Quantity and quality in linguistics

(Or: Tackling a microvariationist's frustrations)

Jeroen van Craenenbroeck

CRISSP/KU Leuven
jeroen.vancraenenbroeck@kuleuven.be

CRISSP 10
Brussels, 15 December 2016

Outline
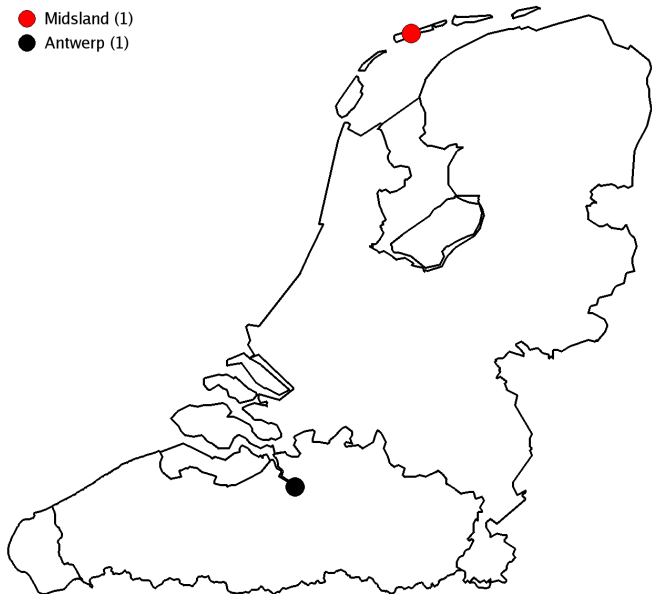
A tale of a village and a city (and then some)

A tale of a village and a city (and then some)

A tale of a village and a city (and then some)

A tale of a village and a city (and then some)

(1) **Midsland**

    a.   Ze weet   niet dat Marie gister   doad gaan is.
         she knows not that Mary yesterday dead gone is
         'She doesn't know Mary died yesterday.'

    b. *Ze weet   niet dat Marie gister   doad is gaan.
         she knows not that Mary yesterday dead is gone

(2) **Antwerp**

    a.   Ge weet nie da   Marie dood is gegaan.
         you know not that Mary dead is gone
         'You don't know Mary died.'

    b. *Ge weet nie da   Marie dood gegaan is.
         you know not that Mary dead gone   is

A tale of a village and a city (and then some)

(3) **Midsland**

    a.   Ik fien dastou   't niet zien mag.
           I find that.you it not see may
           'I feel you shouldn't see it.'

    b. *Ik fien dastou   't niet mag zien.
           I find that.you it not may see

(4) **Antwerp**

    a.   Ik vin dagij   dat nie meug zien.
           I find that.you that not may   see
           'I feel you shouldn't see that.'

    b. *Ik vin dagij   dat nie zien meug.
           I find that.you that not see may

A tale of a village and a city (and then some)

(3) **Midsland**

    a.   Ik fien dastou   't niet zien mag.
          I find that.you it not see may
          'I feel you shouldn't see it.'

    b. *Ik fien dastou   't niet mag zien.
          I find that.you it not may see

(4) **Antwerp**

    a.   Ik vin dagij   dat nie meug zien.
          I find that.you that not may   see
          'I feel you shouldn't see that.'

    b. *Ik vin dagij   dat nie zien meug.
          I find that.you that not see may

- Midsland Dutch is consistently head-final in its (two-verb) clusters, while Antwerp Dutch is consistently head-initial

A tale of a village and a city (and then some)

- let's assume all orders are derived from an underlying (universal) head-initial structure

A tale of a village and a city (and then some)

- let's assume all orders are derived from an underlying (universal) head-initial structure
- orders that are not strictly head-initial are derived via VP-intraposition (Barbiers 2005), implemented via an [EPP]-feature on V
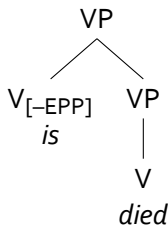
A tale of a village and a city (and then some)

- let's assume all orders are derived from an underlying (universal) head-initial structure
- orders that are not strictly head-initial are derived via VP-intraposition (Barbiers 2005), implemented via an [EPP]-feature on V
- the parameter distinguishing Midsland Dutch from Antwerp Dutch is the feature specification of V: [+EPP] in Midsland, [-EPP] in Antwerp
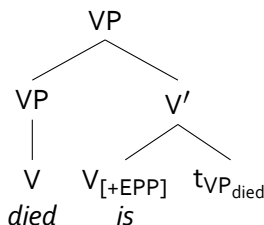
A tale of a village and a city (and then some)

- let's assume all orders are derived from an underlying (universal) head-initial structure
- orders that are not strictly head-initial are derived via VP-intraposition (Barbiers 2005), implemented via an [EPP]-feature on V
- the parameter distinguishing Midsland Dutch from Antwerp Dutch is the feature specification of V: [+EPP] in Midsland, [-EPP] in Antwerp

(5) **Antwerp**

```
        VP
       /  \
V_{[−EPP]}  VP
   is        |
             V
            died
```

(6) **Midsland**

```
          VP
         /  \
       VP    V'
        |   /   \
        V  V_{[+EPP]}  t_{VP_{died}}
      died    is
```

A tale of a village and a city (and then some)

	$V_{[-EPP]}$	$V_{[+EPP]}$
IS_DIED	✓	*
DIED_IS	*	✓
MAY_SEE	✓	*
SEE_MAY	*	✓
DIALECT	Antwerp	Midsland

A tale of a village and a city (and then some)

	$V_{[-EPP]}$	$V_{[+EPP]}$
IS_DIED	✓	*
DIED_IS	*	✓
MAY_SEE	✓	*
SEE_MAY	*	✓
DIALECT	Antwerp	Midsland

however this neat picture breaks down once more dialects are taken into account

A tale of a village and a city (and then some)

	$V_{[-EPP]}$	$V_{[+EPP]}$
IS_DIED	✓	*
DIED_IS	*	✓
MAY_SEE	✓	*
SEE_MAY	*	✓
DIALECT	Antwerp	Midsland

however this neat picture breaks down once more dialects are taken into account

IS_DIED	*
DIED_IS	✓
MAY_SEE	✓
SEE_MAY	*
DIALECT	Lovendegem

A tale of a village and a city (and then some)

	$V_{[-EPP]}$	$V_{[+EPP]}$
IS_DIED	✓	*
DIED_IS	*	✓
MAY_SEE	✓	*
SEE_MAY	*	✓
DIALECT	Antwerp	Midsland

however this neat picture breaks down once more dialects are taken into account

IS_DIED	*	✓
DIED_IS	✓	✓
MAY_SEE	✓	✓
SEE_MAY	*	*
DIALECT	Lovendegem	Hoek

A tale of a village and a city (and then some)

	$V_{[-EPP]}$	$V_{[+EPP]}$
IS_DIED	✓	*
DIED_IS	*	✓
MAY_SEE	✓	*
SEE_MAY	*	✓
DIALECT	Antwerp	Midsland

however this neat picture breaks down once more dialects are taken into account

IS_DIED	*	✓	*
DIED_IS	✓	✓	✓
MAY_SEE	✓	✓	✓
SEE_MAY	*	*	✓
DIALECT	Lovendegem	Hoek	Sliedrecht

A tale of a village and a city (and then some)

	$V_{[-EPP]}$	$V_{[+EPP]}$
IS_DIED	✓	*
DIED_IS	*	✓
MAY_SEE	✓	*
SEE_MAY	*	✓
DIALECT	Antwerp	Midsland

however this neat picture breaks down once more dialects are taken into account

IS_DIED	*	✓	*	✓
DIED_IS	✓	✓	✓	✓
MAY_SEE	✓	✓	✓	*
SEE_MAY	*	*	✓	✓
DIALECT	Lovendegem	Hoek	Sliedrecht	Monster

A tale of a village and a city (and then some)

	$V_{[-EPP]}$	$V_{[+EPP]}$
IS_DIED	✓	*
DIED_IS	*	✓
MAY_SEE	✓	*
SEE_MAY	*	✓
DIALECT	Antwerp	Midsland

however this neat picture breaks down once more dialects are taken into account

IS_DIED	*	✓	*	✓	✓
DIED_IS	✓	✓	✓	✓	✓
MAY_SEE	✓	✓	✓	*	✓
SEE_MAY	*	*	✓	✓	✓
DIALECT	Lovendegem	Hoek	Sliedrecht	Monster	Amsterdam

A tale of a village and a city (and then some)

- ▶ and the picture becomes even more messy in the case of three-verb clusters

A tale of a village and a city (and then some)

- ▶ and the picture becomes even more messy in the case of three-verb clusters

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

A tale of a village and a city (and then some)

- ▶ and the picture becomes even more messy in the case of three-verb clusters

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

MUST_CAN_SWIM
MUST_SWIM_CAN
CAN_MUST_SWIM
CAN_SWIM_MUST
SWIM_MUST_CAN
SWIM_CAN_MUST

A tale of a village and a city (and then some)

- and the picture becomes even more messy in the case of three-verb clusters

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

MUST_CAN_SWIM	✓
MUST_SWIM_CAN	*
CAN_MUST_SWIM	*
CAN_SWIM_MUST	*
SWIM_MUST_CAN	*
SWIM_CAN_MUST	*
DIALECT	Antwerp

A tale of a village and a city (and then some)

- ▶ and the picture becomes even more messy in the case of three-verb clusters

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

MUST_CAN_SWIM	✓	*
MUST_SWIM_CAN	*	*
CAN_MUST_SWIM	*	*
CAN_SWIM_MUST	*	*
SWIM_MUST_CAN	*	*
SWIM_CAN_MUST	*	✓
DIALECT	Antwerp	Lies

A tale of a village and a city (and then some)

- ▶ and the picture becomes even more messy in the case of three-verb clusters

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

MUST_CAN_SWIM	✓	*	*
MUST_SWIM_CAN	*	*	✓
CAN_MUST_SWIM	*	*	*
CAN_SWIM_MUST	*	*	*
SWIM_MUST_CAN	*	*	*
SWIM_CAN_MUST	*	✓	✓
DIALECT	Antwerp	Lies	Midsland

A tale of a village and a city (and then some)

