
Introduction: verbs, word order, and linguistic theory
Introduction: verbs, word order, and linguistic theory

- in Dutch (like in many Germanic languages) verbs tend to group together at the right edge of the (embedded) clause:
Introduction: verbs, word order, and linguistic theory

- In Dutch (like in many Germanic languages) verbs tend to group together at the right edge of the (embedded) clause:

(1) dat hij gisteren tijdens de les gelachen heeft.
that he yesterday during the class laughed has
‘that he laughed yesterday during class.’
Introduction: verbs, word order, and linguistic theory

- in Dutch (like in many Germanic languages) verbs tend to group together at the right edge of the (embedded) clause:

(1) dat hij gisteren tijdens de les *gelachen heeft.*
    that he yesterday during the class laughed has
    ‘that he laughed yesterday during class.’

- moreover, such verbal clusters typically show a certain degree of freedom in their word order:
Introduction: verbs, word order, and linguistic theory

• in Dutch (like in many Germanic languages) verbs tend to group together at the right edge of the (embedded) clause:

(1) dat hij gisteren tijdens de les *gelachen* heeft.
    that he yesterday during the class laughed has
    ‘that he laughed yesterday during class.’

• moreover, such verbal clusters typically show a certain degree of freedom in their word order:

(2) dat hij gisteren tijdens de les *heeft gelachen*.
    that he yesterday during the class had laughed
    ‘that he laughed yesterday during class.’
Introduction: verbs, word order, and linguistic theory

- In Dutch (like in many Germanic languages) verbs tend to group together at the right edge of the (embedded) clause:

  (1) dat hij gisteren tijdens de les *gelachen heeft.*
  that he yesterday during the class laughed has
  ‘that he laughed yesterday during class.’ (21)

- Moreover, such verbal clusters typically show a certain degree of freedom in their word order:

  (2) dat hij gisteren tijdens de les *heeft gelachen.*
  that he yesterday during the class had laughed
  ‘that he laughed yesterday during class.’ (12)
• this word order freedom is typically a source of interdialectal variation:
• this word order freedom is typically a source of interdialectal variation:

(3) **Ferwerd Dutch**

a.  dasto  it ook net *zien* meist.
    that.you it also not see  may
    ‘that you’re also not allowed to see it.’  

b.  *dasto  it ook net *meist* zien.
    that.you it also not may  see
    ‘that you’re also not allowed to see it.’  

(✓21)  
(*12)
• this word order freedom is typically a source of interdialectal variation:

(4) **Gendringen Dutch**

a. dat ee et ook nie zien mag.
   that you it also not see may
   ‘that you’re also not allowed to see it.’  (✓21)

b. dat ee et ook nie mag zien.
   that you it also not may see
   ‘that you’re also not allowed to see it.’  (✓12)
• this word order freedom is typically a source of interdialectal variation:

(5) **Poelkapelle Dutch**

a. *dajtgie ook nie zien meug.
   that.it.you also not see may
   ‘that you’re also not allowed to see it.’ (*21)

b. dajtgie ook nie meug zien.
   that.it.you also not may see
   ‘that you’re also not allowed to see it.’ (✓12)
• and the more complex the verbal cluster, the more variation there is: in verbal clusters consisting of two modal auxiliaries and one main verb, out of the six orders that are theoretically possible, four are attested in Dutch dialects:
and the more complex the verbal cluster, the more variation there is: in verbal clusters consisting of two modal auxiliaries and one main verb, out of the six orders that are theoretically possible, four are attested in Dutch dialects:

\[(6) \quad \text{Ik vind dat iedereen moet kunnen zwemmen.} \]
\[\quad \text{I find that everyone must can swim} \]
\[\quad \text{‘I think everyone should be able to swim.’} \quad (123)\]
• and the more complex the verbal cluster, the more variation there is: in verbal clusters consisting of two modal auxiliaries and one main verb, out of the six orders that are theoretically possible, four are attested in Dutch dialects:

(6)  Ik vind dat iedereen moet kunnen zwemmen.

  I find that everyone must can swim
  ‘I think everyone should be able to swim.’  (123)

(7)  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)
• but once again, it is not the case that each of the four allowed orders is attested in all dialects:
• but once again, it is not the case that each of the four allowed orders is attested in all dialects:

(8)  

_Midsland Dutch_

   that everyone must can swim
   ‘that everyone should be able to swim.’  (*123)

b. dat elkeen mot zwemme kanne.  (✓132)

c. *dat elkeen zwemme mot kanne.  (*312)

d. dat elkeen zwemme kanne mot.  (✓321)

e. *dat elkeen kanne zwemme mot.  (*231)

f. *dat elkeen kanne mot zwemme.  (*213)
but once again, it is not the case that each of the four allowed orders is attested in all dialects:

(9) *Langelo Dutch*

a. dat iedereen moet kunnen zwemmen.
   that everyone must can swim
   ‘that everyone should be able to swim.’  (✓123)

b. *dat iedereen mot zwemmen kunnen.

