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

KU LEUVEN

1 – Outline

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.'

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.'

- 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)

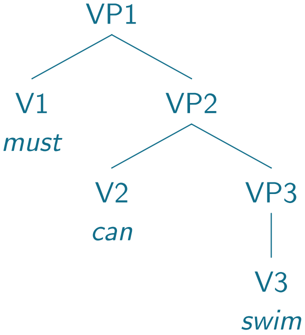
1 – Introduction & central research question

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

1 – Introduction & central research question

- 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)



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)

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.' (*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)

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.

(5) *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)

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:

example dialect	123	132	321	312
Beetgum	✓	✓	✓	✓
Hippolytushoef	✓	✓	✓	*
Warffum	✓	✓	*	*
Oosterend	✓	*	*	*
Schermerhorn	✓	✓	*	✓
Visvliet	✓	*	✓	✓
Kollum	✓	*	✓	*
Langelo	✓	*	*	✓
Midsland	*	✓	✓	*
Lies	*	*	✓	*
Bakkeveen	*	*	✓	✓
Waskemeer	*	✓	*	*

1 – Introduction & central research question

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

1 – Introduction & central research question

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

1 – Introduction & central research question

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

(6) **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:**

1 – Introduction & central research question

- **broader issues:**
  1. the grammatical nature of microvariation:

KU LEUVEN

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?

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

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:

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?

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 – The data

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

KU LEUVEN

2 – The data

- all data in this talk stem from the SAND-project:
  - *Syntactic Atlas of the Dutch Dialects* (2000–2004)

2 – The data

- all data in this talk stem from the SAND-project:
  - *Syntactic Atlas of the Dutch Dialects* (2000–2004)
  - universities of Amsterdam, Leiden, Antwerp, Brussels, Ghent, and the Meertensinstitute

2 – The data

- all data in this talk stem from the SAND-project:
  - *Syntactic Atlas of the Dutch Dialects* (2000–2004)
  - 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:
  - *Syntactic Atlas of the Dutch Dialects* (2000–2004)
  - 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

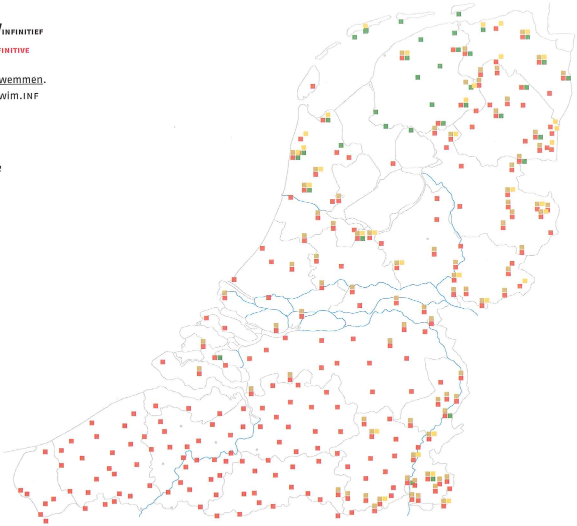
- all data in this talk stem from the SAND-project:
  - *Syntactic Atlas of the Dutch Dialects* (2000–2004)
  - universities of Amsterdam, Leiden, Antwerp, Brussels, Ghent, and the Meertensinstitute
  - four themes: left periphery of the clause, right periphery of the clause, pronominal reference, negation & quantification
  - 267 dialect locations in Belgium and the Netherlands
  - has yielded two atlas volumes

1.3.2.1 **Modaal$_{\text{FINIET}}$ • Modaal$_{\text{INFINITIEF}}$ • V$_{\text{INFINITIEF}}$**
Modal$_{\text{FINITE}}$ • Modal$_{\text{INFINITIVE}}$ • V$_{\text{INFINITIVE}}$

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

'I think that everyone should be able to swim.'

■	V1-V2-V3	(moet kunnen zwemmen)	242
■	V1-V3-V2	(moet zwemmen kunnen)	34
■	V3-V1-V2	(zwemmen moet kunnen)	83
■	V3-V2-V1	(zwemmen kunnen moet)	37



KU LEUVEN

2 – The data

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

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)

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)

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 – 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:
  1. 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

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.)

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

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)

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

3 – A dialectometric analysis: An MDS-analysis

- starting point: a 267×67 matrix with one row per location and one column per linguistic variable

	row.names	1_2_14a	2_1_14a	1_2_14b	2_1_14b	1_2_15a	2_1_15a	1_2_15b	2_1_15b	kunnen_z
1	A001p	0	1	0	1	0	1	0	1	0
2	A001q	0	1	0	1	0	1	0	1	0
3	A002p	0	1	0	1	0	1	1	1	0
4	A006p	NA	NA	0	1	0	1	0	1	0
5	B001a	0	1	NA	NA	0	1	0	1	0
6	B004p	0	1	0	1	0	1	0	1	0
7	B007p	0	1	0	1	0	1	0	1	0
8	B013b	0	1	0	1	0	1	0	1	0
9	B035p	0	1	0	1	0	1	0	1	0
10	B041p	0	1	0	1	0	1	1	1	0
11	B046b	0	1	0	1	0	1	0	1	NA
12	B052a	0	1	NA	NA	NA	NA	NA	NA	0
13	B062p	0	1	0	1	NA	NA	0	1	0
14	B085c	0	1	NA	NA	NA	NA	NA	NA	0
15	B127p	0	1	0	1	0	1	0	1	0
16	B128a	0	1	0	1	0	1	0	1	0
17	C023p	0	1	0	1	0	1	0	1	0
18	C029p	0	1	0	1	0	1	0	1	0

3 – A dialectometric analysis: An MDS-analysis

- starting point: a 267×67 matrix with one row per location and one column per linguistic variable
- step 1: convert the table into a 267×267 (symmetric) distance matrix, whereby for each pair of locations a distance between them is calculated based on the linguistic features they share

