



# Checking the adequacy of second-order vector space models of meaning

Mariana Montes, Dirk Geeraerts, Dirk Speelman, Kris Heylen



RU Quantitative Lexicology and Variational Linguistics

# On the menu

- narrowly defined:  
examine the effect of ppmi weighting of context words (and other parameter settings) in token-based vector space models
- more broadly:  
introduce the Nephological Semantics project to the CL community

# Nephological Semantics

- a research project within the QLV team of the university of Leuven, aimed at developing and testing vector space models for semantic and lexical analysis
- staff members: Dirk Geeraerts, Dirk Speelman, Benedikt Smzrecsanyi, Stefania Marzo
- postdocs: Kris Heylen, Weiwei Zhang, Karlien Franco
- PhD students: Thomas Wielfaert, Stefano De Pascale, Mariana Montes
- informatician: Tao Chen

# Nephological Semantics

- theoretically, a corpus-based scaling up of the categorization model introduced in Geeraerts et al. 1994, [The Structure of Lexical Variation](#) (CLR 5)
  - > branches:
    - semasiology: sense identification (this talk)
    - formal onomasiology: synonym identification and lexical lectometry (Weiwei Zhang's talk on Tuesday)
    - conceptual onomasiology: construal

# Nephological Semantics

why ‘nephology’?

the study of the shape of clouds

here:

- represent word usages as points in a semantic space  
(so that semantically similar tokens belong together)
- see what the shape of the resulting cloud of points tells  
us about the polysemy of the word

# Steps to take

- an introduction to the essentials of type-based and token-based distributional semantics (and the variety of relevant context features)
- a hands-on demonstration of the tool developed in the Nephological Semantics project for the visual analytics of vector models

The Workflow

# DISTRIBUTIONAL SEMANTICS



# Type level vectors

	protestant	tower	cloud	faculty	sovereign	beyond	join	mission
member								
state								
go								



# Type level vectors

	protestant	tower	cloud	faculty	sovereign	beyond	join	mission
member	0.0	0.0	0.0	4.42	0.0	0.0	0.0	1.2
state	0.0	1.60	0.0	0.0	3.59	0.0	1.08	0.78
go	1.59	0.0	0.99	0.0	0.0	2.81	1.12	1.29

# Type level vectors

	protestant	tower	cloud	faculty	sovereign	beyond	join	mission
member	0.0	0.0	0.0	4.42	0.0	0.0	0.0	1.2
state	0.0	1.60	0.0	0.0	3.59	0.0	1.08	0.78
go	1.59	0.0	0.99	0.0	0.0	2.81	1.12	1.29

## PARAMETERS

- Window size
- Corpus
- Association measure (PPMI?)
- Number of context words
- ...

# Token level vectors

	member	steeple	state	go	separation	school
church1						
church2						
church3						
church4						
church5						

# Token level vectors

	member	steeple	state	go	separation	school
church1	0	0	1	0	1	0
church2	0	0	1	0	0	0
church3	0	0	0	0	0	1
church4	1	0	0	0	0	0
church5	0	0	0	1	0	0

# Token level vectors

	member	steeple	state	go	separation	school
church1	0	0	0.0	0	3.27	0
church2	0	0	0.0	0	0	0
church3	0	0	0	0	0	0.65
church4	2.41	0	0	0	0	0
church5	0	0	0	0.29	0	0

# Token level vectors

## PARAMETERS

- Window size
- Association measure (PPMI?)
- Number of context words
- Part of speech
- ...

	member	steeple	state	go	separation	school
church1	0	0	0.0	0	3.27	0
church2	0	0	0.0	0	0	0
church3	0	0	0	0	0	0.65
church4	2.41	0	0	0	0	0
church5	0	0	0	0.29	0	0

## Second order vectors

	protestant	tower	cloud	faculty	sovereign	beyond	join	mission
member	0.0	0.0	0.0	4.42	0.0	0.0	0.0	1.2
state	0.0	1.60	0.0	0.0	3.59	0.0	1.08	0.78
go	1.59	0.0	0.99	0.0	0.0	2.81	1.12	1.29

	member	steeple	state	go	separation	school
church1	0	0	1	0	1	0
church2	0	0	1	0	0	0
church3	0	0	0	0	0	1
church4	1	0	0	0	0	0
church5	0	0	0	1	0	0

# Second order vectors

	protestant	tower	cloud	faculty	sovereign	beyond	join	mission
member	0.0	0.0	0.0	4.42	0.0	0.0	0.0	1.2
state	0.0	1.60	0.0	0.0	3.59	0.0	1.08	0.78
go	1.59	0.0	0.99	0.0	0.0	2.81	1.12	1.29

	member	steeple	state	go	separation	school
church1	0	0	1	0	1	0
church2	0	0	1	0	0	0
church3	0	0	0	0	0	1
church4	1	0	0	0	0	0
church5	0	0	0	1	0	0



# Second order vectors

	protestant	tower	cloud	faculty	sovereign	beyond	join	mission
member	0.0	0.0	0.0	4.42	0.0	0.0	0.0	1.2
state	0.0	1.60	0.0	0.0	3.59	0.0	1.08	0.78
go	1.59	0.0	0.99	0.0	0.0	2.81	1.12	1.29

	member	steeple	state	go	separation	school
church1	0	0	0.0	0	3.27	0
church2	0	0	0.0	0	0	0
church3	0	0	0	0	0	0.65
church4	2.41	0	0	0	0	0
church5	0	0	0	0.29	0	0