   (*132)

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)
more generally, the four possible cluster orders yield a total of 16 possible combinations, of which 12 are attested in Dutch dialects:
more generally, the four possible cluster orders yield a total of 16 possible combinations, of which 12 are attested in Dutch dialects:

<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>
in order to get a more complete picture of the variation, we can look at the results from the SAND-project:
• in order to get a more complete picture of the variation, we can look at the results from the SAND-project:
• in order to get a more complete picture of the variation, we can look at the results from the SAND-project:
  • dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
• in order to get a more complete picture of the variation, we can look at the results from the SAND-project:
  • dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
• the SAND-questionnaire contained eight questions on word order in verb clusters for a total of 31 cluster orders
• in order to get a more complete picture of the variation, we can look at the results from the SAND-project:
  • dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
• the SAND-questionnaire contained eight questions on word order in verb clusters for a total of 31 cluster orders
• if we map, for each of the 267 SAND-dialects, which dialect has which combination of cluster orders, we find 137 different combinations of verb cluster orders
• in order to get a more complete picture of the variation, we can look at the results from the SAND-project:
  • dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
• the SAND-questionnaire contained eight questions on word order in verb clusters for a total of 31 cluster orders
• if we map, for each of the 267 SAND-dialects, which dialect has which combination of cluster orders, we find 137 different combinations of verb cluster orders
• in other words, there are 137 different types of dialects when it comes to word order in verbal clusters
• **question**: how can we make sense of this massive variation from the point of view of theoretical linguistics?
• **question:** how can we make sense of this massive variation from the point of view of theoretical linguistics?

• e.g. Principles & Parameters: natural language is the result of the interplay between:
• **question:** how can we make sense of this massive variation from the point of view of theoretical linguistics?

• e.g. Principles & Parameters: natural language is the result of the interplay between:

  1. Principles: innate properties that are invariant across all languages
• **question**: how can we make sense of this massive variation from the point of view of theoretical linguistics?

• e.g. Principles & Parameters: natural language is the result of the interplay between:
  
  1. Principles: innate properties that are invariant across all languages
  2. Parameters: simple, often binary choices (‘switches’) which are responsible for interlinguistic differences, and which determine the space of variation in natural language
• **question:** how can we make sense of this massive variation from the point of view of theoretical linguistics?

• e.g. Principles & Parameters: natural language is the result of the interplay between:
  
  1. Principles: innate properties that are invariant across all languages  
  2. Parameters: simple, often binary choices (‘switches’) which are responsible for interlinguistic differences, and which determine the space of variation in natural language

• **so:**
question: how can we make sense of this massive variation from the point of view of theoretical linguistics?

- e.g. Principles & Parameters: natural language is the result of the interplay between:
  1. Principles: innate properties that are invariant across all languages
  2. Parameters: simple, often binary choices (‘switches’) which are responsible for interlinguistic differences, and which determine the space of variation in natural language

so:

- what are the parameters of word order variation in verb clusters?
question: how can we make sense of this massive variation from the point of view of theoretical linguistics?

e.g. Principles & Parameters: natural language is the result of the interplay between:

1. Principles: innate properties that are invariant across all languages
2. Parameters: simple, often binary choices (‘switches’) which are responsible for interlinguistic differences, and which determine the space of variation in natural language

so:

- what are the parameters of word order variation in verb clusters?
- is this variation even parameter-related? how much noise is there in these data? is some of the variation extra-grammatical (cf. Barbiers (2005))?
• **question:** how can we make sense of this massive variation from the point of view of theoretical linguistics?

• e.g. Principles & Parameters: natural language is the result of the interplay between:

1. Principles: innate properties that are invariant across all languages
2. Parameters: simple, often binary choices (‘switches’) which are responsible for interlinguistic differences, and which determine the space of variation in natural language

• **so:**

• what are the parameters of word order variation in verb clusters?
• is this variation even parameter-related? how much noise is there in these data? is some of the variation extra-grammatical (cf. Barbiers (2005))?
• related methodological question: how do we go about finding those parameters?
• **in this talk** I argue that a quantitative-statistical analysis of the data enriched with insights from formal-theoretical linguistics can separate the wheat from the chaff
• in this talk I argue that a quantitative-statistical analysis of the data enriched with insights from formal-theoretical linguistics can separate the wheat from the chaff

• more specifically, I will argue that roughly 80% of the variation found in Dutch verb cluster orders can be reduced to three grammatical parameters
Dialect variation and quantitative methods: dialectometry
Dialect variation and quantitative methods: dialectometry

