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)

Two case studies
  Case study #1: Verb clusters
  Case study #2: Microvariation in C and D

Conclusion & future prospects
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
(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

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

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td><strong>DIALECT</strong></td>
<td><strong>Antwerp</strong></td>
<td><strong>Midsland</strong></td>
</tr>
</tbody>
</table>
A tale of a village and a city (and then some)

<table>
<thead>
<tr>
<th></th>
<th>$V_{-EPP}$</th>
<th>$V_{+EPP}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>✓</td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>DIALECT</td>
<td>Antwerp</td>
<td>Midsland</td>
</tr>
</tbody>
</table>

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

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>*</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
</tr>
<tr>
<td>DIALECT</td>
<td>Lovendegem</td>
</tr>
</tbody>
</table>
A tale of a village and a city (and then some)

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>✓</td>
</tr>
</tbody>
</table>

| DIALECT | Antwerp | Midsland |

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

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

| DIALECT | Lovendegem | Hoek |

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

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

| DIALECT | Lovendegem | Hoek |

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

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>DIALECT</td>
<td>Antwerp</td>
<td>Midsland</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>*</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>DIALECT</td>
<td>Lovendegem</td>
<td>Hoek</td>
<td>Sliedrecht</td>
</tr>
</tbody>
</table>
A tale of a village and a city (and then some)

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>DIALECT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Antwerp</td>
<td>Midsland</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
<th>$V_{[+EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>*</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DIALECT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lovendegem</td>
<td>Hoek</td>
<td>Sliedrecht</td>
<td>Monster</td>
</tr>
</tbody>
</table>
A tale of a village and a city (and then some)

<table>
<thead>
<tr>
<th></th>
<th>$V_{[-EPP]}$</th>
<th>$V_{[+EPP]}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td><strong>DIALECT</strong></td>
<td><strong>Antwerp</strong></td>
<td><strong>Midsland</strong></td>
</tr>
</tbody>
</table>

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

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>*</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>*</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>DIALECT</strong></td>
<td><strong>Lovendegem</strong></td>
<td><strong>Hoek</strong></td>
<td><strong>Sliedrecht</strong></td>
<td><strong>Monster</strong></td>
<td><strong>Amsterdam</strong></td>
</tr>
</tbody>
</table>
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.’

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST_CAN_SWIM</td>
<td>✓</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>*</td>
</tr>
<tr>
<td>CAN_MUST_SWIM</td>
<td>*</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_MUST_CAN</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_CAN_MUST</td>
<td>*</td>
</tr>
<tr>
<td>DIALECT</td>
<td>Antwerp</td>
</tr>
</tbody>
</table>
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.’

<table>
<thead>
<tr>
<th>DIALECT</th>
<th>Antwerp</th>
<th>Lies</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST_CAN_SWIM</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CAN_MUST_SWIM</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_MUST_CAN</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_CAN_MUST</td>
<td>* ✓</td>
<td>*</td>
</tr>
</tbody>
</table>
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.

\[ \text{I find that everyone must can swim.} \]

\[ \text{‘I think everyone should be able to swim.’} \]

<table>
<thead>
<tr>
<th>DIALECT</th>
<th>Antwerp</th>
<th>Lies</th>
<th>Midsland</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST_CAN_SWIM</td>
<td>✓</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>*</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>CAN.Must_SWIM</td>
<td>*</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>SWIM_MUST_CAN</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>SWIM_CAN.Must</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
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.’

<table>
<thead>
<tr>
<th>DIALECT</th>
<th>Antwerp</th>
<th>Lies</th>
<th>Midsland</th>
<th>Bakkeveen</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST_CAN_SWIM</td>
<td>✓</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>*</td>
<td>*</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>CAN_MUST_SWIM</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_MUST_CAN</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>SWIM_CAN_MUST</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
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.’