	A001p	A001q	A002p	A006p	B001a	B004p	B007p	B013b	B035p	B041p	B046b	B052a	B062p
A001p	0.000	0.500	0.333	0.706	0.250	0.647	0.357	0.250	0.611	0.650	0.533	0.545	0.500
A001q	0.500	0.000	0.444	0.750	0.588	0.375	0.471	0.563	0.444	0.444	0.632	0.714	0.500
A002p	0.333	0.444	0.000	0.789	0.429	0.667	0.286	0.429	0.632	0.600	0.500	0.500	0.429
A006p	0.706	0.750	0.789	0.000	0.706	0.765	0.737	0.538	0.563	0.600	0.600	0.727	0.813
B001a	0.250	0.588	0.429	0.706	0.000	0.667	0.167	0.000	0.625	0.714	0.462	0.500	0.500
B004p	0.647	0.375	0.667	0.765	0.667	0.000	0.625	0.667	0.400	0.556	0.706	0.750	0.571
B007p	0.357	0.471	0.286	0.737	0.167	0.625	0.000	0.182	0.588	0.682	0.308	0.333	0.333
B013b	0.250	0.563	0.429	0.538	0.000	0.667	0.182	0.000	0.571	0.625	0.417	0.556	0.500
B035p	0.611	0.444	0.632	0.563	0.625	0.400	0.588	0.571	0.000	0.353	0.625	0.643	0.429
B041p	0.650	0.444	0.600	0.600	0.714	0.556	0.682	0.625	0.353	0.000	0.588	0.500	0.667
B046b	0.533	0.632	0.500	0.600	0.462	0.706	0.308	0.417	0.625	0.588	0.000	0.167	0.571
B052a	0.545	0.714	0.500	0.727	0.500	0.750	0.333	0.556	0.643	0.500	0.167	0.000	0.500
B062p	0.500	0.500	0.429	0.813	0.500	0.571	0.333	0.500	0.429	0.667	0.571	0.500	0.000
B085c	0.692	0.667	0.583	0.846	0.545	0.667	0.400	0.600	0.571	0.692	0.500	0.455	0.222
B127p	0.400	0.500	0.438	0.706	0.385	0.563	0.357	0.385	0.438	0.579	0.533	0.545	0.385
B128a	0.438	0.526	0.556	0.818	0.500	0.588	0.471	0.533	0.471	0.652	0.588	0.667	0.429
C023p	0.500	0.412	0.611	0.810	0.563	0.357	0.529	0.600	0.333	0.636	0.706	0.667	0.385
C029p	0.563	0.438	0.667	0.737	0.625	0.429	0.588	0.643	0.400	0.652	0.600	0.636	0.571
C041a	0.667	0.652	0.739	0.550	0.773	0.650	0.739	0.722	0.389	0.455	0.667	0.571	0.684
C108p	0.714	0.682	0.714	0.636	0.783	0.762	0.800	0.778	0.471	0.476	0.684	0.714	0.737
C123p	0.650	0.682	0.650	0.652	0.773	0.762	0.739	0.722	0.556	0.368	0.647	0.615	0.667
C146t	0.727	0.524	0.739	0.652	0.792	0.650	0.760	0.647	0.550	0.500	0.700	0.824	0.810
C148p	0.773	0.636	0.727	0.600	0.818	0.556	0.783	0.789	0.600	0.500	0.611	0.714	0.800

3 – A dialectometric analysis: An MDS-analysis

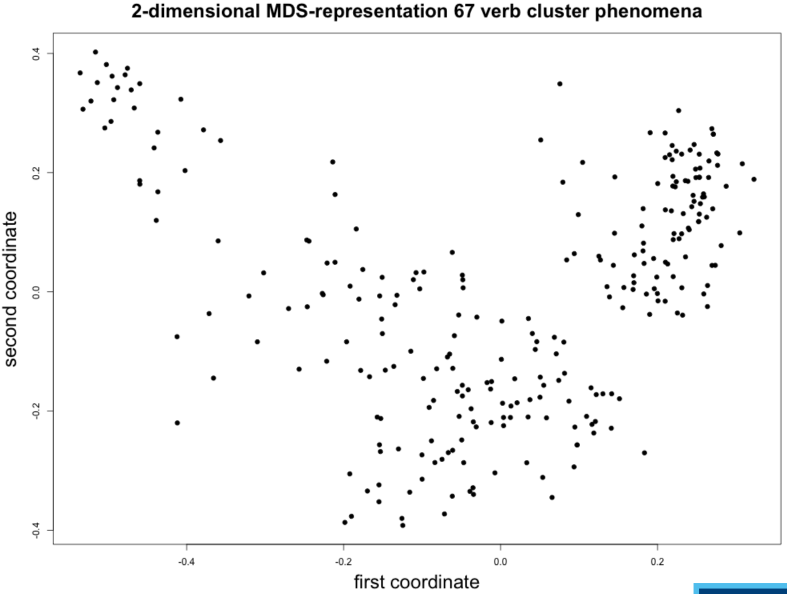
- starting point: a 267×67 matrix with one row per location and one column per linguistic variable
- step 1: convert the table into a 267×267 (symmetric) distance matrix, whereby for each pair of locations a distance between them is calculated based on the linguistic features they share
- step 2: apply multidimensional scaling (MDS) to the distance matrix

3 – A dialectometric analysis: An MDS-analysis

- starting point: a 267×67 matrix with one row per location and one column per linguistic variable
- step 1: convert the table into a 267×267 (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

3 – A dialectometric analysis: An MDS-analysis

- starting point: a 267×67 matrix with one row per location and one column per linguistic variable
- step 1: convert the table into a 267×267 (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)

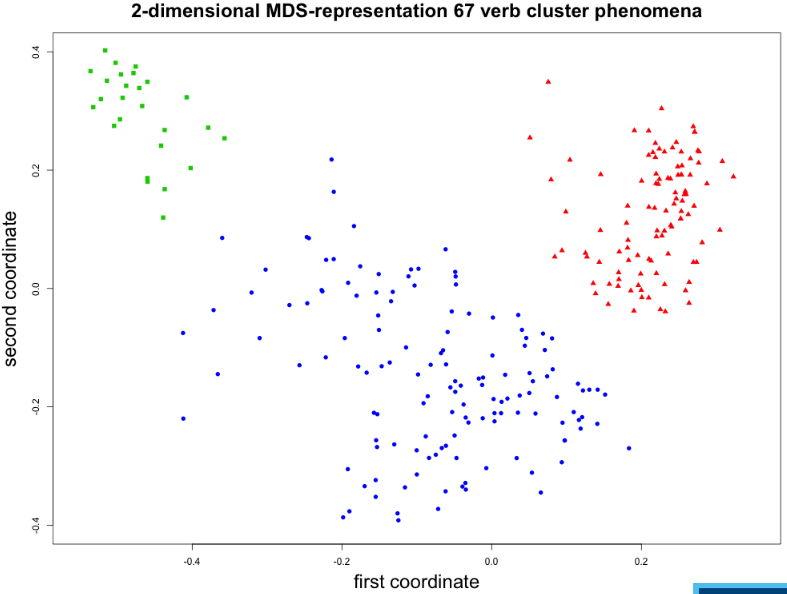

2-dimensional MDS-representation 67 verb cluster phenomena
second coordinate
first coordinate
-0.4
-0.2
0.0
0.2
0.4

