

# Times Change

## Tracking Latin Time Metaphors with Distributional Semantic Models

Krzysztof Nowak

*Department of Medieval Latin*  
*Institute of Polish Language*  
Polish Academy of Sciences

September 20 2017

# Times Change

~~Tracking Latin Time Metaphors with Distributional Semantic Models~~  
A Poor Historical Lexicographer's Road to Distributional Semantic Models

Krzysztof Nowak

*Department of Medieval Latin*  
*Institute of Polish Language*  
Polish Academy of Sciences

September 20 2017

# Motivation and Outline

## Motivation

- general: assessing interpretability of distributional semantics for everyday lexicography of an under-resourced language
- specific: tracking change in metaphorical patterns

## Outline

- 1 What is distributional semantics ...
- 2 ... and why do we need it so badly?
- 3 Corpora
- 4 CMT and the study of metaphor
- 5 Lat. *tempus* 'time': clustering collocates
- 6 Conclusions

# Method: Distributional Semantics

## Procedure (Harispe 2017)

- choose your context (doc, par, sentence ..)
- weight raw frequencies (PMI...)
- reduce sparse matrix (SVD...)
- measure distance (cosine ...)

## Distributional Hypothesis

"the degree of semantic similarity between two words (or other linguistic units) can be modelled ... as a function of the degree of overlap among their linguistic contexts"  
(Baroni and Lenci 2010)

	sacer 'saint'	lego 'to read'
liber 'book'	73	422
uolumen 'volume'	234	158
bellum 'battle'	38	16

Table: Cooccurrence Matrix

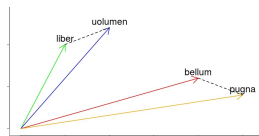
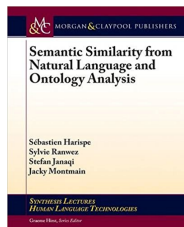
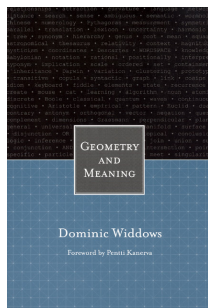


Figure: Measuring Distance

# More on DSMs

## General Introductions

- general: Turney and Pantel 2010
- accessible: Widdows 2004



## Survey

- Harispe et al. 2015 (cf. 2017)

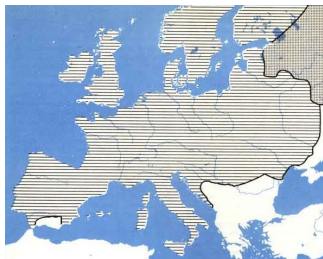
## DSMs in Linguistics

- Peirsman/Heylen et al. 2008-

# Why this approach?

## Poor resources

- a (potentially) huge corpus
- underserved with regard to the NLP tools
- no representative corpus at the moment



## No sound evaluation framework

- no semantically annotated corpora
- no similar tools to the TOEFL dataset etc.
- no native speakers
- visual inspection relies on subjective assessment that the DSM was expected to eliminate

# The Method

## Workflow

### 1 frequency table

Term	Property	Freq
doceo	iudex	4
periculum	repello	3
visus	forma	4
addo	membrum	2

### 2 wordspace package by Stefan Evert:

- weighting raw counts
- matrix reduction

### 3 word similarity functions

- `pair.distances()`
- `nearest.neighbours()`

## Corpora

### Classical

- time: 2 BC - 6 AD
- size: 8M

### Early Christian

- time: 1 AD - 6 AD
- size: 22M

### EC & Medieval

- time: 1 AD - 13 AD
- size: 100M

# The Problem: TIME and its metaphors

CMT (Lakoff & Johnson 1980, 1989, 1999)

- more than stylistic device
- source domain → target domain
- abstracts regularly comprehended in more concrete terms

Evans (2006, 2013) on time

- metaphorical expression → lexical concept
- activation: formal and semantic properties
- ex.: *to loose one's time* (mass noun, interest-related verb) → COMMODITY Sense

## Procedure

- 1 wide collocational window  
vos tot milibus civium Romanorum<sup>5L</sup> uno<sup>4L</sup> nuntio<sup>3L</sup> atque<sup>2L</sup> uno<sup>1L</sup> tempore necatis<sup>1R</sup> quo<sup>2R</sup> tandem<sup>3R</sup> animo<sup>4R</sup> esse<sup>5R</sup> debetis
- 2 association strength
  - Dice:  $\frac{2 \cdot f_{AB}}{f_A + f_B}$
- 3 empirical collocations (Evert 2009) → clustering