<table>
<thead>
<tr>
<th>DIALECT</th>
<th>Hippolytushoef</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST_CAN_SWIM</td>
<td>✓</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>✓</td>
</tr>
<tr>
<td>CAN_MUST_SWIM</td>
<td>*</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_MUST_CAN</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_CAN_MUST</td>
<td>✓</td>
</tr>
</tbody>
</table>
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.’

<table>
<thead>
<tr>
<th>DIALECT</th>
<th>Hippolytushoef</th>
<th>Langelo</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST_CAN_SWIM</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>✓</td>
<td>*</td>
</tr>
<tr>
<td>CAN_MUST_SWIM</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_MUST_CAN</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>SWIM_CAN_MUST</td>
<td>✓</td>
<td>*</td>
</tr>
</tbody>
</table>
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.’

<table>
<thead>
<tr>
<th>DIALECT</th>
<th>Hippolytushoef</th>
<th>Langelo</th>
<th>Beetgum</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST_CAN_SWIM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>CAN_MUST_SWIM</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_MUST_CAN</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SWIM_CAN_MUST</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
</tr>
</tbody>
</table>
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.’

<table>
<thead>
<tr>
<th>MUST_CAN_SWIM</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST_SWIM_CAN</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>CAN_MUST_SWIM</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>SWIM_MUST_CAN</td>
<td>*</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SWIM_CAN_MUST</td>
<td>✓</td>
<td>*</td>
<td>✓</td>
</tr>
</tbody>
</table>

| 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)
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
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

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>✓</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>*</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>*</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>✓</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>DIALECT</th>
<th>MUST_CAN_SWIM</th>
<th>MUST_SWIM_CAN</th>
<th>CAN_MUST_SWIM</th>
<th>CAN_SWIM_MUST</th>
<th>SWIM_MUST_CAN</th>
<th>SWIM_CAN_MUST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th></th>
<th>MUST_CAN_SWIM</th>
<th>MUST_SWIM_CAN</th>
<th>CAN_MUST_SWIM</th>
<th>CAN_SWIM_MUST</th>
<th>SWIM_MUST_CAN</th>
<th>SWIM_CAN_MUST</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIALECT</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

- 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))
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

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

Two case studies
  Case study #1: Verb clusters
  Case study #2: Microvariation in C and D

Conclusion & future prospects
Case study #1: Verb clusters

Step #1

Correspondence Analysis

▶ Correspondence Analysis is 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

<table>
<thead>
<tr>
<th></th>
<th>IS_DIED</th>
<th>no</th>
<th>no</th>
<th>no</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIED_IS</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>…</td>
</tr>
<tr>
<td>HAS_TOLD</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>…</td>
</tr>
<tr>
<td>TOLD_HAS</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>…</td>
</tr>
<tr>
<td>MAY_SEE</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>…</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>…</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>…</td>
</tr>
<tr>
<td>MUST_CAN_SWIM</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>…</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>…</td>
</tr>
</tbody>
</table>

…”
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

<table>
<thead>
<tr>
<th></th>
<th>Midsland</th>
<th>Lies</th>
<th>West-Terschelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>HAS_TOLD</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>TOLD_HAS</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>MUST_CAN_SWIM</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Case study #1: Verb clusters

step #1  Correspondence Analysis

Α which is converted into a distance matrix
Case study #1: Verb clusters

**step #1  Correspondence Analysis**

- which is converted into a distance matrix

<table>
<thead>
<tr>
<th></th>
<th>IS_DIED</th>
<th>DIED_IS</th>
<th>HAS_TOLD</th>
<th>TOLD_HAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIED_IS</td>
<td>14.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAS_TOLD</td>
<td>8.24</td>
<td>13.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOLD_HAS</td>
<td>14.42</td>
<td>8.36</td>
<td>15.68</td>
<td></td>
</tr>
<tr>
<td>MAYSEE</td>
<td>12.60</td>
<td>7.68</td>
<td>11.53</td>
<td>10.72</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>10.77</td>
<td>12.16</td>
<td>11.04</td>
<td>12.08</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>8.42</td>
<td>16.03</td>
<td>9.84</td>
<td>13.74</td>
</tr>
<tr>
<td>MUST_CAN_SWIM</td>
<td>13.34</td>
<td>5.09</td>
<td>12.32</td>
<td>9.79</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>9.27</td>
<td>15.03</td>
<td>10.19</td>
<td>13.26</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Case study #1: Verb clusters