- and the picture becomes even more messy in the case of three-verb clusters

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

MUST_CAN_SWIM	✓	*	*	*
MUST_SWIM_CAN	*	*	✓	*
CAN_MUST_SWIM	*	*	*	*
CAN_SWIM_MUST	*	*	*	*
SWIM_MUST_CAN	*	*	*	✓
SWIM_CAN_MUST	*	✓	✓	✓
DIALECT	Antwerp	Lies	Midsland	Bakkeveen

A tale of a village and a city (and then some)

▶ and the picture becomes even more messy in the case of three-verb clusters

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

MUST_CAN_SWIM	✓
MUST_SWIM_CAN	✓
CAN_MUST_SWIM	*
CAN_SWIM_MUST	*
SWIM_MUST_CAN	*
SWIM_CAN_MUST	✓
DIALECT	Hippolytushoef

A tale of a village and a city (and then some)

- ▶ and the picture becomes even more messy in the case of three-verb clusters

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

MUST_CAN_SWIM	✓	✓
MUST_SWIM_CAN	✓	*
CAN_MUST_SWIM	*	*
CAN_SWIM_MUST	*	*
SWIM_MUST_CAN	*	✓
SWIM_CAN_MUST	✓	*
DIALECT	Hippolytushoef	Langelo

A tale of a village and a city (and then some)

- and the picture becomes even more messy in the case of three-verb clusters

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

MUST_CAN_SWIM	✓	✓	✓
MUST_SWIM_CAN	✓	*	✓
CAN_MUST_SWIM	*	*	*
CAN_SWIM_MUST	*	*	*
SWIM_MUST_CAN	*	✓	✓
SWIM_CAN_MUST	✓	*	✓
DIALECT	Hippolytushoef	Langelo	Beetgum

A tale of a village and a city (and then some)

- and the picture becomes even more messy in the case of three-verb clusters

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

MUST_CAN_SWIM	✓	✓	✓	...
MUST_SWIM_CAN	✓	*	✓	...
CAN_MUST_SWIM	*	*	*	...
CAN_SWIM_MUST	*	*	*	...
SWIM_MUST_CAN	*	✓	✓	...
SWIM_CAN_MUST	✓	*	✓	...
DIALECT	Hippolytushoef	Langelo	Beetgum	...

A tale of a village and a city (and then some)

- in the SAND-project (Barbiers et al. (2006), Barbiers et al. (2008)) we looked at 31 cluster orders in 267 dialects of Dutch

A tale of a village and a city (and then some)

- in the SAND-project (Barbiers et al. (2006), Barbiers et al. (2008)) we looked at 31 cluster orders in 267 dialects of Dutch
- when taken together these data look as follows:

A tale of a village and a city (and then some)
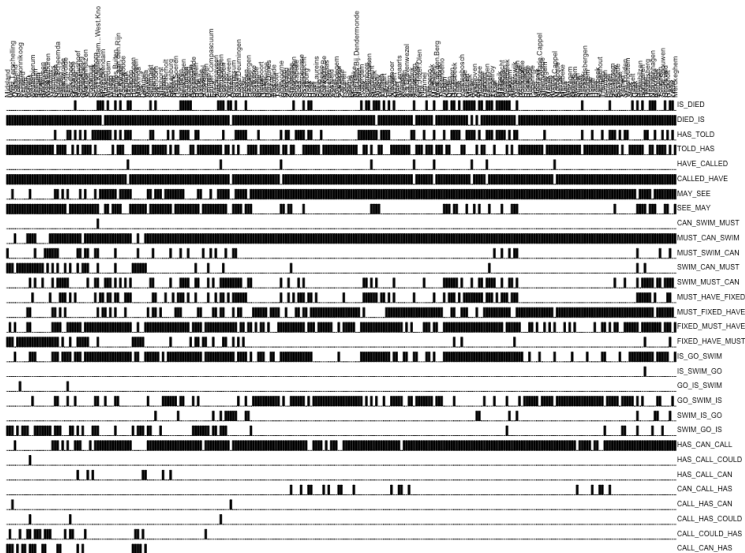
A tale of a village and a city (and then some)

> **The microvariationist's frustration**
> Recent dialect projects have made available large amounts of microvariational data, but our current theoretical tools and methodologies are ill-suited to make sense of those data.

A tale of a village and a city (and then some)

The microvariationist's frustration
Recent dialect projects have made available large amounts of microvariational data, but our current theoretical tools and methodologies are ill-suited to make sense of those data.

- ► Should we just give up?

A tale of a village and a city (and then some)

> **The microvariationist's frustration**
> Recent dialect projects have made available large amounts of microvariational data, but our current theoretical tools and methodologies are ill-suited to make sense of those data.

- **Should we just give up?**
  - **No**: even a cursory glance reveals that there are patterns in the data, that it's not just random chaos.

A tale of a village and a city (and then some)

> **The microvariationist's frustration**
> Recent dialect projects have made available large amounts of microvariational data, but our current theoretical tools and methodologies are ill-suited to make sense of those data.

- **Should we just give up?**
  - **No**: even a cursory glance reveals that there are patterns in the data, that it's not just random chaos.
- e.g. in the case of two-verb clusters, there is no dialect that is obligatorily head-final in modal+infinitive, but obligatorily head-initial in the case of auxiliary+participle

A tale of a village and a city (and then some)

> **The microvariationist's frustration**
> Recent dialect projects have made available large amounts of microvariational data, but our current theoretical tools and methodologies are ill-suited to make sense of those data.

- ▶ **Should we just give up?**
  - ▶ **No**: even a cursory glance reveals that there are patterns in the data, that it's not just random chaos.
- ▶ e.g. in the case of two-verb clusters, there is no dialect that is obligatorily head-final in modal+infinitive, but obligatorily head-initial in the case of auxiliary+participle

IS_DIED	✓
DIED_IS	*
MAY_SEE	*
SEE_MAY	✓
DIALECT	

A tale of a village and a city (and then some)

- ▶ and in three-verb clusters of the type "must can swim", two orders are systematically lacking:

A tale of a village and a city (and then some)

- and in three-verb clusters of the type "must can swim", two orders are systematically lacking:

MUST_CAN_SWIM			
MUST_SWIM_CAN			
CAN_MUST_SWIM	✓	*	✓
CAN_SWIM_MUST	*	✓	✓
SWIM_MUST_CAN			
SWIM_CAN_MUST			
DIALECT			

A tale of a village and a city (and then some)

- and in three-verb clusters of the type "must can swim", two orders are systematically lacking:

MUST_CAN_SWIM			
MUST_SWIM_CAN			
CAN_MUST_SWIM	✓	*	✓
CAN_SWIM_MUST	*	✓	✓
SWIM_MUST_CAN			
SWIM_CAN_MUST			
DIALECT			

- and even a zoomed in version of the Bertin-plot suggests that there is more structure to the data than there seems to be at first glance

A tale of a village and a city (and then some)

A tale of a village and a city (and then some)

- ▶ The field of linguistic dialect research is bifurcated:

A tale of a village and a city (and then some)

- ▶ The field of linguistic dialect research is bifurcated:
  1. **quantitative work** on dialectal variation that is not very theoretical in nature (e.g. Heeringa (2004), Spruit (2008), Heeringa and Nerbonne (2013), Wieling and Nerbonne (2015))

A tale of a village and a city (and then some)

- The field of linguistic dialect research is bifurcated:
  1. **quantitative work** on dialectal variation that is not very theoretical in nature (e.g. Heeringa (2004), Spruit (2008), Heeringa and Nerbonne (2013), Wieling and Nerbonne (2015))
  2. **theoretical work** on dialectal variation that is not very quantitative in nature (e.g. Bayer (1984), Haegeman (1992), Hoekstra (1993), Penner (1994), Poletto (2000), Benincà and Poletto (2004))

A tale of a village and a city (and then some)

- The field of linguistic dialect research is bifurcated:
  1. **quantitative work** on dialectal variation that is not very theoretical in nature (e.g. Heeringa (2004), Spruit (2008), Heeringa and Nerbonne (2013), Wieling and Nerbonne (2015))
  2. **theoretical work** on dialectal variation that is not very quantitative in nature (e.g. Bayer (1984), Haegeman (1992), Hoekstra (1993), Penner (1994), Poletto (2000), Benincà and Poletto (2004))

The goal of the current research project
To bring these two traditions together, and analyze dialectal variation from a perspective that is both quantitative and qualitative in nature.

Outline

Case study #1: Verb clusters

Case study #1: Verb clusters

step #1 **Correspondence Analysis**

Case study #1: Verb clusters

step #1 **Correspondence Analysis**

- ▶ = a technique for exploring and visualizing categorical data, "useful for identification of systematic relationships between variables and capturing the main tendencies" (Levshina (2015:369))

Case study #1: Verb clusters

step #1 **Correspondence Analysis**
- ▶ = a technique for exploring and visualizing categorical data, "useful for identification of systematic relationships between variables and capturing the main tendencies" (Levshina (2015:369))
- ▶ starting point: raw data table

Case study #1: Verb clusters

step #1 **Correspondence Analysis**

- ► = a technique for exploring and visualizing categorical data, "useful for identification of systematic relationships between variables and capturing the main tendencies" (Levshina (2015:369))
- ► starting point: raw data table

	Midsland	Lies	West-Terschelling	…
IS_DIED	no	no	no	…
DIED_IS	yes	yes	yes	…
HAS_TOLD	no	no	no	…
TOLD_HAS	yes	yes	yes	…
MAY_SEE	no	no	yes	…
SEE_MAY	yes	yes	yes	…
CAN_SWIM_MUST	no	no	no	…
MUST_CAN_SWIM	no	no	yes	…
MUST_SWIM_CAN	yes	no	no	…
…	…	…	…	

Case study #1: Verb clusters

step #1 **Correspondence Analysis**
- ▶ which is converted into a distance matrix

Case study #1: Verb clusters

step #1 **Correspondence Analysis**

- ▶ which is converted into a distance matrix

	IS_DIED	DIED_IS	HAS_TOLD	TOLD_HAS
DIED_IS	14.28			
HAS_TOLD	8.24	13.26		
TOLD_HAS	14.42	8.36	15.68	
MAY_SEE	12.60	7.68	11.53	10.72
SEE_MAY	10.77	12.16	11.04	12.08
CAN_SWIM_MUST	8.42	16.03	9.84	13.74
MUST_CAN_SWIM	13.34	5.09	12.32	9.79
MUST_SWIM_CAN	9.27	15.03	10.19	13.26
...	...	...	...	...

