Quantifying dialect Dutch verb clusters

Jeroen van Craenenbroeck

KU Leuven/CRISSP

SyHD-workshop

Dialect syntax – the state of the art

December 5–6, 2014

This talk in one slide

This talk in one slide

- **main topic:** interaction between formal-theoretical and quantitative-statistical linguistics

This talk in one slide

- **main topic:** interaction between formal-theoretical and quantitative-statistical linguistics
- **starting point:** the massive amount of variation attested in Dutch verb clusters necessitates a collaboration between formal and quantitative approaches

This talk in one slide

- **main topic:** interaction between formal-theoretical and quantitative-statistical linguistics
- **starting point:** the massive amount of variation attested in Dutch verb clusters necessitates a collaboration between formal and quantitative approaches
- **traditional dialectometry** measures (dis)similarities between dialect locations based on their linguistic profile

This talk in one slide

- **main topic:** interaction between formal-theoretical and quantitative-statistical linguistics
- **starting point:** the massive amount of variation attested in Dutch verb clusters necessitates a collaboration between formal and quantitative approaches
- **traditional dialectometry** measures (dis)similarities between dialect locations based on their linguistic profile
- **reverse dialectometry** measures (dis)similarities between *linguistic constructions* based on their geographical spread and maps these results against formal-theoretical parameters

This talk in one slide

- **main topic:** interaction between formal-theoretical and quantitative-statistical linguistics
- **starting point:** the massive amount of variation attested in Dutch verb clusters necessitates a collaboration between formal and quantitative approaches
- **traditional dialectometry** measures (dis)similarities between dialect locations based on their linguistic profile
- **reverse dialectometry** measures (dis)similarities between *linguistic constructions* based on their geographical spread and maps these results against formal-theoretical parameters
- **result:** a method that can detect and identify grammatical parameters in a large and highly varied data set

Introduction: variation in verb clusters

Introduction: variation in verb clusters

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

Introduction: variation in verb clusters

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

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

(2) dat hij gisteren tijdens de les heeft gelachen.
that he yesterday during the class had laughed
'that he laughed yesterday during class.'

Introduction: variation in verb clusters

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

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

(2) dat hij gisteren tijdens de les heeft gelachen.
that he yesterday during the class had laughed
'that he laughed yesterday during class.'

- this freedom in word order is a source of massive interdialectal variation

- e.g. the SAND-project:

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:
  - three two-verb clusters of the form AUXILIARY-PARTICIPLE

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:
  - three two-verb clusters of the form AUXILIARY-PARTICIPLE
  - one two-verb cluster of the form MODAL-INFINITIVE

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:
  - three two-verb clusters of the form AUXILIARY-PARTICIPLE
  - one two-verb cluster of the form MODAL-INFINITIVE
  - four three-verb clusters:

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:
  - three two-verb clusters of the form AUXILIARY-PARTICIPLE
  - one two-verb cluster of the form MODAL-INFINITIVE
  - four three-verb clusters:
    - MODAL-MODAL-INFINITIVE

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:
  - three two-verb clusters of the form AUXILIARY-PARTICIPLE
  - one two-verb cluster of the form MODAL-INFINITIVE
  - four three-verb clusters:
    - MODAL-MODAL-INFINITIVE
    - MODAL-AUXILIARY-PARTICIPLE

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:
  - three two-verb clusters of the form AUXILIARY-PARTICIPLE
  - one two-verb cluster of the form MODAL-INFINITIVE
  - four three-verb clusters:
    - MODAL-MODAL-INFINITIVE
    - MODAL-AUXILIARY-PARTICIPLE
    - AUXILIARY-AUXILIARY-INFINITIVE

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:
  - three two-verb clusters of the form AUXILIARY-PARTICIPLE
  - one two-verb cluster of the form MODAL-INFINITIVE
  - four three-verb clusters:
    - MODAL-MODAL-INFINITIVE
    - MODAL-AUXILIARY-PARTICIPLE
    - AUXILIARY-AUXILIARY-INFINITIVE
    - AUXILIARY-MODAL-INFINITIVE

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:
  - three two-verb clusters of the form AUXILIARY-PARTICIPLE
  - one two-verb cluster of the form MODAL-INFINITIVE
  - four three-verb clusters:
    - MODAL-MODAL-INFINITIVE
    - MODAL-AUXILIARY-PARTICIPLE
    - AUXILIARY-AUXILIARY-INFINITIVE
    - AUXILIARY-MODAL-INFINITIVE
  - for a total of 31 cluster orders

- e.g. the SAND-project:
  - Syntactic Atlas of the Dutch Dialects (2000–2004)
  - dialect interviews in 267 dialect locations in Belgium, France, and the Netherlands
- the SAND-questionnaire contained eight questions on word order in verb clusters:
  - three two-verb clusters of the form AUXILIARY-PARTICIPLE
  - one two-verb cluster of the form MODAL-INFINITIVE
  - four three-verb clusters:
    - MODAL-MODAL-INFINITIVE
    - MODAL-AUXILIARY-PARTICIPLE
    - AUXILIARY-AUXILIARY-INFINITIVE
    - AUXILIARY-MODAL-INFINITIVE
  - 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

- this state of affairs raises fundamental questions for parameter theory:

- this state of affairs raises fundamental questions for parameter theory:
    - are there really grammatical (micro)parameters distinguishing between all of these 137 dialect types?