- **dialectometry** is a subdiscipline of linguistics that uses computational and quantitative techniques in dialectology (Nerbonne and Kretzschmar Jr., 2013)
Dialect variation and quantitative methods: dialectometry

- **dialectometry** is a subdiscipline of linguistics that uses computational and quantitative techniques in dialectology (Nerbonne and Kretzschmar Jr., 2013)
- in a typical dialectometric analysis locations are used as individuals and linguistic phenomena as variables → we’re measuring similarities and differences between dialect locations based on their linguistic profile
Dialect variation and quantitative methods: dialectometry

- **dialectometry** is a subdiscipline of linguistics that uses computational and quantitative techniques in dialectology (Nerbonne and Kretzschmar Jr., 2013)
- in a typical dialectometric analysis locations are used as individuals and linguistic phenomena as variables → we’re measuring similarities and differences between dialect locations based on their linguistic profile
- often used method: Multidimensional Scaling (MDS)
Dialect variation and quantitative methods: dialectometry

- **dialectometry** is a subdiscipline of linguistics that uses computational and quantitative techniques in dialectology (Nerbonne and Kretzschmar Jr., 2013)
- in a typical dialectometric analysis locations are used as individuals and linguistic phenomena as variables → we’re measuring similarities and differences between dialect locations based on their linguistic profile
- often used method: Multidimensional Scaling (MDS)
- starting point: data table with dialects in rows and cluster orders in columns
<table>
<thead>
<tr>
<th></th>
<th>AUX1(be.sg)-PART2</th>
<th>PART2-AUX1(be.sg)</th>
<th>AUX1(have.sg)-PART2</th>
<th>PART2-AUX1(have.sg)</th>
<th>AUX1(have.pl)-PART2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midsland / Midslân</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Lies</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>West-Terschelling</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Oosterend</td>
<td>NA</td>
<td>NA</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Hollum</td>
<td>no</td>
<td>yes</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Schiermonnikoog</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Ferwerd / Ferwert</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Anjum / Eanjum</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Kollum</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Visvliet</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Oosterbierum / Ee</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Beetgum / Bitgum</td>
<td>no</td>
<td>yes</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Bergum / Burgum</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Jorwerd / Jorwert</td>
<td>no</td>
<td>yes</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Bakkeveen / Bakkeveen</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Waskemeer / De Waskemeer</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Kloosterburen</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Warffum</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Leermens</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Groningen</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>Nieuw-Scheemda</td>
<td>NA</td>
<td>NA</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Langelo</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>
• step 1: convert the data 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
<table>
<thead>
<tr>
<th></th>
<th>Midsland</th>
<th>Lies</th>
<th>West-Tersch</th>
<th>Oosterend</th>
<th>Hollum</th>
<th>Schiermonnikoog</th>
<th>Ferwerd</th>
<th>Anjum / Eanjuin</th>
<th>Kollum</th>
<th>Visvliet</th>
<th>Oosterbierum</th>
<th>Beetgum</th>
<th>Bergum / Buren</th>
<th>Jorwerd / Jourwerd</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midsland / Mid</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>
<td>0.692</td>
<td></td>
</tr>
<tr>
<td>Lies</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>
<td>0.667</td>
<td></td>
</tr>
<tr>
<td>West-Terschelling</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>
<td>0.583</td>
<td></td>
</tr>
<tr>
<td>Oosterend</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>
<td>0.846</td>
<td></td>
</tr>
<tr>
<td>Hollum</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>
<td>0.545</td>
<td></td>
</tr>
<tr>
<td>Schiermonnikoog</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>0.182</td>
<td>0.000</td>
<td>0.571</td>
<td>0.417</td>
<td>0.556</td>
<td>0.500</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>Ferwerd / Ferwerd</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>0.000</td>
<td>0.571</td>
<td>0.417</td>
<td>0.556</td>
<td>0.500</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>Anjum / Eanjuin</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>
<td>0.600</td>
<td></td>
</tr>
<tr>
<td>Kollum</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>0.588</td>
<td>0.588</td>
<td>0.643</td>
<td>0.429</td>
<td>0.571</td>
<td></td>
</tr>
<tr>
<td>Visvliet</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>0.500</td>
<td>0.167</td>
<td>0.000</td>
<td>0.455</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>Oosterbierum</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>0.500</td>
<td>0.455</td>
<td></td>
</tr>
<tr>
<td>Beetgum / Beetgum</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>
<td>0.455</td>
<td></td>
</tr>
<tr>
<td>Bergum / Buren</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>0.455</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>Jorwerd / Jourwerd</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>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Bakkeveen / Bisbeek</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>
<td>0.583</td>
<td></td>
</tr>
<tr>
<td>Waskmeer / Warffum</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>0.677</td>
<td>0.429</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>Kloosterburen</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>
<td>0.583</td>
<td></td>
</tr>
<tr>
<td>Warffum</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>0.943</td>
<td>0.400</td>
<td>0.652</td>
<td>0.600</td>
<td>0.636</td>
<td>0.571</td>
<td>0.750</td>
<td></td>
</tr>
<tr>
<td>Leermens</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>
<td>0.755</td>
<td></td>
</tr>
<tr>
<td>Groningen</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>
<td>0.786</td>
<td></td>
</tr>
<tr>
<td>Nieuw-Scheemuiden</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>
<td>0.786</td>
<td></td>
</tr>
<tr>
<td>Langelo</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>
<td>0.950</td>
<td></td>
</tr>
</tbody>
</table>
• step 2: reduce this 267-dimensional matrix to a two- or three-dimensional one, so that it can easily be visualized
Introduction

