Comparing constructicons:
A cluster analysis of the causative constructions with *doen*
in Netherlandic and Belgian Dutch

Natalia Levshina

Philipps Universität Marburg

DGKL 5

Outline

1. Dutch causative Cx with *doen*

2. Data and method

3. Quantitative analyses:

     - Netherlandic *doen*

     - Belgian *doen*

4. Is it done with *doen*?

Causative *doen* in Dutch

Haar stem *deed* *het glas* *barsten.*

her voice made the glass break

Semantic variation

- physical causation

- affective causation

- volitional causation

- inducive causation (?)

(Verhagen & Kemmer 1997)

Semantic variation

- physical causation

- affective causation

- volitional causation

- inducive causation (?)
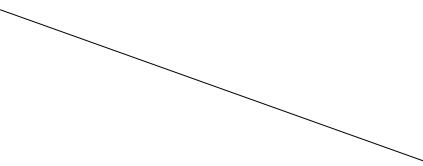
Haar stem deed het glas barsten.

(Verhagen & Kemmer 1997)

Semantic variation

- physical causation

- affective causation

- volitional causation

- inducive causation (?)

Haar stem deed
me aan Lady Gaga
denken.

(Verhagen & Kemmer 1997)

Semantic variation

- physical causation

- affective causation

- volitional causation

- inducive causation (?)
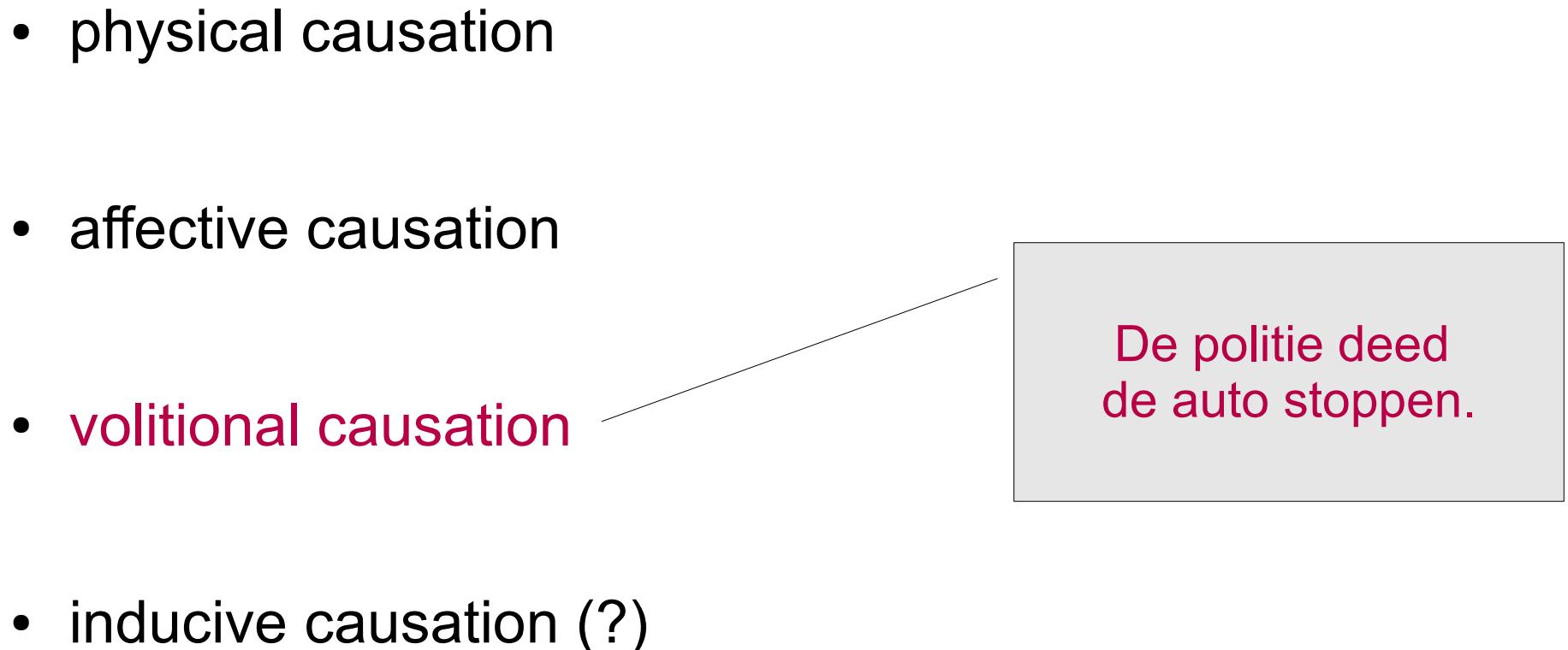
De politie deed
de auto stoppen.

(Verhagen & Kemmer 1997)

Semantic variation

- physical causation

- affective causation

- volitional causation

- inducive causation (?)

Ze strooien poeder op je vel en doen je slapen op bevel.

(Verhagen & Kemmer 1997)

Lectal variation

- *doen* is more common in Belgian Dutch than in Netherlandic Dutch (Speelman & Geeraerts 2009)

- *doen* is more common in formal/written than in informal/spoken Dutch (Levshina 2011)

Frequency of causative doen, per million words
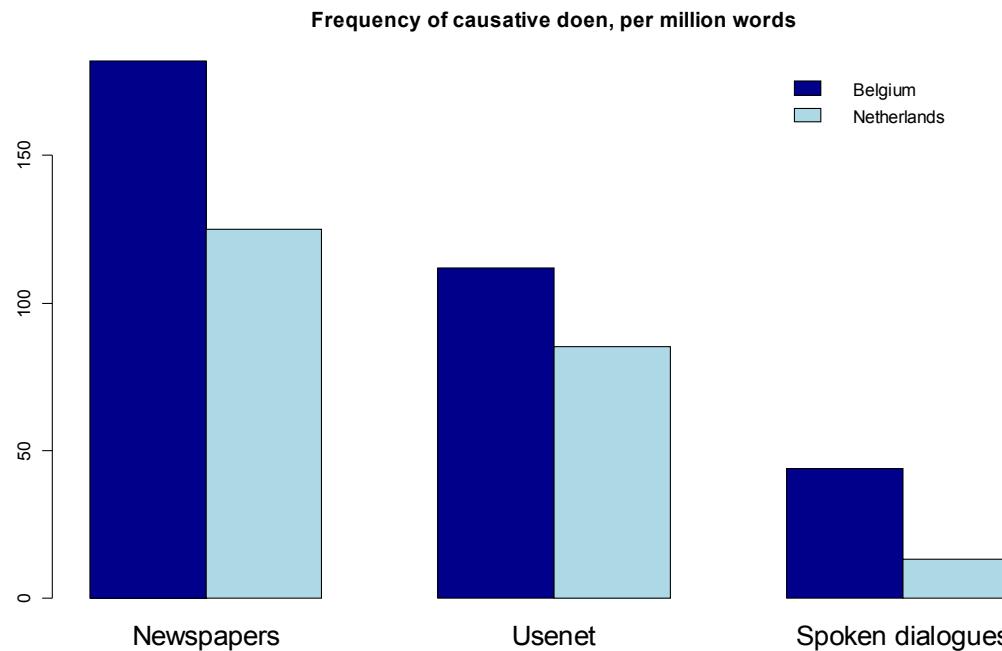
Are there also differences in the semantic structure of the regional varieties?

Outline

1. Dutch causative Cx with *doen*

2. Data and method

3. Quantitative analyses:

     - Netherlandic *doen*

     - Belgian *doen*

4. Is it done with *doen*?

Corpus data

Netherlands

- Twente News Corpus

- Usenet.nl

- CGN-NL spontaneous conversations

Total: 66 observations

Belgium

- Leuven News Corpus

- Usenet.be

- CGN-BE spontaneous conversations

Total: 66 observations

Variables

- Causer, Causee, Affectee: sem. class, person, number, definiteness, POS, synt. expression

- Effected Predicate: transitivity, prepositional complements, semantic class of the caused event, lemma

- Coreferentiality and possession relations btw. the participants

- Causee only: intentionality, semantic role

- Negation, adverbial modifiers

- Mood, tense, type of the clause and sentence

35 categorical variables in total

Analytical Procedure

Data frame with observations (rows)
and variables (columns)

Analytical Procedure

Data frame with observations (rows)
and variables (columns)

Matrix of distances between the observations
based on Gower's distance metric

Analytical Procedure

Data frame with observations (rows)
and variables (columns)

Matrix of distances between the observations
based on Gower's distance metric

Multivariate analysis (hclust)
to explore the semantic structure

Analytical Procedure

Data frame with observations (rows)
and variables (columns)

↓

Matrix of distances between the observations
based on Gower's distance metric

↓

Multivariate analyses (hclust)
to explore the semantic structure

↓

Comparison of NL and BE
clustering solutions
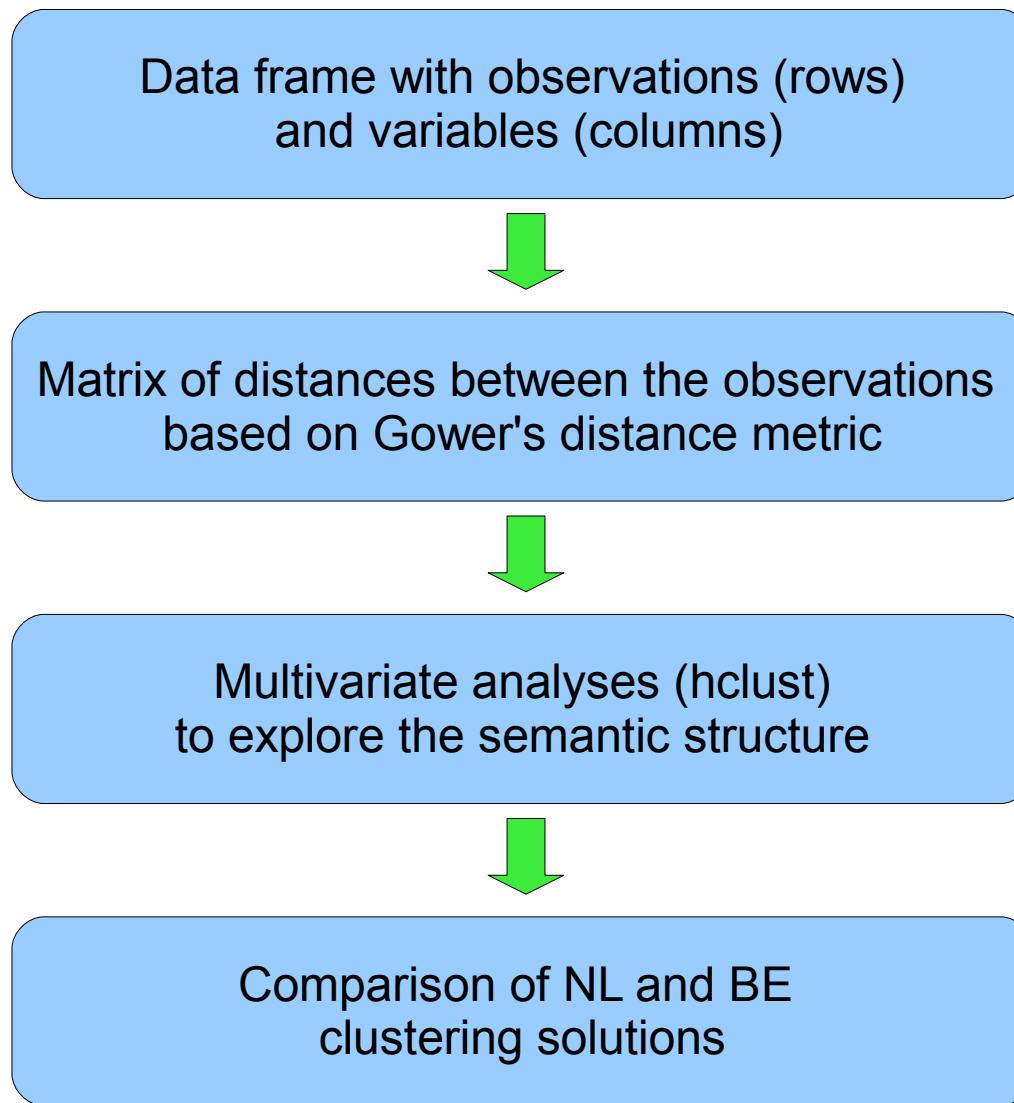
Cluster validation