- this state of affairs raises fundamental questions for parameter theory:
  - are there really grammatical (micro)parameters distinguishing between all of these 137 dialect types?
  - if there are, what are they and how can we detect them?

- this state of affairs raises fundamental questions for parameter theory:
  - are there really grammatical (micro)parameters distinguishing between all of these 137 dialect types?
  - if there are, what are they and how can we detect them?
  - more generally, how can we distinguish between the signal and the noise in such large and highly variable datasets?

- this state of affairs raises fundamental questions for parameter theory:
    - are there really grammatical (micro)parameters distinguishing between all of these 137 dialect types?
    - if there are, what are they and how can we detect them?
    - more generally, how can we distinguish between the signal and the noise in such large and highly variable datasets?
- **in this talk** I use statistical methods to detect and identify grammatical microparameters regulating (a large part of) the variation found in Dutch verb clusters

A dialectometric analysis

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

A dialectometric analysis

- **dialectometry** is a subdiscipline of linguistics that uses computational and quantitative techniques in dialectology (Nerbonne and Kretzschmar Jr., 2013)
- a typical dialectometric analysis measures similarities and differences between dialect locations based on their linguistic profile

A dialectometric analysis

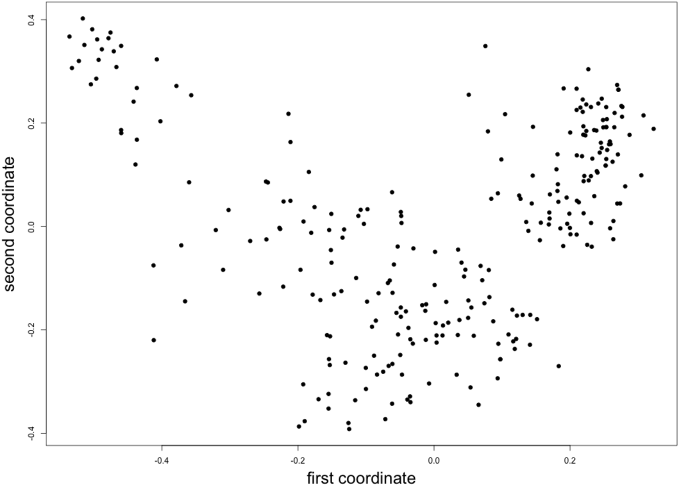
- **dialectometry** is a subdiscipline of linguistics that uses computational and quantitative techniques in dialectology (Nerbonne and Kretzschmar Jr., 2013)
- a typical dialectometric analysis measures similarities and differences between dialect locations based on their linguistic profile
- starting point: data table with dialects in rows and cluster orders in columns

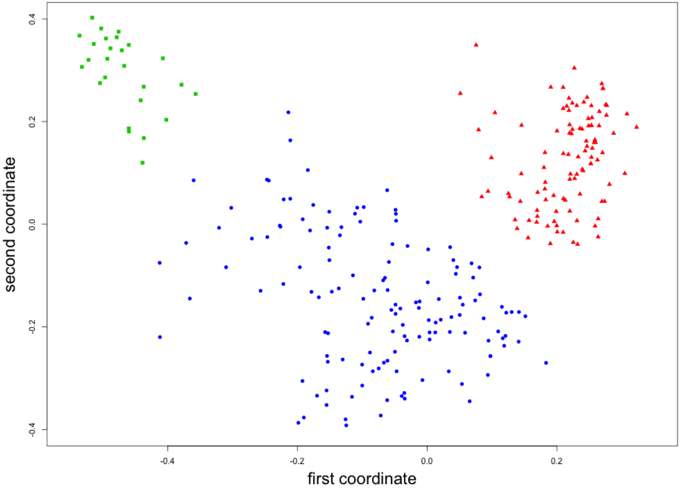
	AUX1(be.sg)-PART2	PART2-AUX1(be.sg)	AUX1(have.sg)-PART2	PART2-AUX1(have.sg)	AUX1(have.pl)-PAF
Midsland / Midslâr	no	yes	no	yes	
Lies	no	yes	no	yes	
West-Terschelling	no	yes	no	yes	
Oosterend	NA	NA	no	yes	
Hollum	no	yes	NA	NA	
Schiermonnikoog	no	yes	no	yes	
Ferwerd / Ferwert	no	yes	no	yes	
Anjum / Eanjum	no	yes	no	yes	
Kollum	no	yes	no	yes	
Visvliet	no	yes	no	yes	
Oosterbierum / Ea	no	yes	no	yes	
Beetgum / Bitgum	no	yes	NA	NA	
Bergum / Burgum	no	yes	no	yes	
Jorwerd / Jorwert	no	yes	NA	NA	
Bakkeveen / Bakk	no	yes	no	yes	
Waskemeer / De \	no	yes	no	yes	
Kloosterburen	no	yes	no	yes	
Warffum	no	yes	no	yes	
Leermens	no	yes	no	yes	
Groningen	no	yes	yes	no	
Nieuw-Scheemda	NA	NA	no	yes	
Langelo	no	yes	no	yes	