Dialectometry

Reverse dialectometry

Conclusion

References
step 3: project back onto a geographical map
• shortcomings of this approach for my current purposes:
shortcomings of this approach for my current purposes:

1. the linguistic constructions themselves play only an indirect role in the outcome of the analysis: we can see when two dialects differ, but we don’t see which cluster orders are responsible for this difference or how they cluster or correlate
• shortcomings of this approach for my current purposes:
  1. the linguistic constructions themselves play only an indirect role in the outcome of the analysis: we can see when two dialects differ, but we don’t see which cluster orders are responsible for this difference or how they cluster or correlate
  2. there is no link between the data that feed into the quantitative analysis and the formal theoretical literature on verb clusters
Reverse dialectometry

- proposal: two changes to the classical dialectometric setup:
  1. cluster orders are individuals rather than variables, i.e. instead of calculating differences between dialect locations, we measure differences between linguistic constructions
  2. Multiple Correspondence Analysis (MCA) instead of Multidimensional Scaling (MDS): involves the same kind of dimension reduction, but applied simultaneously to individuals and variables → will allow for the inclusion of formal theoretical variables alongside geographical ones

- starting point: a data table with cluster orders as rows and dialect locations as columns
Reverse dialectometry

- **proposal**: two changes to the classical dialectometric setup:
Reverse dialectometry

• **proposal:** two changes to the classical dialectometric setup:
  1. cluster orders are *individuals* rather than variables, i.e. instead of calculating differences between dialect locations, we measure differences between linguistic constructions
Reverse dialectometry

• proposal: two changes to the classical dialectometric setup:
  1. cluster orders are *individuals* rather than variables, i.e. instead of calculating differences between dialect locations, we measure differences between linguistic constructions
  2. Multiple Correspondence Analysis (MCA) instead of Multidimensional Scaling (MDS): involves the same kind of dimension reduction, but applied simultaneously to individuals and variables → will allow for the inclusion of formal theoretical variables alongside geographical ones
Reverse dialectometry

**proposal**: two changes to the classical dialectometric setup:

1. **cluster orders are *individuals*** rather than variables, i.e. instead of calculating differences between dialect locations, we measure differences between linguistic constructions

2. **Multiple Correspondence Analysis (MCA)** instead of **Multidimensional Scaling (MDS)**: involves the same kind of dimension reduction, but applied simultaneously to individuals and variables → will allow for the inclusion of formal theoretical variables alongside geographical ones

**starting point**: a data table with cluster orders as rows and dialect locations as columns
<table>
<thead>
<tr>
<th></th>
<th>Midsland</th>
<th>Lies</th>
<th>West.Tersch</th>
<th>Oosterend</th>
<th>Hollum</th>
<th>Schiermonnikoer</th>
<th>Ferwerd</th>
<th>Anjum</th>
<th>Kollum</th>
<th>Visuileil</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUX1(be.sg)-PART2</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>NA</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>PART2-AUX1(be.sg)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>NA</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>AUX1(have.sg)-PART2</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>PART2-AUX1(have.sg)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>AUX1(have.pl)-PART2</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>PART2-AUX1(have.pl)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>MOD1(sg)-INF2</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>INF2-MOD1(sg)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>MOD2-INF3-MOD1(sg)</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>MOD1(sg)-MOD2-INF3</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>MOD1(sg)-INF3-MOD2</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>INF3-MOD2-MOD1(sg)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>INF3-MOD1(sg)-MOD2</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>MOD1(sg)-AUX2(have)-PART3</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>MOD1(sg)-PART3-AUX2(have)</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>PART3-MOD1(sg)-AUX2(have)</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>PART3-AUX2(have)-MOD1(sg)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>AUX1(be.sg)-AUX2(go)-INF3</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>NA</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>AUX1(be.sg)-INF3-AUX2(go)</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>NA</td>
<td>no</td>
</tr>
<tr>
<td>AUX2(go)-AUX1(be.sg)-INF3</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>NA</td>
<td>no</td>
</tr>
<tr>
<td>AUX2(go)-INF3-AUX1(be.sg)</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>NA</td>
<td>no</td>
</tr>
<tr>
<td>INF3-AUX1(be.sg)-AUX2(go)</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>NA</td>
<td>no</td>
</tr>
<tr>
<td>INF3-AUX2(go)-AUX1(be.sg)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>NA</td>
<td>no</td>
</tr>
<tr>
<td>AUX1(have.sg)-MOD2(inf)-INF3</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>AUX1(have.sg)-INF3-MOD2(part)</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>
• transform to a distance matrix and reduce its dimensionality
Two-dimensional representation of the 31 SAND-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
• **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 parameters, then verb cluster orders that are ‘closeby’ should be the result of the same parameter setting, i.e. parameters create **natural classes** of 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