# Second order vectors

	protestant	tower	close	sovereign	beyond	join	mission
member	0.0	0.0	0.0	0.0	0.0	0.0	1.2
state	0.0	1.60	0.0	3.59	0.0	1.08	0.78
go	1.59	0.0	0.99	0.0	2.81	1.12	1.29

**S OCC**

	member	state	separation	school
church1	0	0	3.27	0
church2	0	0	0	0
church3	0	0	0	0.65
church4	2.41	0	0	0
church5	0	0	0.29	0

**F OCC**

S OCC

F OCC

## PARAMETERS

- Window size
- Association measure (PPMI?)
- Number of context words
- Part of speech
- ...

## PARAMETERS

- Window size
- Association measure (PPMI?)
- Number of context words
- Part of speech
- ...

# So, our challenge:

Which **context words** should we use to represent a **token**?

Youth obeyed when commanded. It went to **church** on Sunday and one Saturday a month went to confession.

Source: File p01 (Semcor), token 222

Which **context words** should we use to represent a **token**?

Youth obeyed when commanded. It went to **church** on Sunday and one Saturday a month went to confession.

Source: File p01 (Semcor), token 222

The whole text?

[...] Youth obeyed when commanded. It went to **church** on Sunday and one Saturday a month went to confession. [...]

Which **context words** should we use to represent a **token**?

Youth obeyed when commanded. It went to **church** on Sunday and one Saturday a month went to confession.

Source: File p01 (Semcor), token 222

A 5:5 window?

[...] **commanded. It went to church on Sunday and one Saturday** [...]

Which **context words** should we use to represent a **token**?

Youth obeyed when commanded. It went to **church** on Sunday and one Saturday a month went to confession.

Source: File p01 (Semcor), token 222

According to their PPMI value?

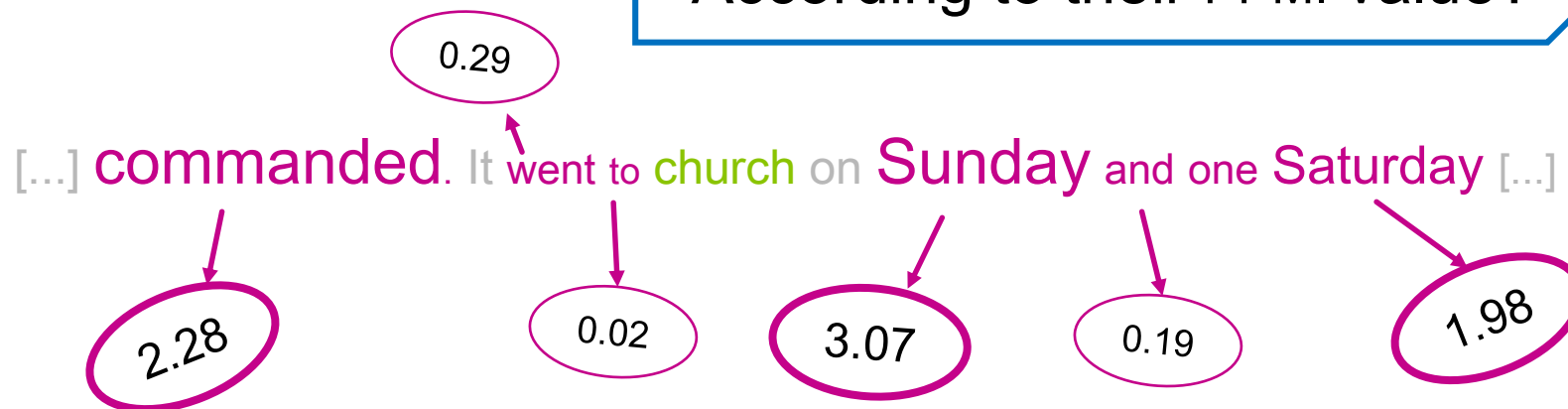
[...] **commanded**. It **went to church** on **Sunday and one Saturday** [...]

Which **context words** should we use to represent a **token**?

Youth obeyed when commanded. It went to **church** on Sunday and one Saturday a month went to confession.

Source: File p01 (Semcor), token 222

According to their PPMI value?





# What does this look like?

<http://delhikiev.github.io/>



# Discussion

- ☁ BETWEEN types, ppmi values are useful
- ☁ WITHIN types, ppmi values might hide the very distinctions we would like to highlight
- ☁ Other parameters should be used to select the first order context words (window size, part-of-speech, dependency relations...)

# Rounding up

we have

- introduced the semasiological workflow of the Nephological Semantics project
- shown how that workflow functions as a diagnostic tool for examining the effect of parameter settings (such as the selection and weighting of context words) in a token-based distributional approach

# Rounding up

next steps to take include

- using the procedures and visualizations illustrated here to optimize parameter settings for semasiological analysis (e.g. in terms of word class)
- further integrate the semasiological and onomasiological lines of enquiry within the project

Thank you!

ご清聴  
ありがとうございました

mariana.montes@kuleuven.be  
dirk.geeraerts@kuleuven.be

<http://delhikiev.github.io>