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

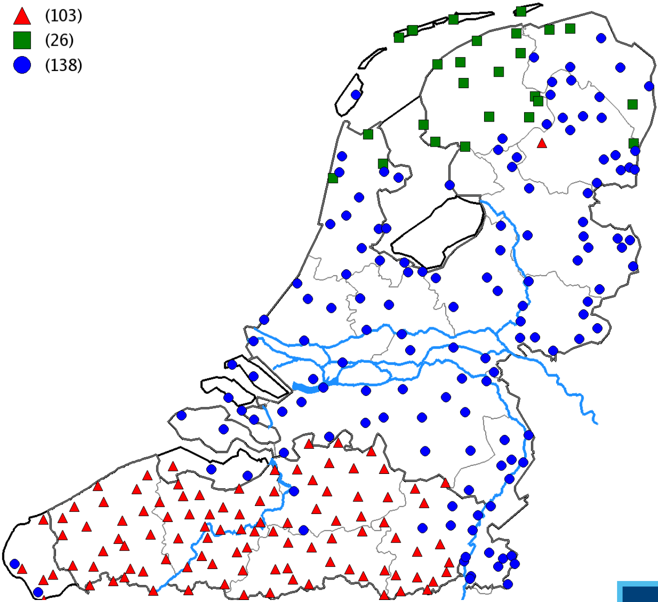
- **note:** the data is not randomly distributed → this suggests that the distribution of verb cluster orderings across the Dutch-speaking area is not random
- we can discern roughly three regions


2-dimensional MDS-representation 67 verb cluster phenomena
second coordinate
first coordinate
-0.4
-0.2
0.0
0.2
0.4

KU LEUVEN

3 – A dialectometric analysis: An MDS-analysis

- **note:** the data is not randomly distributed $\rightarrow$ 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 $\rightarrow$ this means we can plot these (three groups of) points against an actual map


(103)
(26)
(138)

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)

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

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

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

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

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

- 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

KU LEUVEN

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

- 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

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

- 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

1. **linguistic profile of the three regions**

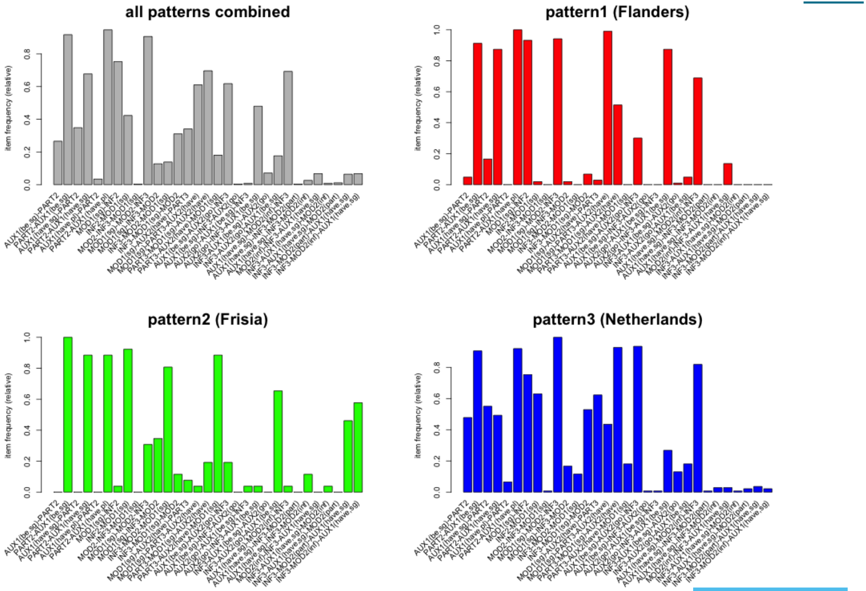
KU LEUVEN

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

- 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

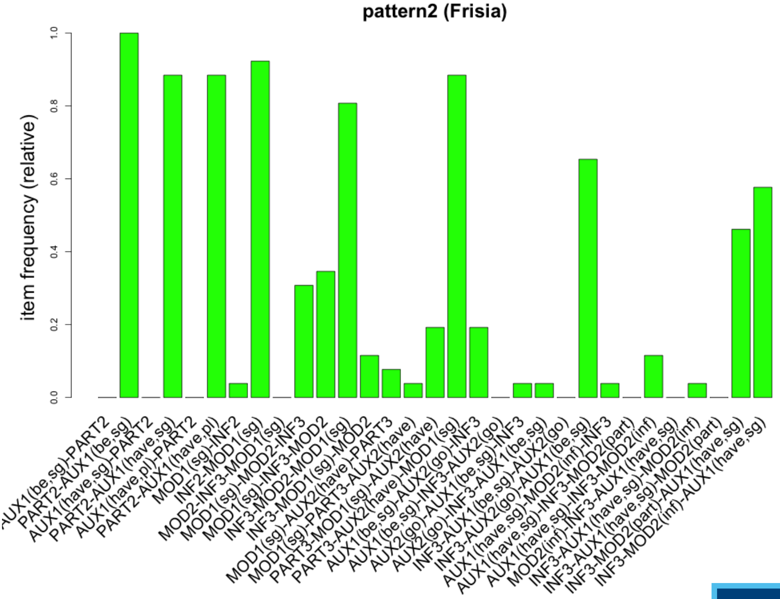
1. **linguistic profile of the three regions**

- 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


all patterns combined
pattern1 (Flanders)
pattern2 (Frisia)
pattern3 (Netherlands)
item frequency (relative)

