Construction Grammar meets Semantic Vector Spaces:

A radically data-driven approach to semantic classification of slot fillers

Natalia Levshina†

Kris Heylen‡

†University of Marburg, RC Deutscher Sprachatlas

‡University of Leuven,
RU Quantitative Lexicology and Variational Linguistics

Acknowledgments

- Most of the presented work was presented when the first author worked at the University of Leuven
- This research was partly funded by an FWO grant awarded to Dirk Geeraerts and Dirk Speelman

Outline

- quantitative approaches to constructional semantics: the problem of semantic classes
- distributional semantic models as a method of semantic classification
- experiments with nominal and verbal classes in Dutch *doen* and *laten* CCx
- discussion and future research

Quantitative Models of Syntactic Variation

semasiological perspective		onomasiological perspective	
collexeme-based	feature-based	collexeme-based	feature-based
collostructional analysis (Stefanowitsch & Gries 2003)	cluster analysis (Levshina 2012)	distinctive collexeme analysis (Gries & Stefanowitsch 2004)	logistic regression (Bresnan et al. 2007), classification trees (Baayen 2008), random forests (Levshina 2011)
		mixed-effect models (Baayen 2008)	

Quantitative Models of Syntactic Variation

semasiological pe…		…logical perspective	
collexeme-b…			…ture-based
collostr… analy… (Stefanowits… Gries 2003)			…ession …al. 2007), …tion trees …en 2008), …ndom forests (Levshina 2011)
		mixed-effect models (Baayen 2008)	

ALL approaches involve semantic classes,
either at the coding stage (feature-based analyses),
or at the interpretation stage (collexemes)

Why do we need semantic classes?

- Theoretically, learning constructions involves learning generalizations, such as semantic classes (cf. Goldberg 2006).
- Epistemologically, we are interested in the most parsimonious explanation.
- This might be a rare case when the interests of both speakers and linguists converge.

Semantic Classes: State of the Art

- as a rule, intuitive and subjective
- ‘standard’ classifications (e.g. Levin 1993, Garretson 2004):

  - not many
  - for English
  - incomplete
  - not tested empirically
- Gries and Stefanowitsch 2010: corpus-driven verb classes, but

  - limited contextual features (18 prepositions)
  - subjective evaluation

Semantic Classes: Desiderata

- data-driven, (potentially) entire vocabulary
- objective validation
- semantic relationships are multidimensional

  ⇨ different criteria of similarity
- varying schematicity of semantic relationships

  ⇨ different levels of granularity

Instead of working with one *a priori* classification,
let's compare different ones and see which works the best

Outline

- quantitative approaches to constructional semantics: the problem of semantic classes
- our proposal: distributional semantic models as a method of semantic classification
- experiments with nominal and verbal classes in Dutch doen and laten CCx
- discussion and future research

Our proposal

- a bottom-up quantitative approach based on distributional Semantic Vector Space models
- task-specific:

  - adjustable criteria of similarity

  - adjustable granularity
- validation in a real data set for near-synonymous doen and laten CCx (onomasiological perspective)

Semantic Vector Space Models

Standard technique in Computational Linguistics:

- corpus based, bottom-up clustering of semantically related words into semantic classes (Turney & Pantel 2010)

Based on the Distributional Hypothesis (Firth 1957):

- *You shall know a word by the company it keeps*
- Words appearing in similar contexts tend to have similar meanings

Method

- each word is assigned a vector stating the word's co-occurrence frequencies with a range of possible contexts
- words with similar context vectors have similar meanings

Semantic Vector Space Models

	gun	psychopat	knife	cruelly	lovingly	mother	lovers	toilet	...
kiss	2	2	0	1	89	56	98	3	
hug	3	1	2	5	77	49	88	0	
kill	10	59	67	69	0	8	12	1	
murder	97	65	58	81	1	9	9	2	
....									

- co-occurrence frequency of target words (rows) with context words (columns)
- High dimensional matrix (only small subset shown)

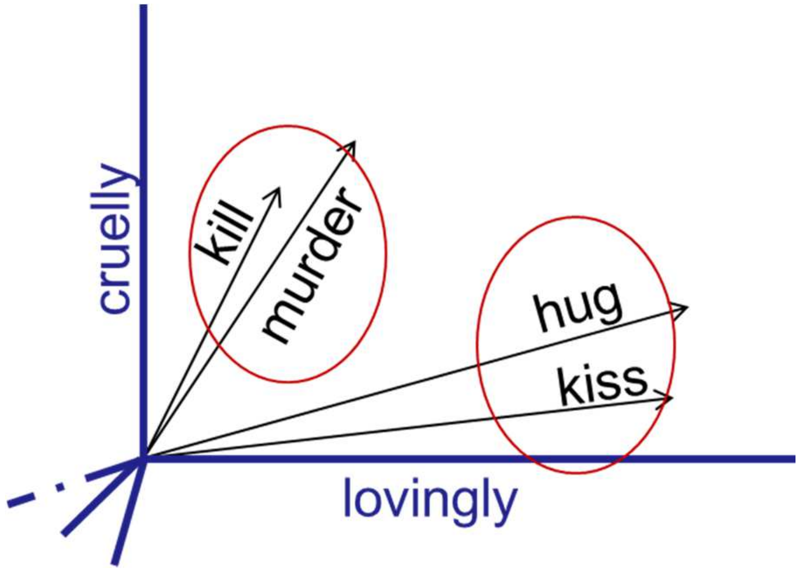
Semantic Vector Space Models

	gun	psychopat	knife	cruelly	lovingly	mother	lovers	toilet	...
kiss	2	2	0	1	89	56	98	3	
hug	3	1	2	5	77	49	88	0	
kill	10	59	67	69	0	8	12	1	
murder	97	65	58	81	1	9	9	2	
....									