- step 1: convert the data 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

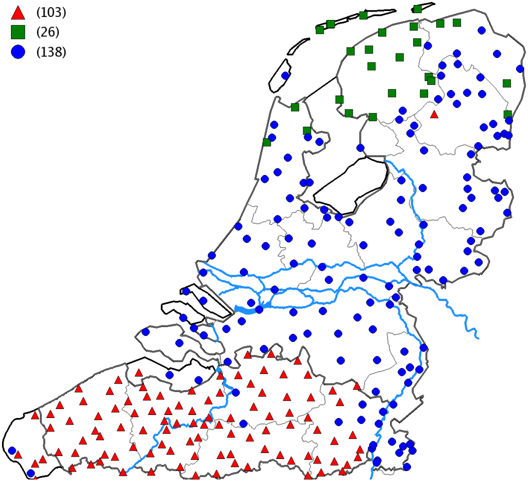
	Midsland	Lies	West-Ter:	Oosteren	Hollum	Schiermo	Ferwerd /	Anjum /	Kollum	Visvliet	Oosterbie	Beetgum	Bergum /	Jorwerd /
Midsland / Mic	0,000	0,500	0,333	0,706	0,250	0,647	0,357	0,250	0,611	0,650	0,533	0,545	0,500	0,692
Lies	0,500	0,000	0,444	0,750	0,588	0,375	0,471	0,563	0,444	0,444	0,632	0,714	0,500	0,667
West-Terschel	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	0,583
Oosterend	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	0,846
Hollum	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	0,545
Schiermonnik	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	0,667
Ferwerd / Fer	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	0,400
Anjum / Eanju	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	0,600
Kollum	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	0,571
Visvliet	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	0,692
Oosterbierum	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	0,500
Beetgum / Bit	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	0,455
Bergum / Bur	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	0,222
Jorwerd / Jorv	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	0,000
Bakkeveen / E	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	0,583
Waskemeer /	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	0,500
Kloosterburen	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	0,583
Warffum	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	0,750
Leermens	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	0,765
Groningen	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	0,786
Nieuw-Scheer	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	0,786
Langelo	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	0,950

- step 2: reduce this 267-dimensional matrix to a two- or three-dimensional one, so that it can easily be visualized


second coordinate
first coordinate
0.4
0.2
0.0
-0.2
-0.4
-0.4
-0.2
0.0
0.2


second coordinate
first coordinate
0.4
0.2
0.0
-0.2
-0.4
-0.4
-0.2
0.0
0.2

- step 3: project back onto a geographical map


(103)
(26)
(138)

- 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

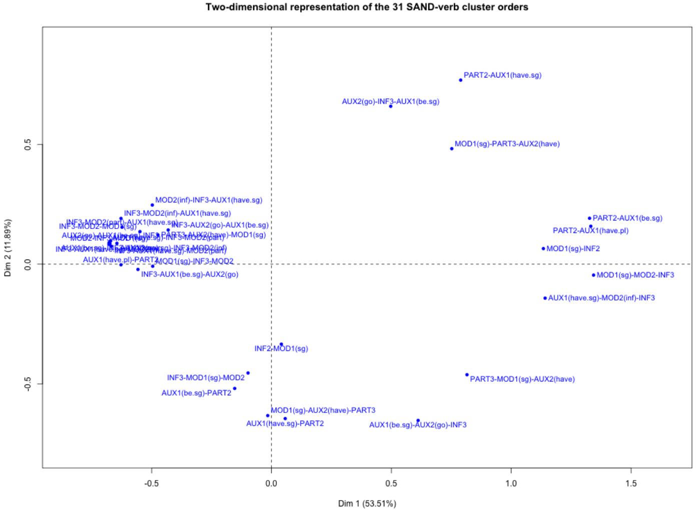
- **proposal:** let's treat cluster orders as *individuals* rather than variables, i.e. instead of calculating differences between dialect locations, we measure differences between linguistic constructions

Reverse dialectometry

- **proposal:** let's treat cluster orders as *individuals* rather than variables, i.e. instead of calculating differences between dialect locations, we measure differences between linguistic constructions
- starting point: a data table with cluster orders as rows and dialect locations as columns