Case study #1: Verb clusters

step #1 Correspondence Analysis

- ▶ which is converted into a distance matrix

	IS_DIED	DIED_IS	HAS_TOLD	TOLD_HAS
DIED_IS	14.28			
HAS_TOLD	8.24	13.26		
TOLD_HAS	14.42	8.36	15.68	
MAY_SEE	12.60	7.68	11.53	10.72
SEE_MAY	10.77	12.16	11.04	12.08
CAN_SWIM_MUST	8.42	16.03	9.84	13.74
MUST_CAN_SWIM	13.34	5.09	12.32	9.79
MUST_SWIM_CAN	9.27	15.03	10.19	13.26
...	...	...	...	...

- ▶ the elements in the distance matrix are represented as points in a lower-dimensional space whereby geographical distance between points corresponds (as closely as possible) to distance recorded in the distance matrix

Case study #1: Verb clusters
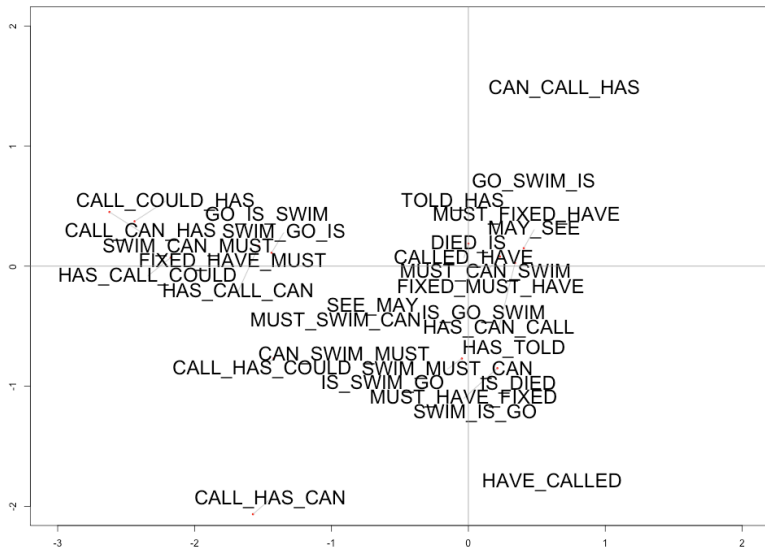
Case study #1: Verb clusters

step #2 **linguistic analyses as supplementary variables**

Case study #1: Verb clusters

step #2 **linguistic analyses as supplementary variables**

- ▶ supplementary variables are additional columns that are added to the data table

Case study #1: Verb clusters

step #2 **linguistic analyses as supplementary variables**

- ▶ supplementary variables are additional columns that are added to the data table
- ▶ they do not contribute to measuring the degree of correspondence between the rows (i.e. cluster orders), but can be used to interpret the data

Case study #1: Verb clusters

step #2 linguistic analyses as supplementary variables

- ▶ supplementary variables are additional columns that are added to the data table
- ▶ they do not contribute to measuring the degree of correspondence between the rows (i.e. cluster orders), but can be used to interpret the data
- ▶ the supplementary variables used in this analysis are decomposed theoretical analyses of verb cluster orders
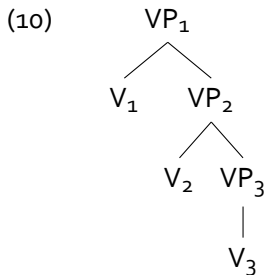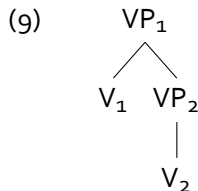
Case study #1: Verb clusters

step #2 **linguistic analyses as supplementary variables**

- ▶ supplementary variables are additional columns that are added to the data table
- ▶ they do not contribute to measuring the degree of correspondence between the rows (i.e. cluster orders), but can be used to interpret the data
- ▶ the supplementary variables used in this analysis are decomposed theoretical analyses of verb cluster orders
- ▶ example: Barbiers (2005)

Case study #1: Verb clusters

- ▶ Barbiers (2005) derives verb cluster orders as follows:

Case study #1: Verb clusters

- Barbiers (2005) derives verb cluster orders as follows:
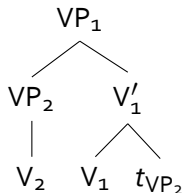   - base order is uniformly head-initial → derives 12 and 123

Case study #1: Verb clusters

- Barbiers (2005) derives verb cluster orders as follows:
  - base order is uniformly head-initial $\rightarrow$ derives 12 and 123

(9)

$$
\begin{array}{c}
VP_1 \\
\diagdown \\
V_1 \quad VP_2 \\
\mid \\
V_2
\end{array}
$$

(10)

$$
\begin{array}{c}
VP_1 \\
\diagdown \\
V_1 \quad VP_2 \\
\diagdown \\
V_2 \quad VP_3 \\
\mid \\
V_3
\end{array}
$$

Case study #1: Verb clusters

- ▶ Barbiers (2005) derives verb cluster orders as follows:
    - ▶ movement is VP-intraposition → derives 21 and 231, 312 and 132, and fails to derive 213
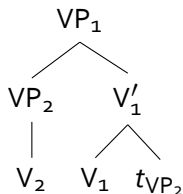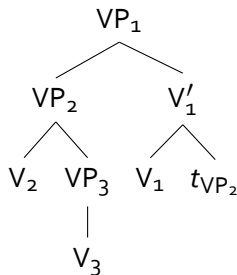
Case study #1: Verb clusters

- Barbiers (2005) derives verb cluster orders as follows:
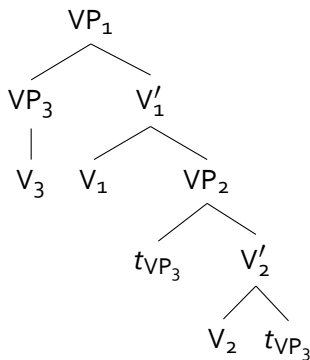   - movement is VP-intraposition → derives 21 and 231, 312 and 132, and fails to derive 213

(11)

$$\begin{array}{c} \text{VP}_1 \\ \diagup \qquad \diagdown \\ \text{VP}_2 \qquad \text{V}'_1 \\ \mid \qquad \diagup \quad \diagdown \\ \text{V}_2 \qquad \text{V}_1 \quad t_{\text{VP}_2} \end{array}$$

Case study #1: Verb clusters

- Barbiers (2005) derives verb cluster orders as follows:
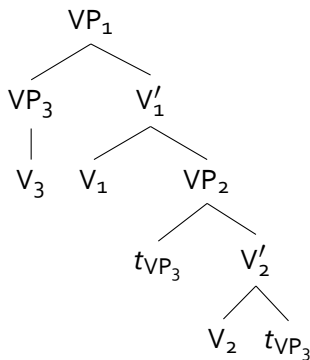 - movement is VP-intraposition → derives 21 and 231, 312 and 132, and fails to derive 213

(11)

$$
\begin{array}{c}
VP_1 \\
\diagdown \\
VP_2 \quad V'_1 \\
| \quad \diagup \diagdown \\
V_2 \quad V_1 \quad t_{VP_2}
\end{array}
$$

(12)

$$
\begin{array}{c}
VP_1 \\
\diagdown \\
VP_2 \quad V'_1 \\
\diagup \diagdown \quad \diagup \diagdown \\
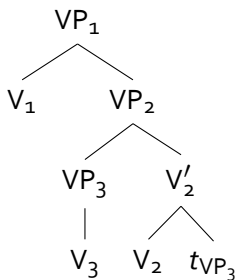V_2 \quad VP_3 \quad V_1 \quad t_{VP_2} \\
| \\
V_3
\end{array}
$$

Case study #1: Verb clusters

- Barbiers (2005) derives verb cluster orders as follows:
  - movement is VP-intraposition $\rightarrow$ derives 21 and 231, 312 and 132, and fails to derive 213

(13)

$$
\begin{array}{c}
\text{VP}_1 \\
\swarrow \qquad \searrow \\
\text{VP}_3 \qquad \text{V}'_1 \\
| \qquad \swarrow \quad \searrow \\
\text{V}_3 \quad \text{V}_1 \qquad \text{VP}_2 \\
\swarrow \qquad \searrow \\
t_{\text{VP}_3} \qquad \text{V}'_2 \\
\swarrow \quad \searrow \\
\text{V}_2 \quad t_{\text{VP}_3}
\end{array}
$$

Case study #1: Verb clusters

- Barbiers (2005) derives verb cluster orders as follows:
  - movement is VP-intraposition → derives 21 and 231, 312 and 132, and fails to derive 213
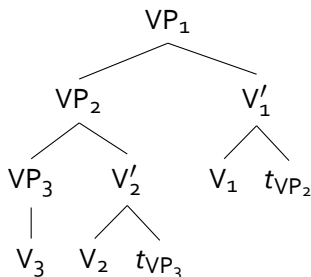

(13)

(14)

Case study #1: Verb clusters

- ▶ Barbiers (2005) derives verb cluster orders as follows:
  - ▶ VP-intraposition can pied-pipe other material → derives 321 (movement of VP3 to specVP1 via specVP2 and with pied-piping of VP2)

Case study #1: Verb clusters

- ▶ Barbiers (2005) derives verb cluster orders as follows:
  - ▶ VP-intraposition can pied-pipe other material $\rightarrow$ derives 321 (movement of VP3 to specVP1 via specVP2 and with pied-piping of VP2)

(15)

$$
\begin{array}{c}
\text{VP}_1 \\
\diagup \qquad \diagdown \\
\text{VP}_2 \qquad \text{V}_1' \\
\diagup \quad \diagdown \qquad \diagup \quad \diagdown \\
\text{VP}_3 \quad \text{V}_2' \qquad \text{V}_1 \quad t_{\text{VP}_2} \\
| \qquad \diagup \quad \diagdown \\
\text{V}_3 \quad \text{V}_2 \quad t_{\text{VP}_3}
\end{array}
$$

Case study #1: Verb clusters

- ▶ from this theoretical account we can distill the following variables:

Case study #1: Verb clusters

- from this theoretical account we can distill the following variables:
  - [BASE-GENERATION]: can the order be base-generated?