- general overview of collactional behaviour and distributional properties of a language's vocabulary
- ≈ behavioural profiles (Divjak etal 2006) but many more features

Semantic Vector Space Models

- weighting of frequencies (pointwise mutual information)
- projection of vectors into a semantic "word space"
- measure proximity of vectors in space (cosine)
- cluster words based on vector proximity



Semantic Vector Space Models

SVS come in many different flavours

- Technical parameters (frequency weighting scheme, similarity measure, dimensionality reduction technique, clustering technique,...)
- Number of clusters → granularity of semantic distinctions
  - dependent on specific application (cf. infra)
- Definition of 'context'→ type of semantics captured
  - Document-based context features: topical relations
  - Window-based, bag-of-words context features: loose associations
  - Dependency-based context features: tight relations (synonymy)
  - Subcategorisation Frame features (verbs only): Levin-like classes

Bag of words model

	gun	psychopat	knife	cruelly	lovingly	mother	lovers	toilet	...
kiss	2	2	0	1	89	56	98	3	
hug	3	1	2	5	77	49	88	0	
kill	10	59	67	69	0	8	12	1	
murder	97	65	58	81	1	9	9	2	
....									

- context feature = co-occurring word in window left and right of target word
- looser, associative semantic relations: e.g. *doctor-hospital*

Bag of words model

	gun	psychopat	knife	cruelly	lovingly	mother	lovers	toilet	...
kiss	2	2	0	1	89	56	98	3	
hug	3	1	2	5	77	49	88	0	

The psychopath killed his victims with a blunt knife. ...

- context feature = co-occurring word in window left and right of target word
- looser, associative semantic relations: e.g. *doctor-hospital*

Dependency based model

	+PP with gun	+SU psychopat	+OBJ psychopat	+PP with knife	+ ADV cruelly	+ADV lovingly	+ SU mother	+SU Lovers	+PP on toilet	...
kiss	2	2	0	0	1	89	56	98	3	
hug	3	1	0	2	5	77	49	88	0	
kill	10	59	2	67	69	0	8	12	1	
murder	97	65	1	58	81	1	9	9	2	
....										

- context feature = word in specific syntactic dependency relation with target word
- tight semantic relations: *hospital - clinic*

Dependency based model

	+PP with gun	+SU psychopat	+OBJ psychopat	+PP with knife	+ ADV cruelly	+ADV lovingly	+ SU mother	+SU Lovers	+PP on toilet	...
kiss	2	2	0	0	1	89	56	98	3	
hug	3	1	0	2	5	77	49	88	0	
...										

The psychopath killed his victims with a blunt knife. ...

SU OBJ PP

- context feature = word in specific syntactic dependency relation with target word
- tight semantic relations: *hospital - clinic*

Subcat frame model

	SU	SU / OBJ	SU / OBJ / ADV	SU / PP	SU / OBJ / PP	SU / OBJ / IOBJ	…
kiss	2	2	0	0	1	89	
hug	3	1	0	2	5	77	
kill	10	59	2	67	69	0	
murder	97	65	1	58	81	1	
….							

- context feature = subcategorization frame co-occurring with target verb (only used for verbs!) (Schulte i.Walde 2006)
- Levin-like verb classes: e.g. *lie, stand, sit, lean*

Subcat frame model

	SU	SU / OBJ	SU / OBJ / ADV	SU / PP	SU / OBJ / PP	SU / OBJ / IOBJ	...
kiss	2	2	0	0	1	89	
....							

$\underbrace{\text{The psychopat}}_{\text{SU}}$ killed $\underbrace{\text{his victims}}_{\text{OBJ}}$ $\underbrace{\text{with a blunt knife}}_{\text{PP}}$. ...

SU / OBJ / PP

- context feature = subcategorization frame co-occurring with target verb (only used for verbs!) (Schulte i.Walde 2006)
- Levin-like verb classes: e.g. *lie, stand, sit, lean*

Semantic Vector Space Models

3 models form a continuum lexical to syntactic

purely lexical distributional information

⇩

lexical and syntactic (dependency) informtion

⇩

purely syntactic (subcat.) distributional properties

more intermediate forms, depending on

- number of dependency relations (e.g. arguments only)
- inclusion of some "lexical" info in subcat frames (e.g. prepositions or semantic noun classes)

Semantic Vector Space Models

3 models form a continuum lexical to syntactic

purely lexical distributional information

⇩

lexical and syntactic (dependency) informtion

The psychopath killed his victims with a blunt knife. ...

SU OBJ PP

more intermed… , depending on

- number of dependency relations (e.g. arguments only)
- inclusion of some "lexical" info in subcat frames (e.g. prepositions or semantic noun classes)

Semantic Vector Space Models

3 models form a continuum lexical to syntactic

purely lexical distributional information

⇩

lexical and syntactic (dependency) informtion

⇩

purely syntactic (subcat.) distributional properties

more intermediate forms, depending on

- number of dependency relations (e.g. arguments only)
- inclusion of some "lexical" info in subcat frames (e.g. prepositions or semantic noun classes)