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

- Frisia (green):


pattern2 (Frisia)
item frequency (relative)
0.0
0.2
0.4
0.6
0.8
1.0
UX1(be,sg)-PART2
PART2-AUX1(be,sg)
AUX1(have,sg)-PART2
PART2-AUX1(have,sg)
AUX1(have,pl)-PART2
PART2-AUX1(have,pl)
MOD1(sg)-INF2
INF2-MOD1(sg)
MOD2-INF3-MOD1(sg)
MOD1(sg)-MOD2-INF3
MOD1(sg)-INF3-MOD2
INF3-MOD2-MOD1(sg)
INF3-MOD1(sg)-MOD2
MOD1(sg)-AUX2(have)-PART3
MOD1(sg)-PART3-AUX2(have)
PART3-MOD1(sg)-AUX2(have)
PART3-AUX2(have)-MOD1(sg)
AUX1(be,sg)-AUX2(go)-INF3
AUX1(be,sg)-INF3-AUX2(go)
AUX2(go)-AUX1(be,sg)-INF3
AUX2(go)-INF3-AUX1(be,sg)
INF3-AUX1(be,sg)-AUX2(go)
INF3-AUX2(go)-AUX1(be,sg)
AUX1(have,sg)-MOD2(inf)-INF3
AUX1(have,sg)-INF3-MOD2(part)
AUX1(have,sg)-INF3-MOD2(inf)
MOD2(inf)-INF3-AUX1(have,sg)
INF3-AUX1(have,sg)-MOD2(inf)
INF3-AUX1(have,sg)-MOD2(part)
INF3-MOD2(part)-AUX1(have,sg)
INF3-MOD2(inf)-AUX1(have,sg)

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

- Frisia (green):

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

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

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

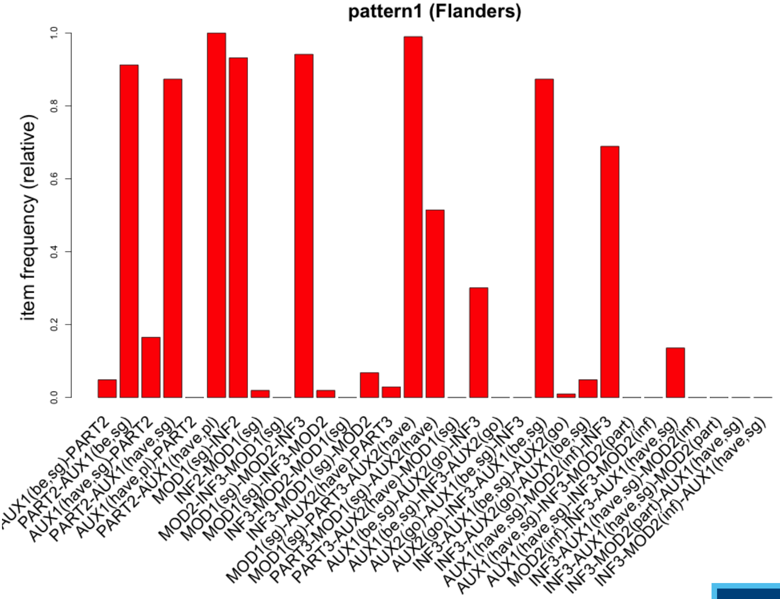
- Frisia (green):
  - consistenly 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)

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

- Frisia (green):
  - consistenly 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):
  - consistenly 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):


pattern1 (Flanders)
item frequency (relative)
0.0
0.2
0.4
0.6
0.8
1.0
AUX1(be,sg)-PART2
PART2-AUX1(be,sg)
AUX1(have,sg)-PART2
PART2-AUX1(have,sg)
AUX1(have,pl)-PART2
PART2-AUX1(have,pl)
MOD1(sg)-INF2
INF2-MOD1(sg)
MOD2-INF3-MOD1(sg)
MOD1(sg)-MOD2-INF3
MOD1(sg)-INF3-MOD2
INF3-MOD2-MOD1(sg)
INF3-MOD1(sg)-MOD2
MOD1(sg)-AUX2(have)-PART3
MOD1(sg)-PART3-AUX2(have)
PART3-MOD1(sg)-AUX2(have)
PART3-AUX2(have)-MOD1(sg)
AUX1(be,sg)-AUX2(go)-INF3
AUX1(be,sg)-INF3-AUX2(go)
AUX2(go)-AUX1(be,sg)-INF3
AUX2(go)-INF3-AUX1(be,sg)
INF3-AUX1(be,sg)-AUX2(go)
INF3-AUX2(go)-AUX1(be,sg)
AUX1(have,sg)-MOD2(inf)-INF3
AUX1(have,sg)-INF3-MOD2(part)
AUX1(have,sg)-INF3-MOD2(inf)
MOD2(inf)-INF3-AUX1(have,sg)
INF3-AUX1(have,sg)-MOD2(inf)
INF3-AUX1(have,sg)-MOD2(part)
INF3-MOD2(part)-AUX1(have,sg)
INF3-MOD2(inf)-AUX1(have,sg)

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

- Frisia (green):
  - consistenly 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

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

- Frisia (green):
  - consistenly 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

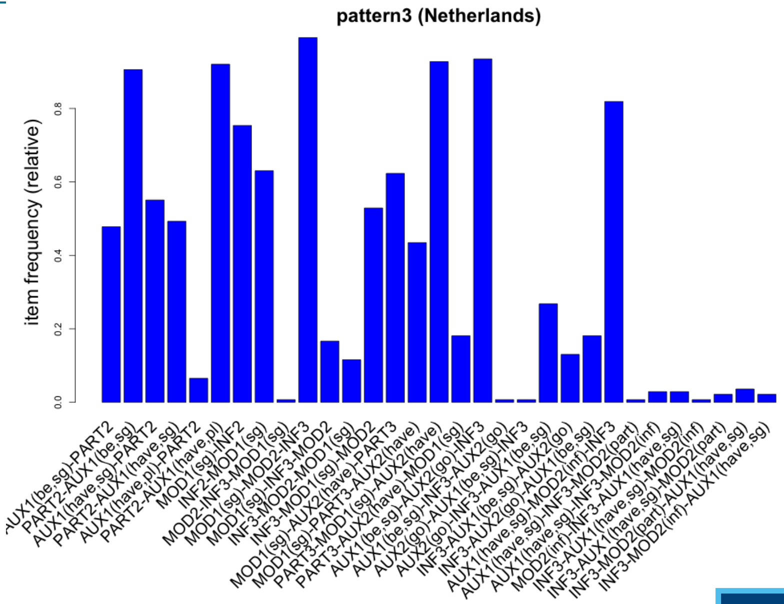
- Frisia (green):
  - consistenly 21 in two-verb clusters (i.e. V-AUX and V-MOD)
  - overwhelmingly 321 in three-verb clusters (MOD-MOD-V, AUX-AUX-V, AUX-MOD-V)
  - some 123 and 132 in MOD-MOD-V
- Flanders (red):
  - consistently 21 in AUX-V-clusters and 12 in MOD-V-clusters
  - consistently 123 in MOD-MOD-V-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