Case study #1: Verb clusters

- from this theoretical account we can distill the following variables:
    - [BASE-GENERATION]: can the order be base-generated?
    - [MOVEMENT]: can the order be derived via movement?

Case study #1: Verb clusters

- from this theoretical account we can distill the following variables:
    - [BASE-GENERATION]: can the order be base-generated?
    - [MOVEMENT]: can the order be derived via movement?
    - [PIED-PIPING]: does the derivation involve pied-piping?

Case study #1: Verb clusters

- ▶ from this theoretical account we can distill the following variables:
    - ▶ [BASE-GENERATION]: can the order be base-generated?
    - ▶ [MOVEMENT]: can the order be derived via movement?
    - ▶ [PIED-PIPING]: does the derivation involve pied-piping?

Case study #1: Verb clusters

- ► from this theoretical account we can distill the following variables:
    - ► [BASE-GENERATION]: can the order be base-generated?
    - ► [MOVEMENT]: can the order be derived via movement?
    - ► [PIED-PIPING]: does the derivation involve pied-piping?

	BASE-GENERATION	MOVEMENT	PIED-PIPING	...
IS_DIED	yesBase	noMvt	noPiedP	...
DIED_IS	noBase	yesMvt	noPiedP	...
HAS_TOLD	yesBase	noMvt	noPiedP	...
TOLD_HAS	noBase	yesMvt	noPiedP	...
MAY_SEE	yesBase	noMvt	noPiedP	...
SEE_MAY	noBase	yesMvt	noPiedP	...
CAN_SWIM_MUST	noBase	yesMvt	noPiedP	...
MUST_CAN_SWIM	yesBase	noMvt	noPiedP	...
...	...	...	...	...

Case study #1: Verb clusters

step #3 interpret the CA-results using the linguistic variables

Case study #1: Verb clusters

step #3 **interpret the CA-results using the linguistic variables**
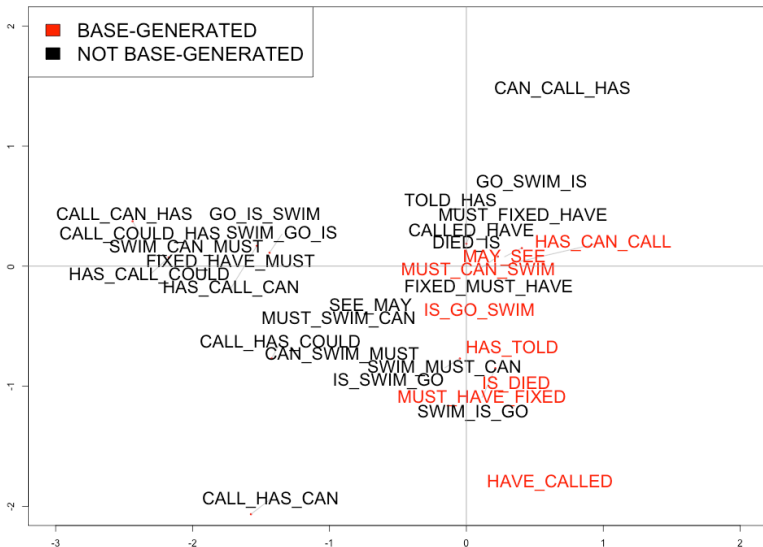
- ▶ the degree of correlation between a supplementary (i.e. linguistic) variable and a dimension of the CA-plot can help to interpret that dimension and hence understand the underlying cause of variation in verb cluster ordering

Case study #1: Verb clusters

step #3 interpret the CA-results using the linguistic variables

- ▶ the degree of correlation between a supplementary (i.e. linguistic) variable and a dimension of the CA-plot can help to interpret that dimension and hence understand the underlying cause of variation in verb cluster ordering
- ▶ there are various ways of measuring/visualizing those correlations:

Case study #1: Verb clusters

step #3 interpret the CA-results using the linguistic variables

- ▶ the degree of correlation between a supplementary (i.e. linguistic) variable and a dimension of the CA-plot can help to interpret that dimension and hence understand the underlying cause of variation in verb cluster ordering

- ▶ there are various ways of measuring/visualizing those correlations:
  - ▶ the plot can be color-coded according to specific variables

Case study #1: Verb clusters

Case study #1: Verb clusters

step #3 interpret the MCA-results using the linguistic variables

- ▶ the degree of correlation between a supplementary (i.e. linguistic) variable and a dimension of the MCA-plot can help to interpret that dimension and hence understand the underlying cause of variation in verb cluster ordering

- ▶ there are various ways of measuring/visualizing those correlations:
  - ▶ the plot can be color-coded according to specific variables
  - ▶ by calculating the squared correlation ratio (η^2):

	dimension 1	dimension 2
Barbiers (2005) base generation	0.159	0.146

Case study #1: Verb clusters

- using this methodology, we can determine which (parts of) analyses of verb clusters provides the best fit for the attested variation

Case study #1: Verb clusters

- ▶ using this methodology, we can determine which (parts of) analyses of verb clusters provides the best fit for the attested variation
- ▶ recent new account of verb cluster variation in Dutch: Barbiers et al. (2016a)

Case study #1: Verb clusters

- ▶ using this methodology, we can determine which (parts of) analyses of verb clusters provides the best fit for the attested variation
- ▶ recent new account of verb cluster variation in Dutch: Barbiers et al. (2016a)
- ▶ Barbiers et al. (2016a) derive verb cluster orders as follows:

Case study #1: Verb clusters

- ▶ using this methodology, we can determine which (parts of) analyses of verb clusters provides the best fit for the attested variation
- ▶ recent new account of verb cluster variation in Dutch: Barbiers et al. (2016a)
- ▶ Barbiers et al. (2016a) derive verb cluster orders as follows:
  1. there are two possible base orders: stricly ascending (12, 123) and strictly descending (21, 321)

Case study #1: Verb clusters

- ▶ using this methodology, we can determine which (parts of) analyses of verb clusters provides the best fit for the attested variation
- ▶ recent new account of verb cluster variation in Dutch: Barbiers et al. (2016a)
- ▶ Barbiers et al. (2016a) derive verb cluster orders as follows:
  1. there are two possible base orders: stricly ascending (12, 123) and strictly descending (21, 321)
  2. participles can be adjectivized or not: if they are, they precede the verb cluster (and hence their selecting verb): $PART_2$-AUX_1, $PART_3$-MOD_1-AUX_2, $INF_{IPP.2}$-INF_3-AUX_1, and (ambiguously) $PART_3$-AUX_2-MOD_1

Case study #1: Verb clusters

- ▶ using this methodology, we can determine which (parts of) analyses of verb clusters provides the best fit for the attested variation
- ▶ recent new account of verb cluster variation in Dutch: Barbiers et al. (2016a)
- ▶ Barbiers et al. (2016a) derive verb cluster orders as follows:
  1. there are two possible base orders: stricly ascending (12, 123) and strictly descending (21, 321)
  2. participles can be adjectivized or not: if they are, they precede the verb cluster (and hence their selecting verb): $PART_2$-AUX_1, $PART_3$-MOD_1-AUX_2, $INF_{IPP.2}$-INF_3-AUX_1, and (ambiguously) $PART_3$-AUX_2-MOD_1
  3. infinitives can be nominalized or not: if they are, they precede the verb cluster (and hence their selecting verb): INF_2-MOD_1, INF_3-MOD_1-MOD_2, and (ambiguously) INF_3-MOD_2-MOD_1

Case study #1: Verb clusters

- using this methodology, we can determine which (parts of) analyses of verb clusters provides the best fit for the attested variation
- recent new account of verb cluster variation in Dutch: Barbiers et al. (2016a)
- Barbiers et al. (2016a) derive verb cluster orders as follows:
  1. there are two possible base orders: stricly ascending (12, 123) and strictly descending (21, 321)
  2. participles can be adjectivized or not: if they are, they precede the verb cluster (and hence their selecting verb): $PART_2$-AUX_1, $PART_3$-MOD_1-AUX_2, $INF_{IPP.2}$-INF_3-AUX_1, and (ambiguously) $PART_3$-AUX_2-MOD_1
  3. infinitives can be nominalized or not: if they are, they precede the verb cluster (and hence their selecting verb): INF_2-MOD_1, INF_3-MOD_1-MOD_2, and (ambiguously) INF_3-MOD_2-MOD_1
  4. dialects do/do not allow for interruption of the cluster by non-verbal material (requires an adjectival participle or a nominal infinitive) $\rightarrow$ yields the order 132

Case study #1: Verb clusters

- ▶ as before, these linguistic properties can be coded as supplementary variables in the analysis:

Case study #1: Verb clusters

- as before, these linguistic properties can be coded as supplementary variables in the analysis:
  - [BASE123]: is the order compatible with an ascending base order?

Case study #1: Verb clusters

- as before, these linguistic properties can be coded as supplementary variables in the analysis:
  - [BASE123]: is the order compatible with an ascending base order?
  - [BASE321]: is the order compatible with a descending base order?

Case study #1: Verb clusters

- as before, these linguistic properties can be coded as supplementary variables in the analysis:
  - [BASE123]: is the order compatible with an ascending base order?
  - [BASE321]: is the order compatible with a descending base order?
  - [ADJPART]: does the order involve an adjectivized participle?

Case study #1: Verb clusters

- as before, these linguistic properties can be coded as supplementary variables in the analysis:
    - [BASE123]: is the order compatible with an ascending base order?
    - [BASE321]: is the order compatible with a descending base order?
    - [ADJPART]: does the order involve an adjectivized participle?
    - [NOMINF]: does the order involve a nominalized infinitive?

Case study #1: Verb clusters

- as before, these linguistic properties can be coded as supplementary variables in the analysis:
  - [BASE123]: is the order compatible with an ascending base order?
  - [BASE321]: is the order compatible with a descending base order?
  - [ADJPART]: does the order involve an adjectivized participle?
  - [NOMINF]: does the order involve a nominalized infinitive?
  - [CLUSTINTERR]: does the order involve cluster interruption?

Case study #1: Verb clusters

- as before, these linguistic properties can be coded as supplementary variables in the analysis:
    - [BASE123]: is the order compatible with an ascending base order?
    - [BASE321]: is the order compatible with a descending base order?
    - [ADJPART]: does the order involve an adjectivized participle?
    - [NOMINF]: does the order involve a nominalized infinitive?
    - [CLUSTINTERR]: does the order involve cluster interruption?

