Static and contextual embeddings for tracing semantic change: the case of Christian Latin

Valentina Lunardi
vlunardi@g.ucla.edu

Barbara McGillivray
barbara.mcgillivray@kcl.ac.uk

Semantic change

WHAT IS IT

- Change in the meaning of words over time
- OE *mete* ‘food’ > PDE *meat*
  - “narrowing”
- Lat. *salārium* ‘(soldier’s) allotment of salt’ > ‘(soldier’s) wages’ > ‘wages’
  - “widening”
- All languages have evidence of this phenomenon

UCLA

KING'S College LONDON

Semantic change

STATUS QUO ANTE

- For a long time, the “black sheep” of historical linguistics:

  “Any attempt at a systematic study of semantic change, in fact, will yield only limited rewards, for two reasons: with rare […] exceptions, ***semantic change is completely patternless*** […]”

  Sihler 2000: 94 (emphasis mine)

- Why?

Semantic change

STATUS QUO ANTE

1) The motivations behind semantic change:
   - Numerous, varied, and often extra-linguistic
2) The method available to trace semantic change:
   - Tracing the meaning of each individual word within a language across time through close-reading of texts
     - Slow!!
   - Each language has tens of thousands of words, making it difficult to find trends of change (and therefore predictability)

Semantic change

WHAT HAS CHANGED

- Availability of digital textual corpora
- Quantitative, statistical, and machine-learning methods have been adopted to analyze digitized textual data
- Methods to detect semantic change within diachronic corpora have been developed under this push (see e.g. Tahmasebi et al. 2021)
- Generally motivated both by the desire to
  1) provide data-driven linguistic knowledge
  2) apply an interesting modelling tool to a complex data type

Semantic change

WHAT HAS CHANGED

- …but these changes might eventually have an impact on the theory of semantic change as well
- Hundreds of texts can be processed in a matter of minutes/hours

  ⇒ Is it achievable to try and find ***trends of change*** ?
- Not our goal per se: this would require comparing change across languages to form conclusions about typological trends
- We can, however, contribute to this in the form of:
  - Tests/refinements of the methodology for historical languages
  - Data about change within a specific language

Introduction to word embeddings

FUNDAMENTAL IDEAS

- The distributional hypothesis:

  "The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear" (Lenci 2008: 3)

- Vector representation of words (Jurafsky and Martin 2023: 106–7)

Introduction to word embeddings

Table 1 Co-occurrence matrix of the words *strawberry, blackberry, blueberry,* and *computer* with the context words *all, eat, nice, office, phone, raspberry, Samsung, software* in the BNC Spoken corpus

	all	eat	nice	office	phone	raspberry	Samsung	software
strawberry	9	6	5	0	0	5	0	0
blackberry	0	0	3	0	3	5	4	0
blueberry	0	5	4	0	0	7	0	0
computer	0	0	0	3	12	0	0	4

McGillivray, B. (2022, Jul 12). How to Use Word Embeddings for Natural Language Processing. SAGE Publications Ltd. https://doi.org/10.4135/9781529609578

Introduction to word embeddings

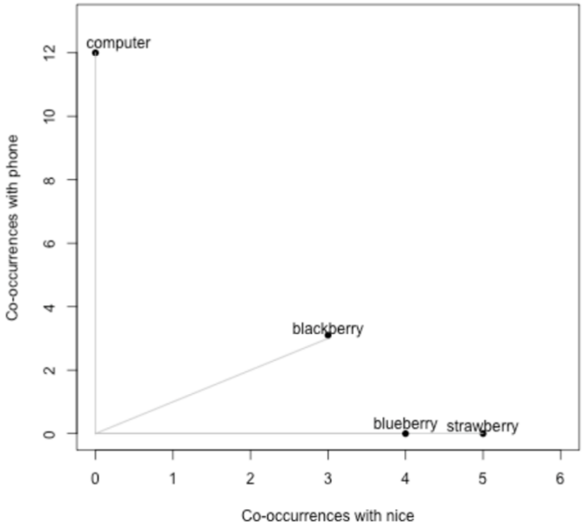


Figure 1. Vector representation of the words *strawberry*, *blackberry*, *blueberry,* and *computer* in a bi-dimensional space.

The horizontal axis corresponds to the concurrence with *nice*, and the vertical axis corresponds to co-occurrences with *phone*. *Strawberry* will have the coordinates (5,0), *blackberry* (3,3), *blueberry* (4,0), and *computer* (0,12).