• if this likeness is the result of grammatical parameters, then verb cluster orders that are ‘closeby’ should be the result of the same parameter setting, i.e. parameters create **natural classes** of verb cluster orders

• in order to find those parameters, we can also encode the cluster orders in terms of their theoretical linguistic analyses
• theoretical accounts differ in which analysis they assign to which cluster order ⇒ cluster orders have their own specific ‘fingerprint’ in each analysis, some of them very similar to one another and others very different
• theoretical accounts differ in which analysis they assign to which cluster order ⇒ cluster orders have their own specific ‘fingerprint’ in each analysis, some of them very similar to one another and others very different
• we can encode the SAND cluster orders in our database in terms of those fingerprints and then compare them to the geographical clustering
• theoretical accounts differ in which analysis they assign to which cluster order ⇒ cluster orders have their own specific ‘fingerprint’ in each analysis, some of them very similar to one another and others very different
• we can encode the SAND cluster orders in our database in terms of those fingerprints and then compare them to the geographical clustering
• e.g. in Barbiers (2005)’s analysis cluster orders can differ from one another on four counts:
• theoretical accounts differ in which analysis they assign to which cluster order \( \Rightarrow \) cluster orders have their own specific ‘fingerprint’ in each analysis, some of them very similar to one another and others very different

• we can encode the SAND cluster orders in our database in terms of those fingerprints and then compare them to the geographical clustering

• e.g. in Barbiers (2005)’s analysis cluster orders can differ from one another on four counts:
  • [±base-generation]: can the order be base-generated?
• theoretical accounts differ in which analysis they assign to which cluster order ⇒ cluster orders have their own specific ‘fingerprint’ in each analysis, some of them very similar to one another and others very different
• we can encode the SAND cluster orders in our database in terms of those fingerprints and then compare them to the geographical clustering
• e.g. in Barbiers (2005)’s analysis cluster orders can differ from one another on four counts:
  • [±base-generation]: can the order be base-generated?
  • [±movement]: can the order be derived via movement?
• theoretical accounts differ in which analysis they assign to which cluster order ⇒ cluster orders have their own specific ‘fingerprint’ in each analysis, some of them very similar to one another and others very different
• we can encode the SAND cluster orders in our database in terms of those fingerprints and then compare them to the geographical clustering
• e.g. in Barbiers (2005)’s analysis cluster orders can differ from one another on four counts:
  • [±base-generation]: can the order be base-generated?
  • [±movement]: can the order be derived via movement?
  • [±pied-piping]: does the derivation involve pied-piping?
• theoretical accounts differ in which analysis they assign to which cluster order ⇒ cluster orders have their own specific ‘fingerprint’ in each analysis, some of them very similar to one another and others very different

• we can encode the SAND cluster orders in our database in terms of those fingerprints and then compare them to the geographical clustering