**step #1** Correspondence Analysis

- which is converted into a distance matrix

<table>
<thead>
<tr>
<th></th>
<th>IS_DIED</th>
<th>DIED_IS</th>
<th>HAS_TOLD</th>
<th>TOLD_HAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIED_IS</td>
<td>14.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAS_TOLD</td>
<td>8.24</td>
<td>13.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOLD_HAS</td>
<td>14.42</td>
<td>8.36</td>
<td>15.68</td>
<td></td>
</tr>
<tr>
<td>MAYSEE</td>
<td>12.60</td>
<td>7.68</td>
<td>11.53</td>
<td>10.72</td>
</tr>
<tr>
<td>SEEMAY</td>
<td>10.77</td>
<td>12.16</td>
<td>11.04</td>
<td>12.08</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>8.42</td>
<td>16.03</td>
<td>9.84</td>
<td>13.74</td>
</tr>
<tr>
<td>MUST_CAN_SWIM</td>
<td>13.34</td>
<td>5.09</td>
<td>12.32</td>
<td>9.79</td>
</tr>
<tr>
<td>MUST_SWIM_CAN</td>
<td>9.27</td>
<td>15.03</td>
<td>10.19</td>
<td>13.26</td>
</tr>
</tbody>
</table>

- 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

example: Barbiers (2005)
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 → derives 12 and 123
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
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

(11) \[
\begin{array}{c}
\text{VP}_1 \\
\text{VP}_2 \quad \text{V}' \quad \text{V}_1 \\
\quad \text{V}_2 \\
\end{array}
\]
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) \[
\begin{array}{c}
\text{VP}_1 \\
\text{VP}_2 \\
V_2 \\
\text{V}_1 \\
t_{\text{VP}_2} \\
\end{array}
\]

(12) \[
\begin{array}{c}
\text{VP}_1 \\
\text{VP}_2 \\
V_2 \\
\text{VP}_3 \\
V_2 \\
\text{V}_1 \\
t_{\text{VP}_2} \\
\text{V}_3 \\
\end{array}
\]
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

(13)
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:
  - 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 → derives 321 (movement of VP3 to specVP1 via specVP2 and with pied-piping of VP2)

(15)

```
     VP1
    /   \
   VP2   V'1
  / \    /   \
VP3  V2 V1   tVP2
   |   |    |
   V3  V2  tVP3
```
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?

<table>
<thead>
<tr>
<th>BASE-GENERATION</th>
<th>MOVEMENT</th>
<th>PIED-PIPING</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS_DIED</td>
<td>yesBase</td>
<td>noMvt</td>
</tr>
<tr>
<td>DIED_IS</td>
<td>noBase</td>
<td>yesMvt</td>
</tr>
<tr>
<td>HAS_TOLD</td>
<td>yesBase</td>
<td>noMvt</td>
</tr>
<tr>
<td>TOLD_HAS</td>
<td>noBase</td>
<td>yesMvt</td>
</tr>
<tr>
<td>MAYSEE</td>
<td>yesBase</td>
<td>noMvt</td>
</tr>
<tr>
<td>SEE_MAY</td>
<td>noBase</td>
<td>yesMvt</td>
</tr>
<tr>
<td>CAN_SWIM_MUST</td>
<td>noBase</td>
<td>yesMvt</td>
</tr>
<tr>
<td>MUST_CAN_SWIM</td>
<td>yesBase</td>
<td>noMvt</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
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 ($\eta^2$):