	Midsland	Lies	West.Tersch	Oosterend	Hollum	Schiermonni	Ferwerd	Anjum	Kollum	Visvliet
AUX1(be.sg)-PART2	no	no	no	NA	no	no	no	no	no	no
PART2-AUX1(be.sg)	yes	yes	yes	NA	yes	yes	yes	yes	yes	yes
AUX1(have.sg)-PART2	no	no	no	no	NA	no	no	no	no	no
PART2-AUX1(have.sg)	yes	yes	yes	yes	NA	yes	yes	yes	yes	yes
AUX1(have.pl)-PART2	no	no	no	no	no	no	no	no	no	no
PART2-AUX1(have.pl)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
MOD1(sg)-INF2	no	no	yes	no	no	no	no	no	no	yes
INF2-MOD1(sg)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
MOD2-INF3-MOD1(sg)	no	no	no	no	no	no	no	no	no	no
MOD1(sg)-MOD2-INF3	no	no	no	yes	no	no	no	no	yes	yes
MOD1(sg)-INF3-MOD2	yes	no	no	no	no	no	no	no	no	no
INF3-MOD2-MOD1(sg)	yes	yes	yes	no	yes	yes	yes	yes	yes	yes
INF3-MOD1(sg)-MOD2	no	no	no	no	no	no	no	no	no	yes
MOD1(sg)-AUX2(have)-PART3	no	no	no	no	no	no	no	NA	no	no
MOD1(sg)-PART3-AUX2(have)	no	no	no	no	no	no	no	NA	yes	yes
PART3-MOD1(sg)-AUX2(have)	no	yes	no	yes	no	no	no	NA	yes	yes
PART3-AUX2(have)-MOD1(sg)	yes	yes	yes	no	yes	yes	yes	NA	yes	yes
AUX1(be.sg)-AUX2(go)-INF3	no	no	no	yes	no	no	no	no	NA	yes
AUX1(be.sg)-INF3-AUX2(go)	no	no	no	no	no	no	no	no	NA	no
AUX2(go)-AUX1(be.sg)-INF3	no	no	no	no	no	yes	no	no	NA	no
AUX2(go)-INF3-AUX1(be.sg)	no	no	no	no	no	no	no	no	NA	no
INF3-AUX1(be.sg)-AUX2(go)	no	no	no	no	no	no	no	no	NA	no
INF3-AUX2(go)-AUX1(be.sg)	yes	yes	yes	no	yes	no	yes	yes	NA	no
AUX1(have.sg)-MOD2(inf)-INF3	no	no	no	yes	no	no	no	no	no	no
AUX1(have.sg)-INF3-MOD2(part)	no	no	no	no	no	no	no	no	no	yes
AUX1(have.sg)-INF3-MOD2(inf)	no	no	no	no	no	no	no	no	no	no
MOD2(inf)-INF3-AUX1(have.sg)	no	no	no	no	no	no	no	no	no	no
INF3-AUX1(have.sg)-MOD2(inf)	no	no	yes	no	no	no	no	no	no	no
INF3-AUX1(have.sg)-MOD2(part)	no	no	no	no	no	no	no	no	no	yes
INF3-MOD2(part)-AUX1(have.sg)	no	yes	no	no	no	yes	no	no	yes	yes
INF3-MOD2(inf)-AUX1(have.sg)	yes	yes	yes	no	yes	no	yes	yes	no	yes

- transform to a distance matrix and reduce its dimensionality


Two-dimensional representation of the 31 SAND-verb cluster orders
PART2-AUX1(have.sg)
AUX2(go)-INF3-AUX1(be.sg)
MOD1(sg)-PART3-AUX2(have)
MOD2(inf)-INF3-AUX1(have.sg)
INF3-MOD2(inf)-AUX1(have.sg)
INF3-AUX2(go)-AUX1(be.sg)
AUX1(have.pl)-PART3
MOD1(sg)-INF3-MOD2
INF3-AUX1(be.sg)-AUX2(go)
PART2-AUX1(be.sg)
PART2-AUX1(have.pl)
MOD1(sg)-INF2
MOD1(sg)-MOD2-INF3
AUX1(have.sg)-MOD2(inf)-INF3
INF2-MOD1(sg)
INF3-MOD1(sg)-MOD2
AUX1(be.sg)-PART2
PART3-MOD1(sg)-AUX2(have)
MOD1(sg)-AUX2(have)-PART3
AUX1(have.sg)-PART2
AUX1(be.sg)-AUX2(go)-INF3
Dim 2 (11.89%)
0.5
0.0
-0.5
-0.5
0.0
0.5
1.0
1.5
Dim 1 (53.51%)

- **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 microparameters, 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 microparameters, 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

- e.g. Barbiers (2005)'s analysis of verb clusters: head-initial base structure, all movements are VP-intrapositions, movement is feature-driven and can pied-pipe VPs other than the one undergoing feature checking

- e.g. Barbiers (2005)'s analysis of verb clusters: head-initial base structure, all movements are VP-intrapositions, movement is feature-driven and can pied-pipe VPs other than the one undergoing feature checking

⇒ this account can be decomposed into the following microparameters:

- e.g. Barbiers (2005)'s analysis of verb clusters: head-initial base structure, all movements are VP-intrapositions, movement is feature-driven and can pied-pipe VPs other than the one undergoing feature checking

$\Rightarrow$ this account can be decomposed into the following microparameters:

  - [$\pm$base-generation]: can the order be base-generated?