Semantic Vector Space Models

3 models form a continuum lexical to syntactic

purely lexical distributional information

lexical and syntactic (dependency) informtion

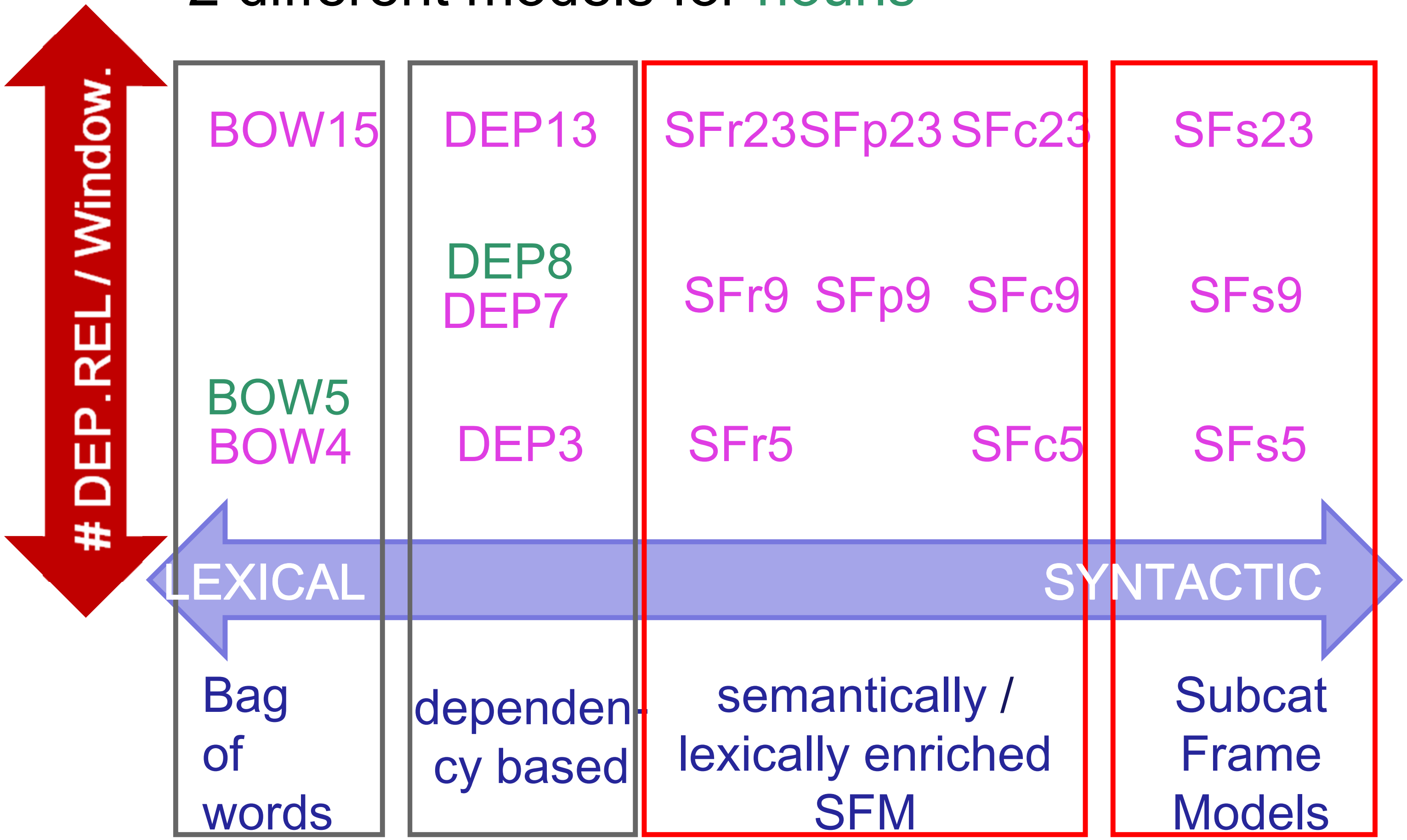
The psychopath killed his victims with a blunt knife. ...

SU_anim / OBJ_anim / PP_with

- numbe ents only)
- inclusion of some "lexical" info in subcat frames (e.g. prepositions or semantic noun classes)

Semantic Vector Space Models

- 16 different models for verbs
- 2 different models for nouns



Outline

- quantitative approaches to constructional semantics: the problem of semantic classes
- distributional semantic models as a method of semantic classification
- experiments with nominal and verbal classes in Dutch doen and laten CCx
- discussion and future research

Dutch causative constructions

Haar stem	***deed***	***het glas***	***barsten.***
her voice	**did**	**the glass**	**break**

Harry	***liet***	***het glas***	***barsten*****.**
Harry	**made**	**the glass**	**break**

Dutch causative constructions

Haar stem	***deed***	***het glas***	***barsten.***
her voice	**did**	**the glass**	**break**
Harry	***liet***	***het glas***	***barsten.***
Harry	**made**	**the glass**	**break**

Causer

Dutch causative constructions

Haar stem	***deed***	***het glas***	***barsten.***
her voice	**did**	**the glass**	**break**

Harry	***liet***	***het glas***	***barsten.***
Harry	**made**	**the glass**	**break**

Auxiliary

Dutch causative constructions

Haar stem	***deed***	***het glas***	***barsten.***
her voice	**did**	**the glass**	**break**

Harry	***liet***	***het glas***	***barsten*****.**
Harry	**made**	**the glass**	**break**

Causee

NWASV Nijmegen 2012

Dutch causative constructions

Haar stem	***deed***	***het glas***	***barsten.***
her voice	**did**	**the glass**	**break**

Harry	***liet***	***het glas***	***barsten.***
Harry	**made**	**the glass**	**break**

Effected Predicate

Classes of Cr and Ce

- data: Twente News Corpus
- 2 models:

  - purely lexical distributional information (bag of words)

  - lexical and syntactic dependency information

   (e.g. noun N as a subject of a verb V)
- granularity: from 2 to 100 classes emerging from hierarchical cluster analysis