McGillivray, B. (2022, Jul 12). How to Use Word Embeddings for Natural Language Processing. SAGE Publications Ltd. https://doi.org/10.4135/9781529609578

Introduction to word embeddings

THE NEWER MODELS

- Vectors still used to represent the meaning of words, but coordinate values are found automatically
  - This is done through “training”
- Training is subject to certain pre-set parameters, which vary according to the type of model used:
  - Window size for co-occurrence
  - Minimum frequency of a word for inclusion in training
  - Epochs
  - Learning rate
  - …

Introduction to word embeddings

THE NEWER MODELS

- In earlier models, coordinate values = frequency counts
- In newer models, coordinate values are not easily interpreted
- One way to think about it is that each value is representative of some property of the word, e.g.
  - ‘blueberry’:
    - high value for the coordinate representing its fruit-like qualities
    - low value for the one representing its tech-relatedness
- Words that appear in similar contexts will have similar values and thus be close in space

Introduction to word embeddings

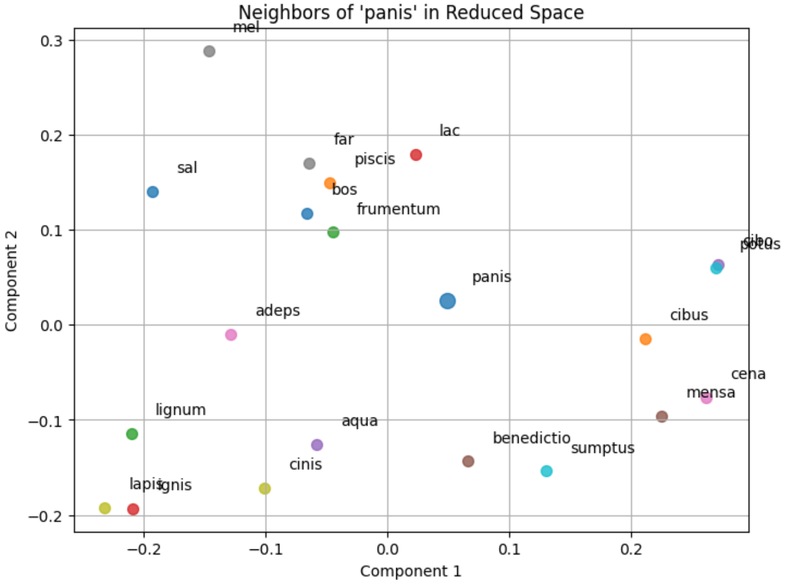
Introduction to word embeddings

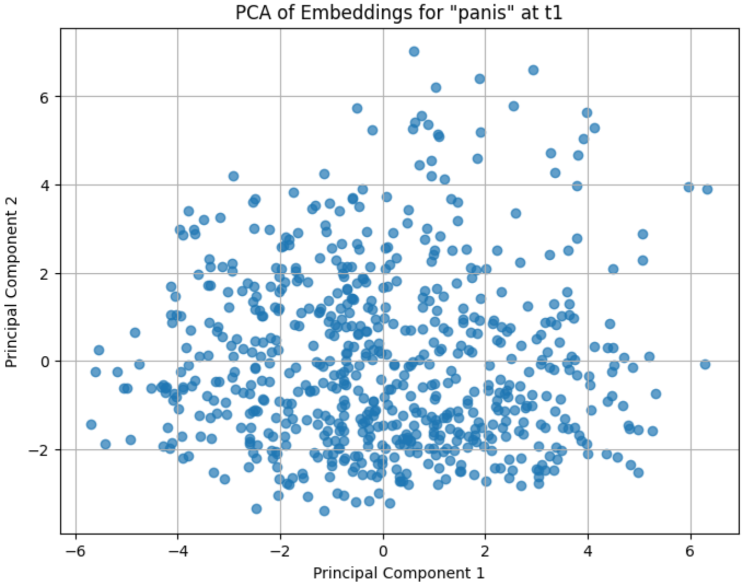
STATIC VS. CONTEXTUAL EMBEDDING MODELS

STATIC	CONTEXTUAL
One vector per word type	One vector per word token
Do not deal with polysemy	Can deal with polysemy
Word2Vec, FastText, etc.	BERT, GPT, etc.
Need a moderate amount of data	Need a large amount of data
Computationally lightweight	Computationally demanding

Introduction to word embeddings

STATIC VS. CONTEXTUAL EMBEDDING MODELS





Introduction to word embeddings

COSINE SIMILARITY (AND OTHER CHANGE FUNCTIONS)

- The spatial representation of words allows us to quantify the difference between two words by comparing their vectors
- A popular function: cosine similarity, with value ranging from 0 to 1
  - 0 indicating no similarity, 1 indicating maximum similarity (Jurafsky and Martin 2023, 112–3)
- Some other functions:
  - Inverted similarity over word prototype (PRT)
  - Average pairwise distance (APD)
  - Jensen-Shannon divergence (JSD)

Introduction to word embeddings

COSINE SIMILARITY (AND OTHER CHANGE FUNCTIONS)