# Collocational stability (Classical corpus)

Attested in 5 periods  
vernus ('springtime')

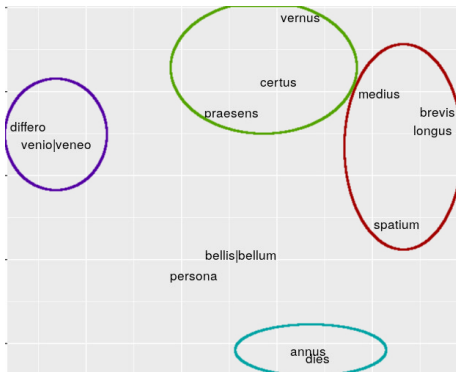
Attested in 4 periods

annus ('year')  
brevis ('short')  
certus ('certain')  
dies ('day')  
differo ('postpone')  
longus ('long')  
praesens ('present')  
spatium ('space')

Attested in 3 periods

bellum ('war')  
medius ('middle')  
persona ('person')  
venio ('to come')

# Stable collocations (Classical Corpus)



## TIME IS SPACE

*longus* 'long', *brevis* 'short', *medius* 'middle', *spatium* 'space'

## Syntagmatic similarity

*certus* 'certain', *praesens* 'present'

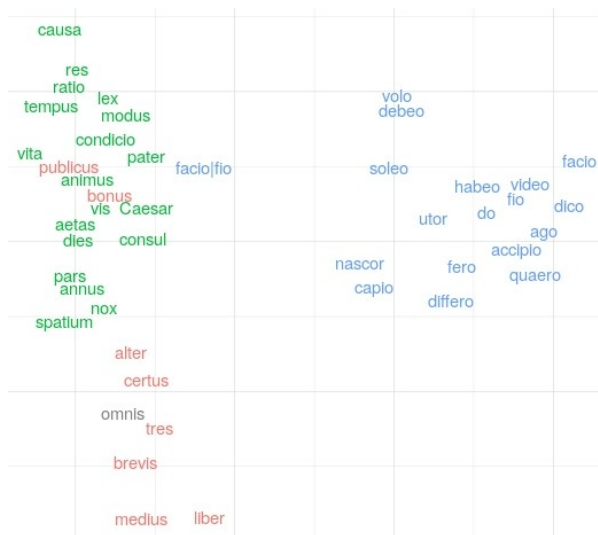
## Paradigmatic similarity: hyponyms

*annus* 'year', *dies* 'day'

## MOVING TIME?

*venio* 'to come', *differo* 'to postpone'

# All collocations (Classical Corpus)



nouns  
verbs  
adjectives

# Nominal collocations (Classical Corpus)



## TIME IS SPACE

*spatium* 'space', *locus* 'place'

Paradigmatic similarity: hyponyms and near synonyms

*annus* 'year', *dies* 'day', *nox* 'night', *aetas* 'epoch'...

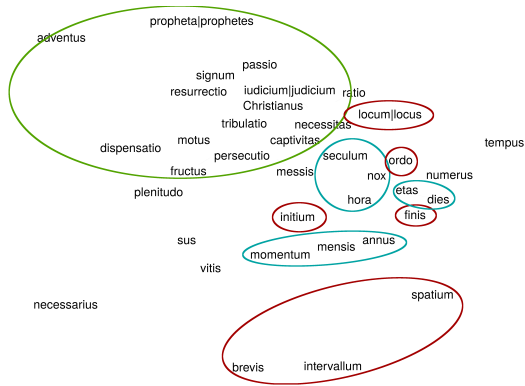
## Time Reference

*Caesar* 'Cesar', *consul* 'consul' (ex. 'under consul X')

## Legal uses

*vis* 'power', *lex* 'law', *condicio* 'condition', *testamentum* 'testament'

# Nominal collocations (Early Christian)



## TIME IS SPACE

*spatium* 'space', *intervallum* (etym. 'between walls'), but also: *initium* 'beginning', *finis* 'end'

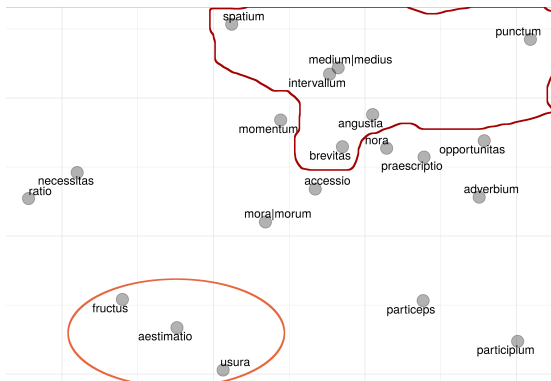
## Paradigmatic similarity: hyponyms

*annus* 'year', *saeculum* 'century', *aetas* 'epoch'...