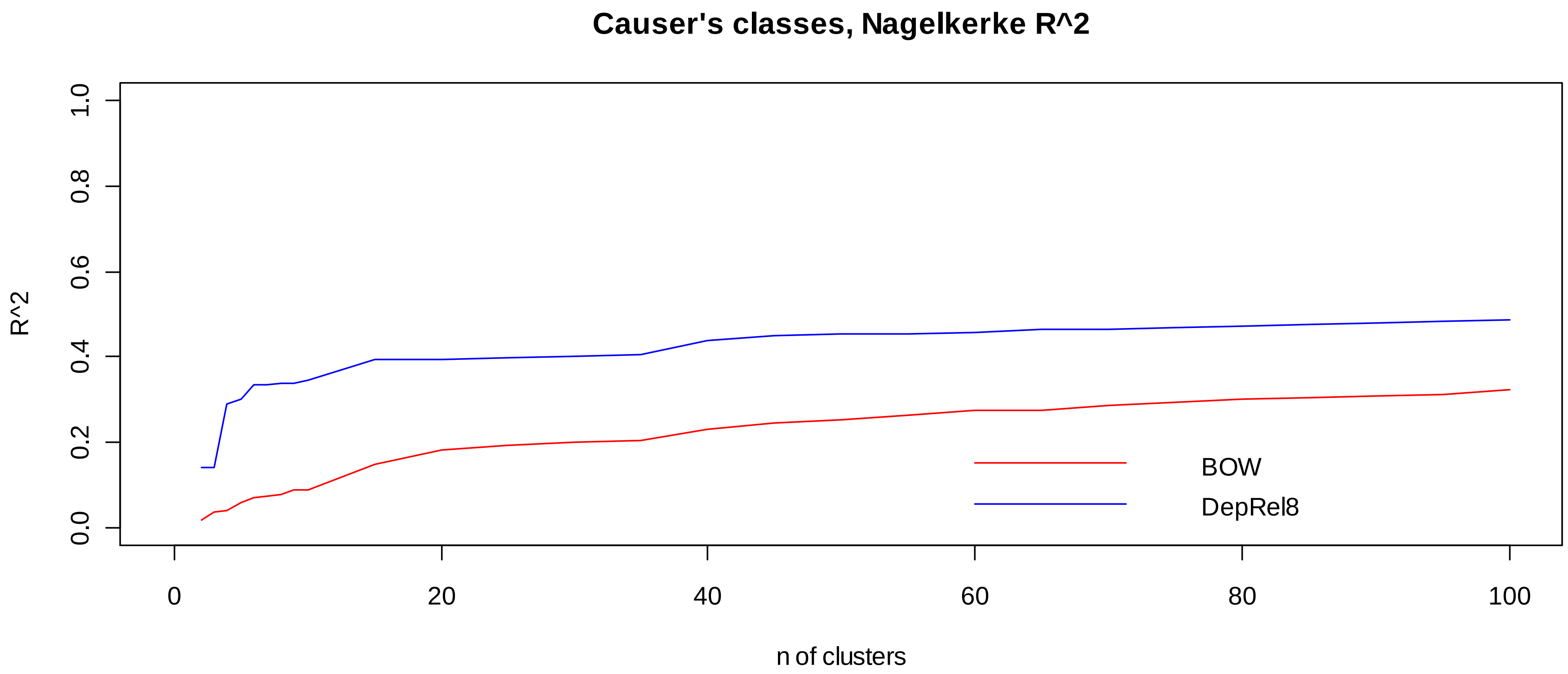
Classes of Effected Predicates

- 16 models, a continuum from purely lexical information to purely syntactic information (subcat frames)
- different granularity (number of clusters in hierarchical cluster analysis): from 5 to 100

Evaluation

- test set: 6800 obs. with causative doen and laten from Dutch newspaper corpora
- objects: all explicit non-pronominal fillers of Causer, Causee and Effected Predicate slots
- criterion: prediction of doen or laten in the observations
- method: logistic regression model, several indicators R^2, C, Somer's D_{xy}, *Gamma*, *Tau*, *AIC*