- `clusterboot{fpc}`, Hennig 2007 (modified)

- random sampling from 66 observations with replacement

- random sampling from 66 original + 33 (66) additional observations

- calculate the Jaccard coefficient for each cluster: how similar a cluster in the original solution to its most similar cluster in the bootstrapping solution

- run it *n* (100) times

- calculate the average Jaccard coefficient

Outline

1. Dutch causative Cx with *doen*

2. Data and method

3. Quantitative analyses:

     - Netherlandic *doen*

     - Belgian *doen*

4. Is it done with *doen*?

Semantic structure of NL *doen*
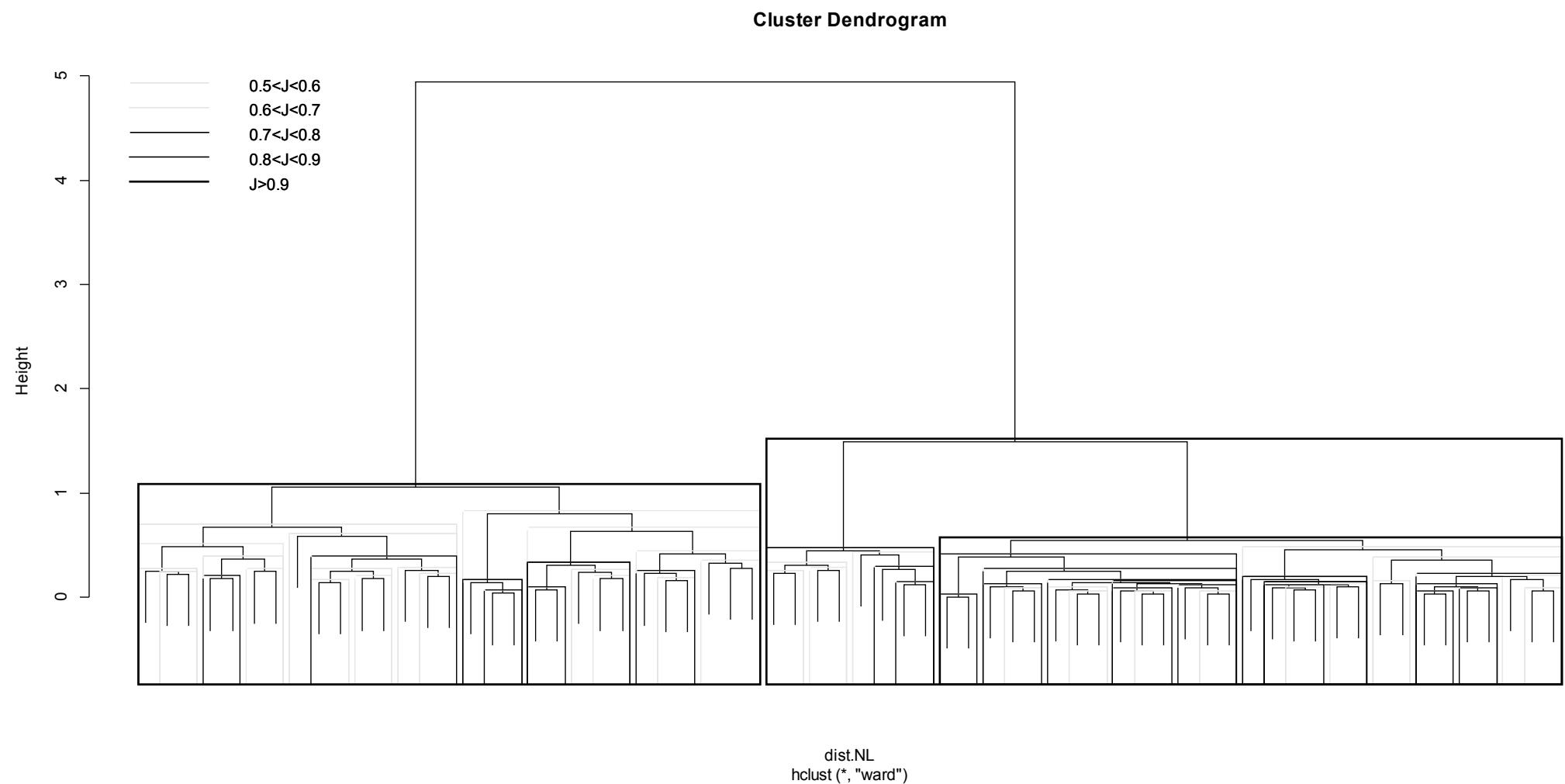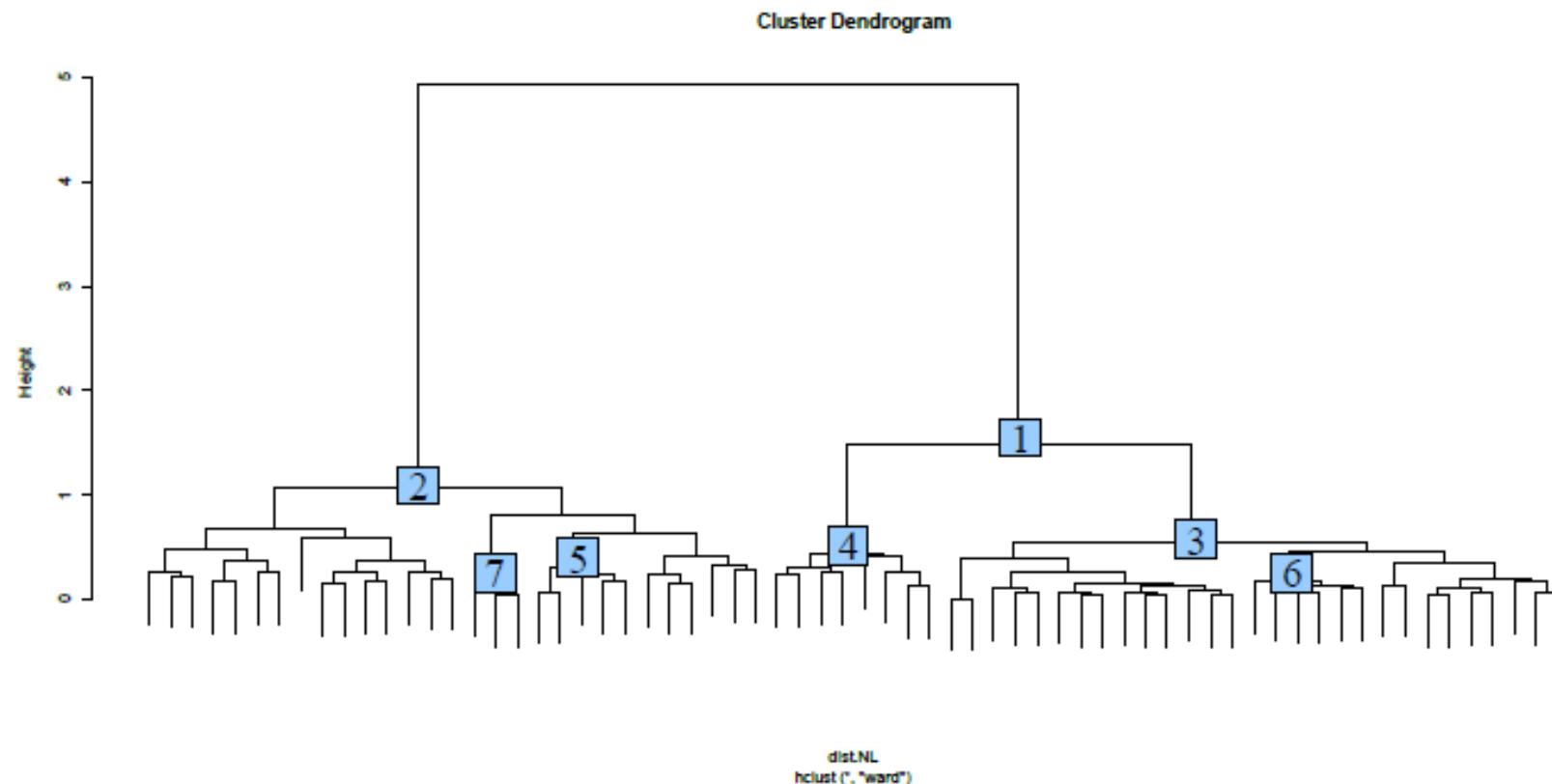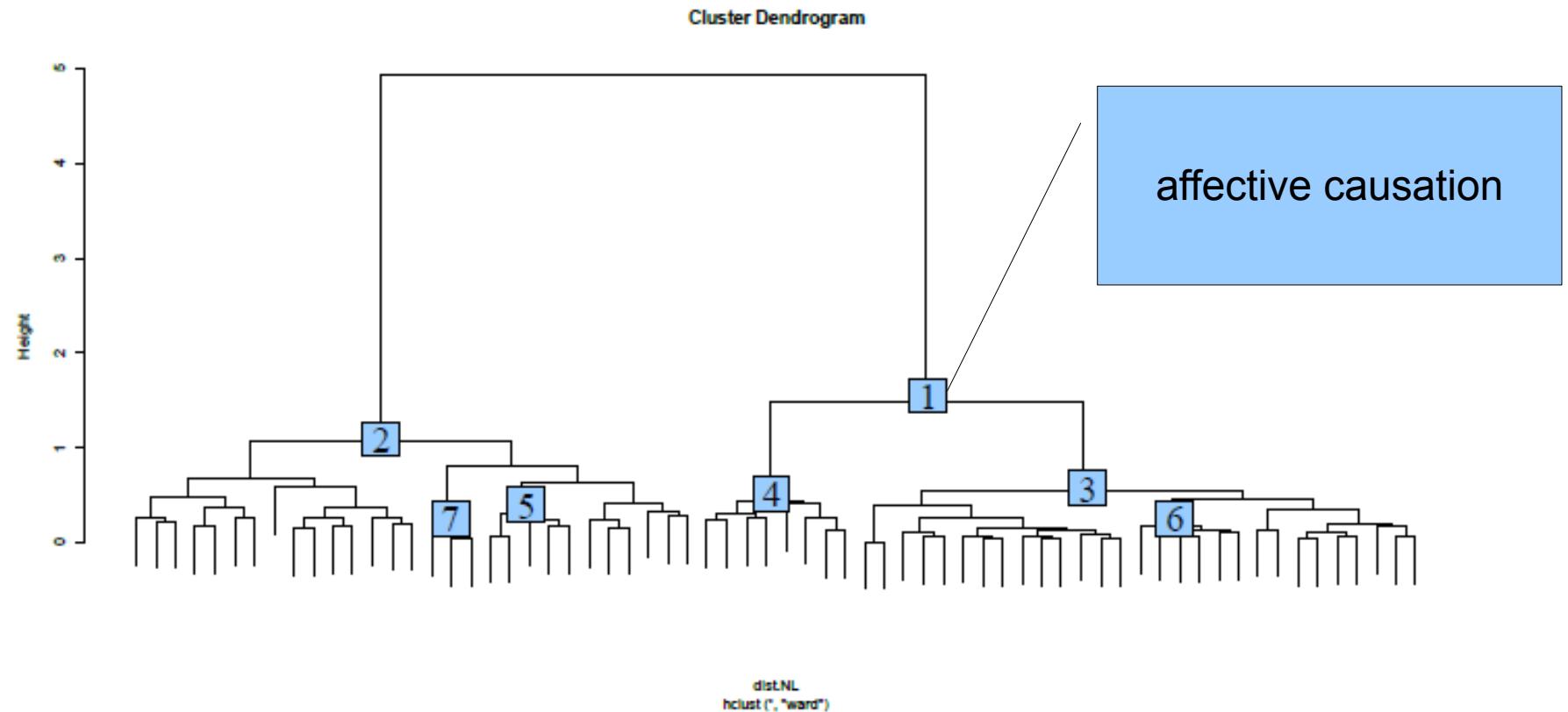
Cluster Dendrogram