• e.g. in Barbiers (2005)’s analysis cluster orders can differ from one another on four counts:
  • [±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?
<table>
<thead>
<tr>
<th>AUX1(be.sg)-PART2</th>
<th>Barbiers-base.generation</th>
<th>Barbiers-movement</th>
<th>Barbiers-spec-pied-piping</th>
<th>Barbiers-feature.checking-failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>PART2-AUX1(be.sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>AUX1(have.sg)-PART2</td>
<td>yesBase</td>
<td>noMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>PART2-AUX1(have.sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>AUX1(have.pl)-PART2</td>
<td>yesBase</td>
<td>noMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>PART2-AUX1(have.pl)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>MOD1(sg)-INF2</td>
<td>yesBase</td>
<td>noMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>INF2-MOD1(sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>MOD2-INF3-MOD1(sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>MOD1(sg)-MOD2-INF3</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>MOD1(sg)-INF3-MOD2</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>INF3-MOD2-MOD1(sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>yesPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>INF3-MOD1(sg)-MOD2</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>MOD1(sg)-AUX2(have)-PART3</td>
<td>yesBase</td>
<td>noMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>MOD1(sg)-PART3-AUX2(have)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>PART3-MOD1(sg)-AUX2(have)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>PART3-AUX2(have)-MOD1(sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>yesPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>AUX1(be.sg)-AUX2(go)-INF3</td>
<td>yesBase</td>
<td>noMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>AUX1(be.sg)-INF3-AUX2(go)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>AUX2(go)-AUX1(be.sg)-INF3</td>
<td>noBase</td>
<td>noMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>INF3-AUX1(be.sg)-AUX2(go)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>INF3-AUX2(go)-AUX1(be.sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>AUX1(have.sg)-MOD2(inf)-INF3</td>
<td>yesBase</td>
<td>noMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>AUX1(have.sg)-INF3-MOD2(part)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>AUX1(have.sg)-INF3-MOD2(inf)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>MOD2(inf)-INF3-AUX1(have.sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>INF3-AUX1(have.sg)-MOD2(inf)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>yesFeatCheckFail</td>
</tr>
<tr>
<td>INF3-AUX1(have.sg)-MOD2(part)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>yesFeatCheckFail</td>
</tr>
<tr>
<td>INF3-MOD2(part)-AUX1(have.sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
<tr>
<td>INF3-MOD2(inf)-AUX1(have.sg)</td>
<td>noBase</td>
<td>yesMvt</td>
<td>noPiedP</td>
<td>noFeatCheckFail</td>
</tr>
</tbody>
</table>
• **in total:** 70 additional variables distilled from the theoretical literature on verb clusters have been added to the data table:
• **in total**: 70 additional variables distilled from the theoretical literature on verb clusters have been added to the data table:
  • the analyses of Barbiers (2005), Barbiers and Bennis (2010), Abels (2011), Haegeman and Riemsdijk (1986), Bader (2012), and Schmid and Vogel (2004)
• **in total:** 70 additional variables distilled from the theoretical literature on verb clusters have been added to the data table:
  - four analyses from Wurmbrand (2005): a head-initial head movement analysis, a head-final head movement analysis, a head-initial XP-movement analysis, a head-final XP-movement analysis
• **in total:** 70 additional variables distilled from the theoretical literature on verb clusters have been added to the data table:

  - four analyses from Wurmbrand (2005): a head-initial head movement analysis, a head-final head movement analysis, a head-initial XP-movement analysis, a head-final XP-movement analysis
  - 17 additional variables based on the theoretical literature, but not linked to a specific analysis
• **in total**: 70 additional variables distilled from the theoretical literature on verb clusters have been added to the data table:
  • the analyses of Barbiers (2005), Barbiers and Bennis (2010), Abels (2011), Haegeman and Riemsdijk (1986), Bader (2012), and Schmid and Vogel (2004)
  • four analyses from Wurmbrand (2005): a head-initial head movement analysis, a head-final head movement analysis, a head-initial XP-movement analysis, a head-final XP-movement analysis
  • 17 additional variables based on the theoretical literature, but not linked to a specific analysis

• in the analysis, these 70 variables are used as **supplementary variables**: they do not contribute to the dimension reduction, but they are mapped against its output, in order to interpret the results
• **recall:** we are trying to determine if the variation in word order in verbal clusters is determined by grammatical parameters, and if so to what extent
• **recall:** we are trying to determine if the variation in word order in verbal clusters is determined by grammatical parameters, and if so to what extent

• this means we need to determine **how many** parameters there are and **what they are**
• **recall**: we are trying to determine if the variation in word order in verbal clusters is determined by grammatical parameters, and if so to what extent

• this means we need to determine **how many** parameters there are and **what they are**

• **proposal (I)**: the number of parameters responsible for the verb cluster variation = the number of dimensions we reduce our data set to
Percentage of variance explained per dimension

Number of dimensions

Percentage of variance explained
• **note:** there seems to be a clear cut-off point after the third dimension
• **note:** there seems to be a clear cut-off point after the third dimension

• together, the first three dimensions account for 78.46% of the variation in the SAND verb cluster data
• **note:** there seems to be a clear cut-off point after the third dimension

• together, the first three dimensions account for 78.46% of the variation in the SAND verb cluster data

• this means that roughly 80% of the variation in verb cluster ordering in SAND can be reduced to three parameters
• **note:** there seems to be a clear cut-off point after the third dimension

• together, the first three dimensions account for 78.46% of the variation in the SAND verb cluster data

• this means that roughly 80% of the variation in verb cluster ordering in SAND can be reduced to three parameters