- How can we use these functions to detect semantic change?
  - Compare vectors of the same word, where the vectors are calculated from separate sets of documents from different timeframes
  - Look at how vectors of a target word relate to other vectors in the semantic space: do relationships change across different sets of documents?
- The latter is useful for qualitative assessment: word embeddings which are “closest” to a target embedding are known as “neighbours”

Christian Latin

BASICS

- The spread of Christianity had an effect on the politics, society, culture, and unsurprisingly also the languages of the western world
- That Christianity had influence on Latin is generally unchallenged
- Many lexical phenomena are a result of influence from Greek:
  - Loanwords: *angelus* ‘angel, herald of God’ ⟵ ἄγγελος
  - Loan translations: *glōrificō* ‘glorify’ ⟵ δοξάζω
  - Sematic loans: *virtūs* ‘miracle’ ⟵ δύναμις

Burton 2011: 489

- …but plenty of native words also acquired a new Christian sense

Christian Latin

THE DEBATE

- "Christian Latin" as a distinct linguistic entity is debated
- *Sondersprache* hypothesis (Schrijnen 1932, Mohrmann 1958–1977):
  - The Christian community developed a unique form of Latin: different morphology, syntax, and lexicon
- Its critics (e.g. Marouzeau 1932, Coleman 1987):
  - *Sondersprache* hypothesis:
    - Exaggerates impact that supposed communal living of early Christians would have had on their language
    - Relies too heavily on evidence from a limited number of educated writers, when Christian writers often differ in style

Christian Latin

HOW TO MOVE FORWARD

- Even the harshest critics of the *Sondersprache* hypothesis acknowledged the existence of a ***unique Christian vocabulary*** (Burton 2011: 487–8)

  ⇒ further study of the lexicon might actually prove useful

Ultimate (ambitious) goals

- Leverage the new methods to contribute to:

  - The theory of semantic change
  - The study of the history of Latin
  - The debate about Christian Latin
  - Further establishment of DH within Historical Linguistics / Classics

Plan for (the rest of) today

- Choose lexemes for analysis + outline meaning change
  - Today: 2 words only, picked to highlight strengths and weaknesses of both methods
- Outline corpus and subcorpora for analysis
- Compare and evaluate results from:
  - Static embeddings
  - Contextual embeddings from fine tuned Latin BERT
- Conclude and discuss future work

Selected lexemes

ŌRĀTIŌ

- Fairly, but not massively frequent in working corpus
- In Classical Latin it can mean
  - ‘speech’, ‘locution’
  - ‘public speech’
  - ‘eloquence’
- In Christian texts (starting already in Tertullian) it means ‘prayer’ (Teßmer 1978)

Selected lexemes

DEUS

- Very frequent in working corpus
- With the advent of Christianity, *deus* began to be used to refer to the Christian god in addition to referring to the Roman gods (Gudeman 1912)
- An even more obvious choice, but good for illustrative purposes

Corpus design

DEFINING THE TIME FRAME

- LatinISE: approx.13 million words (McGillivray and Kilgarriff 2013)
- The corpus size is reduced to use texts from 300 BCE to 600 CE, for a total of 6.8 million words:
  - The end date depends on the willingness to research Latin while it was a living language
  - The start date was chosen to make the pre- and post-Christian subcorpora chronologically balanced, with the split coinciding with the first attestations of Christian texts

Corpus design

DEFINING THE SUBCORPORA

- Two chronological subcorpora, with the split set to 150 CE:
  - First attestations of Christian texts approx. late 2nd cent. CE
  - Comparison of representations for the same words across the two timeframes
- Two subcorpora contained in the second timeframe, one containing exclusively Christian texts, the other all non-Christian ones
  - Comparison of representations for the same words across different sets of texts within the same timeframe
- The subcorpora are fairly balanced with their counterparts in terms of number of tokens

Static embeddings: model setup

- Starting point: code by McGillivray (2023)
- Choice of model and values for parameters conforming largely to the findings of Sprugnoli, Passarotti, and Moretti (2019), Sprugnoli, Moretti, and Passarotti (2020), and Ribary and McGillivray (2020)
- Lemmatised corpus to reduce variability
- FastText picked for suitability for morphologically rich languages, given its use of n-grams (i.e. subwords) during training
- Some tested parameters were:
  - Window size: both 5 and 10
  - Minimum frequency: 50 for high freq., 5 for low freq. words
  - Exclusion / inclusion of subwords during training

Static embeddings: semantic similarity

- Remember: one vector per lemma in each subcorpus
- Simple cosine similarity calculation
- Compare:
  - Word embedding in first time slice w equivalent in second time slice
  - Word embedding in Christian subcorpus w equivalent in non-Christian subcorpus
  - Target lemma in each subcorpus with its nearest neighbours
- Results that follow obtained with:
  - Window size: 5; minimum frequency: 50; subwords excluded