0.5<J<0.6
0.6<J<0.7
0.7<J<0.8
0.8<J<0.9
J>0.9

Height

dist.NL
hclust (*, "ward")

# Highly stable clusters J > 0.8

Cluster Dendrogram

dist.NL
hclust (*, "ward")

# Highly stable clusters J > 0.8

Highly stable clusters J > 0.8

Highly stable clusters J > 0.8

Cluster Dendrogram

implicit Causee
Het nummer doet met zijn bombast aan zijn oude band denken,

Highly stable clusters J > 0.8

Highly stable clusters J > 0.8

Cluster Dendrogram

non-affective causation

dist.NL
hclust (*, "ward")

Highly stable clusters J > 0.8

Cluster Dendrogram

Een verhaal in De Telegraaf deed het aandeel met 18.3 procent duikelen.

dist.NL
hclust (", "ward")
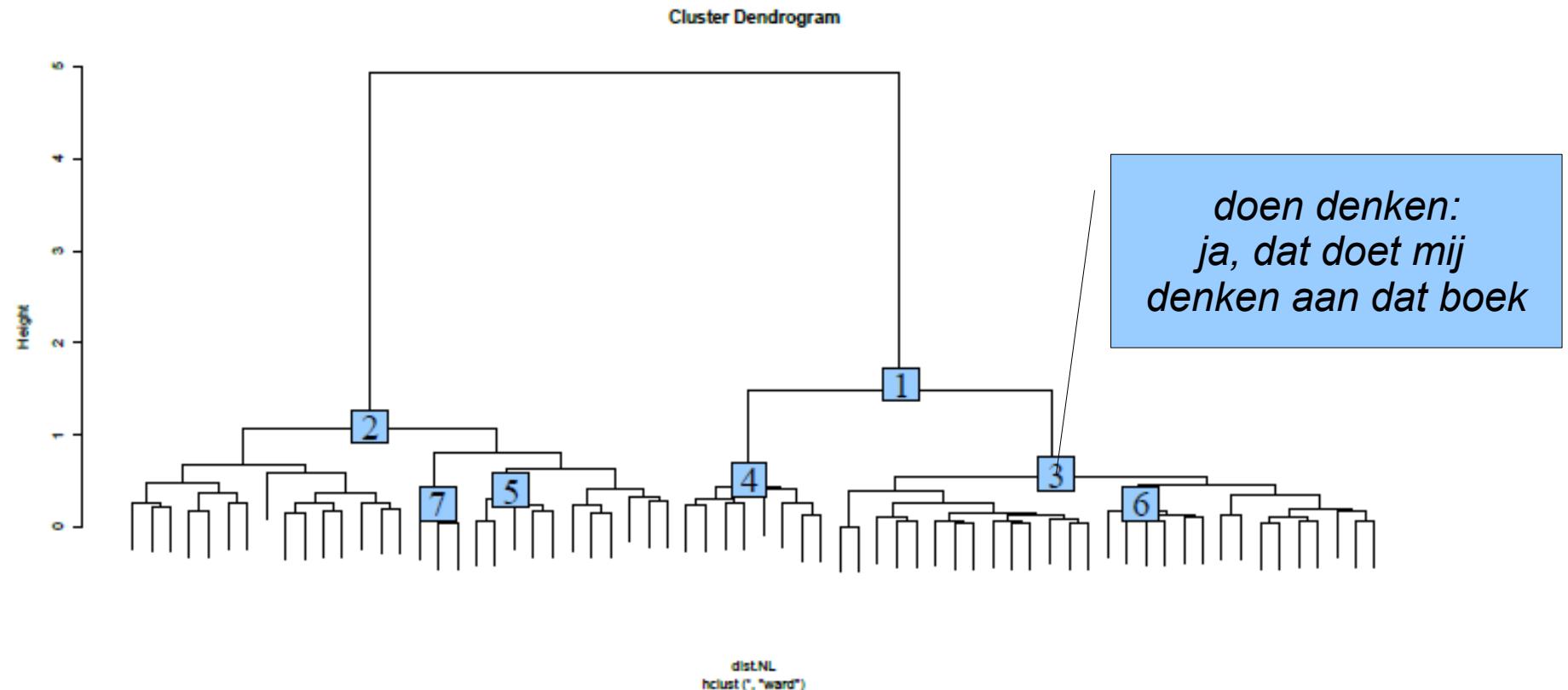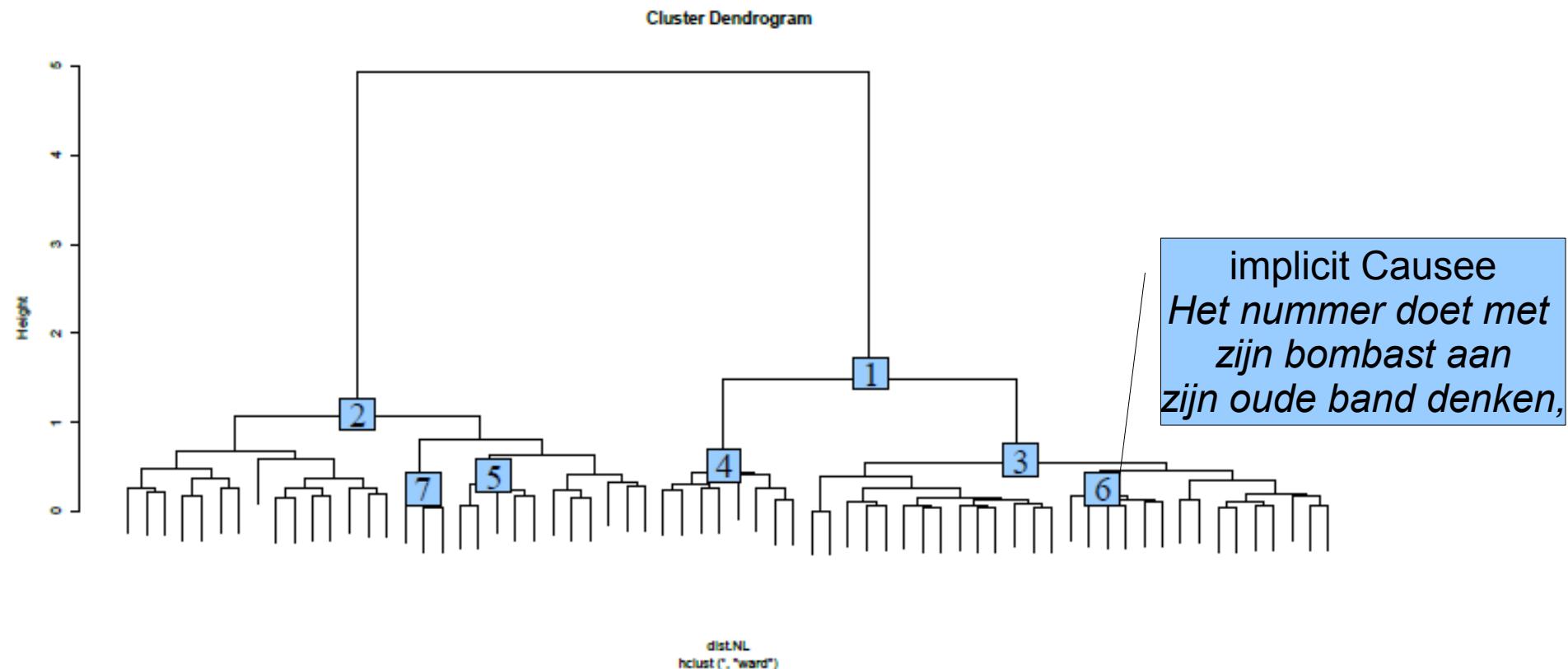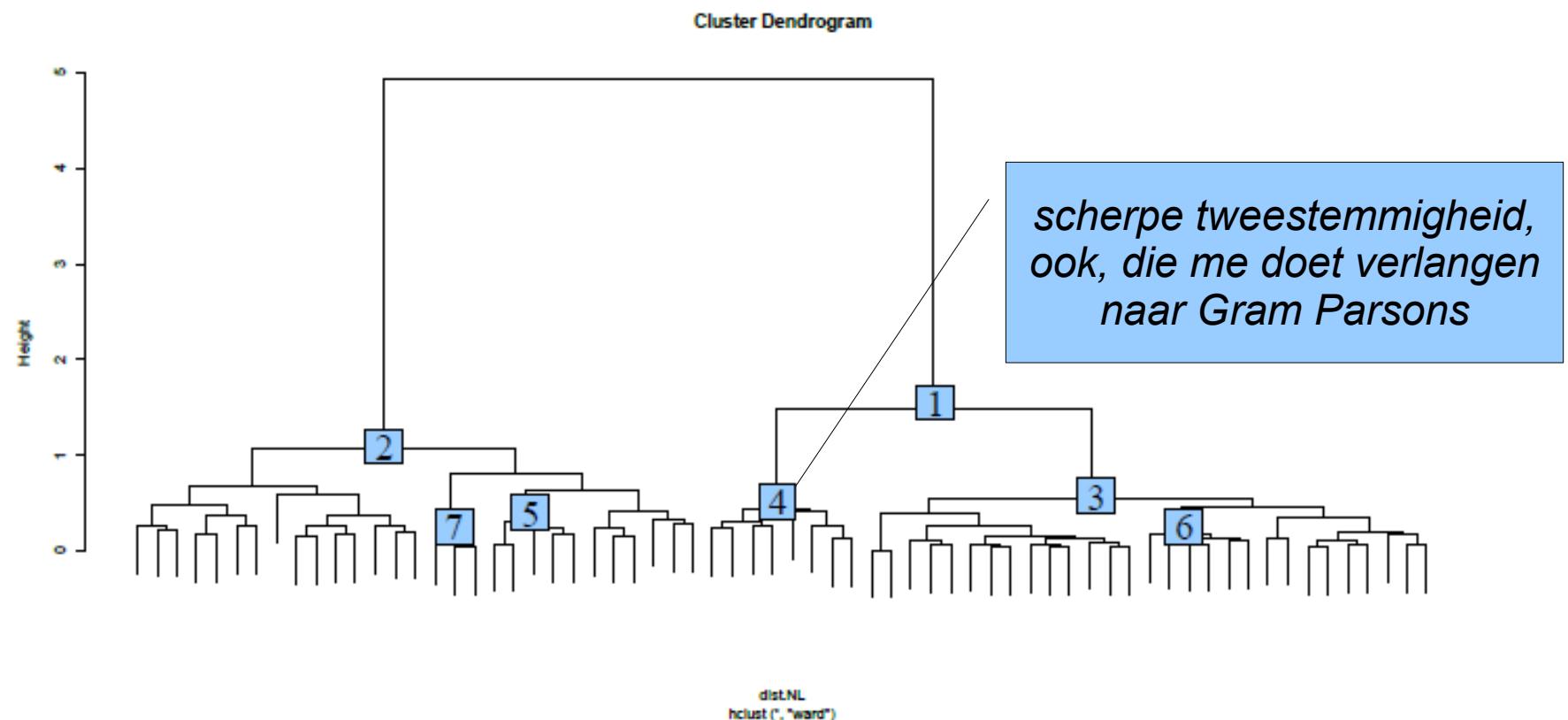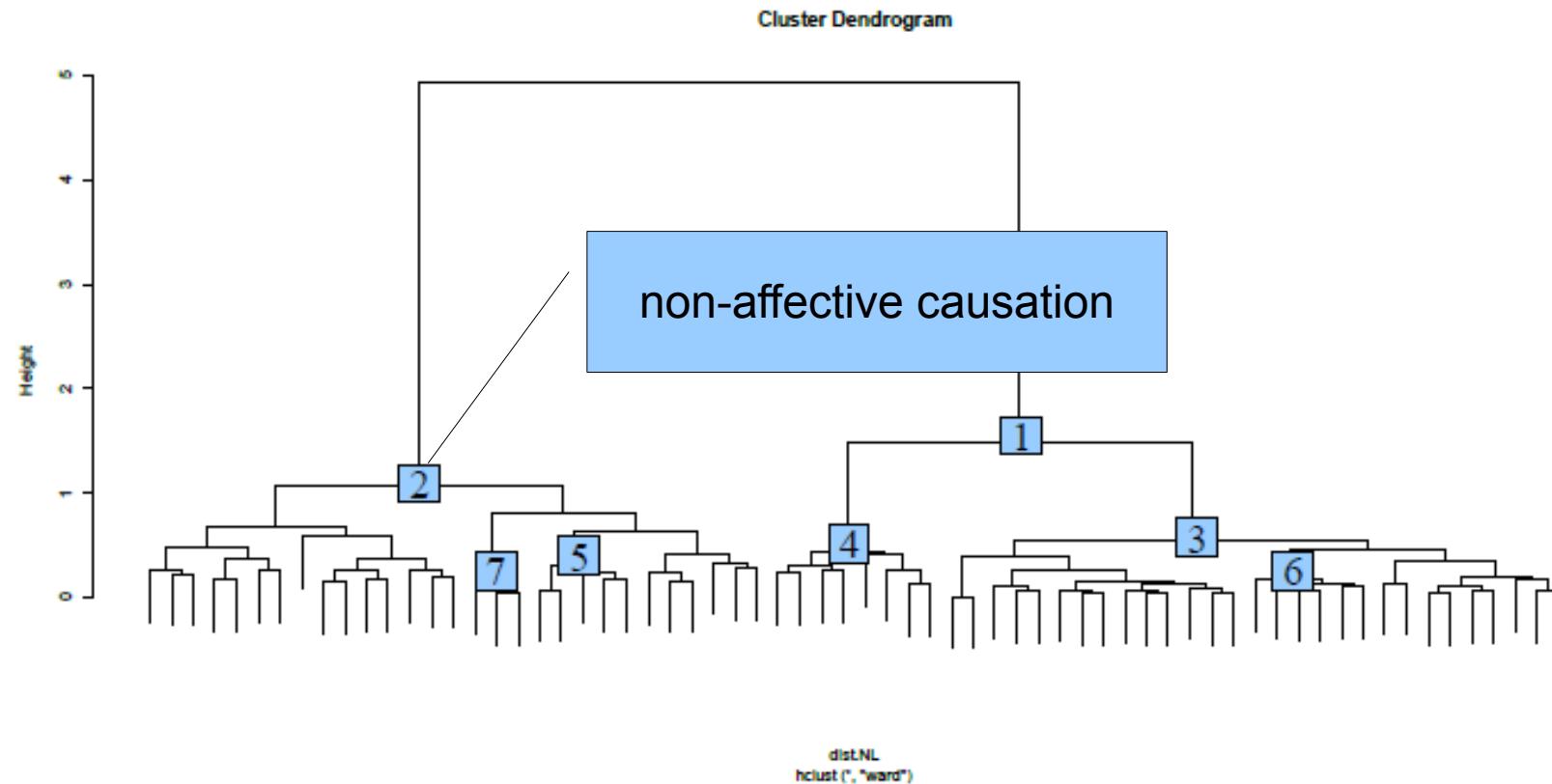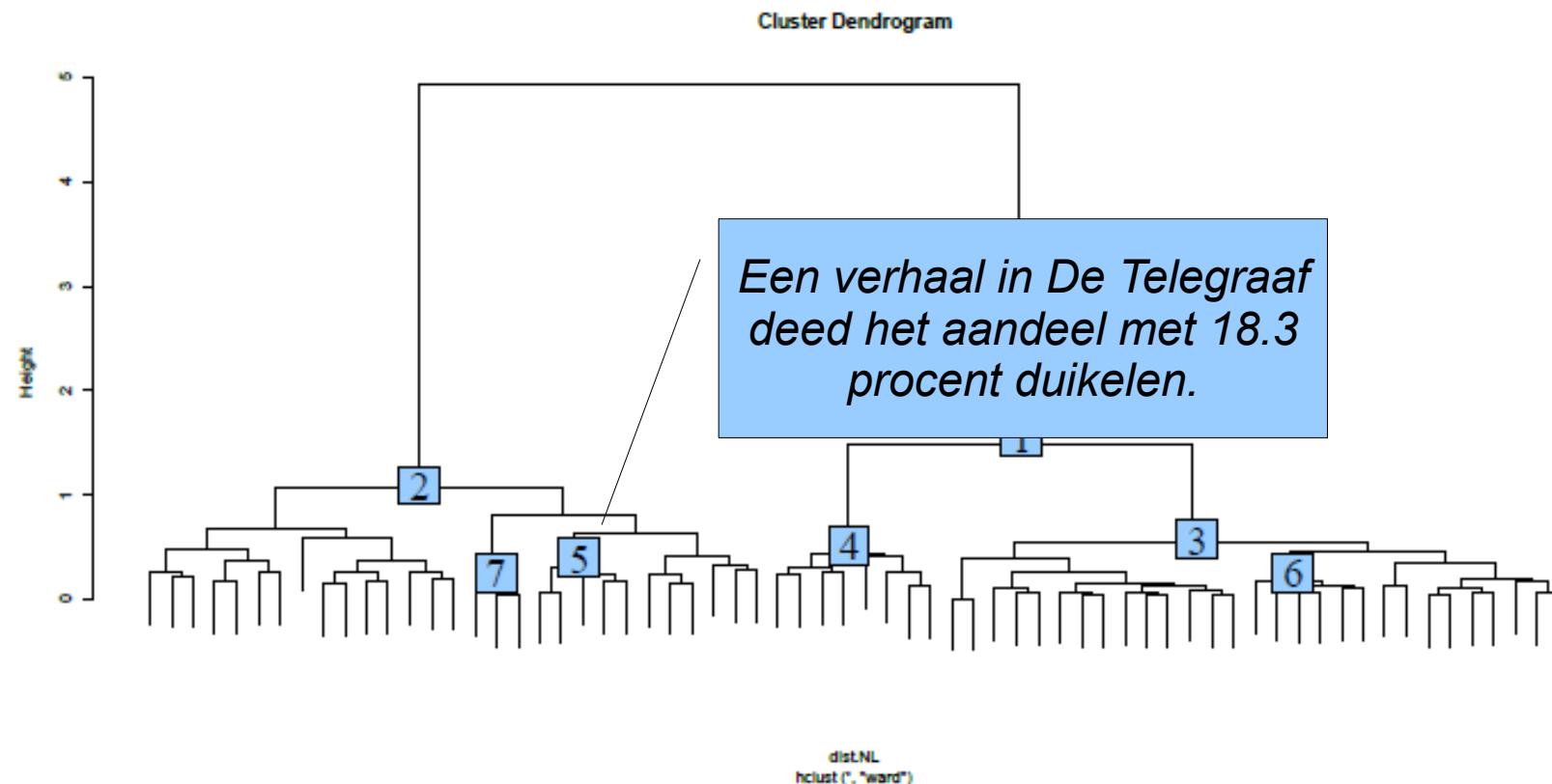
Highly stable clusters J > 0.8
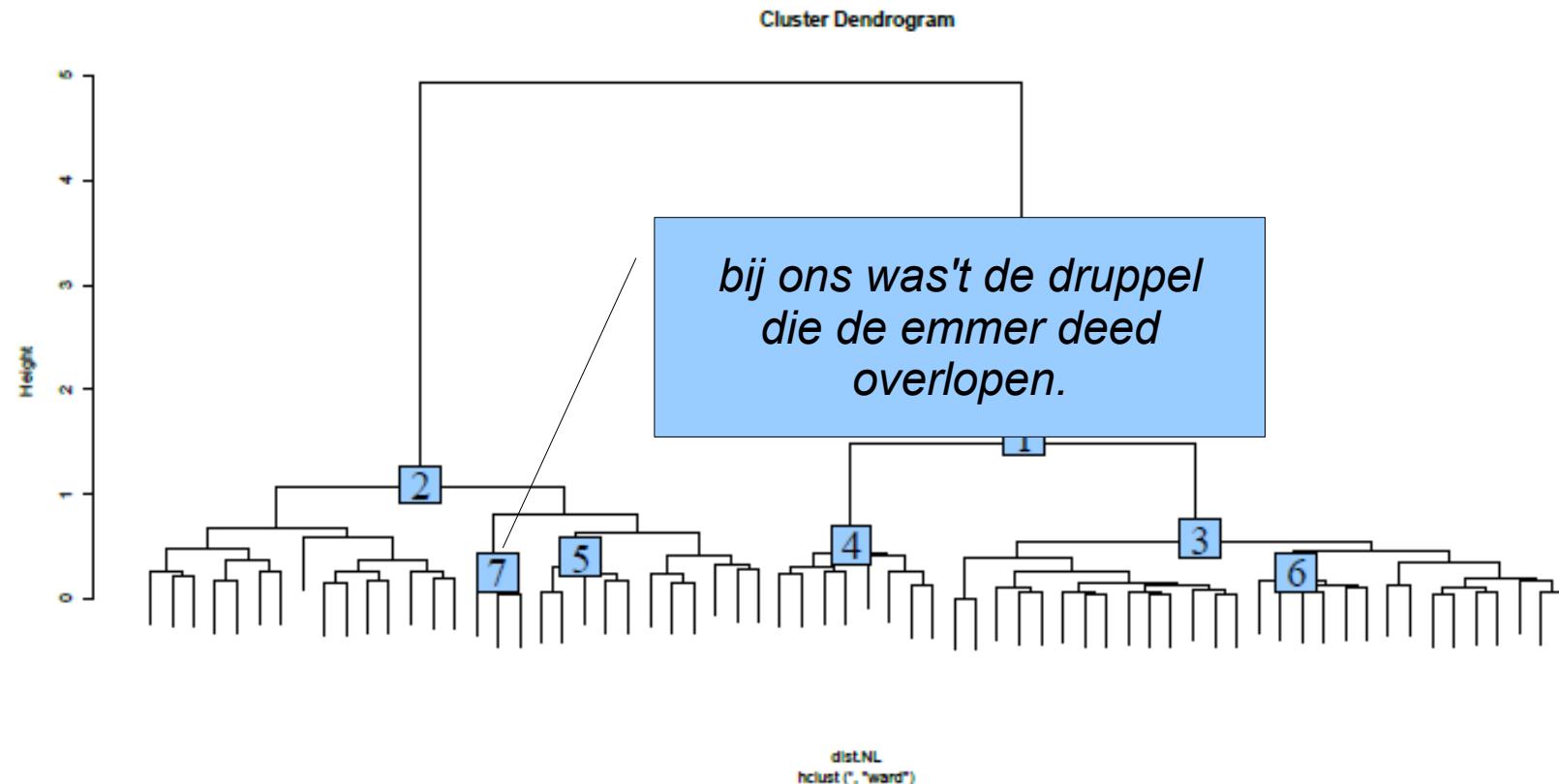
Outline

1. Dutch causative Cx with *doen*

2. Data and method

3. Quantitative analyses:

      - Netherlandic *doen*

      - Belgian *doen*

4. Is it done with *doen*?

Semantic structure of BE *doen*
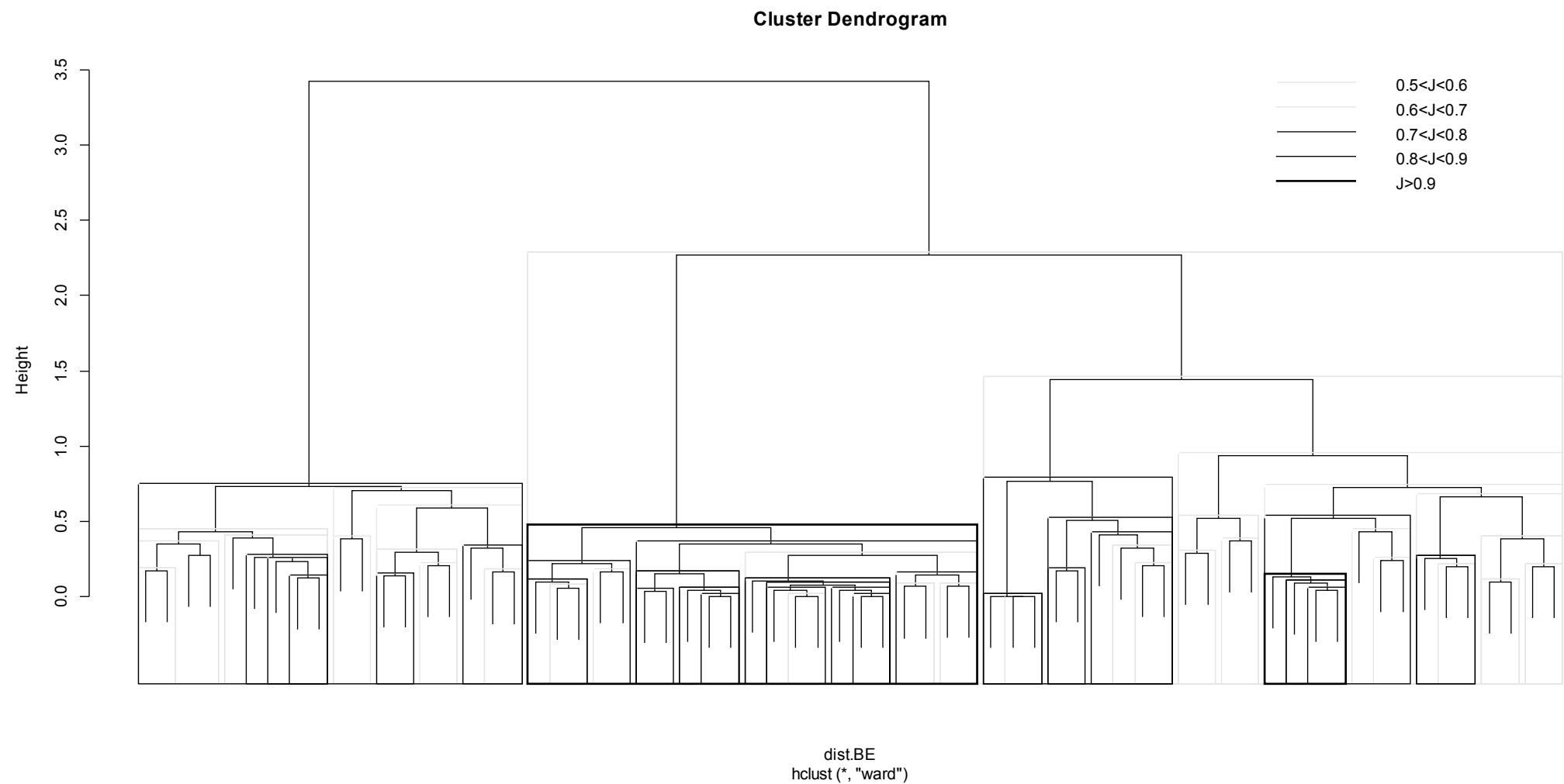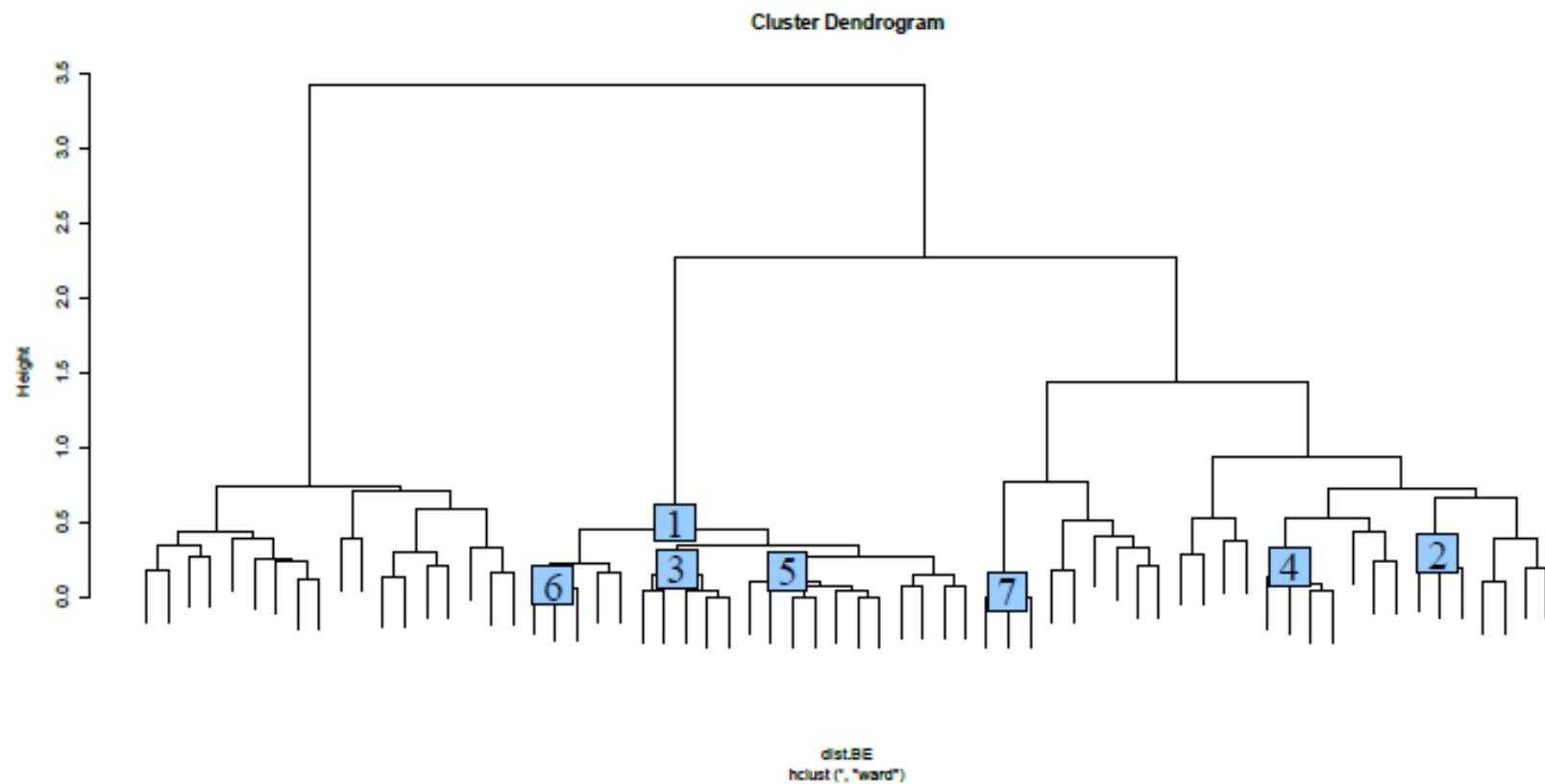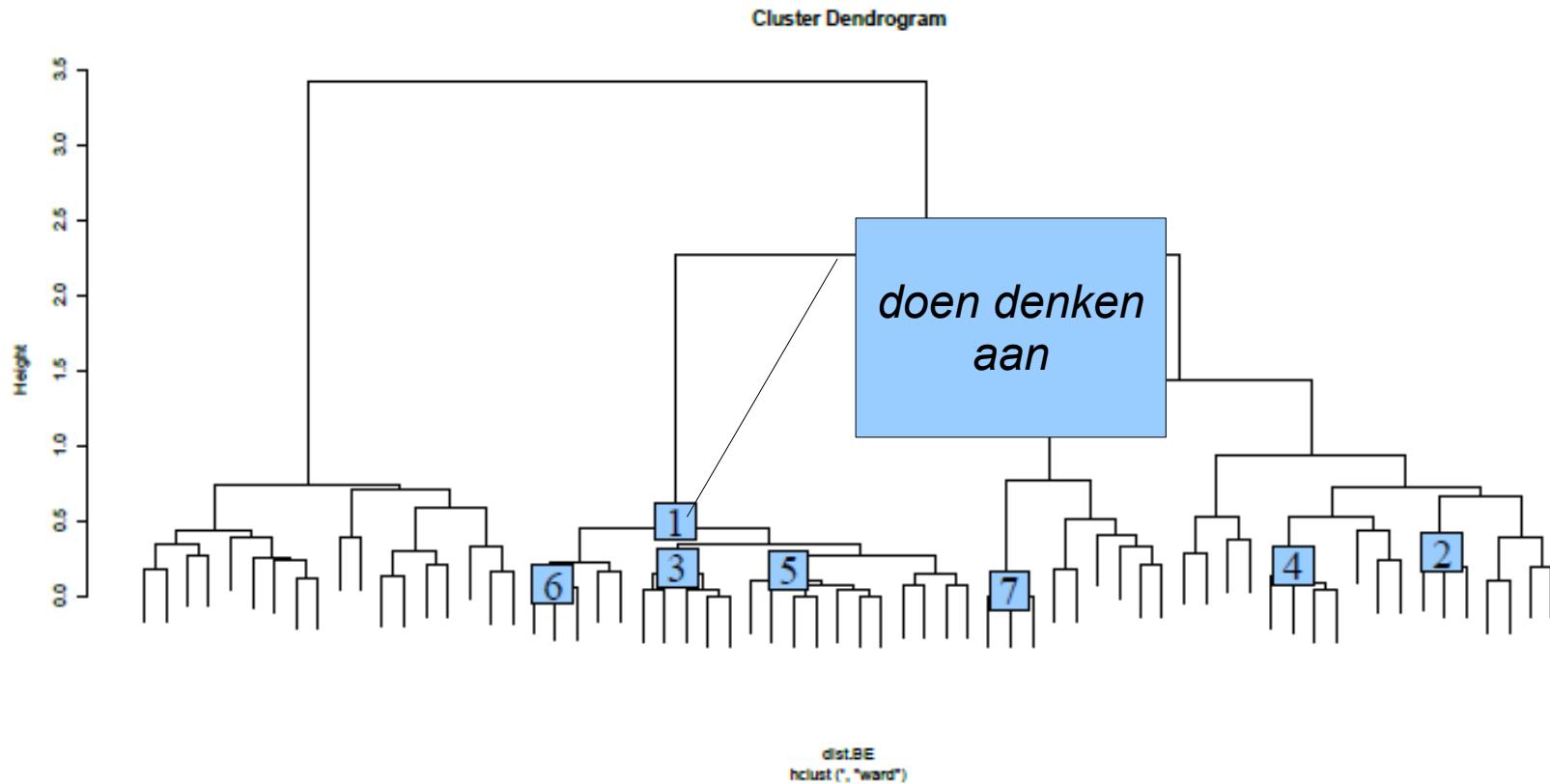
Cluster Dendrogram

0.5<J<0.6
0.6<J<0.7
0.7<J<0.8
0.8<J<0.9
J>0.9

Height

dist.BE
hclust (*, "ward")

Highly stable clusters (J > 0.80)

Cluster Dendrogram

dist.BE
hclust (", "ward")

Highly stable clusters (J > 0.80)

Highly stable clusters (J > 0.80)

Highly stable clusters (J > 0.80)

Cluster Dendrogram

Dat doet mij eventjes denken aan een anekdote

Highly stable clusters (J > 0.80)

Highly stable clusters (J > 0.80)


Cluster Dendrogram

lexically specific clusters:
doen vermoeden,
hoop doet leven, etc.
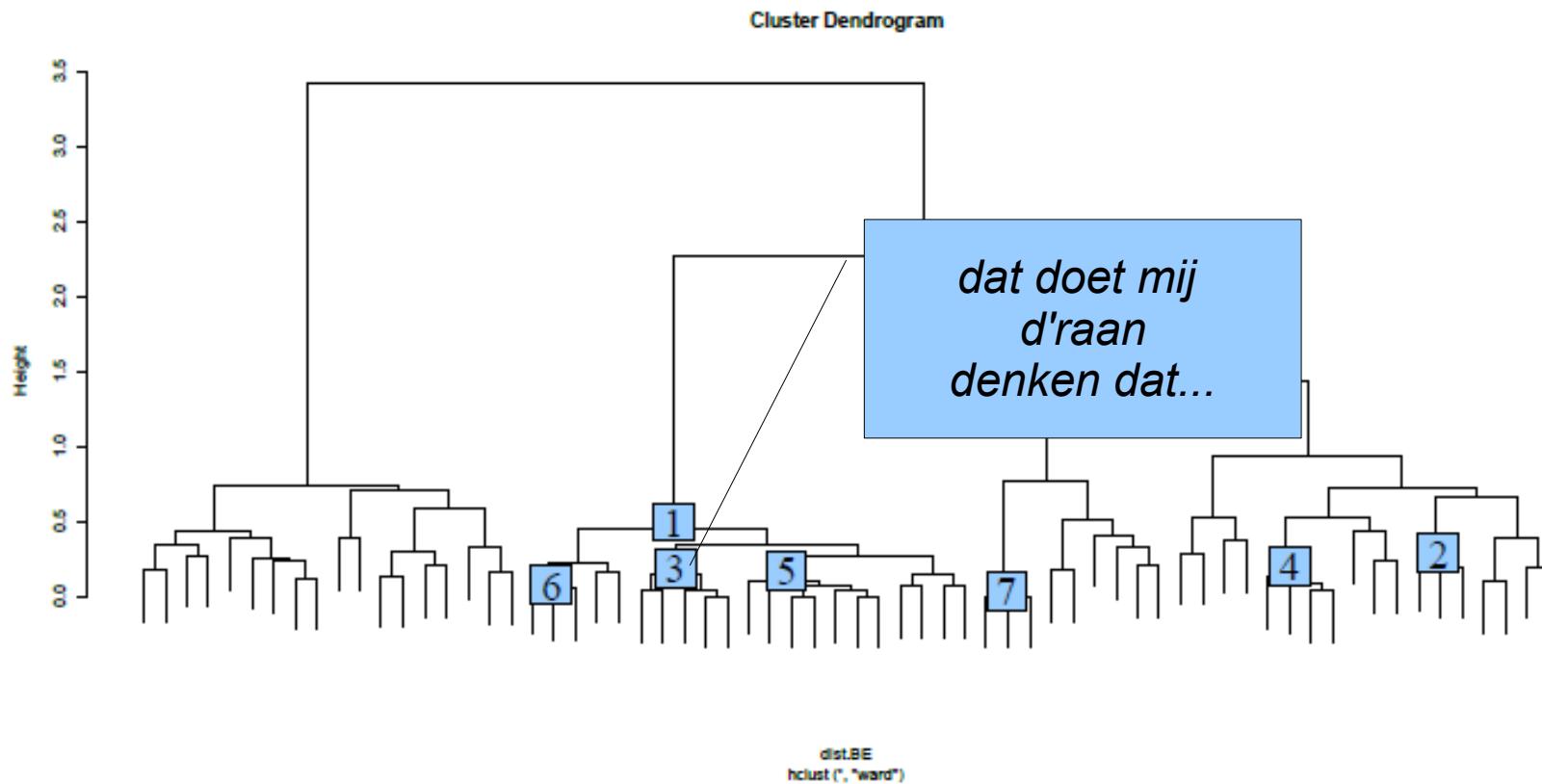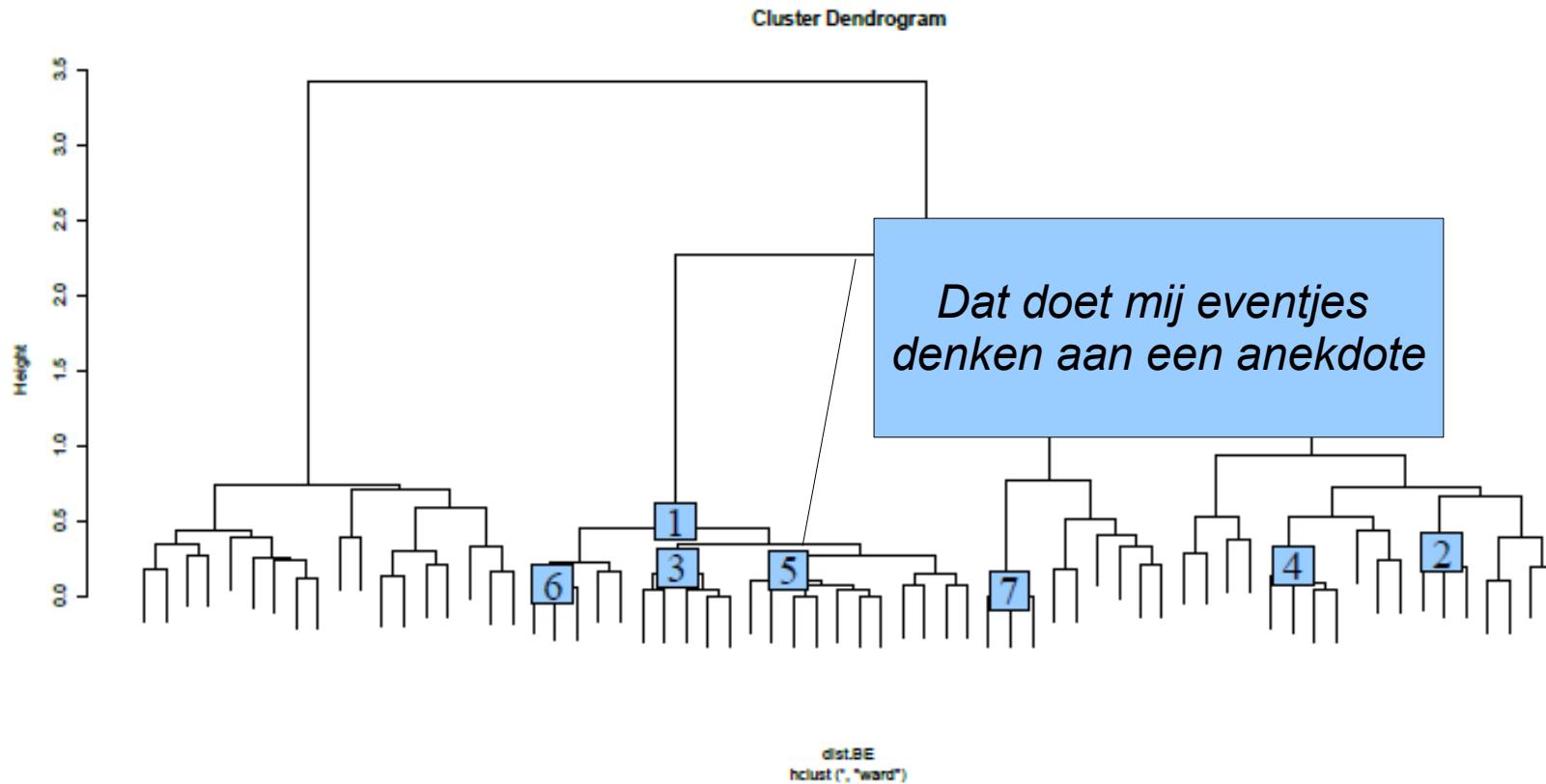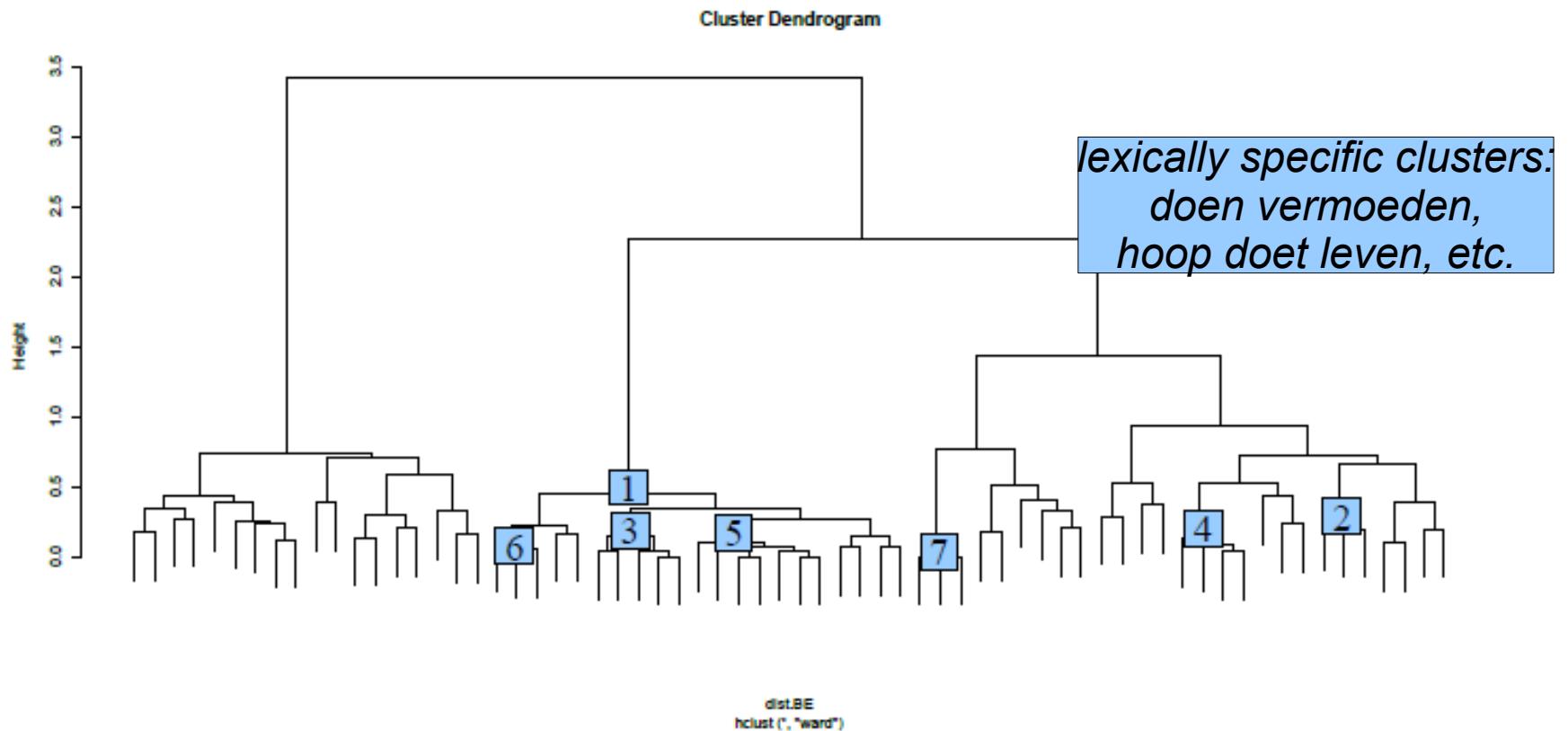
Interim conclusions

- previous classifications of causation events (Verhagen & Kemmer 1997) find some support only in the case of the NL *doen*

- the Belgian sample yields a weaker cluster structure than the Netherlandic one

- more evidence?

**Agglomerative coefficients**
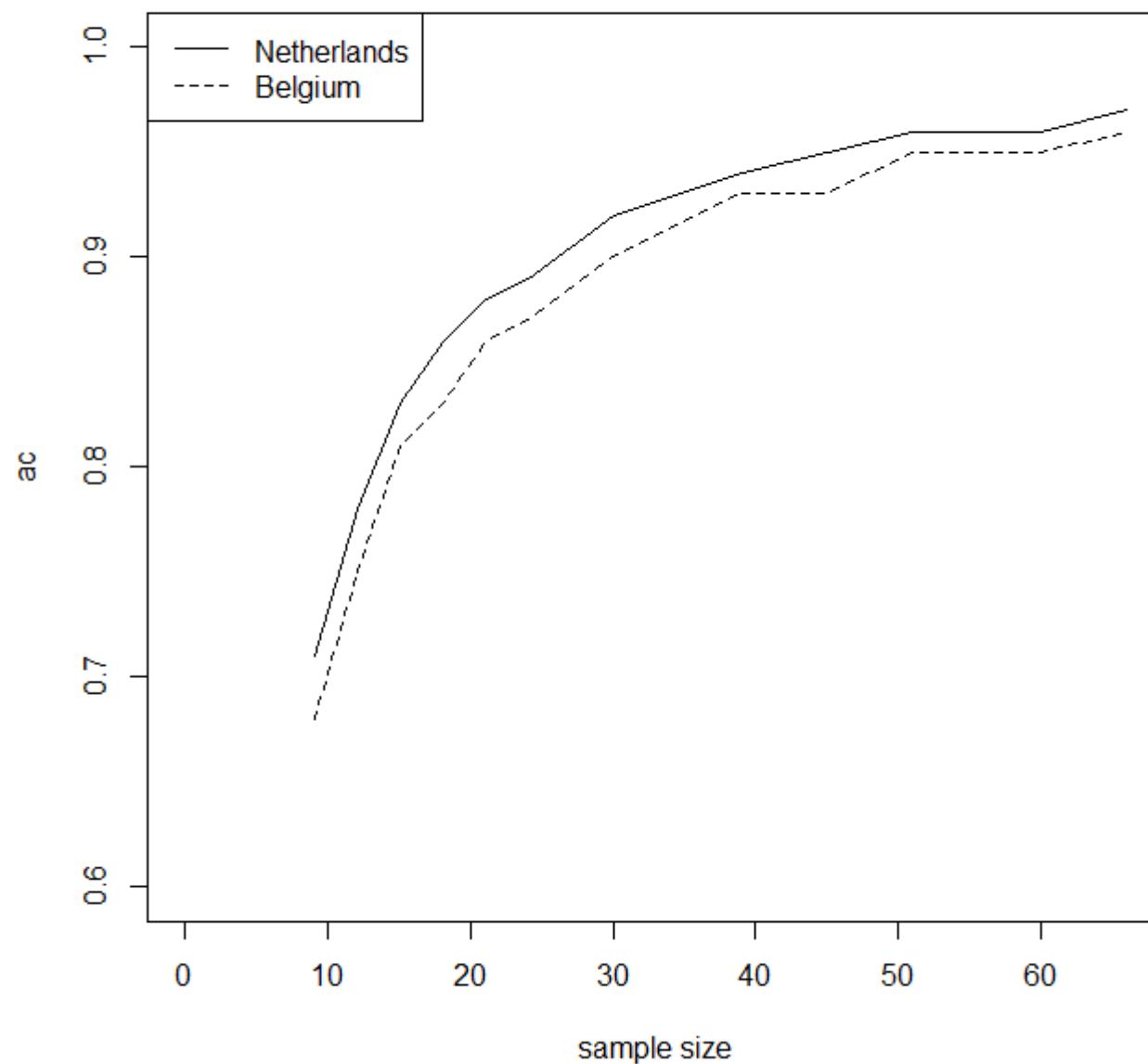
Average silhouette widths
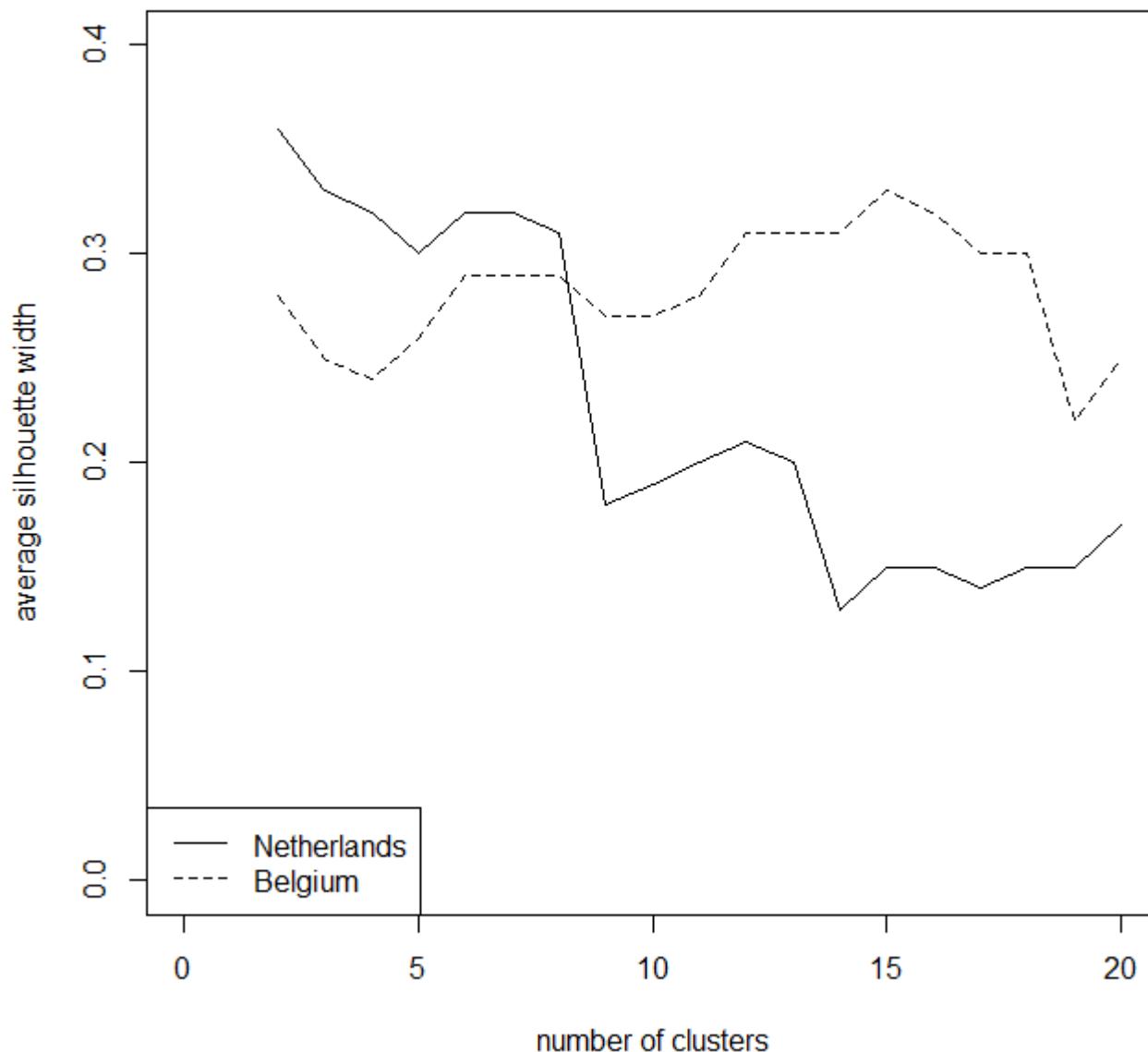
Neighbour-joining algorithm
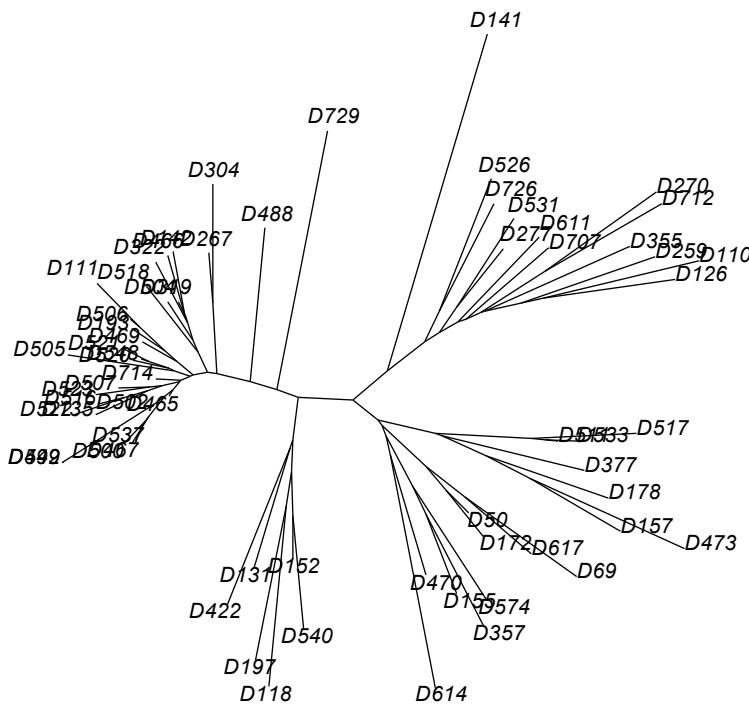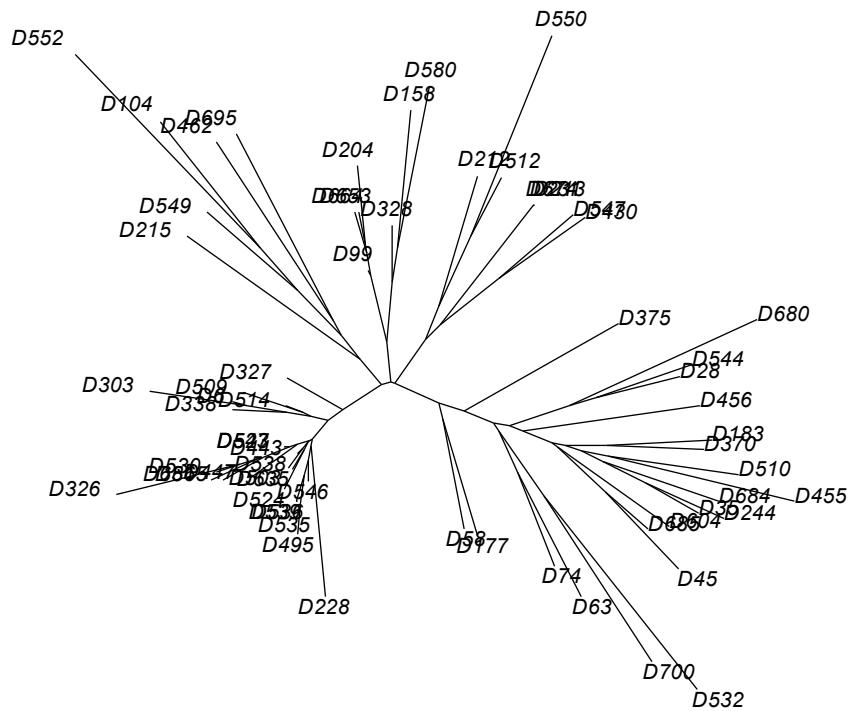
Outline