• in order to know what those parameters are, we need to *interpret* the first three dimensions
• **proposal (I):** the number of parameters responsible for the verb cluster variation = the number of dimensions we reduce our data set to
• **proposal (I):** the number of parameters responsible for the verb cluster variation = the number of dimensions we reduce our data set to

• **proposal (II):** the identity of those parameters = the interpretation of the dimensions
• **proposal (I):** the number of parameters responsible for the verb cluster variation = the number of dimensions we reduce our data set to

• **proposal (II):** the identity of those parameters = the interpretation of the dimensions

• the degree of similarity/correlation between a dimension and a linguistic variable can be determined by:
  1. visual inspection of a color-coded map
• **proposal (I):** the number of parameters responsible for the verb cluster variation = the number of dimensions we reduce our data set to

• **proposal (II):** the identity of those parameters = the interpretation of the dimensions

• the degree of similarity/correlation between a dimension and a linguistic variable can be determined by:
  1. visual inspection of a color-coded map
  2. calculating the squared correlation ratio ($\eta^2$): value between 0 and 1 indicating the strength of the link between a dimension and a particular categorical variable; can be interpreted as the percentage of variation on the dimension that can be explained by that categorical variable
Dimension 1
Dimension 1

• is related to the **morphological form** of the verb: infinitive (*will see*) or auxiliary (*have seen*)
Dimension 1

- is related to the **morphological form** of the verb: infinitive (*will see*) or auxiliary (*have seen*)
- this dimension separates dialects where the infinitive follows the auxiliary it combines with (*will see*) and the participle precedes the auxiliary it combines with (*seen have*) from dialects where at least one of those orders differs
Dimension 1

• is related to the **morphological form** of the verb: infinitive (*will see*) or auxiliary (*have seen*)

• this dimension separates dialects where the infinitive follows the auxiliary it combines with (*will see*) and the participle precedes the auxiliary it combines with (*seen have*) from dialects where at least one of those orders differs

• more specifically, the variable InfMod.AuxPart:
Dimension 1

- is related to the **morphological form** of the verb: infinitive (*will see*) or auxiliary (*have seen*)
- this dimension separates dialects where the infinitive follows the auxiliary it combines with (*will see*) and the participle precedes the auxiliary it combines with (*seen have*) from dialects where at least one of those orders differs
- more specifically, the variable **InfMod.AuxPart**:  
  - set to ‘no’ when the modal precedes the infinitive (when present) and the participle precedes the auxiliary (when present)
Dimension 1

- is related to the **morphological form** of the verb: infinitive (*will see*) or auxiliary (*have seen*)
- this dimension separates dialects where the infinitive follows the auxiliary it combines with (*will see*) and the participle precedes the auxiliary it combines with (*seen have*) from dialects where at least one of those orders differs
- more specifically, the variable **InfMod.AuxPart**:  
  - set to ‘no’ when the modal precedes the infinitive (when present) and the participle precedes the auxiliary (when present)  
  - set to ‘yes’ when at least one of these conditions is not met
Dimension 1

- is related to the morphological form of the verb: infinitive (will see) or auxiliary (have seen)
- this dimension separates dialects where the infinitive follows the auxiliary it combines with (will see) and the participle precedes the auxiliary it combines with (seen have) from dialects where at least one of those orders differs
- more specifically, the variable \texttt{INFMod.AuxPart}:
  - set to ‘no’ when the modal precedes the infinitive (when present) and the participle precedes the auxiliary (when present)
  - set to ‘yes’ when at least one of these conditions is not met
- this variable has a $\eta^2$ of 0.6142
Dimension 2

- is related to the ‘slope’ of the cluster: ascending (e.g. 1 ↗ 2 ↗ 3) or descending (e.g. 3 ↘ 2 ↘ 1)
Dimension 2

• is related to the ‘slope’ of the cluster:
  ascending (e.g. 1 ↗ 2 ↗ 3) or descending (e.g. 3 ↘ 2 ↘ 1)
• more specifically, the variable `FINAL_DESCENT`:
Dimension 2

- is related to the ‘slope’ of the cluster:
  - ascending (e.g. $1 \uparrow 2 \uparrow 3$) or descending (e.g. $3 \downarrow 2 \downarrow 1$)
- more specifically, the variable $\text{FINAL DESCENT}$:
  - set to ‘yes’ if the cluster ends in a descending order
Dimension 2

- is related to the ‘slope’ of the cluster:
  ascending (e.g. $1 \uparrow 2 \uparrow 3$) or descending (e.g. $3 \downarrow 2 \downarrow 1$)
- more specifically, the variable $\text{FINALDESCENT}$:
  - set to ‘yes’ if the cluster ends in a descending order
  - set to ‘no’ if it ends in an ascending order
Dimension 2