Static embeddings: results

DEUS: NEIGHBOURS

300 BCE - 150 CE (freq. 5408)		150 – 600 CE (freq. 15086)		Christian (freq. 13928)		Non-Christian (freq. 1158)	
immortalis	0.723	*dominus*	0.712	*pater*	0.622	*numen*	0.856
numen	0.718	*omnipotens*	0.661	*omnipotens*	0.599	*sanctus*	0.842
superi	0.694	*Christus*	0.625	*dominus*	0.590	*Christus*	0.831
dea	0.656	*creator*	0.622	*maiestas*	0.586	*pietas*	0.817
Iuppiter	0.652	*iustitia*	0.611	*creator*	0.579	*pius*	0.815
propitius	0.598	*factor*	0.586	*exalto*	0.569	*o*	0.780

Static embeddings: results

DEUS

- Between the first and second timeframe, there is a visible difference in terms of neighbours, with clear associations with:
  - Roman religion in the first
  - Christian religion in the second
- Cosine similarity between time slices: 0.618
- Within the Christian subcorpus the associations with Christianity are confirmed
- Within the non-Christian one, we have a mix, signaling that the change in meaning of *deus* affected the language as a whole

Static embeddings: results

ŌRĀTIŌ: NEIGHBOURS

300 BCE - 150 CE (freq. 2295)		150 – 600 CE (freq. 950)		Christian (freq. 629)		Non-Christian (freq. 321)	
disputatio	0.718	*lectio*	0.593	*hymnus*	0.704	*litterae*	0.815
sermo	0.673	*scriptum*	0.437	*consuetudo*	0.642	*Graecus*	0.809
actio	0.670	*epistula*	0.434	*dominicus*	0.621	*vocabulum*	0.809
sententia	0.642	*interrogatio*	0.430	*evangelium*	0.618	*Varro*	0.808
lectio	0.636	*confessio*	0.428	*lectio*	0.614	*Cicero*	0.777
verbum	0.623	*verbum*	0.423	*ecclesia*	0.612	*Cato*	0.769

Static embeddings: results

ŌRĀTIŌ

- Between the first and second timeframe, neighbours show associations with:
  - Speech, public speech, legal action in the first
  - Different means of communication in the second
  - No strong Christian meaning in second
- Cosine similarity between time slices: 0.545
- Within the Christian subcorpus: perhaps a loose association with prayer, but definitely an association with Christianity
- Within the non-Christian one: interestingly, *ōrātiō* connected with famous orators or educational themes

Contextual embeddings: model setup

- Starting point: Latin BERT (Bamman and Burns 2020)
- There are 3 LLMs trained on Latin: Latin BERT is trained on the largest amount of Latin among those (approx. 650M tokens)
- (To my knowledge) only one study using Latin BERT for semantic change detection (Liu et al. 2025, for Medieval Latin charters)
- Rather than training from scratch, the idea for LLMs is to perform domain fine tuning (i.e. further training on a smaller corpus)
- Lemmatized corpus to reduce variability
- Some tested parameters for fine tuning were:
  - Learning rate: 5e-5, 3e-5, 1e-5
  - Epochs: 2, 3, 4

Contextual embeddings: semantic similarity

- Remember: one vector per token
- Consequences:
  - Comparison across the subcorpora is not as easy
  - Assessment of different senses/usages of a word now possible
- Comparison solutions:
  - Averaging:
    - End up with one vector per lemma, again
    - Easy, but defeats the point of using contextual embeddings
  - Clustering:
    - End up with vector clusters representing different senses
    - Needs to be performed on target words
    - Basic units to compare are less clear

Contextual embeddings: semantic similarity

- Some similarity/change measurement functions:
  - Averaging:
    - Cosine similarity / cosine distance (CS / CD)
    - Inverted similarity over word prototypes (PRT)
  - Clustering:
    - First, (normally) use *k-means* and *silhouette score* to find optimal number of clusters
    - Jensen-Shannon divergence (JSD)
    - Cosine distance between cluster distribution (CDCD)
    - Cosine distance between semantic prototypes (PDIS)
  - Other:
    - Average pairwise distance (APD)

Contextual embeddings: semantic similarity

- What we will look at:
  - Clustering:
    - Use *k-means* and *silhouette score*
    - Average each cluster to get one embedding per sense
    - Compare each sense embedding to averaged embeddings, essentially answering the question “which words is the ‘god’ sense of *dominus* close to?”
    - (a little unconventional, but we believe the most useful qualitative assessment)
- Results that follow obtained with:
  - Learning rate: 1e-5; epochs: 2; batch size: 32