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

- Frisia (green):
  - consistenly 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):


pattern3 (Netherlands)
item frequency (relative)
0.0
0.2
0.4
0.6
0.8
UX1(be,sg)-PART2
PART2-AUX1(be,sg)
AUX1(have,sg)-PART2
PART2-AUX1(have,sg)
AUX1(have,pl)-PART2
PART2-AUX1(have,pl)
MOD1(sg)-INF2
INF2-MOD1(sg)
MOD2-INF3-MOD1(sg)
MOD1(sg)-MOD2-INF3
MOD1(sg)-INF3-MOD2
INF3-MOD2-MOD1(sg)
INF3-MOD1(sg)-MOD2
MOD1(sg)-AUX2(have)-PART3
MOD1(sg)-PART3-AUX2(have)
PART3-MOD1(sg)-AUX2(have)
PART3-AUX2(have)-MOD1(sg)
AUX1(be,sg)-AUX2(go)-INF3
AUX1(be,sg)-INF3-AUX2(go)
AUX2(go)-AUX1(be,sg)-INF3
AUX2(go)-INF3-AUX1(be,sg)
INF3-AUX1(be,sg)-AUX2(go)
INF3-AUX2(go)-AUX1(be,sg)
AUX1(have,sg)-MOD2(inf)-INF3
AUX1(have,sg)-INF3-MOD2(part)
AUX1(have,sg)-INF3-MOD2(inf)
MOD2(inf)-INF3-AUX1(have,sg)
INF3-AUX1(have,sg)-MOD2(inf)
INF3-AUX1(have,sg)-MOD2(part)
INF3-MOD2(part)-AUX1(have,sg)
INF3-MOD2(inf)-AUX1(have,sg)

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

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

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

2. **data per province**

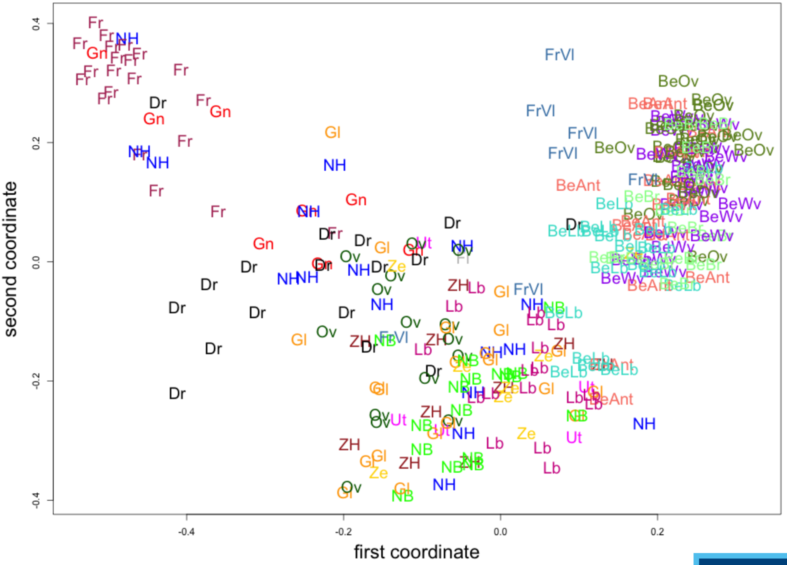
3 – A dialectometric analysis: A closer look at the data

- **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

2. **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)



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

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

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

3. **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

3. **scree plot**

- a scree plot maps the Stress of the MDS-representation against its dimensionality

- **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

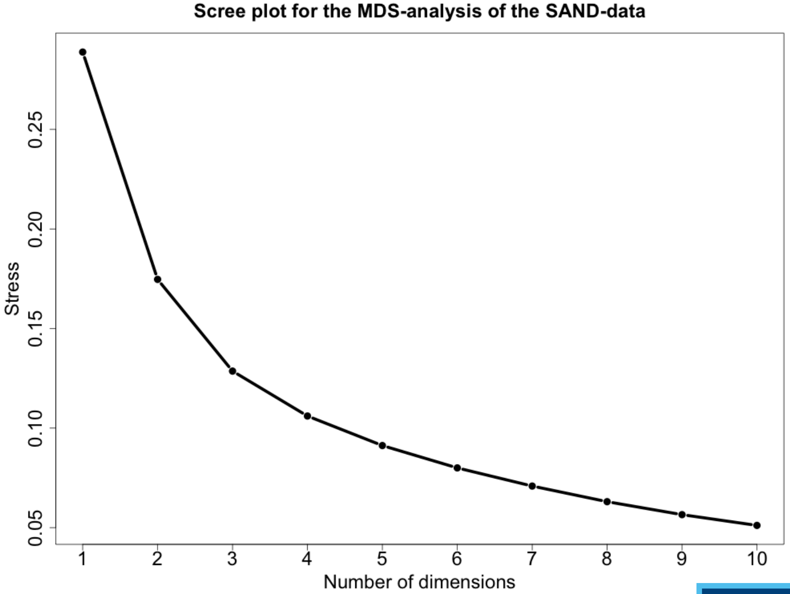
3. **scree plot**

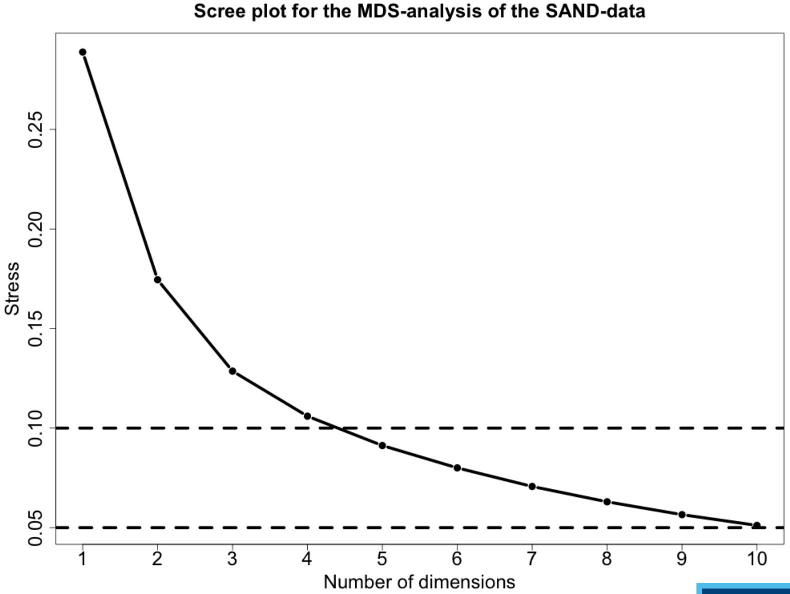
- a scree plot maps the Stress of the MDS-representation against its dimensionality
- Stress is a measure of ‘badness of fit’: how poorly does the MDS-analysis represent the actual distance matrix?