## Time Reference

*resurrectio* 'resurrection', *persecutio* 'prosecution', *captivitas* 'captivity'...

# Syntactic constraints: *temporis* 'of time'



## TIME IS SPACE

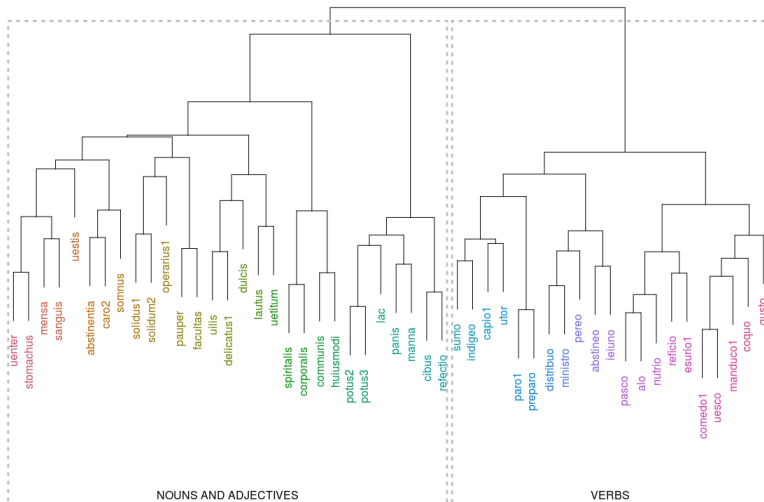
*spatium* 'space', *intervallum* 'interval'  
(etym. 'between walls'), *angustia*  
'narrowness', *brevitas* 'shortness'

## TIME IS MONEY

*fructus* 'income', *aestimatio*  
'evaluation', *usura* 'profit'

# Semantic constraints: FOOD terminology

Sense Clusters of 50 CIBUS Collocations





# Conclusions

- testing, testing, testing...
- building evaluation framework
- towards automatic comparison of vectors between corpora
- building tools
  - dictionary
  - testing framework



## Our projects

### eCorpus of ML



[scriptores.pl/en/efontes](http://scriptores.pl/en/efontes)

### eDictionary of ML



[scriptores.pl/elexicon](http://scriptores.pl/elexicon)

# References

- Baroni, M., & Lenci, A. (2010). Distributional Memory: A General Framework for Corpus-based Semantics. *Computational Linguistics*, 36(4), 673–721.
- Evans, V. 2003. *The structure of time: language, meaning, and temporal cognition*. Amsterdam - Philadelphia: John Benjamins.
- Evans, V. 2013. *Language and time: a cognitive linguistics approach*. Cambridge: Cambridge University Press.
- Evert, S. (2014). Distributional Semantics in R with the `wordspace` Package. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations* (pp. 110–114). Dublin: Dublin City University and Association for Computational Linguistics.
- Harispe, S., Ranwez, S., Janaqi, S., & Montmain, J. (2015). *Semantic similarity from natural language and ontology analysis*. San Rafael: Morgan & Claypool.
- Heylen, K., Peirsman, Y., Geeraerts, D., & Speelman, D. (2008). Modelling word similarity: an evaluation of automatic synonymy extraction algorithms. In *Proceedings of the sixth international language resources and evaluation (LREC 2008)* (pp. 3243–3249).
- Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, G., & Turner, M. (1989). *More than cool reason: a field guide to poetic metaphor*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh. The embodied mind and its challenge to Western thought*. New York: Basic Books.
- Turney, P. D., & Pantel, P. (2010). *From Frequency to Meaning: Vector Space Models of Semantics*. ArXiv:1003.1141 [Cs].
- Widdows, D. (2004). *Geometry and meaning*. Stanford: CSLI Publications.