- this analysis turns out to line up very nicely with the CA-based analysis of the data set:
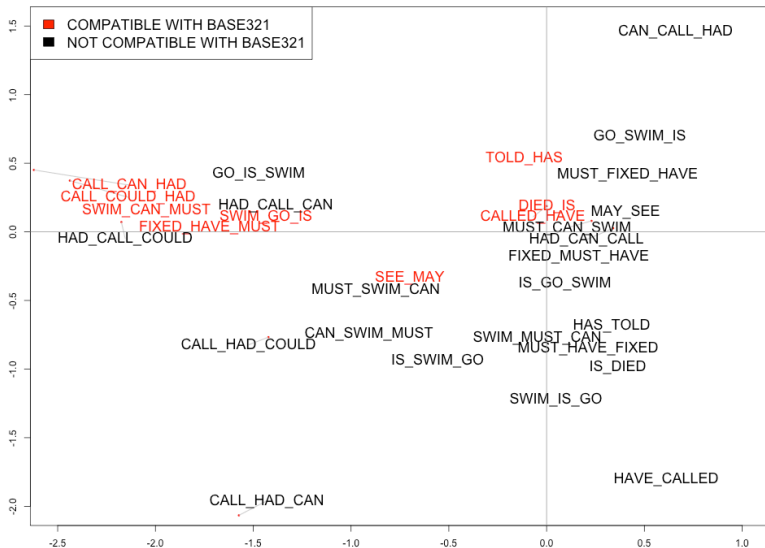
Case study #1: Verb clusters

- ► as before, these linguistic properties can be coded as supplementary variables in the analysis:
    - ► [BASE123]: is the order compatible with an ascending base order?
    - ► [BASE321]: is the order compatible with a descending base order?
    - ► [ADJPART]: does the order involve an adjectivized participle?
    - ► [NOMINF]: does the order involve a nominalized infinitive?
    - ► [CLUSTINTERR]: does the order involve cluster interruption?

- ► this analysis turns out to line up very nicely with the CA-based analysis of the data set:

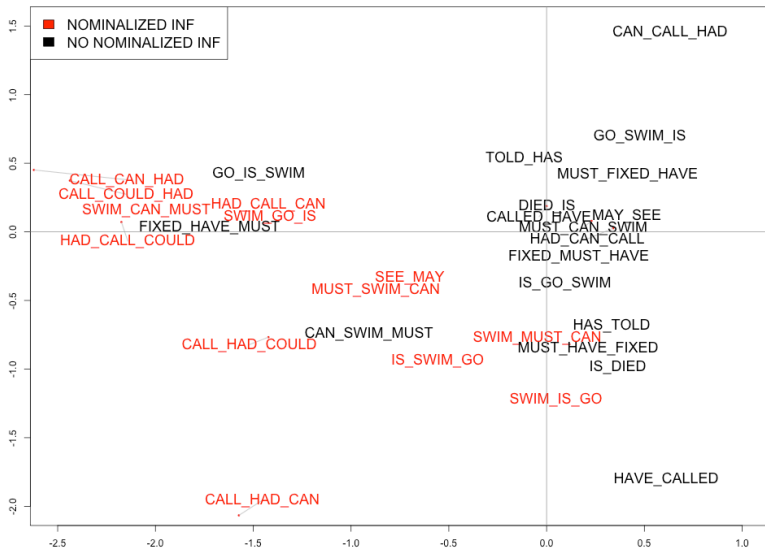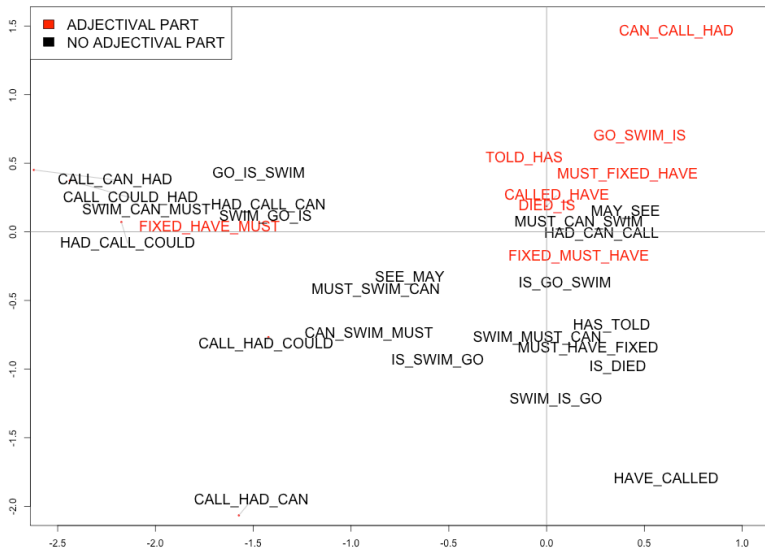| η^2	dimension #1	dimension #2
BASE123	0.706	0.009
BASE321	0.312	0.096
ADJPART	0.007	0.321
NOMINF	0.454	0.073
CLUSTINTERR	0.003	0.028

Case study #1: Verb clusters
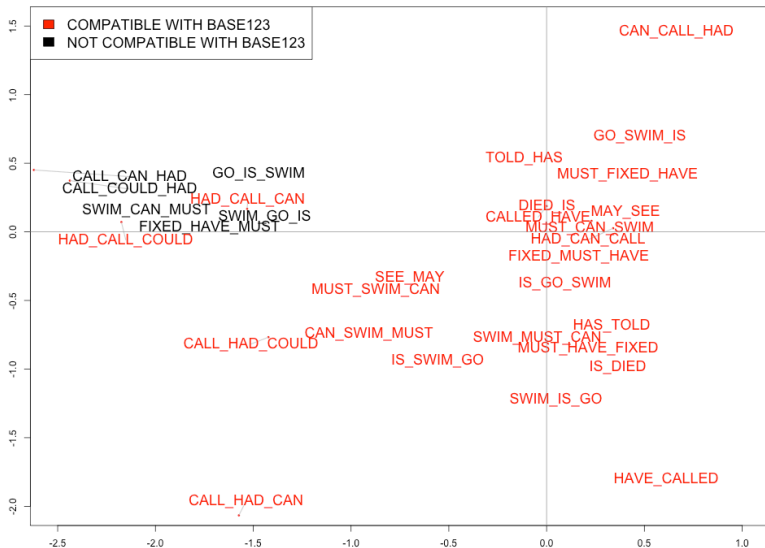
Case study #1: Verb clusters

Case study #1: Verb clusters

Case study #1: Verb clusters

Case study #1: Conclusion

- by integrating a quantitative-statistical methodology with formal-theoretical analyses, we can (a) make sense of a large and varied dataset, while at the same time (b) going beyond mere number crunching

Case study #1: Conclusion

- by integrating a quantitative-statistical methodology with formal-theoretical analyses, we can (a) make sense of a large and varied dataset, while at the same time (b) going beyond mere number crunching
- clearly, many questions remain, e.g.:

Case study #1: Conclusion

- ▶ by integrating a quantitative-statistical methodology with formal-theoretical analyses, we can (a) make sense of a large and varied dataset, while at the same time (b) going beyond mere number crunching
- ▶ clearly, many questions remain, e.g.:
  - ▶ one would like to see more evidence for the adjectival/nominal status of preverbal participles/infinitives

Case study #1: Conclusion

- by integrating a quantitative-statistical methodology with formal-theoretical analyses, we can (a) make sense of a large and varied dataset, while at the same time (b) going beyond mere number crunching
- clearly, many questions remain, e.g.:
    - one would like to see more evidence for the adjectival/nominal status of preverbal participles/infinitives
    - what about the cluster interruption parameter of Barbiers et al. (2016a)?

Case study #1: Conclusion

- by integrating a quantitative-statistical methodology with formal-theoretical analyses, we can (a) make sense of a large and varied dataset, while at the same time (b) going beyond mere number crunching
- clearly, many questions remain, e.g.:
    - one would like to see more evidence for the adjectival/nominal status of preverbal participles/infinitives
    - what about the cluster interruption parameter of Barbiers et al. (2016a)?
    - the implicational relation between pre-auxiliary participles and pre-modal infinitives discussed above doesn't yet follow from Barbiers et al. (2016a)'s analysis

Case study #2: Microvariation in C and D

Case study #2: Microvariation in C and D

note: this entire subsection is based on joint work with Marjo van Koppen

Case study #2: Microvariation in C and D

note: this entire subsection is based on joint work with Marjo van Koppen

- **starting point:** ten dialect phenomena in 267 dialects of Dutch that have roughly the same geographical distribution

Case study #2: Microvariation in C and D

note: this entire subsection is based on joint work with Marjo van Koppen

- **starting point:** ten dialect phenomena in 267 dialects of Dutch that have roughly the same geographical distribution
- **complementizer agreement (CA)**

Case study #2: Microvariation in C and D

note: this entire subsection is based on joint work with Marjo van Koppen

- ▶ **starting point:** ten dialect phenomena in 267 dialects of Dutch that have roughly the same geographical distribution

- ▶ **complementizer agreement (CA)**

(16) **O-n** Bart en Lieske in t paradijs levn
if-PL Bart and Lieske in the paradise live
'If Bart and Lieske are living in paradise, ...' (Gistel)

Case study #2: Microvariation in C and D

note: this entire subsection is based on joint work with Marjo van Koppen

- **starting point:** ten dialect phenomena in 267 dialects of Dutch that have roughly the same geographical distribution

- **complementizer agreement (CA)**

  (16)   **O-n** Bart en   Lieske in t   paradijs levn
         if-PL Bart and Lieske in the paradise live
         'If Bart and Lieske are living in paradise, ...'   (Gistel)

- **clitic doubling (CD)**

Case study #2: Microvariation in C and D

note: this entire subsection is based on joint work with Marjo van Koppen

- **starting point:** ten dialect phenomena in 267 dialects of Dutch that have roughly the same geographical distribution

- **complementizer agreement (CA)**

    (16)   **O-n** Bart en Lieske in t   paradijs levn
           if-PL Bart and Lieske in the paradise live
           'If Bart and Lieske are living in paradise, ...'   (Gistel)

- **clitic doubling (CD)**

    (17)   da-**ze**   **zaaile**   lachen.
           that-they$_{CLITIC}$ they$_{STRONG}$ laugh
           'that they are laughing.'   (Wambeek)