- e.g. Barbiers (2005)'s analysis of verb clusters: head-initial base structure, all movements are VP-intrapositions, movement is feature-driven and can pied-pipe VPs other than the one undergoing feature checking

$\Rightarrow$ this account can be decomposed into the following microparameters:

  - [$\pm$base-generation]: can the order be base-generated?
  - [$\pm$movement]: can the order be derived via movement?

- e.g. Barbiers (2005)'s analysis of verb clusters: head-initial base structure, all movements are VP-intrapositions, movement is feature-driven and can pied-pipe VPs other than the one undergoing feature checking

$\Rightarrow$ this account can be decomposed into the following microparameters:

  - [$\pm$base-generation]: can the order be base-generated?
  - [$\pm$movement]: can the order be derived via movement?
  - [$\pm$pied-piping]: does the derivation involve pied-piping?

- e.g. Barbiers (2005)'s analysis of verb clusters: head-initial base structure, all movements are VP-intrapositions, movement is feature-driven and can pied-pipe VPs other than the one undergoing feature checking

$\Rightarrow$ this account can be decomposed into the following microparameters:

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

- e.g. Barbiers (2005)'s analysis of verb clusters: head-initial base structure, all movements are VP-intrapositions, movement is feature-driven and can pied-pipe VPs other than the one undergoing feature checking

⇒ this account can be decomposed into 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 our 31 cluster orders can be encoded in terms of these microparameters

	Barbiers-base.generation	Barbiers-movement	Barbiers-spec-pied-piping	Barbiers-feature.checking-failure
AUX1(be.sg)-PART2	yesBase	noMvt	noPiedP	noFeatCheckFail
PART2-AUX1(be.sg)	noBase	yesMvt	noPiedP	noFeatCheckFail
AUX1(have.sg)-PART2	yesBase	noMvt	noPiedP	noFeatCheckFail
PART2-AUX1(have.sg)	noBase	yesMvt	noPiedP	noFeatCheckFail
AUX1(have.pl)-PART2	yesBase	noMvt	noPiedP	noFeatCheckFail
PART2-AUX1(have.pl)	noBase	yesMvt	noPiedP	noFeatCheckFail
MOD1(sg)-INF2	yesBase	noMvt	noPiedP	noFeatCheckFail
INF2-MOD1(sg)	noBase	yesMvt	noPiedP	noFeatCheckFail
MOD2-INF3-MOD1(sg)	noBase	yesMvt	noPiedP	yesFeatCheckFail
MOD1(sg)-MOD2-INF3	yesBase	noMvt	noPiedP	noFeatCheckFail
MOD1(sg)-INF3-MOD2	noBase	yesMvt	noPiedP	noFeatCheckFail
INF3-MOD2-MOD1(sg)	noBase	yesMvt	yesPiedP	noFeatCheckFail
INF3-MOD1(sg)-MOD2	noBase	yesMvt	noPiedP	noFeatCheckFail
MOD1(sg)-AUX2(have)-PART3	yesBase	noMvt	noPiedP	noFeatCheckFail
MOD1(sg)-PART3-AUX2(have)	noBase	yesMvt	noPiedP	noFeatCheckFail
PART3-MOD1(sg)-AUX2(have)	noBase	yesMvt	noPiedP	noFeatCheckFail
PART3-AUX2(have)-MOD1(sg)	noBase	yesMvt	yesPiedP	noFeatCheckFail
AUX1(be.sg)-AUX2(go)-INF3	yesBase	noMvt	noPiedP	noFeatCheckFail
AUX1(be.sg)-INF3-AUX2(go)	noBase	yesMvt	noPiedP	noFeatCheckFail
AUX2(go)-AUX1(be.sg)-INF3	noBase	noMvt	noPiedP	noFeatCheckFail
AUX2(go)-INF3-AUX1(be.sg)	noBase	yesMvt	noPiedP	noFeatCheckFail
INF3-AUX1(be.sg)-AUX2(go)	noBase	yesMvt	noPiedP	yesFeatCheckFail
INF3-AUX2(go)-AUX1(be.sg)	noBase	yesMvt	noPiedP	noFeatCheckFail
AUX1(have.sg)-MOD2(inf)-INF3	yesBase	noMvt	noPiedP	noFeatCheckFail
AUX1(have.sg)-INF3-MOD2(part)	noBase	yesMvt	noPiedP	noFeatCheckFail
AUX1(have.sg)-INF3-MOD2(inf)	noBase	yesMvt	noPiedP	noFeatCheckFail
MOD2(inf)-INF3-AUX1(have.sg)	noBase	yesMvt	noPiedP	noFeatCheckFail
INF3-AUX1(have.sg)-MOD2(inf)	noBase	yesMvt	noPiedP	yesFeatCheckFail
INF3-AUX1(have.sg)-MOD2(part)	noBase	yesMvt	noPiedP	yesFeatCheckFail
INF3-MOD2(part)-AUX1(have.sg)	noBase	yesMvt	noPiedP	noFeatCheckFail
INF3-MOD2(inf)-AUX1(have.sg)	noBase	yesMvt	noPiedP	noFeatCheckFail