<table>
<thead>
<tr>
<th></th>
<th>dimension 1</th>
<th>dimension 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbiers (2005) base generation</td>
<td>0.159</td>
<td>0.146</td>
</tr>
</tbody>
</table>
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

Barbiers et al. (2016a) derive verb cluster orders as follows:

1. there are two possible base orders: strictly 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): $P\_\text{ART}-\text{AUX}_1$, $P\_\text{ART}_2-\text{MOD}_1-\text{AUX}_2$, $\text{INF}_{2/3}-\text{INF}$, and (ambiguously) $P\_\text{ART}_3-\text{AUX}_2-\text{MOD}_1$

3. infinitives can be nominalized or not: if they are, they precede the verb cluster (and hence their selecting verb): $\text{INF}_2-\text{MOD}_1$, $\text{INF}_3-\text{MOD}_1-\text{MOD}_2$, and (ambiguously) $\text{INF}_3-\text{MOD}_2-\text{MOD}_1$

4. dialects do/do not allow for interruption of the cluster by non-verbal material (requires an adjectival participle or a nominal infinitive) yields the order 132
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: strictly 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: strictly 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): \( \text{PART}_2\text{-AUX}_1, \text{PART}_3\text{-MOD}_1\text{-AUX}_2, \text{INF}_{\text{IPP},2}\text{-INF}_3\text{-AUX}_1 \), and (ambiguously) \( \text{PART}_3\text{-AUX}_2\text{-MOD}_1 \)
  3. infinitives can be nominalized or not: if they are, they precede the verb cluster (and hence their selecting verb): \( \text{INF}_2\text{-MOD}_1, \text{INF}_3\text{-MOD}_1\text{-MOD}_2 \), and (ambiguously) \( \text{INF}_3\text{-MOD}_2\text{-MOD}_1 \)
  4. dialects do/do not allow for interruption of the cluster by non-verbal material (requires an adjectival participle or a nominal infinitive) yields the order 132
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: strictly 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: strictly 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): \( \text{PART}_2-\text{AUX}_1, \text{PART}_3-\text{MOD}_1-\text{AUX}_2, \text{INF}_{\text{IPP}.2}-\text{INF}_3-\text{AUX}_1 \), and (ambiguously) \( \text{PART}_3-\text{AUX}_2-\text{MOD}_1 \)
  3. infinitives can be nominalized or not: if they are, they precede the verb cluster (and hence their selecting verb): \( \text{INF}_2-\text{MOD}_1, \text{INF}_3-\text{MOD}_1-\text{MOD}_2 \), and (ambiguously) \( \text{INF}_3-\text{MOD}_2-\text{MOD}_1 \)
  4. dialects do/do not allow for interruption of the cluster by non-verbal material (requires an adjectival participle or a nominal infinitive) → 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?

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:
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:

<table>
<thead>
<tr>
<th></th>
<th>dimension #1</th>
<th>dimension #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASE123</td>
<td>0.706</td>
<td>0.009</td>
</tr>
<tr>
<td>BASE321</td>
<td>0.312</td>
<td>0.096</td>
</tr>
<tr>
<td>ADJPART</td>
<td>0.007</td>
<td>0.321</td>
</tr>
<tr>
<td>NOMINF</td>
<td>0.454</td>
<td>0.073</td>
</tr>
<tr>
<td>CLUSTINTERR</td>
<td>0.003</td>
<td>0.028</td>
</tr>
</tbody>
</table>
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

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) 
If Bart and Lieske are living in paradise, … (Gistel)

Clitic doubling (CD)

(17) 
that they are laughing. (Wambeek)
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 leven
    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 zaaille lachen. that-they_{CLITIC} they_{STRONG} laugh ‘that they are laughing.’ (Wambeek)
Case study #2: Microvariation in C and D

- short do replies (SDR)

  A: He won't come. B: Y es, he will.' (Berlare)

- negative clitic (NEG)

  'I'm not going to school.' (Tielt)

- clitics on yes and no (CYN)

  A: Do you want some more coffee, Jan? B: Y es.' (Malderen)
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)

    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)

    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)

   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)

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

- *it* as *there*-expletive (EXPL-T)

There will be no dancing. (Brugge)

She thinks you'll be home sooner than me. (Oostkerke)

The obligatory use of *there* in embedded clauses and inverted main clauses (ER.OBL)

*that a boy worked in the factory (Lapscheure, Haegeman (1986:3))
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

- \textit{t ‘it’} as \textit{there}-expletive (EXPL-T)

\quad (21) \quad T \text{ en } goa \text{ niemand nie dansn.} \\
\quad \text{‘There will be no dancing.’} \\
\quad \text{(Brugge)}

- \textit{if} as a comparative complementizer (CMPR-IF)

\quad (22) \quad Zie peist \text{ dat } \text{ eer } \text{ ga thuis zijn of ik.} \\
\quad \text{‘She thinks you’ll be home sooner than me.’} \\
\quad \text{(Oostkerke)}
Case study #2: Microvariation in C and D

- *t ‘it’ as there-expletive (EXPL-T)*

  
  (21)  
  \[T \text{ en } \text{ goa niemand nie dansn.} \]
  \("it \text{ NEG goes no. one not dance}\"
  \(\text{‘There will be no dancing.’}\)  
  (Brugge)

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

  
  (22)  
  \[\text{Zie peist daj eer ga thuis zijn of ik}.\]
  \("\text{she thinks that you sooner go home be if I}\"
  \(\text{‘She thinks you’ll be home sooner than me.’}\)  
  (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))
Case study #2: Microvariation in C and D

- determiner+demonstrative in NP-ellipsis (THE+THAT)
Case study #2: Microvariation in C and D

- determiner+demontstrative in NP-ellipsis (THE+THAT)

(24) **De die zou ik wiln op eetn.**

the those would I_{CLITIC} I_{STRONG} want up eat

‘I would like to eat those.’

(Merelbeke)
Case study #2: Microvariation in C and D

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

(24) De die zou 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)
Case study #2: Microvariation in C and D

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

(24) De die zou 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:

<table>
<thead>
<tr>
<th></th>
<th>Brugge</th>
<th>Hulst</th>
<th>Dirksland</th>
<th>Ossendrecht</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>CD</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>SDR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>NEG</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>CYN</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>EXPL-T</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>CMPR-IF</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>ER.OBL</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>THE+THAT</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>GO-GET</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
</tbody>
</table>
Case study #2: Microvariation in C and D

- which is converted into a distance matrix:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>–</td>
<td>11.40</td>
<td>10.00</td>
<td>10.63</td>
<td>10.04</td>
<td>10.72</td>
</tr>
<tr>
<td>B</td>
<td>11.40</td>
<td>–</td>
<td>10.14</td>
<td>7.28</td>
<td>10.14</td>
<td>10.56</td>
</tr>
<tr>
<td>C</td>
<td>10.00</td>
<td>10.14</td>
<td>–</td>
<td>6.48</td>
<td>6.30</td>
<td>6.63</td>
</tr>
<tr>
<td>D</td>
<td>10.63</td>
<td>7.28</td>
<td>6.48</td>
<td>–</td>
<td>6.08</td>
<td>6.63</td>
</tr>
<tr>
<td>E</td>
<td>10.04</td>
<td>10.14</td>
<td>6.30</td>
<td>6.08</td>
<td>–</td>
<td>5.56</td>
</tr>
<tr>
<td>F</td>
<td>10.72</td>
<td>10.56</td>
<td>6.63</td>
<td>6.63</td>
<td>5.56</td>
<td>–</td>
</tr>
</tbody>
</table>

- which is in turn converted into a lower-dimensional (here: 3D) representation.
Case study #2: Microvariation in C and D

- which is converted into a distance matrix:

<table>
<thead>
<tr>
<th></th>
<th>CA</th>
<th>CD</th>
<th>SDR</th>
<th>CYN</th>
<th>NEG</th>
<th>EXPL-T</th>
<th>CMPR-IF</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td></td>
<td>11.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDR</td>
<td>10.14</td>
<td></td>
<td>7.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYN</td>
<td>10.00</td>
<td>6.48</td>
<td></td>
<td>4.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEG</td>
<td>10.63</td>
<td>6.08</td>
<td>4.69</td>
<td></td>
<td>5.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPL-T</td>
<td>10.04</td>
<td>8.30</td>
<td>4.24</td>
<td>5.56</td>
<td>6.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMPR-IF</td>
<td>10.72</td>
<td>8.54</td>
<td>4.69</td>
<td>5.91</td>
<td>6.63</td>
<td>4.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>THE+THAT</td>
<td>10.77</td>
<td>5.83</td>
<td>6.70</td>
<td>6.63</td>
<td>6.40</td>
<td>7.68</td>
<td>8.06</td>
<td></td>
</tr>
<tr>
<td>ER.OBL</td>
<td>10.34</td>
<td>8.06</td>
<td>4.24</td>
<td>5.38</td>
<td>6.00</td>
<td>4.00</td>
<td>4.69</td>
<td>...</td>
</tr>
<tr>
<td>GO-GET</td>
<td>10.72</td>
<td>8.30</td>
<td>4.89</td>
<td>5.91</td>
<td>6.32</td>
<td>5.29</td>
<td>5.09</td>
<td>...</td>
</tr>
</tbody>
</table>
Case study #2: Microvariation in C and D

- which is converted into a distance matrix:

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

- which is in turn converted into a lower-dimensional (here: 3D) representation
Case study #2: Microvariation in C and D
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:

\[ \text{the AgrC-parameter:} \]

\[ C \{\text{does/does not}\} \text{ have unvalued } \phi \text{-features.} \]
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):
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):

\[
\begin{align*}
\text{(27)} & \quad \text{FP} \\
& \quad \text{CLITIC} \quad \text{F'} \\
& \quad \quad \text{F} \\
& \quad \quad \quad \text{DP} \\
& \quad \quad \quad \quad \phi \text{P} \\
& \quad \quad \quad \quad \quad \phi \text{ NP} \\
\quad \rightarrow \text{STRONG} \\
\text{(28)} & \quad \text{FP} \\
& \quad \text{THE} \\
& \quad \quad \text{F'} \\
& \quad \quad \quad \text{F} \\
& \quad \quad \quad \quad \text{DP} \\
& \quad \quad \quad \quad \quad \phi \text{P} \\
& \quad \quad \quad \quad \quad \quad \phi \text{ NP} \\
\quad \rightarrow \text{THE} \\
\end{align*}
\]
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)

ER.OBL: the expletive ‘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

- 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.

we can now examine the interaction between the three parameters we proposed
Case study #2: Microvariation in C and D
Case study #2: Microvariation in C and D

<table>
<thead>
<tr>
<th>+AgrC</th>
<th>+SPLIT C</th>
<th>−SPLIT C</th>
</tr>
</thead>
<tbody>
<tr>
<td>+split D</td>
<td>East &amp; West Flanders (N=59)</td>
<td>Nieuwmoer, Sint Lenaarts, Moerdijk (N=3)</td>
</tr>
<tr>
<td>−split D</td>
<td>Opglabbeek, Sliedrecht, Hoek (N=3)</td>
<td>Holland, Limburg, Friesland, Groningen (N=83)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>−AgrC</th>
<th>+SPLIT C</th>
<th>−SPLIT C</th>
</tr>
</thead>
<tbody>
<tr>
<td>+split D</td>
<td>Flemish Brabant &amp; Antwerp (N=23)</td>
<td>North Brabant (N=21)</td>
</tr>
<tr>
<td>−split D</td>
<td>Borgloon (N=1)</td>
<td>Drenthe, Utrecht (N=67)</td>
</tr>
</tbody>
</table>
Case study #2: Microvariation in C and D