- **short *do* replies (SDR)**

Case study #2: Microvariation in C and D

- **short *do* replies (SDR)**

(18)   A: IJ zal nie komen. B: **IJ doet**.
         he will not come   he does
       'A: He won't come. B: Yes, he will.'   (Berlare)

Case study #2: Microvariation in C and D

- **short *do* replies (SDR)**

(18) A: IJ zal nie komen. B: **IJ doet**.
         he will not come   he does
      'A: He won't come. B: Yes, he will.'   (Berlare)

- **negative clitic (NEG)**

Case study #2: Microvariation in C and D

- **short *do* replies (SDR)**

  (18)   A: IJ   zal   nie   komen. B: **IJ   doet**.
           he will not come   he does
           'A: He won't come. B: Yes, he will.'   (Berlare)

- **negative clitic (NEG)**

  (19)   K **en**   goa nie noar schole.
           I   NEG go   not to   school
           'I'm not going to school.'   (Tielt)

Case study #2: Microvariation in C and D

- **short *do* replies (SDR)**

  (18)   A: IJ zal nie komen. B: **IJ doet**.
           he will not come   he does
          'A: He won't come. B: Yes, he will.'   (Berlare)

- **negative clitic (NEG)**

  (19)   K **en** goa nie noar schole.
           I NEG go not to   school
          'I'm not going to school.'   (Tielt)

- **clitics on *yes* and *no* (CYN)**

Case study #2: Microvariation in C and D

- **short *do* replies (SDR)**

  (18) A: IJ zal nie komen. B: **IJ doet**.
       he will not come   he does
       'A: He won't come. B: Yes, he will.'   (Berlare)

- **negative clitic (NEG)**

  (19) K **en** goa nie noar schole.
       I NEG go not to school
       'I'm not going to school.'   (Tielt)

- **clitics on *yes* and *no* (CYN)**

  (20) A: Wilde nog koffie, Jan? B: **Ja-k**.
       want.you PART coffee Jan   Yes-I
       'A: Do you want some more coffee, Jan? B: Yes.'
       (Malderen)

Case study #2: Microvariation in C and D

- *t* 'it' as *there*-expletive (EXPL-T)

Case study #2: Microvariation in C and D

- ▶ **_t_ 'it' as _there_-expletive (EXPL-T)**

(21) **T** en goa niemand nie dansn.
      it NEG goes no.one   not dance
      'There will be no dancing.'   (Brugge)

Case study #2: Microvariation in C and D

- ***t* 'it' as *there*-expletive (EXPL-T)**

(21) **T** en goa niemand nie dansn.
      it NEG goes no.one not dance
      'There will be no dancing.'   (Brugge)

- ***if* as a comparative complementizer (CMPR-IF)**

Case study #2: Microvariation in C and D

- ▶ *t* 'it' as *there*-expletive (EXPL-T)

   (21) **T** en goa niemand nie dansn.
         it NEG goes no.one not dance
         'There will be no dancing.'   (Brugge)

- ▶ *if* as a comparative complementizer (CMPR-IF)

   (22) Zie peist daj   eer   ga thuis zijn **of** ik.
         she thinks that.you sooner go home be if I
         'She thinks you'll be home sooner than me.' (Oostkerke)

Case study #2: Microvariation in C and D

- **_t_ 'it' as _there_-expletive (EXPL-T)**

  (21) **T** en   goa niemand nie dansn.
         it NEG goes no.one   not dance
         'There will be no dancing.'   (Brugge)

- **_if_ as a comparative complementizer (CMPR-IF)**

  (22) Zie peist daj   eer   ga thuis zijn **of** ik.
         she thinks that.you sooner go home be   if I
         'She thinks you'll be home sooner than me.' (Oostkerke)

- **the obligatory use of expletive _there_ in embedded clauses and inverted main clauses (ER.OBL)**

Case study #2: Microvariation in C and D

- ▶ **_t_ 'it' as _there_-expletive (EXPL-T)**

  (21) **T** en goa niemand nie dansn.
        it NEG goes no.one not dance
        'There will be no dancing.' (Brugge)

- ▶ **_if_ as a comparative complementizer (CMPR-IF)**

  (22) Zie peist daj eer ga thuis zijn **of** ik.
        she thinks that.you sooner go home be if I
        'She thinks you'll be home sooner than me.' (Oostkerke)

- ▶ **the obligatory use of expletive _there_ in embedded clauses
  and inverted main clauses (ER.OBL)**

  (23) dat ***(er)** in de fabrieke nen jongen werkte
        that there in the factory a boy worked
        'that a boy worked in the factory (Lapscheure,
        Haegeman (1986:3))

- ► determiner+demonstrative in NP-ellipsis (THE+THAT)

Case study #2: Microvariation in C and D

▶ **determiner+demonstrative in NP-ellipsis (THE+THAT)**

(24)   **De die** zou k ik wiln op eetn.
the those would I_{CLITIC} I_{STRONG} want up eat
'I would like to eat those.'   (Merelbeke)

- **determiner+demonstrative in NP-ellipsis (THE+THAT)**

  (24)   **De die** zou k   ik   wiln op eetn.
         the those would I$_{\text{CLITIC}}$ I$_{\text{STRONG}}$ want up eat
         'I would like to eat those.'   (Merelbeke)

- ***go get* in imperatives (GO-GET)**

Case study #2: Microvariation in C and D

- **determiner+demonstrative in NP-ellipsis (THE+THAT)**

  (24) **De die** zou k ik wiln op eetn.
       the those would I_{CLITIC} I_{STRONG} want up eat
       'I would like to eat those.'   (Merelbeke)

- ***go get*** **in imperatives (GO-GET)**

  (25) **Gon haalt** die bestelling ne keer!
       go_{INF} get_{IMP} that order a time
       'Go get that order!'   (Ghent)

Case study #2: Microvariation in C and D

Case study #2: Microvariation in C and D

- ▶ first technique used: Correspondence Analysis

Case study #2: Microvariation in C and D

- first technique used: Correspondence Analysis
- raw data table:

Case study #2: Microvariation in C and D

- first technique used: Correspondence Analysis
- raw data table:

	Brugge	Hulst	Dirksland	Ossendrecht	...
CA	1	1	1	0	...
CD	1	1	0	1	...
SDR	0	0	0	0	...
NEG	1	0	0	0	...
CYN	1	1	0	0	...
EXPL-T	1	0	0	0	...
CMPR-IF	0	1	0	0	...
ER.OBL	1	0	0	0	...
THE+THAT	1	0	0	1	...
GO-GET	1	0	0	1	...

- which is converted into a distance matrix:

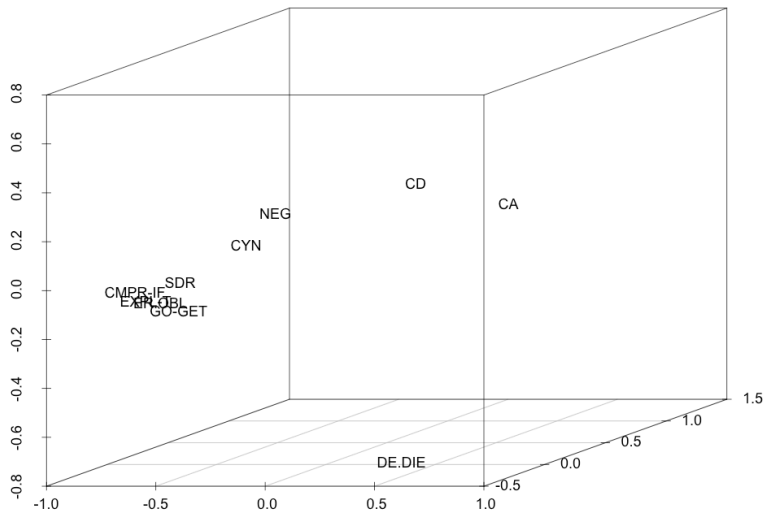# Case study #2: Microvariation in C and D

- which is converted into a distance matrix:

	CA	CD	SDR	CYN	NEG	EXPL-T	CMPR-IF	...
CD	11.40							
SDR	10.14	7.28						
CYN	10.00	6.48	4.58					
NEG	10.63	6.08	4.69	5.56				
EXPL-T	10.04	8.30	4.24	5.56	6.16			
CMPR-IF	10.72	8.54	4.69	5.91	6.63	4.47		
THE+THAT	10.77	5.83	6.70	6.63	6.40	7.68	8.06	
ER.OBL	10.34	8.06	4.24	5.38	6.00	4.00	4.69	...
GO-GET	10.72	8.30	4.89	5.91	6.32	5.29	5.09	...

Case study #2: Microvariation in C and D

- ▶ which is converted into a distance matrix:

	CA	CD	SDR	CYN	NEG	EXPL-T	CMPR-IF	...
CD	11.40							
SDR	10.14	7.28						
CYN	10.00	6.48	4.58					
NEG	10.63	6.08	4.69	5.56				
EXPL-T	10.04	8.30	4.24	5.56	6.16			
CMPR-IF	10.72	8.54	4.69	5.91	6.63	4.47		
THE+THAT	10.77	5.83	6.70	6.63	6.40	7.68	8.06	
ER.OBL	10.34	8.06	4.24	5.38	6.00	4.00	4.69	...
GO-GET	10.72	8.30	4.89	5.91	6.32	5.29	5.09	...