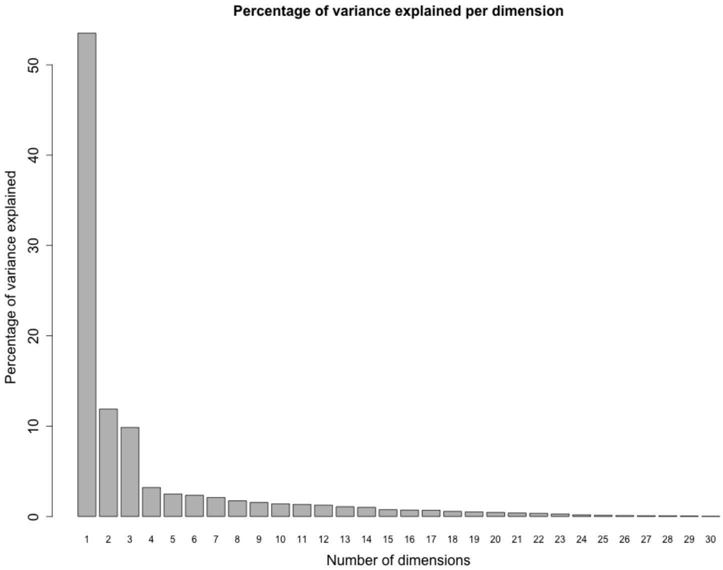
- **in total:** 70 additional variables distilled from the theoretical literature on verb clusters:

- **in total:** 70 additional variables distilled from the theoretical literature on verb clusters:
    - 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:
  - the analyses of Barbiers (2005), Barbiers and Bennis (2010), Abels (2011), Haegeman and Riemsdijk (1986), Bader (2012), and Schmid and Vogel (2004)
  - a head-initial head movement analysis, a head-final head movement analysis, a head-initial XP-movement analysis, a head-final XP-movement analysis (all based on Wurmbrand (2005))

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

- **proposal (I):** the number of microparameters responsible for the verb cluster variation = the number of dimensions we reduce our data set to


Percentage of variance explained per dimension
Percentage of variance explained
0
10
20
30
40
50
Number of dimensions
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

- **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 microparameters

- **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 microparameters
- in order to know what those microparameters are, we need to *interpret* the first three dimensions

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

- **proposal (I):** the number of microparameters responsible for the verb cluster variation = the number of dimensions we reduce our data set to
- **proposal (II):** the identity of those microparameters = 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 (η^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

- is related to the position of infinitives and participles *vis-à-vis* their selecting verbs (modals and auxiliaries respectively)

Dimension 1

- is related to the position of infinitives and participles *vis-à-vis* their selecting verbs (modals and auxiliaries respectively)
- more specifically, the variable InfMod.AuxPart:

Dimension 1

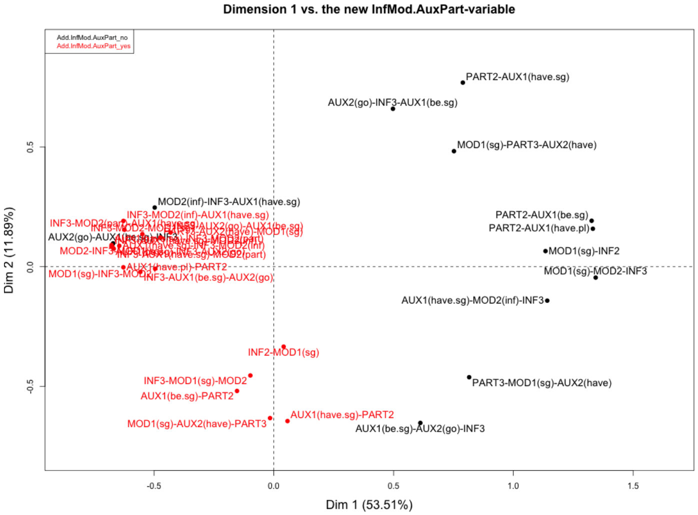
- is related to the position of infinitives and participles *vis-à-vis* their selecting verbs (modals and auxiliaries respectively)
- more specifically, the variable InfMod.AuxPart:
  - set to 'no' when the modal precedes the infinitive (when present) and the participle precedes the auxiliary (when present)

Dimension 1

- is related to the position of infinitives and participles *vis-à-vis* their selecting verbs (modals and auxiliaries respectively)
- 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 position of infinitives and participles *vis-à-vis* their selecting verbs (modals and auxiliaries respectively)
- 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
- this variable has a η^2 of 0.6142