Prediction by Causer Classes

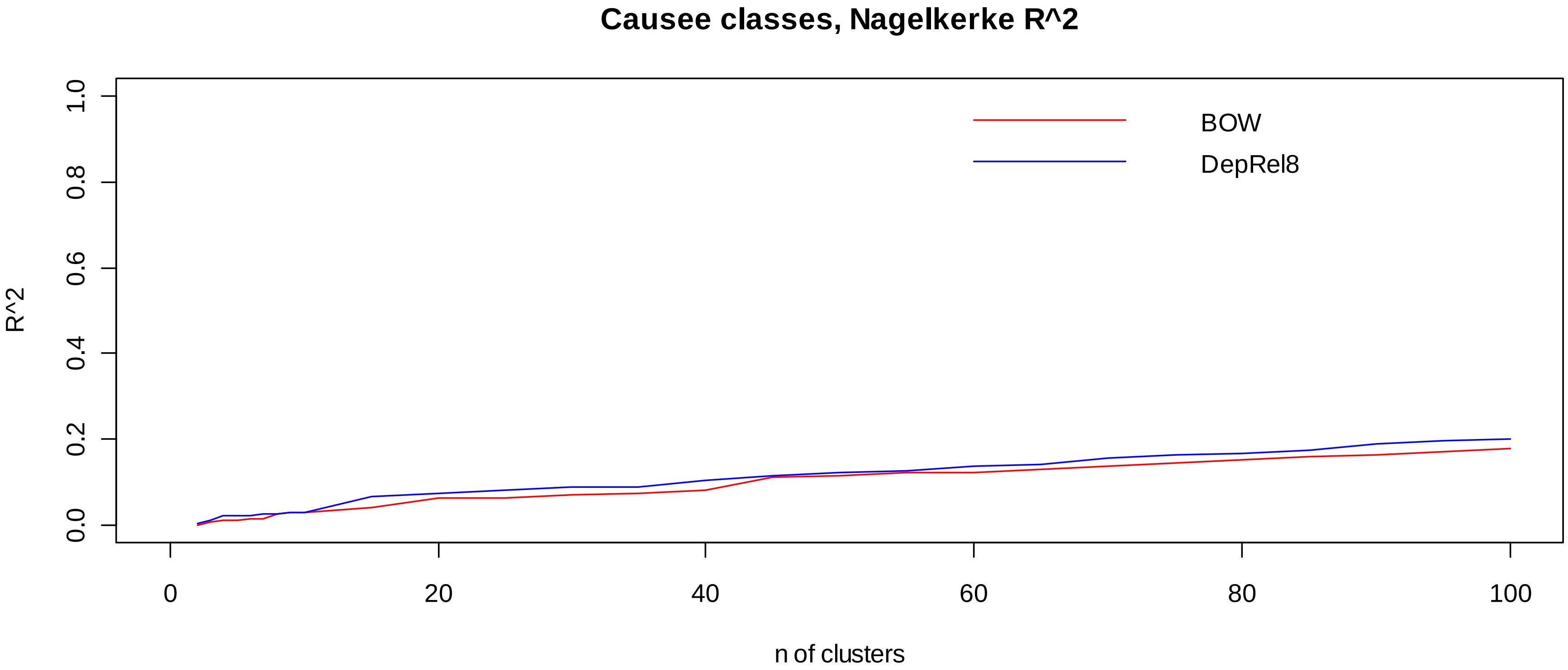


NWASV Nijmegen 2012

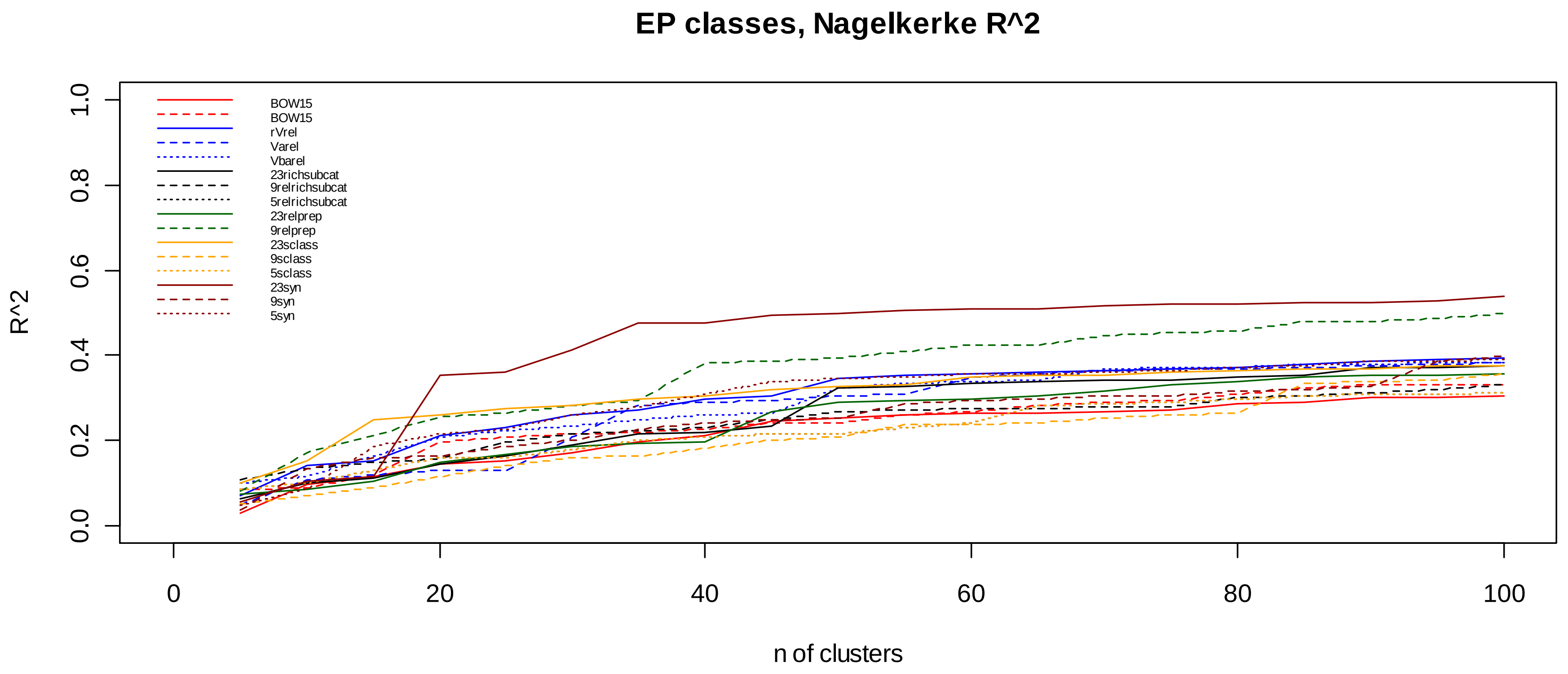
6 Clusters of Causer

No	Top freq nouns	doen or laten
1	cd, cijfer, plaat, herstel, stem, aanslag, afwezigheid, rentree, resultaat, aanpak	doen
2	Feyenoord, PSV, dirigent, speler, beurs, orkest, doelman, zanger, Van Gaal, componist	laten
3	Gergiev, Van Hecke, gemeente bestuur, Harnoncourt, Morissette, Pollini, secretaris generaal, AH To Go, alleskunner, Ax	laten
4	Verenigde Staten, VS, Amerika, Europa, Washington, geheel, tempo, Duitsland, Engels, India	laten
5	regering, minister, bedrijf, trainer, president, belegger, muziek, ploeg, premier, man	laten
6	Mahlerstem, zaal technicus, beroepskader, Neal Evans, Bastuba en contrabasclarinet, Tuomarila, Wanderlied	NA