- ▶ which is in turn converted into a lower-dimensional (here: 3D) representation

Case study #2: Microvariation in C and D

- **First dimension:** sets apart complementizer agreement (CA) from all other phenomena:

Case study #2: Microvariation in C and D

- **First dimension:** sets apart complementizer agreement (CA) from all other phenomena:
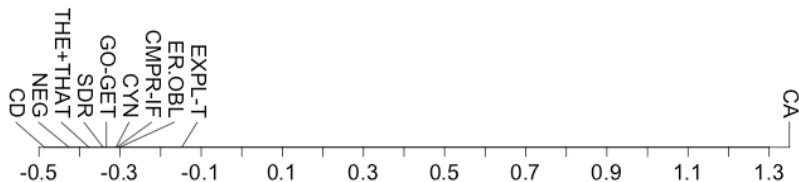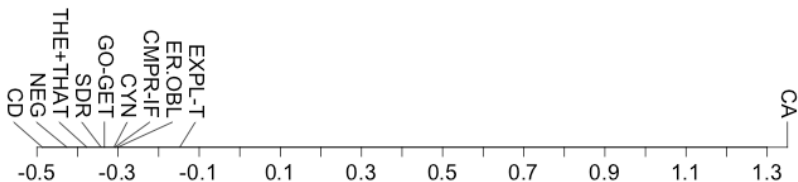
Case study #2: Microvariation in C and D

▶ **First dimension:** sets apart complementizer agreement (CA) from all other phenomena:

▶ following a.o. Haegeman and Koppen (2012), van Koppen (2005) we analyze CA as involving a phi-agreement probe on C and propose the following parameter:
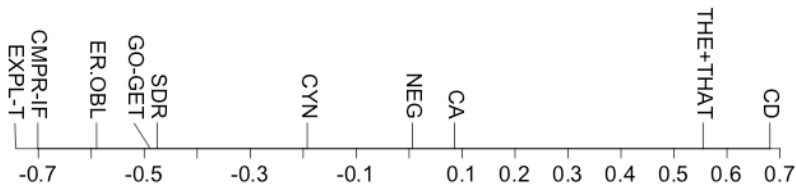
Case study #2: Microvariation in C and D

- **First dimension:** sets apart complementizer agreement (CA) from all other phenomena:

- following a.o. Haegeman and Koppen (2012), van Koppen (2005) we analyze CA as involving a phi-agreement probe on C and propose the following parameter:

(26) **the AgrC-parameter:**
C {does/does not} have unvalued ϕ-features.

- **Second dimension:** sets apart CD and THE+THAT from all other phenomena (with CA no longer playing a role):

Case study #2: Microvariation in C and D

- **Second dimension:** sets apart CD and THE+THAT from all other phenomena (with CA no longer playing a role):

Case study #2: Microvariation in C and D

- **Second dimension:** sets apart CD and THE+THAT from all other phenomena (with CA no longer playing a role):
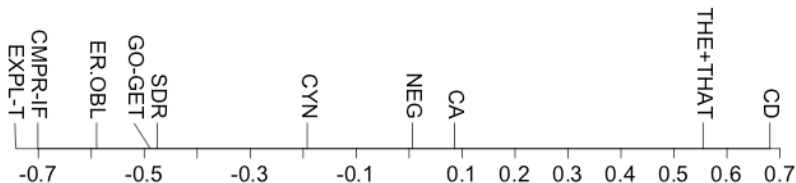
- we propose that CD and THE+THAT are indicative of a **split D-domain**, while EXPL-T, CMPR-IF, ER.OBL, GO-GET, SDR, CYN, and NEG are indicative of a **split C-domain**
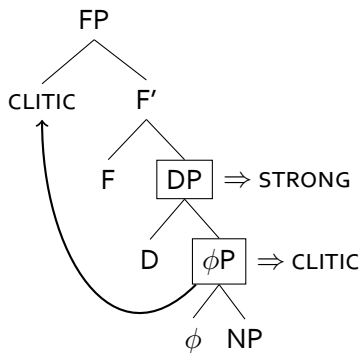
Case study #2: Microvariation in C and D

- ▶ both CD and THE+THAT involve raising of a subpart of DP into the extended left periphery of D (van Craenenbroeck and van Koppen 2008, Barbiers et al. 2016b):
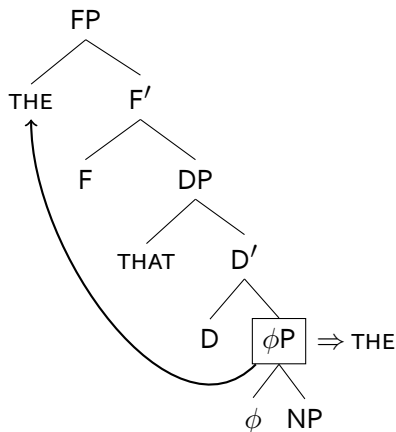
Case study #2: Microvariation in C and D

- both CD and THE+THAT involve raising of a subpart of DP into the extended left periphery of D (van Craenenbroeck and van Koppen 2008, Barbiers et al. 2016b):

(27) (28)

Case study #2: Microvariation in C and D

- ▶ SDR, NEG, and CYN involve the activation of a high left-peripheral Pol(arity)-head (van Craenenbroeck 2010)

Case study #2: Microvariation in C and D

- ► SDR, NEG, and CYN involve the activation of a high left-peripheral Pol(arity)-head (van Craenenbroeck 2010)
- ► EXPL-T spells out a left-peripheral C-head (van Craenenbroeck 2011)

Case study #2: Microvariation in C and D

- ▶ SDR, NEG, and CYN involve the activation of a high left-peripheral Pol(arity)-head (van Craenenbroeck 2010)
- ▶ EXPL-T spells out a left-peripheral C-head (van Craenenbroeck 2011)
- ▶ ER.OBL: the expletive *er* 'there' can only be elided when C is sufficiently close to the subject, i.e. when the C-domain is not split

Case study #2: Microvariation in C and D

- ▶ SDR, NEG, and CYN involve the activation of a high left-peripheral Pol(arity)-head (van Craenenbroeck 2010)
- ▶ EXPL-T spells out a left-peripheral C-head (van Craenenbroeck 2011)
- ▶ ER.OBL: the expletive *er* 'there' can only be elided when C is sufficiently close to the subject, i.e. when the C-domain is not split
- ▶ GO-GET: *go* in imperatives spells out a high left-peripheral functional projection (cf. also McCloskey (1997:214))

Case study #2: Microvariation in C and D

- SDR, NEG, and CYN involve the activation of a high left-peripheral Pol(arity)-head (van Craenenbroeck 2010)

- EXPL-T spells out a left-peripheral C-head (van Craenenbroeck 2011)

- ER.OBL: the expletive *er* 'there' can only be elided when C is sufficiently close to the subject, i.e. when the C-domain is not split

- GO-GET: *go* in imperatives spells out a high left-peripheral functional projection (cf. also McCloskey (1997:214))

- CMPR-IF: dialects with CMPR-IF differ from dialects without CMPR-IF in that they have an unique form for the conditional complementizer → CMPR-IF dialects have two separate C-layers to express conditional and comparative information, whereas the other dialects bundle both features on one single head

Case study #2: Microvariation in C and D

(29) **the D-parameter:**
DP {does/does not} have an extended left periphery.

(30) **the C-parameter**
CP {does/does not} have an extended left periphery.

Case study #2: Microvariation in C and D

> (29) **the D-parameter:**
> DP {does/does not} have an extended left periphery.

> (30) **the C-parameter**
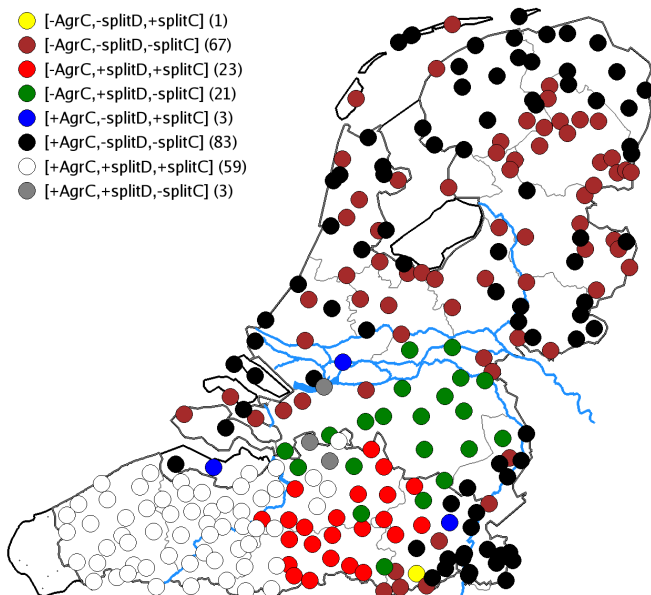> CP {does/does not} have an extended left periphery.

$\rightarrow$ we can now examine the interaction between the three parameters we proposed

Case study #2: Microvariation in C and D

Legend:
- [-AgrC,-splitD,+splitC] (1) — yellow
- [-AgrC,-splitD,-splitC] (67) — dark red
- [-AgrC,+splitD,+splitC] (23) — red
- [-AgrC,+splitD,-splitC] (21) — green
- [+AgrC,-splitD,+splitC] (3) — blue
- [+AgrC,-splitD,-splitC] (83) — black
- [+AgrC,+splitD,+splitC] (59) — white
- [+AgrC,+splitD,-splitC] (3) — grey

Case study #2: Microvariation in C and D

+AgrC		
	+split C	−split C
+split D	East & West Flanders (N=59)	Nieuwmoer, Sint Lenaarts, Moerdijk (N=3)
−split D	Opglabbeek, Sliedrecht, Hoek (N=3)	Holland, Limburg, Friesland, Groningen (N=83)

−AgrC		
	+split C	−split C
+split D	Flemish Brabant & Antwerp (N=23)	North Brabant (N=21)
−split D	Borgloon (N=1)	Drenthe, Utrecht (N=67)

Digging a little deeper:

Case study #2: Microvariation in C and D

Digging a little deeper:

1. To what extent are these accidental geographical patterns?

Case study #2: Microvariation in C and D

Digging a little deeper:

1. To what extent are these accidental geographical patterns?
   - people who live close together tend to speak alike, and different from people who live further away

Case study #2: Microvariation in C and D

Digging a little deeper:

1. To what extent are these accidental geographical patterns?
   - people who live close together tend to speak alike, and different from people who live further away
   - cf. Nerbonne and Kleiweg (2007)'s **Fundamental Dialectological Postulate**: geographically proximate varieties tend to be more similar (linguistically) than distant ones

Case study #2: Microvariation in C and D

Digging a little deeper:

1. To what extent are these accidental geographical patterns?
   - people who live close together tend to speak alike, and different from people who live further away
   - cf. Nerbonne and Kleiweg (2007)'s **Fundamental Dialectological Postulate**: geographically proximate varieties tend to be more similar (linguistically) than distant ones
   - we can now quantify exactly how much of the variation is purely due to geographical distance

Case study #2: Microvariation in C and D

- ▶ we start from an actual (geographical) distance matrix:

▶ we start from an actual (geographical) distance matrix:

	Midsland	Lies	W. Terschelling	Oosterend
Lies	4.49			
W. Terschelling	8.15	12.65		
Oosterend	56.35	60.60	48.77	
Hollum	39.71	35.25	47.82	93.54
Schiermonnikoog	97.27	92.85	105.33	148.91
Ferwerd	59.98	55.80	67.68	109.13
…	…	…	…	…

- we start from an actual (geographical) distance matrix:

	Midsland	**Lies**	**W. Terschelling**	**Oosterend**
Lies	4.49			
W. Terschelling	8.15	12.65		
Oosterend	56.35	60.60	48.77	
Hollum	39.71	35.25	47.82	93.54
Schiermonnikoog	97.27	92.85	105.33	148.91
Ferwerd	59.98	55.80	67.68	109.13
...	...	...	...	...

- and measure the correlation between geographical distances and linguistic distances (using a Mantel test)

Case study #2: Microvariation in C and D

- we start from an actual (geographical) distance matrix:

	Midsland	Lies	W. Terschelling	Oosterend
Lies	4.49			
W. Terschelling	8.15	12.65		
Oosterend	56.35	60.60	48.77	
Hollum	39.71	35.25	47.82	93.54
Schiermonnikoog	97.27	92.85	105.33	148.91
Ferwerd	59.98	55.80	67.68	109.13
...	...	...	...	...

- and measure the correlation between geographical distances and linguistic distances (using a Mantel test)

- r = 0.315

Case study #2: Microvariation in C and D

- we start from an actual (geographical) distance matrix:

	Midsland	Lies	W. Terschelling	Oosterend
Lies	4.49			
W. Terschelling	8.15	12.65		
Oosterend	56.35	60.60	48.77	
Hollum	39.71	35.25	47.82	93.54
Schiermonnikoog	97.27	92.85	105.33	148.91
Ferwerd	59.98	55.80	67.68	109.13
...	...	...	...	...

- and measure the correlation between geographical distances and linguistic distances (using a Mantel test)
- r = 0.315
- i.e. there is only a fairly weak correlation between linguistic distance and geographical distance in our data set

Digging a little deeper:

2. What is the relation between the 6 phenomena that are indicative of a split C-domain?

Case study #2: Microvariation in C and D

Digging a little deeper:

2. What is the relation between the 6 phenomena that are indicative of a split C-domain?

   ▸ we can explore the internal relations between these phenomena using association rule data mining (Spruit 2008, Piatetsky-Shapiro 1991)

Case study #2: Microvariation in C and D
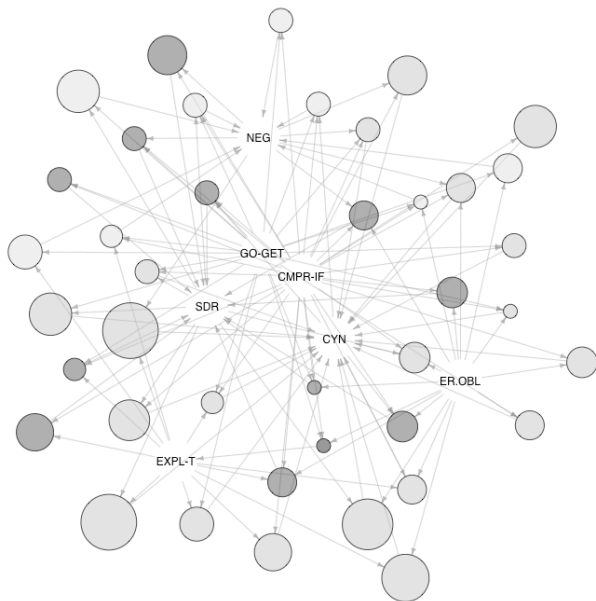
Case study #2: Microvariation in C and D

- all arrows point towards either CYN, NEG, or SDR, i.e. to the polarity-related phenomena

Case study #2: Microvariation in C and D

- all arrows point towards either CYN, NEG, or SDR, i.e. to the polarity-related phenomena
- **IF** a dialect has one or more of the split C-phenomena, **THEN** it also always has one of the polarity-related phenomena

Case study #2: Microvariation in C and D

- all arrows point towards either CYN, NEG, or SDR, i.e. to the polarity-related phenomena
- **IF** a dialect has one or more of the split C-phenomena, **THEN** it also always has one of the polarity-related phenomena
- we take this to mean that polarity acts as a cue for the language learner that she is acquiring a split C-dialect

Outline

Conclusion & future prospects

- combining quantitative (statistical) and qualitative (formal-theoretical) methods provides a way of tackling the microvariationist's frustration, in that it makes possible a theoretical analysis of large and highly varied dialect data sets

Conclusion & future prospects

- combining quantitative (statistical) and qualitative (formal-theoretical) methods provides a way of tackling the microvariationist's frustration, in that it makes possible a theoretical analysis of large and highly varied dialect data sets
- future prospects:

Conclusion & future prospects

- combining quantitative (statistical) and qualitative (formal-theoretical) methods provides a way of tackling the microvariationist's frustration, in that it makes possible a theoretical analysis of large and highly varied dialect data sets
- future prospects:
    - move from micro- to meso- and macrovariation: Dutch dialects vs. Swiss German dialects, Germanic dialects vs. Romance dialects, dialect variation vs. language variation (WALS)

Conclusion & future prospects

- combining quantitative (statistical) and qualitative (formal-theoretical) methods provides a way of tackling the microvariationist's frustration, in that it makes possible a theoretical analysis of large and highly varied dialect data sets
- future prospects:
  - move from micro- to meso- and macrovariation: Dutch dialects vs. Swiss German dialects, Germanic dialects vs. Romance dialects, dialect variation vs. language variation (WALS)
  - find more ways of translating/adapting statistical methods (in)to theoretically relevant notions

References I

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. Amsterdam: John Benjamins.

Barbiers, Sjef, Johan van der Auwera, Hans Bennis, Eefje Boef, Gunther De Vogelaer, and Margreet van der Ham. 2008. *Syntactische atlas van de Nederlandse dialecten. Deel II*. Amsterdam: Amsterdam University Press.

Barbiers, Sjef, Hans Bennis, and Lotte Hendriks. 2016a. Merging verb cluster variation. Ms. Meertens Institute.

Barbiers, Sjef, Marjo van Koppen, Hans Bennis, and Norbert Corver. 2016b. Microcomparative MOrphosyntactic REsearch (MIMORE): Mapping partial grammars of Flemish, Brabantish and Dutch. *Lingua* 178:5–31.

Barbiers, Sjef, et al. 2006. *Dynamische syntactische atlas van de Nederlandse dialecten (dynasand)*. Meertens Institute. www.meertens.knaw.nl/sand/.

Bayer, Josef. 1984. COMP in Bavarian syntax. *The Linguistic Review* 3:209–274.

References II

Benincà, Paola, and Cecilia Poletto. 2004. Topic, focus, and V2: Defining the CP sublayers. In *The structure of CP and IP*, ed. Luigi Rizzi, 52–75. Oxford: Oxford University Press.

van Craenenbroeck, Jeroen. 2010. *The syntax of ellipsis. Evidence from Dutch dialects*. New York: OUP.

van Craenenbroeck, Jeroen. 2011. Germanic expletives revisited. In pursuit of Kayne's dream. Handout for an invited talk at the 26th Comparative Germanic Syntax Workshop.

van Craenenbroeck, Jeroen, and Marjo van Koppen. 2008. Pronominal doubling in Dutch dialects: big DPs and coordinations. In *Microvariation in syntactic doubling.*, ed. Sjef Barbiers, Olaf Koeneman, Marika Lekakou, and Margreet van der Ham, volume 36 of *Syntax and Semantics*, 207–249. Bingley: Emerald.

Haegeman, Liliane. 1986. *Er*-sentences in West-Flemish. Ms. Université de Genève.

Haegeman, Liliane. 1992. *Theory and description in generative syntax*. Cambridge: Cambridge University Press.

Haegeman, Liliane, and Marjo van Koppen. 2012. Complementizer agreement and the relation between T and C. *Linguistic Inquiry* 43:441–454.

Heeringa, Wilbert. 2004. Measuring dialect pronunciation differences using Levenshtein distance. Doctoral Dissertation, University of Groningen.

References III

Heeringa, Wilbert, and John Nerbonne. 2013. Dialectometry. In *Language and Space. An International Handbook of Linguistic Variation. Volume 3: Dutch*, ed. Frans Hinskens and Johan Taeldeman, volume 30 of *Handbooks of Linguistics and Communication Science*, 624–645. Berlin/Boston: De Gruyter.

Hoekstra, Jarich. 1993. Dialectal variation inside CP as parametric variation. In *Dialektsyntax*, ed. Werner Abraham and Josef Bayer, volume 5 of *Linguistische Berichte/Sonderheft*, 161–179. Opladen: Westdeutscher Verlag.

van Koppen, Marjo. 2005. One probe, two goals: aspects of agreement in dutch dialects. Doctoral Dissertation, Universiteit Leiden.

Levshina, Natalia. 2015. *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam: John Benjamins.

McCloskey, James. 1997. Subjecthood and subject positions. In *Elements of grammar*, ed. Liliane Haegeman, 197–235. Dordrecht: Kluwer Academic Publishers.

Nerbonne, John, and Peter Kleiweg. 2007. Toward a dialectological yardstick. *Journal of Quantitative Linguistics* 14:148–166.

References IV

Penner, Zvi. 1994. Asking questions without CPs? On the acquisition of root wh-questions in Bernese Swiss German and Standard German. In *Language acquisition studies in generative grammar*, ed. Teun Hoekstra and Bonnie D. Schwartz, 177–214. Amsterdam: John Benjamins Publishing Company.

Piatetsky-Shapiro, G. 1991. Discovery, analysis, and presentation of strong rules. In *Knowledge discovery in databases*, ed. G. Piatetsky-Shapiro and W. Frawley, 229–248. Cambridge, Massachusetts: AAAI/The MIT Press.

Poletto, Cecilia. 2000. *The higher functional field: Evidence from Northern Italian dialects*. Oxford University Press.

Spruit, Marco René. 2008. Quantitative perspectives on syntactic variation in Dutch dialects. Doctoral Dissertation, Universiteit van Amsterdam.

Wieling, Martijn, and John Nerbonne. 2015. Advances in dialectometry. *Annual Review of Linguistics* 1:243–264.