Dimension 1 vs. the new InfMod.AuxPart-variable
Add.InfMod.AuxPart_no
Add.InfMod.AuxPart_yes
PART2-AUX1(have.sg)
AUX2(go)-INF3-AUX1(be.sg)
MOD1(sg)-PART3-AUX2(have)
MOD2(inf)-INF3-AUX1(have.sg)
INF3-MOD2(inf)-AUX1(have.sg)
PART2-AUX1(be.sg)
PART2-AUX1(have.pl)
MOD1(sg)-INF2
MOD1(sg)-MOD2-INF3
AUX1(have.sg)-MOD2(inf)-INF3
AUX1(have.pl)-PART2
INF3-AUX1(be.sg)-AUX2(go)
INF2-MOD1(sg)
INF3-MOD1(sg)-MOD2
AUX1(be.sg)-PART2
PART3-MOD1(sg)-AUX2(have)
AUX1(have.sg)-PART2
MOD1(sg)-AUX2(have)-PART3
AUX1(be.sg)-AUX2(go)-INF3
Dim 1 (53.51%)
Dim 2 (11.89%)
-0.5
0.0
0.5
1.0
1.5

Dimension 2

- is related to the ‘slope’ of the cluster: ascending or descending

Dimension 2

- is related to the ‘slope’ of the cluster: ascending or descending
- more specifically, the variable FINALDESCENT:

Dimension 2

- is related to the 'slope' of the cluster: ascending or descending
- more specifically, the variable FINALDESCENT:
    - set to 'yes' if the cluster ends in a descending order

Dimension 2

- is related to the 'slope' of the cluster: ascending or descending
- 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

Dimension 2

- is related to the 'slope' of the cluster: ascending or descending
- 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

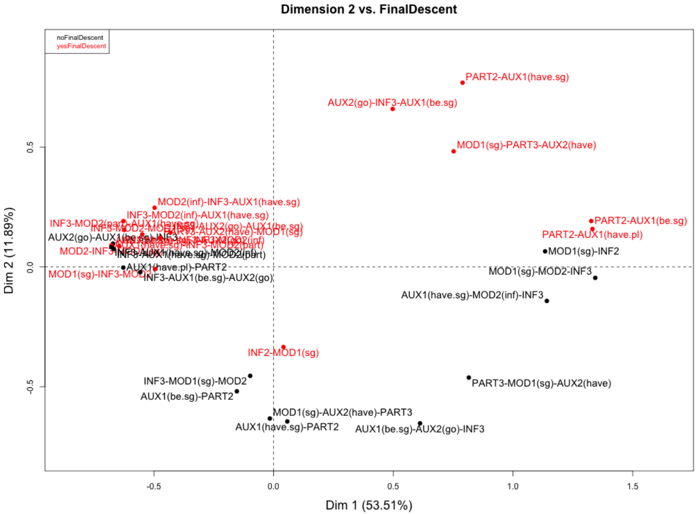
FinalDescent_yes	**FinalDescent_no**
21	12
132	123
321	312
231	213

Dimension 2

- is related to the 'slope' of the cluster: ascending or descending
- 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

FinalDescent_yes	FinalDescent_no
21	12
132	123
321	312
231	213

- this variable has a η^2 of 0.382

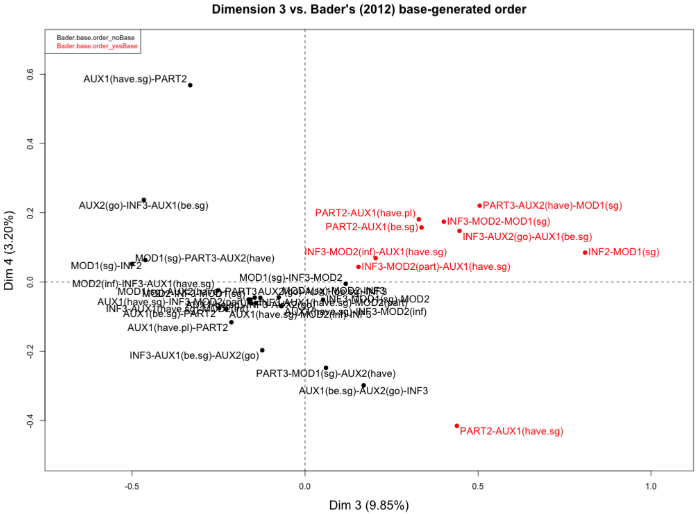

Dimension 2 vs. FinalDescent
noFinalDescent
yesFinalDescent
PART2-AUX1(have.sg)
AUX2(go)-INF3-AUX1(be.sg)
MOD1(sg)-PART3-AUX2(have)
MOD2(inf)-INF3-AUX1(have.sg)
INF3-MOD2(inf)-AUX1(have.sg)
PART2-AUX1(be.sg)
PART2-AUX1(have.pl)
MOD1(sg)-INF2
AUX2(go)-AUX1(be.sg)-INF3
AUX1(have.pl)-PART2
MOD1(sg)-MOD2-INF3
INF3-AUX1(be.sg)-AUX2(go)
AUX1(have.sg)-MOD2(inf)-INF3
INF2-MOD1(sg)
INF3-MOD1(sg)-MOD2
PART3-MOD1(sg)-AUX2(have)
AUX1(be.sg)-PART2
MOD1(sg)-AUX2(have)-PART3
AUX1(have.sg)-PART2
AUX1(be.sg)-AUX2(go)-INF3
Dim 2 (11.89%)
Dim 1 (53.51%)
0.5
0.0
-0.5
1.0
1.5