- **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

3. **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


Scree plot for the MDS-analysis of the SAND-data
Stress
0.05
0.10
0.15
0.20
0.25
1
2
3
4
5
6
7
8
9
10
Number of dimensions


Scree plot for the MDS-analysis of the SAND-data
Stress
0.05
0.10
0.15
0.20
0.25
1
2
3
4
5
6
7
8
9
10
Number of dimensions

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

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
- 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

3 – A dialectometric analysis: Conclusion

- 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

KU LEUVEN

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

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 31×267 matrix with one row per **verb cluster order** and one column per location

	row.names	A001p	A001q	A002p	A006p	B001a	B004p	B007p	B013b	B035p	B
1	1_2_14a	0	0	0	NA	0	0	0	0	0	0
2	2_1_14a	1	1	1	NA	1	1	1	1	1	1
3	1_2_14b	0	0	0	0	NA	0	0	0	0	0
4	2_1_14b	1	1	1	1	NA	1	1	1	1	1
5	1_2_15a	0	0	0	0	0	0	0	0	0	0
6	2_1_15a	1	1	1	1	1	1	1	1	1	1
7	1_2_15b	0	0	1	0	0	0	0	0	0	1
8	2_1_15b	1	1	1	1	1	1	1	1	1	1
9	kunnen_zwemmen_moet	0	0	0	0	0	0	0	0	0	0
10	moet_kunnen_zwemmen	0	0	0	1	0	0	0	0	1	1
11	moet_zwemmen_kunnen	1	0	0	0	0	0	0	0	0	0
12	zwemmen_kunnen_moet	1	1	1	0	1	1	1	1	1	1
13	zwemmen_moet_kunnen	0	0	0	0	0	0	0	0	0	1
14	123_moet_hebben_gemaakt	0	0	0	0	0	0	0	NA	0	0
15	132_moet_gemaakt_hebben	0	0	0	0	0	0	0	NA	1	1
16	312_gemaakt_moet_hebben	0	1	0	1	0	0	0	NA	1	1
17	321_gemaakt_hebben_moet	1	1	1	0	1	1	1	NA	1	1
18	v1v2v3	0	0	0	1	0	0	0	0	NA	1

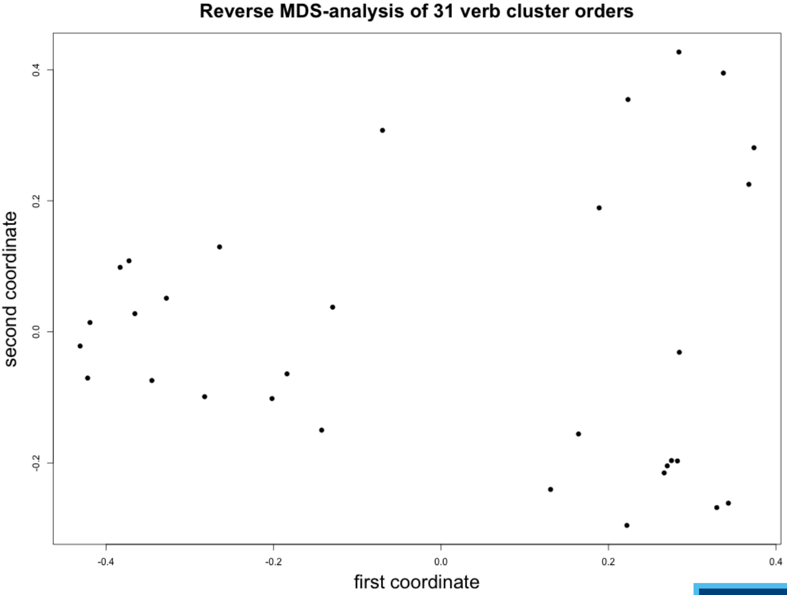
4 – Reversing the perspective: A reverse MDS-analysis

- a traditional dialectometric MDS-analysis allows us to approach the main research question only inditectly, 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 31×267 matrix with one row per **verb cluster order** and one column per location
- step 1: convert the table into a 31×31 (symmetric) distance matrix, whereby for each pair of verb cluster orders a distance between them is calculated based on their geographical spread

	1_2_14a	2_1_14a	1_2_14b	2_1_14b	1_2_15a	2_1_15a	1_2_15b	2_1_15b	ku
1_2_14a	0,000	0.755905511	0.580357142	0.876190476	0.916666666	0.738775510	0.702020202	0.75	0.
2_1_14a	0.755905511	0,000	0.648760330	0.280851063	0.971074380	0.040816326	0.231092436	0.551282051	0.
1_2_14b	0.580357142	0.648760330	0,000	0.925490196	0.934065934	0.661354581	0.600985221	0.727272727	0.
2_1_14b	0.876190476	0.280851063	0.925490196	0,000	0.989071038	0.276422764	0.440677966	0.647342995	
1_2_15a	0.916666666	0.971074380	0.934065934	0.989071038	0,000	0.992307692	0.959595959	0.974137931	
2_1_15a	0.738775510	0.040816326	0.661354581	0.276422764	0.992307692	0,000	0.219512195	0.543568464	0.
1_2_15b	0.702020202	0.231092436	0.600985221	0.440677966	0.959595959	0.219512195	0,000	0.744	
2_1_15b	0.75	0.551282051	0.727272727	0.647342995	0.974137931	0.543568464	0.744	0,000	0.
kunnen_zwe	0.985915492	0.995798319	0.989130434	1,000	1,000	0.995918367	1,000	0.990990990	
moet_kunne	0.691304347	0.105263157	0.603448275	0.356275303	0.962025316	0.087301587	0.149779735	0.607438016	0.
moet_zwem	0.892473118	0.870292887	0.891891891	0.885869565	1,000	0.869387755	0.918660287	0.756521739	0.
zwemmen_k	0.950980392	0.848739495	0.932203389	0.850828729	1,000	0.865306122	0.959090909	0.684684684	0.
zwemmen_m	0.654867256	0.681818181	0.645161290	0.803827751	0.940476190	0.689516129	0.695238095	0.563909774	0.
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	0.
312_gemaak	0.644808743	0.318548387	0.578125	0.530864197	0.961748633	0.303149606	0.355263157	0.562814070	0.
321_gemaak	0.955357142	0.808333333	0.896	0.812834224	1,000	0.823293172	0.920704845	0.619469026	0.
v1v2v3	0.608433734	0.381147540	0.554913294	0.593220338	0.944099378	0.388888888	0.438053097	0.519553072	0.