- is related to the ‘slope’ of the cluster: ascending (e.g. 1 ↗ 2 ↗ 3) or descending (e.g. 3 ↘ 2 ↘ 1)
- more specifically, the variable **FinalDescent**:
  - set to ‘yes’ if the cluster ends in a descending order
  - set to ‘no’ if it ends in an ascending order

<table>
<thead>
<tr>
<th>FinalDescent_yes</th>
<th>FinalDescent_no</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>132</td>
<td>123</td>
</tr>
<tr>
<td>321</td>
<td>312</td>
</tr>
<tr>
<td>231</td>
<td>213</td>
</tr>
</tbody>
</table>
Dimension 2

- is related to the ‘slope’ of the cluster: ascending (e.g. $1 \uparrow 2 \uparrow 3$) or descending (e.g. $3 \downarrow 2 \downarrow 1$)
- more specifically, the variable $\text{FINALDescent}$:
  - set to ‘yes’ if the cluster ends in a descending order
  - set to ‘no’ if it ends in an ascending order

<table>
<thead>
<tr>
<th>$\text{FinalDescent_yes}$</th>
<th>$\text{FinalDescent_no}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>132</td>
<td>123</td>
</tr>
<tr>
<td>321</td>
<td>312</td>
</tr>
<tr>
<td>231</td>
<td>213</td>
</tr>
</tbody>
</table>

- this variable has a $\eta^2$ of 0.382
Dimension 3

- is again related to the slope of the cluster (and strongly so)
Dimension 3

- is again related to the slope of the cluster (and strongly so)
- it separates the strictly descending orders (i.e. 21 and 321) from all the others (12, 123, 132, 312, 213, 231): $\eta^2 = 0.686$
Dimension 3 vs. Bader’s (2012) base-generated order
Combining the dimensions into a theoretical analysis
Combining the dimensions into a theoretical analysis

• the quantitative-statistical analysis thus yields three ingredients which theoretical linguists can use to base their analysis on
Combining the dimensions into a theoretical analysis

- the quantitative-statistical analysis thus yields three ingredients which theoretical linguists can use to base their analysis on
- for example, a possible parametrized analysis of verb clusters:
Combining the dimensions into a theoretical analysis

- the quantitative-statistical analysis thus yields three ingredients which theoretical linguists can use to base their analysis on
- for example, a possible parametrized analysis of verb clusters:
  1. a head-final base order
Combining the dimensions into a theoretical analysis

• the quantitative-statistical analysis thus yields three ingredients which theoretical linguists can use to base their analysis on

• for example, a possible parametrized analysis of verb clusters:
  1. a head-final base order
  2. which dialects can diverge from or not: $\pm$Movement
     (dimension 3)
Combining the dimensions into a theoretical analysis

• the quantitative-statistical analysis thus yields three ingredients which theoretical linguists can use to base their analysis on

• for example, a possible parametrized analysis of verb clusters:
  1. a head-final base order
  2. which dialects can diverge from or not: [±Movement] (dimension 3)
  3. those that diverge can diverge strongly or not: Economy of Movement (dimension 2)
Combining the dimensions into a theoretical analysis

- the quantitative-statistical analysis thus yields three ingredients which theoretical linguists can use to base their analysis on
- for example, a possible parametrized analysis of verb clusters:
  1. a head-final base order
  2. which dialects can diverge from or not: [±Movement] (dimension 3)
  3. those that diverge can diverge strongly or not: Economy of Movement (dimension 2)
  4. above and beyond all this, a headedness parameter regulates the order of infinitives and participles vis-à-vis their selecting verbs: [±ModInf&PartAux] (dimension 1)
Conclusion

• roughly 80% of the variation found in Dutch verb cluster orders can be reduced to three grammatical parameters by applying a statistical analysis to the data
Conclusion

- roughly 80% of the variation found in Dutch verb cluster orders can be reduced to three grammatical parameters by applying a statistical analysis to the data
- more generally, there is room for fruitful collaboration between formal-theoretical and quantitative-statistical linguistics:
Conclusion

• roughly 80% of the variation found in Dutch verb cluster orders can be reduced to three grammatical parameters by applying a statistical analysis to the data

• more generally, there is room for fruitful collaboration between formal-theoretical and quantitative-statistical linguistics:
  • the former can guide the interpretation of results from the latter
Conclusion

- roughly 80% of the variation found in Dutch verb cluster orders can be reduced to three grammatical parameters by applying a statistical analysis to the data
- more generally, there is room for fruitful collaboration between formal-theoretical and quantitative-statistical linguistics:
  - the former can guide the interpretation of results from the latter
  - the latter can help evaluate and test hypotheses of the former
References I


References II