Dimension 3

- is strongly correlated with head-finality

Dimension 3

- is strongly correlated with head-finality
- a variable like HEADFINALBASEORDER that separates strictly head-final orders from all others has a η^2 of 0.686


Dimension 3 vs. Bader's (2012) base-generated order
Bader.base.order_noBase
Bader.base.order_yesBase
AUX1(have.sg)-PART2
AUX2(go)-INF3-AUX1(be.sg)
MOD1(sg)-PART3-AUX2(have)
MOD1(sg)-INF2
MOD2(inf)-INF3-AUX1(have.sg)
MOD1(sg)-INF3-MOD2
AUX1(have.pl)-PART2
INF3-AUX1(be.sg)-AUX2(go)
PART3-MOD1(sg)-AUX2(have)
AUX1(be.sg)-AUX2(go)-INF3
PART3-AUX2(have)-MOD1(sg)
PART2-AUX1(have.pl)
INF3-MOD2-MOD1(sg)
PART2-AUX1(be.sg)
INF3-AUX2(go)-AUX1(be.sg)
INF3-MOD2(inf)-AUX1(have.sg)
INF2-MOD1(sg)
INF3-MOD2(part)-AUX1(have.sg)
PART2-AUX1(have.sg)
Dim 4 (3.20%)
Dim 3 (9.85%)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.5
0.0
0.5
1.0

A possible analysis

- based on these three parameters, a rough, parametrized analysis of verb clusters can be constructed:

A possible analysis

- based on these three parameters, a rough, parametrized analysis of verb clusters can be constructed:
    1. a head-final base order

A possible analysis

- based on these three parameters, a rough, parametrized analysis of verb clusters can be constructed:
  1. a head-final base order
  2. which dialects can diverge from or not: [±MOVEMENT] (dimension 3)

A possible analysis

- based on these three parameters, a rough, parametrized analysis of verb clusters can be constructed:
  1. a head-final base order
  2. which dialects can diverge from or not: [$\pm$MOVEMENT] (dimension 3)
  3. those that diverge can diverge strongly or not: ECONOMY OF MOVEMENT (dimension 2)

A possible analysis

- based on these three parameters, a rough, parametrized analysis of verb clusters can be constructed:
  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 microparameters 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 microparameters 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 microparameters 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 microparameters 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

Bonus: headedness

- the technique developed here can shed new light on old verb cluster chestnuts such as headedness:

Bonus: headedness

- the technique developed here can shed new light on old verb cluster chestnuts such as headedness:

	dimension 1	**dimension 2**	**dimension 3**
head-initial	0.126	0.309	0.130
head-final	0.006	0.101	0.686
mixed_1 (Barbiers and Bennis (2010))	0.146	0.039	0.193
mixed_2 (Abels (2011))	0.044	0.027	0.014

Bonus: headedness

- the technique developed here can shed new light on old verb cluster chestnuts such as headedness:

	dimension 1	**dimension 2**	**dimension 3**
head-initial	0.126	0.309	0.130
head-final	0.006	0.101	0.686
mixed_1 (Barbiers and Bennis (2010))	0.146	0.039	0.193
mixed_2 (Abels (2011))	0.044	0.027	0.014

- basically, any theoretical proposal that predicts certain data patterns to co-occur can be put to the test with this method

References I

Abels, Klaus. 2011. Hierarchy-order relations in the germanic verb cluster and in the noun phrase. *GAGL* 53:1–28.

Bader, Markus. 2012. Verb-cluster variations: a harmonic grammar analysis. Handout of a talk presented at "New ways of analyzing syntactic variation", November 2012.

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.

Barbiers, Sjef, and Hans Bennis. 2010. De plaats van het werkwoord in zuid en noord. In *Voor Magda. Artikelen voor Magda Devos bij haar afscheid van de Universiteit Gent*, ed. Johan De Caluwe and Jacques Van Keymeulen, 25–42. Gent: Academia.

References II

Haegeman, Liliane, and Henk van Riemsdijk. 1986. Verb projection raising, scope, and the typology of verb movement rules. *Linguistic Inquiry* 17:417–466.

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

Schmid, Tanja, and Ralf Vogel. 2004. Dialectal variation in German 3-Verb clusters. *The Journal of Comparative Germanic Linguistics* 7:235–274.

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.