Contextual embeddings: results

ŌRĀTIŌ: 300 BCE – 150 CE SENSE CLUSTER NEIGHBOURS

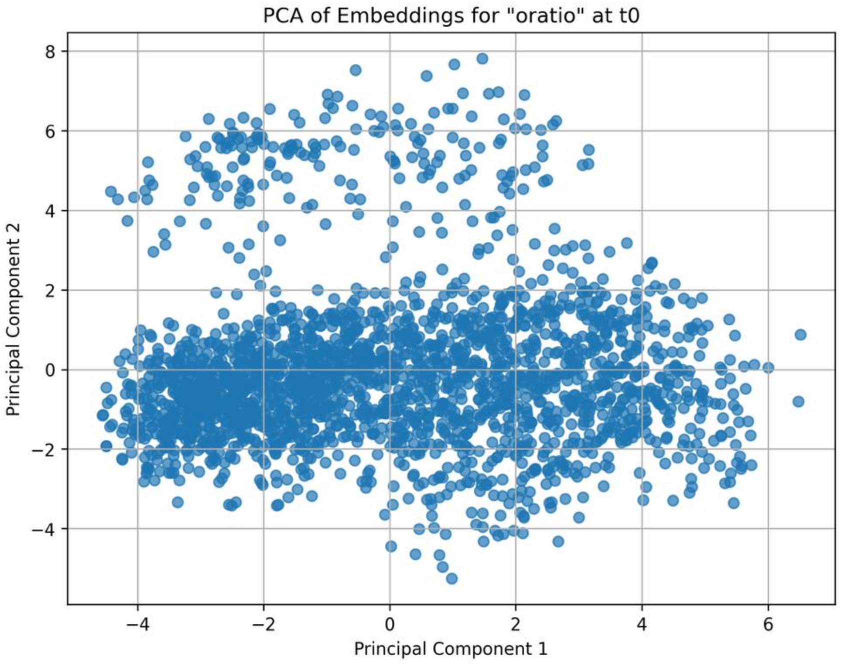
CLUSTER 1		CLUSTER 2		CLUSTER 3	
locutio	0.846	*orator*	0.893	*sermo*	0.902
sermo	0.813	*sermo*	0.891	*accusatio*	0.885
actio	0.802	*eloquentia*	0.887	*disputatio*	0.876
dictio	0.800	*disputatio*	0.872	*actio*	0.870
eloquentia	0.793	*actio*	0.871	*interrogatio*	0.865
compositio	0.790	*accusatio*	0.870	*eloquentia*	0.862

K-means: 3 (i.e. optimal number of clusters)

Silhouette score: 0.105 (i.e. how well defined the clusters are)

Contextual embeddings: results

ŌRĀTIŌ: 300 BCE – 150 CE SENSE CLUSTERS



Contextual embeddings: results

ŌRĀTIŌ: 150 – 600 CE SENSE CLUSTER NEIGHBOURS

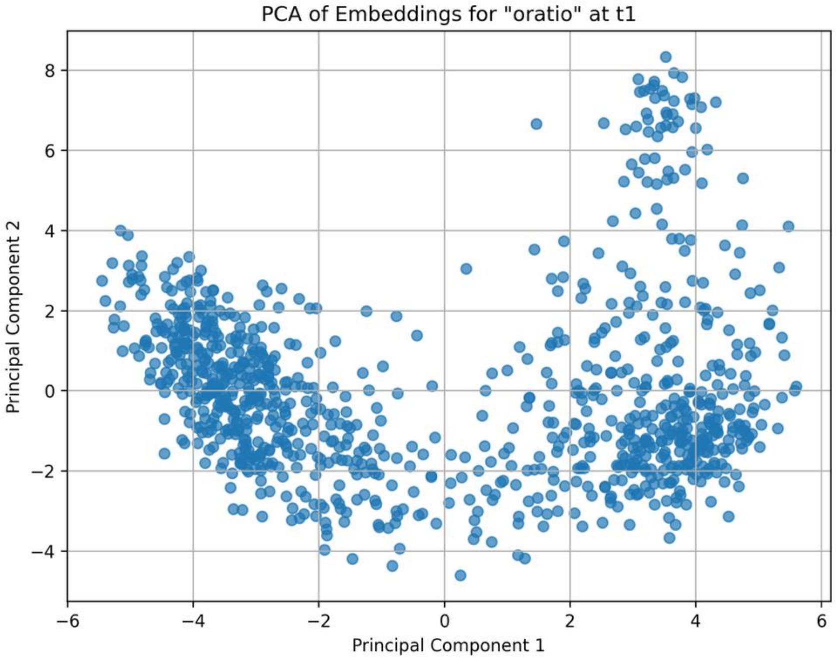
CLUSTER 1		CLUSTER 2	
dictio	0.901	*oblatio*	0.867
orator	0.887	*hymnus*	0.862
disputatio	0.885	*precatio*	0.857
sermo	0.879	*sacrificium*	0.856
locutio	0.875	*lectio*	0.853
narratio	0.874	*oro*	0.853

K-means: 2 (i.e. optimal number of clusters)