4 – Reversing the perspective: A reverse MDS-analysis

- a traditional dialectometric MDS-analysis allows us to approach the main research question only inditectly, 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 31×267 matrix with one row per **verb cluster order** and one column per location
- step 1: convert the table into a 267×367 (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
second coordinate
0.4
0.2
0.0
-0.2
first coordinate
-0.4
-0.2
0.0
0.2
0.4

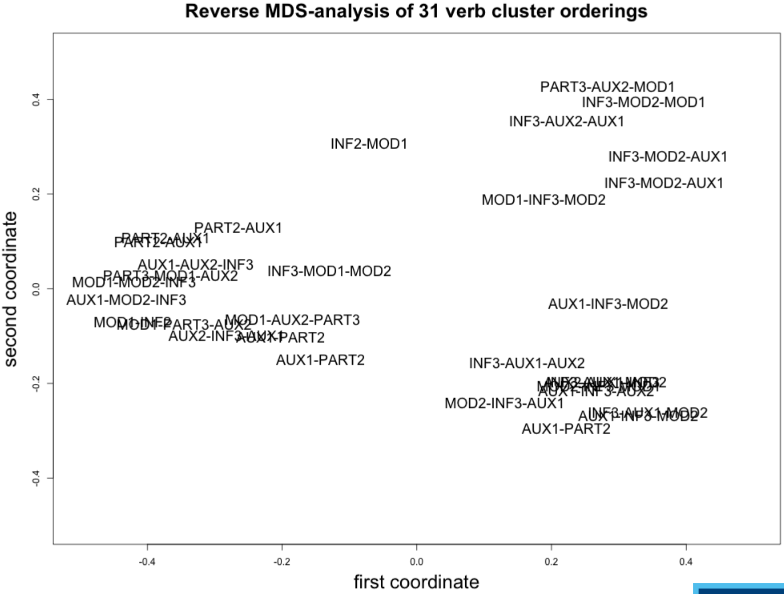
KU LEUVEN

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

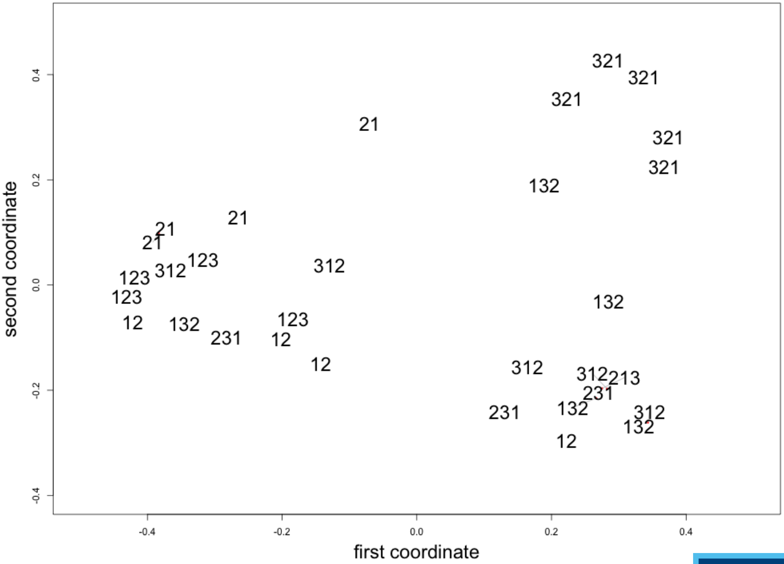
4 – Reversing the perspective: A reverse MDS-analysis

- **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
PART3-AUX2-MOD1
INF3-MOD2-MOD1
INF3-AUX2-AUX1
INF2-MOD1
INF3-MOD2-AUX1
INF3-MOD2-AUX1
MOD1-INF3-MOD2
PART2-AUX1
PART2-AUX1
PART2-AUX1
AUX1-AUX2-INF3
PART3-MOD1-AUX2
INF3-MOD1-MOD2
MOD1-MOD2-INF3
AUX1-MOD2-INF3
AUX1-INF3-MOD2
MOD1-INF2
MOD1-AUX2-PART3
AUX1-PART2
AUX1-PART2
INF3-AUX1-AUX2
AUX1-INF3-AUX2
MOD2-INF3-AUX1
INF3-AUX1-MOD2
AUX1-INF3-MOD2
AUX1-PART2
second coordinate
first coordinate
0.4
0.2
0.0
-0.2
-0.4

KU LEUVEN

Reverse MDS-analysis of 31 verb cluster orderings



KU LEUVEN

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

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 $\rightarrow$ 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)

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)
  - 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

4 – Reversing the perspective: A reverse MDS-analysis

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

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?

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?

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?
  - [±pied-piping]: does the derivation involve pied-piping?

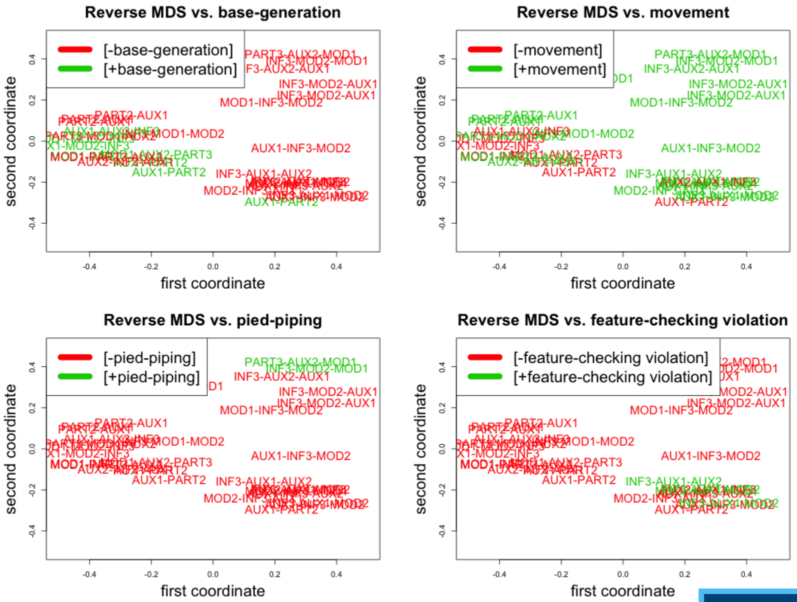
4 – Reversing the perspective: A reverse MDS-analysis

- 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?
  - [±pied-piping]: does the derivation involve pied-piping?
  - [±feature-checking violation]: does the order involve a feature checking violation?