Prediction by Causee Classes



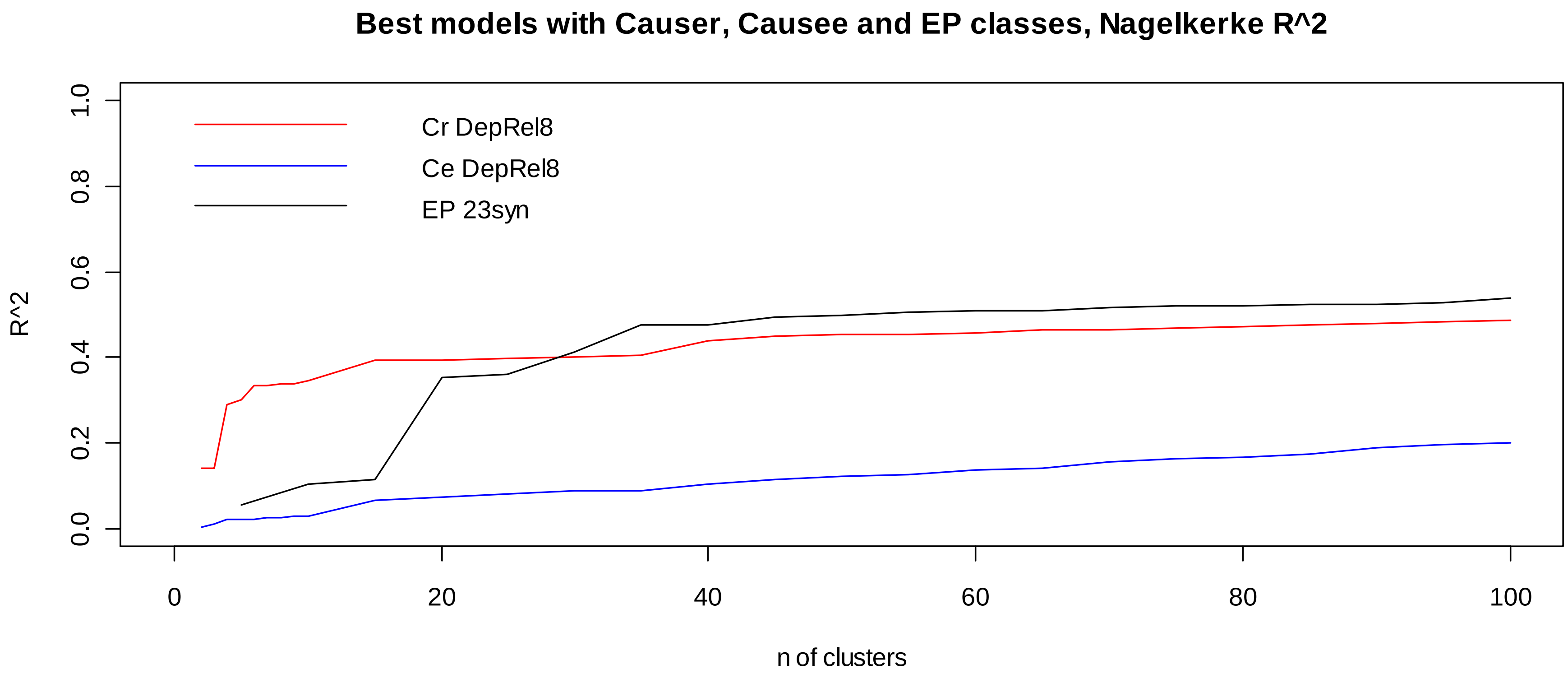
Prediction by Effected Predicate



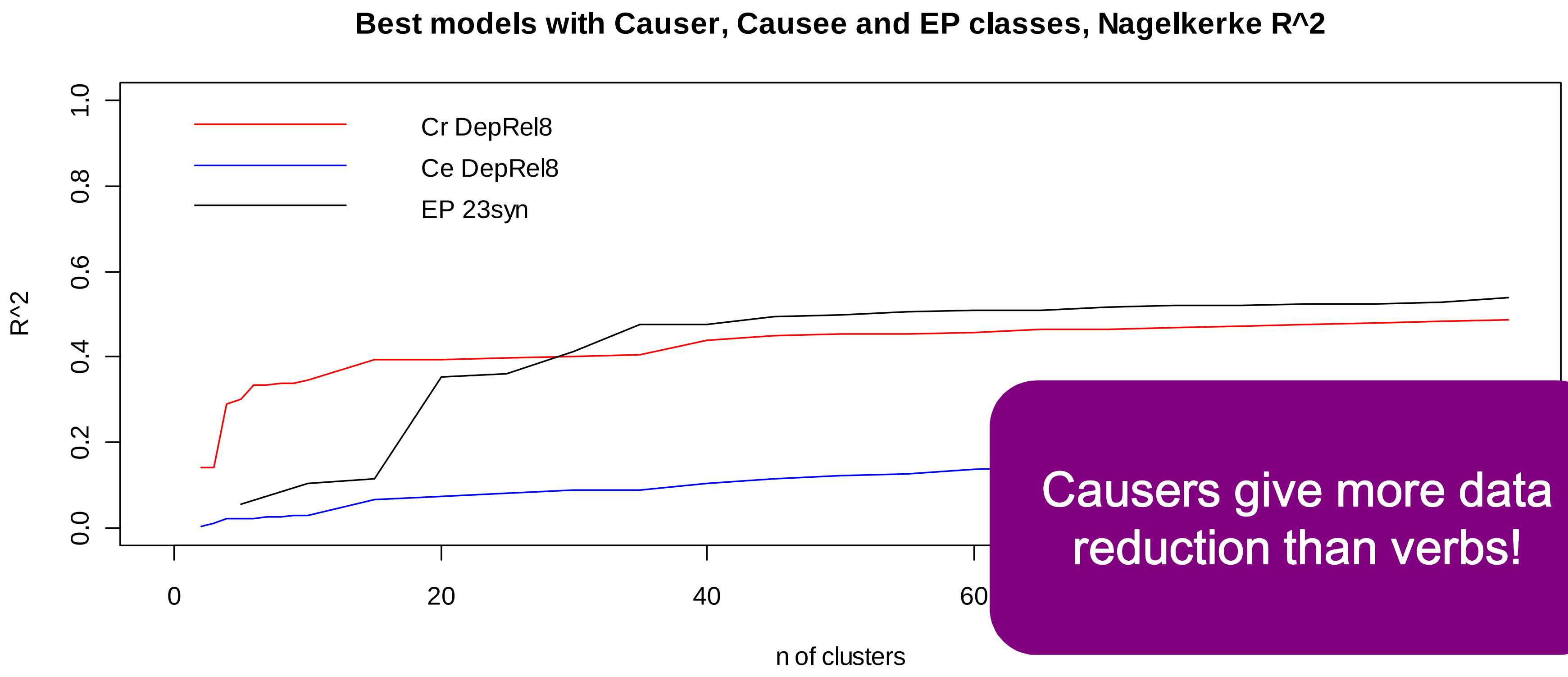
35 Clusters for EP (Examples)

No	Top freq verbs	doen or laten
3	herleven, kantelen, versmelten, beven, samensmelten, verslappen, sidderen, smelten, instorten, vervagen	doen
24	stijgen, zakken, dalen, groeien, overlopen, rijzen, terugzakken, wankelen, daveren, overvloeien	doen
25	denken, vermoeden, geloven, besluiten, vrezen, hopen, verzuchten, toegeven, betogen, concluderen	doen
23	zien, weten, leiden, maken, doen, blijken, voelen, kennen, worden, schijnen	laten
12	horen, verstaan, betrappen, afleiden, merken, delen, rijmen, verkiezen, verwachten, associëren	laten
34	liggen, vallen, gaan, komen, lopen, staan, zitten	laten

The best models for 3 Slots

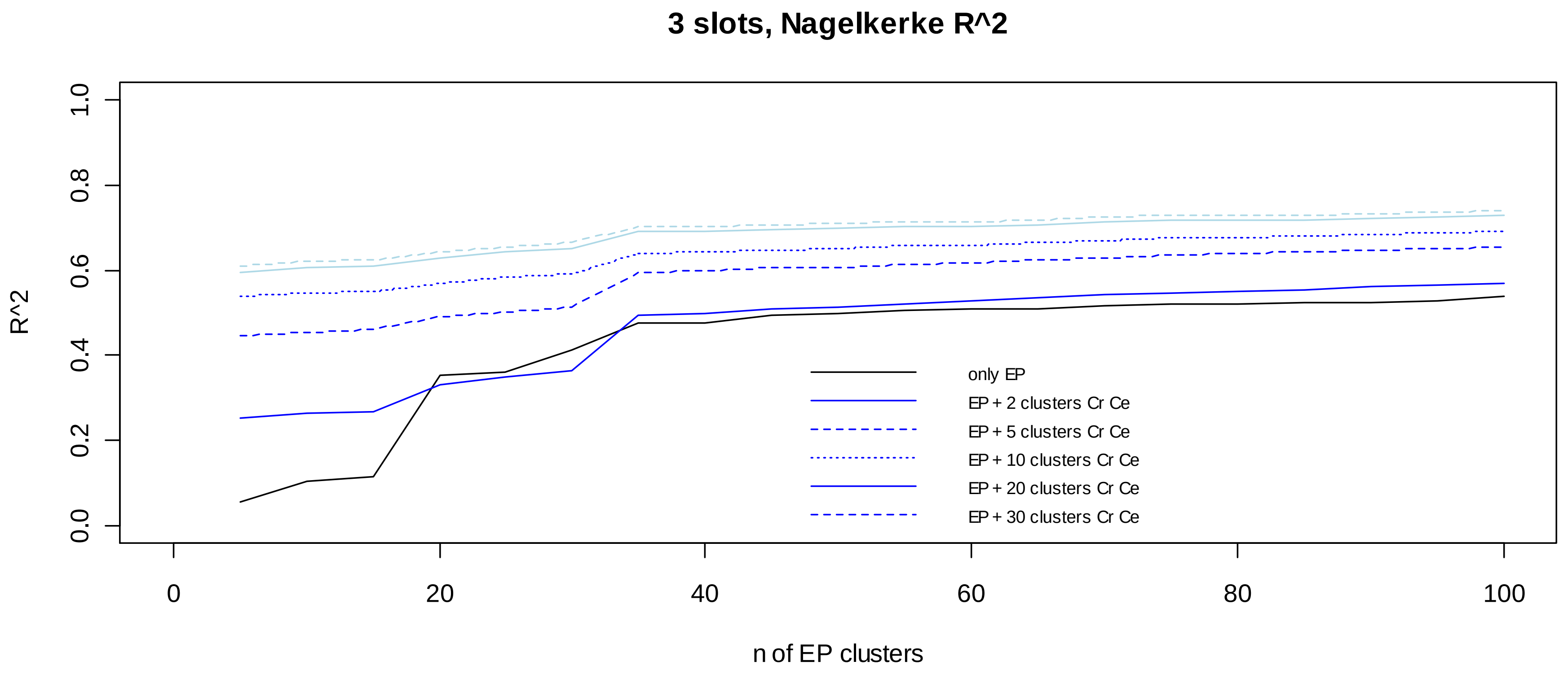


The best models for 3 Slots

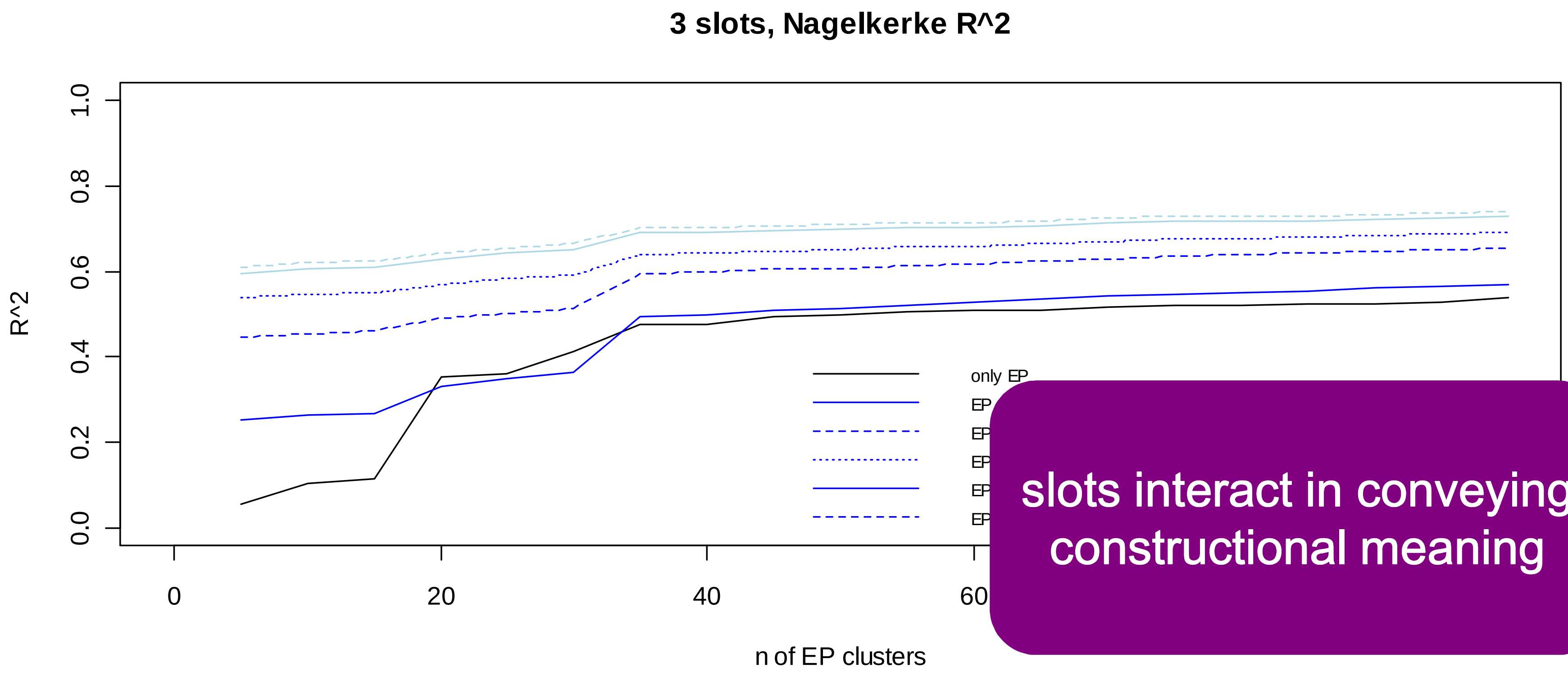


Causers give more data reduction than verbs!

All three slots in one model



All three slots in one model



slots interact in conveying constructional meaning

Outline

- quantitative approaches to constructional semantics: the problem of semantic classes
- distributional semantic models as a method of semantic classification
- experiments with nominal and verbal classes in Dutch doen and laten CCx
- discussion and future research

Desiderata Revisited

- data-driven, (potentially) entire vocabulary
- objective validation
- semantic relationships are multidimensional
  ⇨ different criteria of similarity
- varying schematicity of semantic relationships
  ⇨ different levels of granularity

Desiderata Revisited

- data-driven, (potentially) entire vocabulary

  bottom-up classes work!

- objective validation

- semantic relationships are multidimensional

  ⇨ different criteria of similarity

- varying schematicity of semantic relationships

  ⇨ different levels of granularity

Desiderata Revisited

- data-driven, (potentially) entire vocabulary

  bottom-up classes work!

- objective validation

  Cr and EP classes perform better than Ce

- semantic relationships are multidimensional

  ⇒ different criteria of similarity

- varying schematicity of semantic relationships

  ⇒ different levels of granularity

Desiderata Revisited

- data-driven, (potentially) entire vocabulary

  bottom-up classes work!

- objective validation

  Cr and EP classes perform better than Ce

- semantic relationships are multidimensional

  ⇒ different criteria of similarity

  syntax-sensitive models perform the best

- varying schematicity of semantic relationships

  ⇒ different levels of granularity

Desiderata Revisited

- data-driven, (potentially) entire vocabulary

  bottom-up classes work!

- objective validation

  Cr and EP classes perform better than Ce

- semantic relationships are multidimensional

  ⇨ different criteria of similarity

  syntax-sensitive models perform the best

- varying schematicity of semantic relationships

  ⇨ different levels of granularity

  nouns ‘need’ less classes than verbs

Future research

- the bottom-up construction-specific classes are similar to the classes found in the literature. Are semantic classes cross-constructionally (cross-linguistically) stable?
  - compare with other constructions
  - compute validity measures for different clustering solutions (e.g. silhouette widths)
- a solution for semasiological studies

Thank you!

for further information:

natalevs@gmail.com

kris.heylen@arts.kuleuven.be