Silhouette score: 0.133 (i.e. how well defined the clusters are)

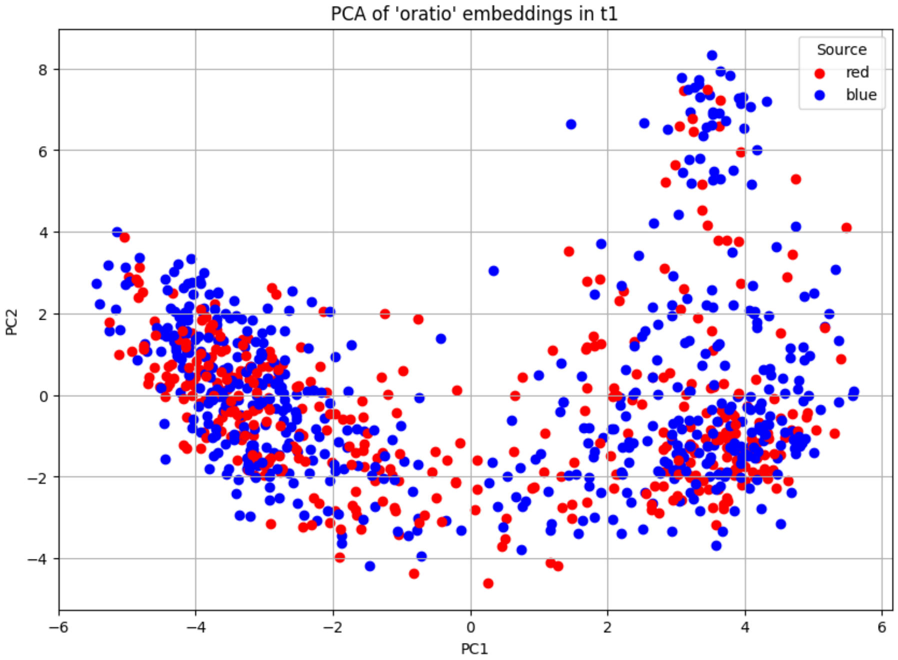
Contextual embeddings: results

ŌRĀTIŌ: 150 – 600 CE SENSE CLUSTERS



Contextual embeddings: results

ŌRĀTIŌ: 150 – 600 CE SENSE CLUSTERS BY TEXT-TYPE



Red dots: embeddings for occurrences of *ōrātiō* in texts labeled as Christian
Blue dots: embeddings for occurrences of *ōrātiō* in texts labeled as non-Christian

Contextual embeddings: results

ŌRĀTIŌ

- Clusters in first timeframe fairly close, yet senses somewhat discernible:
  - Cluster 1: speech, speech ability
  - Cluster 2: public speech, oration
  - Cluster 3: legal context/purpose of public speech
- Clusters in second timeframe slightly more discernible:
  - Cluster 1: speech, oration
  - Cluster 2: prayer, offering
- However, the sense distribution does not appear to be directly proportional to text type, according to the plot

Contextual embeddings: results

DEUS: 300 BCE – 150 CE SENSE CLUSTER NEIGHBOURS

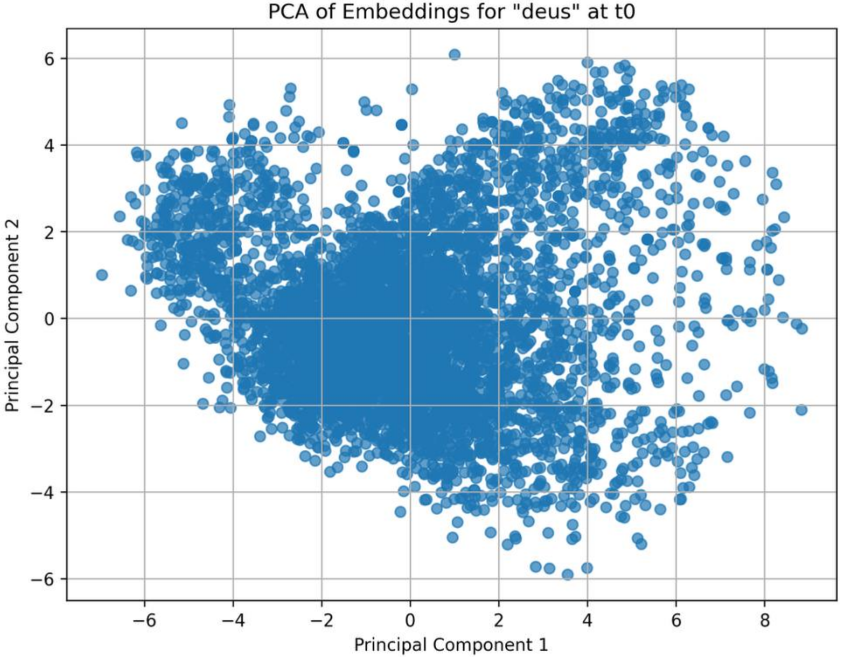
CLUSTER 1		CLUSTER 2		CLUSTER 3	
dea	0.900	*dea*	0.925	*iuppiter*	0.867
iuppiter	0.871	*iuppiter*	0.916	*dea*	0.856
apollo	0.819	*numen*	0.874	*numen*	0.817
numen	0.813	*apollo*	0.858	*divus*	0.812
senex	0.807	*divus*	0.848	*divinus*	0.791
diana	0.800	*semideus*	0.842	*apollo*	0.788