KU LEUVEN

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?
  - [±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

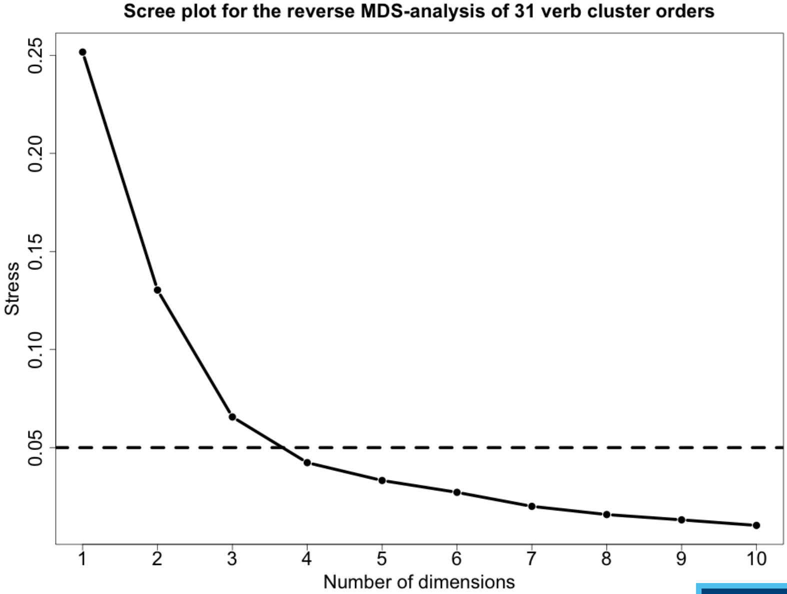

Reverse MDS vs. base-generation
[-base-generation]
[+base-generation]
Reverse MDS vs. movement
[-movement]
[+movement]
Reverse MDS vs. pied-piping
[-pied-piping]
[+pied-piping]
Reverse MDS vs. feature-checking violation
[-feature-checking violation]
[+feature-checking violation]
second coordinate
first coordinate
PART3-AUX2-MOD1
INF3-MOD2-MOD1
INF3-AUX2-AUX1
INF3-MOD2-AUX1
MOD1-INF3-MOD2
PART2-AUX1
AUX1-INF3-MOD2
AUX1-PART2
INF3-AUX1-AUX2

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?
  - [±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

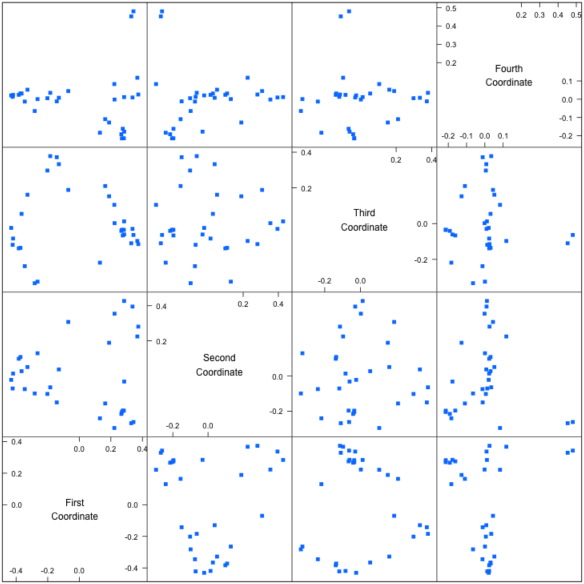

Scree plot for the reverse MDS-analysis of 31 verb cluster orders
Stress
0.05
0.10
0.15
0.20
0.25
1
2
3
4
5
6
7
8
9
10
Number of dimensions

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
- 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

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
- 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



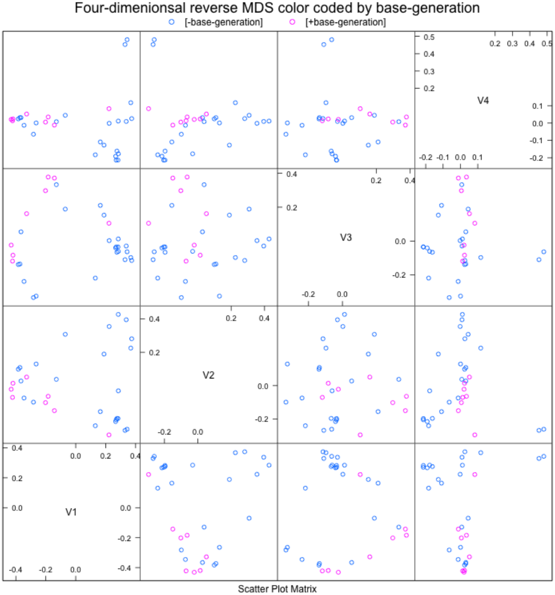
Four-dimensional reverse MDS-analysis of 31 verb cluster orderings

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
- 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

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
- the plot suggests that our data is four-dimensional $\rightarrow$ 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 [$\pm$base-generation]-parameter doesn't seem to correspond to any of the four dimensions


Four-dimenionsal reverse MDS color coded by base-generation
[-base-generation]
[+base-generation]
V4
V3
V2
V1
Scatter Plot Matrix

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

KU LEUVEN

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?

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

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
- **broader issues:**

KU LEUVEN

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
- **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

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
- **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

6 – References

Barbiers, Sjef. 2005. Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics. In *Syntax and variation. Reconciling the biological and the social*, ed. Leonie Cornips and Karen P. Corrigan, volume 265 of *Current issues in linguistic theory*, 233–264. John Benjamins.

Borg, Ingwer, and Patrick J.F. Groenen. 2005. *Modern Multidimensional Scaling. Theory and applications.*. Springer, 2nd edition.

Nerbonne, John, and William A. Kretzschmar Jr. 2013. Dialectometry++. *Literary and Linguistic Computing* 28:2–12.

Wurmbrand, Susanne. 2005. Verb clusters, verb raising, and restructuring. In *The Blackwell Companion to Syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume V, chapter 75, 227–341. Oxford: Blackwell.