1. Dutch causative Cx with *doen*

2. Data and method

3. Quantitative analyses:

   - Netherlandic *doen*

   - Belgian *doen*

4. Is it done with *doen*?

Entrenchment of *doen*-schema
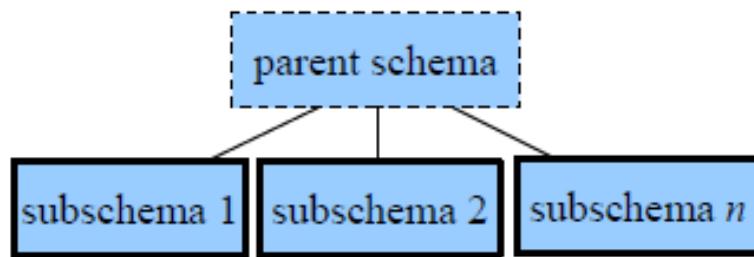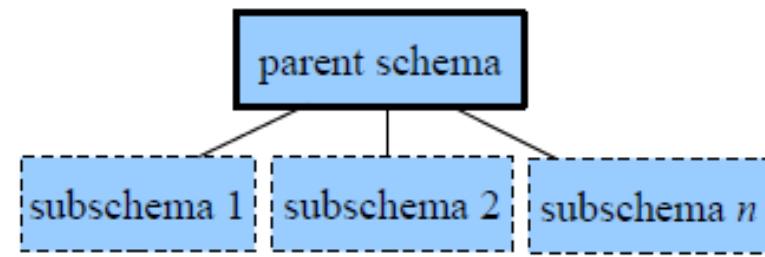
Productivity = entrenchment
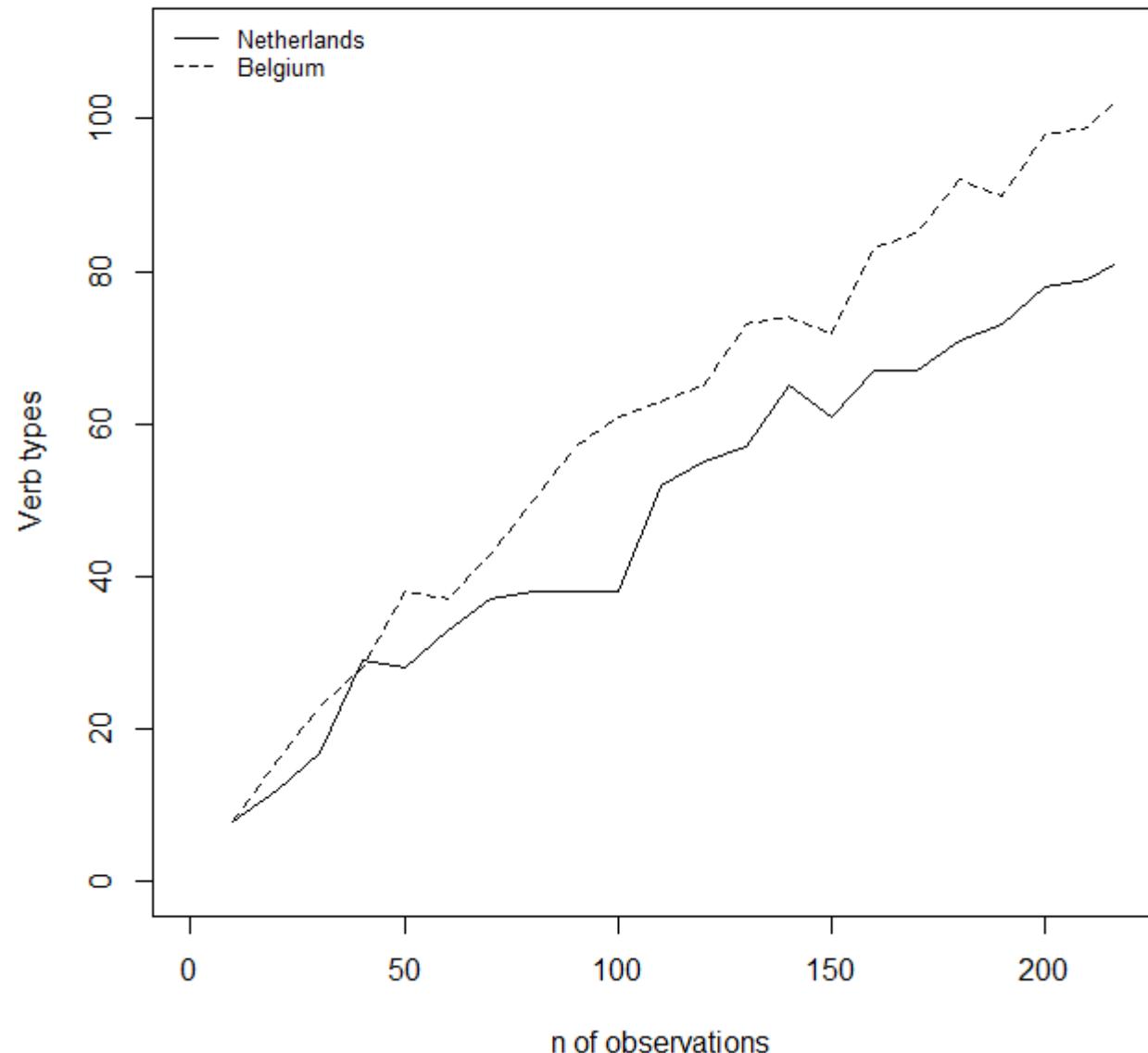
Diachronic evidence

- semantic shrinking of *doen* (Duinhoven 1994; Verhagen 2000)

- e.g. the inducive use (interpersonal causation) – more common in the 18th century, now *laten* is preferred

- *doen* is more frequent in formal texts (more archaic features)

- *doen* is more frequent in BE (a more archaic variety)

- is the weaker schema of *doen* in NL yet another symptom that it is done with *doen*?

Conclusions

- socially and geographically uniform taxonomic networks of constructions are rather an exception than a rule

- the global constructicon of a language is a dynamic heterogeneous network of networks (cf. the Internet)

- an analysis of lectal variation in the use of constructions should also include a comparison of the constructional networks

- this procedure can be a tool for tracking down ongoing constructional changes

- experimental support is needed (weights of variables)

Thank you!

natalevs@gmail.com