K-means: 3 (i.e. optimal number of clusters)

Silhouette score: 0.088 (i.e. how well defined the clusters are)

Contextual embeddings: results

DEUS: 300 BCE – 150 CE SENSE CLUSTERS



Contextual embeddings: results

DEUS: 150 – 600 CE SENSE CLUSTER NEIGHBOURS

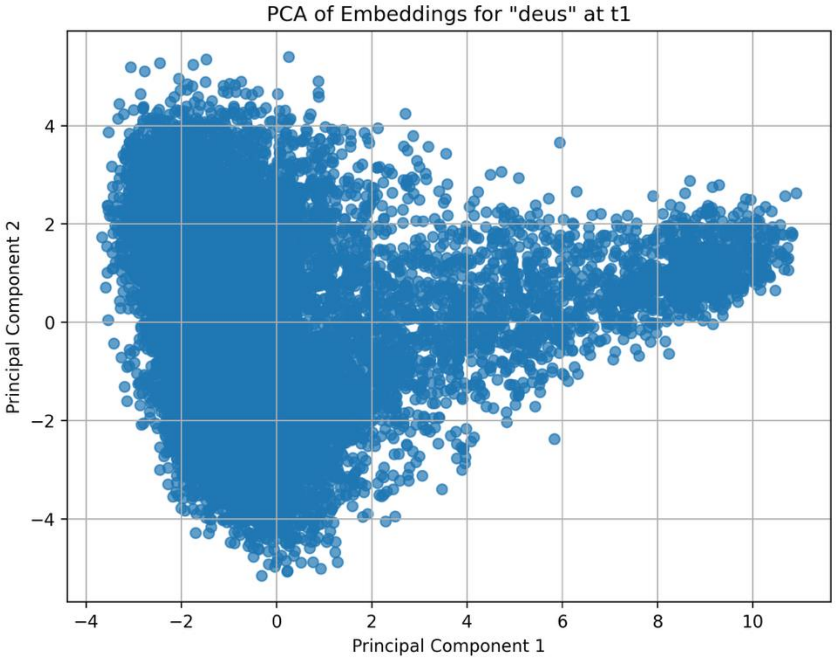
CLUSTER 1		CLUSTER 2	
dea	0.901	*divus*	0.867
iuppiter	0.887	*iuppiter*	0.862
dominus	0.885	*iesus*	0.857
numen	0.879	*dominus*	0.856
divinus	0.875	*dea*	0.853
christus	0.874	*deitas*	0.853

K-means: 2 (i.e. optimal number of clusters)

Silhouette score: 0.182 (i.e. how well defined the clusters are)