Digging a little deeper:

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
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:

\[ r = 0.315 \]

e.g. there is only a fairly weak correlation between linguistic distance and geographical distance in our data set.
Case study #2: Microvariation in C and D

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

<table>
<thead>
<tr>
<th></th>
<th>Midsland</th>
<th>Lies</th>
<th>W. Terschelling</th>
<th>Oosterend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lies</td>
<td>4.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W. Terschelling</td>
<td>8.15</td>
<td>12.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oosterend</td>
<td>56.35</td>
<td>60.60</td>
<td>48.77</td>
<td></td>
</tr>
<tr>
<td>Hollum</td>
<td>39.71</td>
<td>35.25</td>
<td>47.82</td>
<td>93.54</td>
</tr>
<tr>
<td>Schiermonnikoog</td>
<td>97.27</td>
<td>92.85</td>
<td>105.33</td>
<td>148.91</td>
</tr>
<tr>
<td>Ferwerd</td>
<td>59.98</td>
<td>55.80</td>
<td>67.68</td>
<td>109.13</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Case study #2: Microvariation in C and D

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

<table>
<thead>
<tr>
<th>Midsland</th>
<th>Lies</th>
<th>W. Terschelling</th>
<th>Oosterend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lies</td>
<td>4.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W. Terschelling</td>
<td>8.15</td>
<td>12.65</td>
<td></td>
</tr>
<tr>
<td>Oosterend</td>
<td>56.35</td>
<td>60.60</td>
<td>48.77</td>
</tr>
<tr>
<td>Hollum</td>
<td>39.71</td>
<td>35.25</td>
<td>47.82</td>
</tr>
<tr>
<td>Schiermonnikoog</td>
<td>97.27</td>
<td>92.85</td>
<td>105.33</td>
</tr>
<tr>
<td>Ferwerd</td>
<td>59.98</td>
<td>55.80</td>
<td>67.68</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- 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:

<table>
<thead>
<tr>
<th></th>
<th>Midsland</th>
<th>Lies</th>
<th>W. Terschelling</th>
<th>Oosterend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lies</td>
<td>4.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W. Terschelling</td>
<td>8.15</td>
<td>12.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oosterend</td>
<td>56.35</td>
<td>60.60</td>
<td>48.77</td>
<td></td>
</tr>
<tr>
<td>Hollum</td>
<td>39.71</td>
<td>35.25</td>
<td>47.82</td>
<td>93.54</td>
</tr>
<tr>
<td>Schiermonnikoog</td>
<td>97.27</td>
<td>92.85</td>
<td>105.33</td>
<td>148.91</td>
</tr>
<tr>
<td>Ferwerd</td>
<td>59.98</td>
<td>55.80</td>
<td>67.68</td>
<td>109.13</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- 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:

<table>
<thead>
<tr>
<th></th>
<th>Midsland</th>
<th>Lies</th>
<th>W. Terschelling</th>
<th>Oosterend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lies</td>
<td>4.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W. Terschelling</td>
<td>8.15</td>
<td>12.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oosterend</td>
<td>56.35</td>
<td>60.60</td>
<td>48.77</td>
<td></td>
</tr>
<tr>
<td>Hollum</td>
<td>39.71</td>
<td>35.25</td>
<td>47.82</td>
<td>93.54</td>
</tr>
<tr>
<td>Schiermonnikoog</td>
<td>97.27</td>
<td>92.85</td>
<td>105.33</td>
<td>148.91</td>
</tr>
<tr>
<td>Ferwerd</td>
<td>59.98</td>
<td>55.80</td>
<td>67.68</td>
<td>109.13</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- 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
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?
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

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

Two case studies
  Case study #1: Verb clusters
  Case study #2: Microvariation in C and D

Conclusion & 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
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


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.


References II


References III


References IV