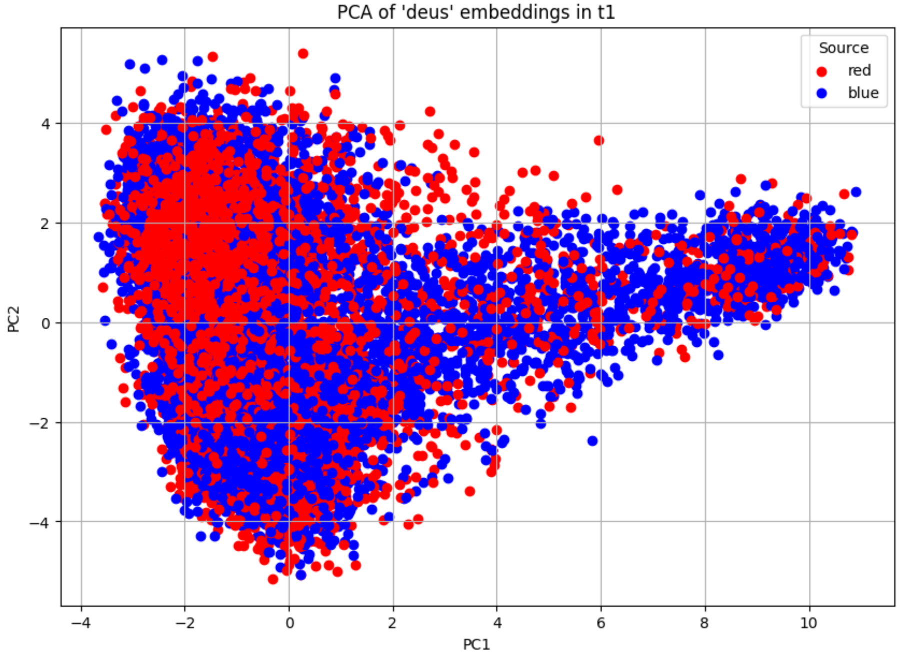
Contextual embeddings: results

DEUS: 150 – 600 CE SENSE CLUSTERS



Contextual embeddings: results

DEUS: 150 – 600 CE SENSE CLUSTERS BY TEXT-TYPE



Red dots: embeddings for occurrences of *deus* in texts labeled as Christian
Blue dots: embeddings for occurrences of *deus* in texts labeled as non-Christian

Contextual embeddings: results

DEUS

- Clusters in both first and second timeframe seem virtually undistinguishable, judging from the nearest neighbours…
- The post-150 AD text-type plot seems to include a slight prevalence of embeddings from Christian texts for one of the two clusters, but this seems irrelevant given the qualitative cluster similarity
- Questions to ask:
  - Does very high frequency make it harder to have well defined clusters for contextual embeddings methods? If so, do we need to scale down?
  - How much does the pre-training influence the results?

Conclusions

- Static embeddings:
  - Good for high frequency words
  - (Not so good if the frequency is lower)
  - Need to further split the corpus into text-types to gather more fine-grained information
  - Otherwise, polysemy is a struggle and only the dominant sense is picked up
  - Results produced quickly, testing of different parameters is easy

Conclusions

- Contextual embeddings:
  - We can investigate whether a word was polysemous, and try to gather information about each of its senses
  - No strict need to rely on text type (in fact, not relying on it might reveal other information)
  - Questions to ask about potentially less good performance with high frequency
  - Medium frequency seems optimal for current setup
  - (Low frequency still slightly problematic)
  - Slow, need supercomputers, testing different configurations takes a long time

Next steps

- Develop benchmark and evaluate results more rigorously
- Better assessment of current shortcomings
  - We want to advance the field by making our job easier
  - But, of course, we want the final product to be good!
- Compare with other distributional methods (e.g., collocational analysis)

Some essential bibliography

Bamman, David and Patrick **Burns**. 2020. “Latin BERT: A Contextual Language Model for Classical Philology.” *CoRR*, abs/2009.10053.

Burton, Philip. 2011. “Christian Latin.” In *A Companion to the Latin Language*, edited by James Clackson, 485–501. Oxford: Wiley-Blackwell.

Coleman, Robert. 1987. “Vulgar Latin and the diversity of Christian Latin.” In *Latin vulgaire – Latin tardif: Actes du 1er Colloque International sur le latin vulgaire et tardif, Pécs, 2-5 septembre 1985*, edited by József Hermann, 37–52. Tübringen: Niemeyer.

Jurafsky, Dan, and James H. **Martin**. 2023. “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.” Available at https://web.stanford.edu/~jurafsky/slp3/.

Liu, Gelila Tilahun, Xinxiang Gao, Qianfeng Wen, and Michael Gervers. 2025. “A Comparative Study of Static and Contextual Embeddings for Analyzing Semantic Changes in Medieval Latin Charters.” In *Proceedings of the First Workshop on Language Models for Low-Resource Languages*: 182–192. Abu Dhabi, UAE: Association for Computational Linguistics.

McGillivray, Barbara. 2023. *Semantic change in LatinISE*. Available at https://github.com/BarbaraMcG/latinise/blob/master/lvlt22/Semantic_change.ipynb, accessed 2023-09-01.

Mohrmann, Christine. 1958–1977. *Études sur le latin des chrétiens*. *4 vols*. Rome: Storia e letteratura.

Ribary, Marton, and Barbara **McGillivray**. 2020. “A Corpus Approach to Roman Law Based on Justinian’s Digest.” *Informatics* 7 (4).

Schrijnen, Jos. 1932. *Charakteristik des altchristlichen Latein*. Nijmegen: Dekker en Van de Vegt & Van Leeuwen.

Sihler, Andrew L. 2000. *Language History : An Introduction*. Amsterdam: John Benjamins Publishing.

Sprugnoli, Rachele, Giovanni **Moretti**, and Marco **Passarotti**. 2020. “Building and Comparing Lemma Embeddings for Latin. Classical Latin versus Thomas Aquinas.” *Italian Journal of Computational Linguistics* 6 (1): 29–45.

Tahmasebi, Nina, Lars **Borin**, and Adam **Jatowt**. 2021. “Survey of Computational Approaches to Diachronic Conceptual Change.” In *Computational approaches to semantic change*, edited by Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, and Simon Hengchen, 1–91. Berlin: Language Science Press.

Thank You

Personal website: