Approaching lexical semantic change detection across many time periods

Bill Noble

CLASP Seminar

November 1, 2023

Lexical semantic change (LSC) detection

- The goal of lexical semantic change detection is to identify changes in conventional word meaning
- Typically LSC detection performed across two distinct time periods: t_1 and t_2
  - Q: Is the conventional meaning of word w in t_2 different from what it was in t_1?
- All usages within a time period are typically treated as synchronic for modeling purposes
- There's Often a gap between t_1 and t_2. E.g. SemEval-2020 Task 1:[1]

English	(CCOHA)	1810–1860	1960–2010
German	(DTA/BZ+ND)	1800–1899	1946–1990
Latin	(LatinISE)	-200–0	0–2000
Swedish	(Kubhist)	1790–1830	1895–1903

[1] Schlechtweg et al., 2020

Sense-aware LSC detection evaluation...

...without explicit word-sense annotation (Schlechtweg et al., 2020).

Subtask 1 Binary classification: for a set of target words, decide which words lost or gained sense(s) between t_1 and t_2, and which ones did not.

Subtask 2 Ranking: rank a set of target words according to their degree of LSC between t_1 and t_2.

Binary change and degree of change are both derived by comparing sense frequency distributions between time periods.

Sense-aware LSC detection evaluation...

...without explicit word-sense annotation (Schlechtweg et al., 2020).

Subtask 1 Binary classification: for a set of target words, decide which words lost or gained sense(s) between t_1 and t_2, and which ones did not.

Subtask 2 Ranking: rank a set of target words according to their degree of LSC between t_1 and t_2.

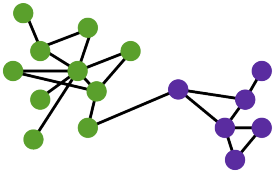
Binary change and degree of change are both derived by comparing sense frequency distributions between time periods.

So how do we get sense frequency distributions without explicit sense annotation?

Word Usage Graphs (WUGs)

A usage graph $G = (U, E, W)$

- U - set of usages: $u_1, u_2, \ldots$
- E - edges between usages (u_i, u_j)
- W - weight of edges: $W(u_i, u_j) \in \mathbb{R}^+$



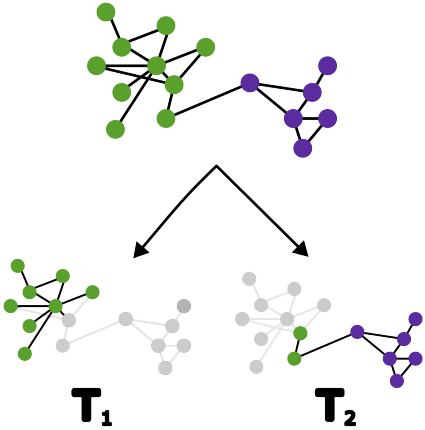
Diachronic Word Usage Graphs (DWUGs)

A usage graph $G = (U, E, W)$

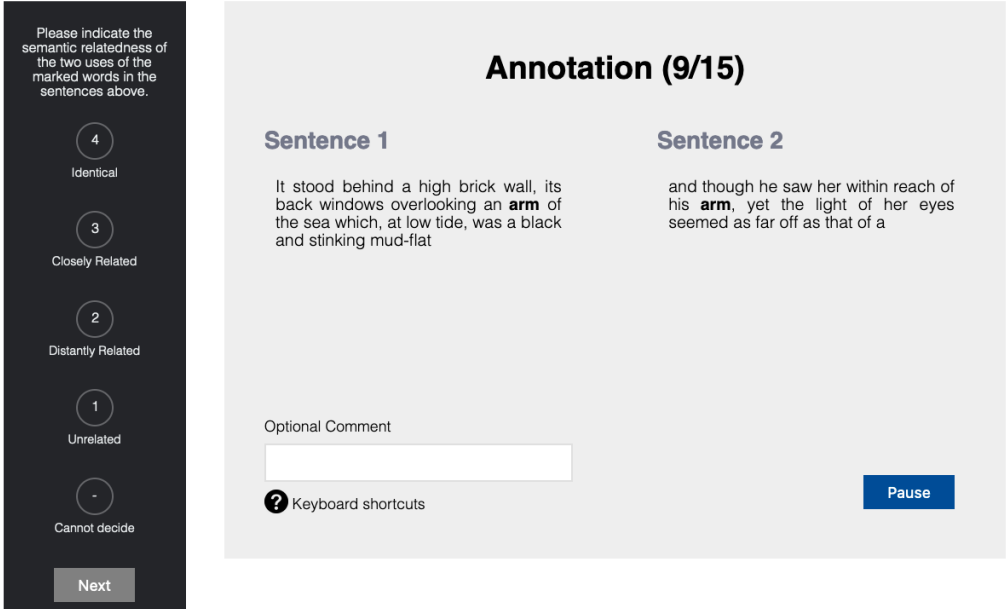
- U - set of usages: $u_1, u_2, \ldots$
- E - edges between usages (u_i, u_j)
- W - weight of edges: $W(u_i, u_j) \in \mathbb{R}^+$

Time periods partition U into the sets of usages falling within the time period.

- $U_1 \cup U_2 = U$
- $U_1 \cap U_2 = \varnothing$

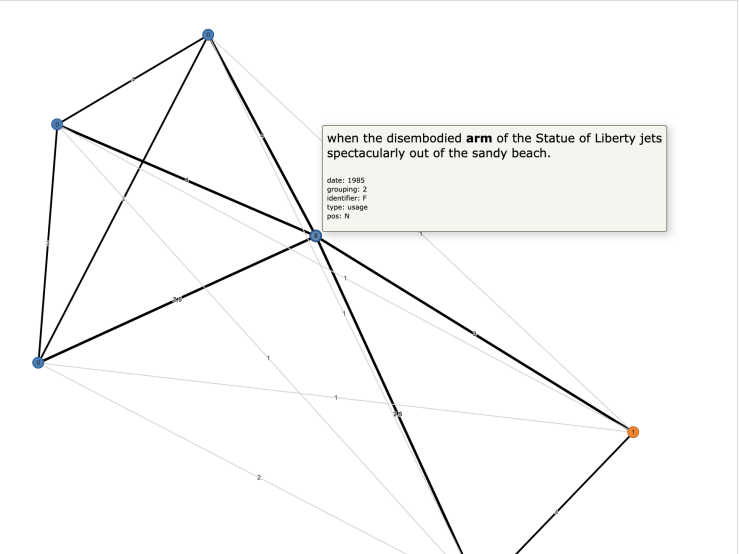


DURel annotation (Schlechtweg et al., 2018, 2021)



DURel annotation

arm (full)



Info:

Node position: spring

Clustering method: correlation

Statistics:

Cluster frequency distribution: [4, 2]

Cluster probability distribution: [0.667, 0.333]

Noise Cluster: [0]

Edge weight mean: 2.46667

Edge weight standard deviation: 1.18977

Edge filters:

Show NaN edges ☑

Min weight: 1

Max weight: 4

Node filters:

Show noise cluster ☑

From date: 1824 to date: 1985 Filter

Grouping: All

Annotator filter (resets all other filters):

PaulineSander ☑ Random ☑ XL-Lexeme ☑ billnoble ☑

Many[2] time period LCD

- In the real world, meaning change is a (more or less) continuous process
- We want to develop methods that don't rely on the artificial assumption of two time periods. Why?
  - Practical applications: LSC needs to be able to detect changes in real time.
  - Historical linguistics: We may want to ask when a change took place.
- New Q for N time periods: Is there a change in the conventional meaning of word w between any pair of time periods in $t_1, ..., t_N$?

[2] I.e., more than 2! (possibly *many* more...)

DWUGS with many time points

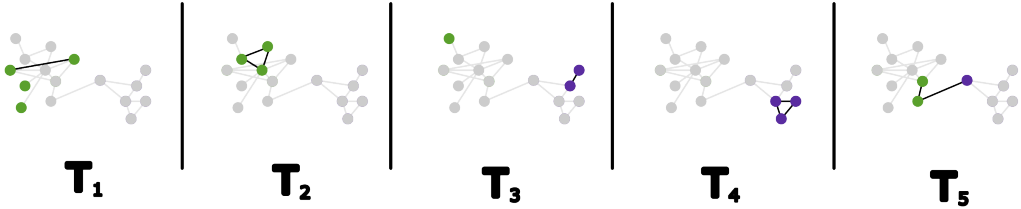
Many time point diachronic usage graphs are basically the same but we partition U further for $t_1, ..., t_N$.

- $\bigcup_{i \leq N} U_i = U$
- $U_i \cap U_j = \varnothing$ for all $i, j \leq N$

DWUGS with many time points

Many time point diachronic usage graphs are basically the same but we partition U further for $t_1, ..., t_N$.

- $\bigcup_{i \leq N} U_i = U$
- $U_i \cap U_j = \varnothing$ for all $i, j \leq N$



What to do?

Challenges:

- The more time periods the sparser our data in any one time period becomes
- Some time periods may end up with very few usages
  - If we want to know if changes have happened between say t_1 and t_5, can we leverage information from usages in t_4?
- This is both a modeling challenge (how do we detect if change has occurred in an unsupervised way?)
- And an annotation challenge (how do we determine if change has occurred given sense-annotated usages?)

What to do?

Challenges:

- The more time periods the sparser our data in any one time period becomes
- Some time periods may end up with very few usages
  - If we want to know if changes have happened between say t_1 and t_5, can we leverage information from usages in t_4?
- This is both a modeling challenge (how do we detect if change has occurred in an unsupervised way?)
- And an annotation challenge (how do we determine if change has occurred given sense-annotated usages?)

Ways forward (annotation & evaluation):

- (How) should the DURel edge sampling heuristics be modified??
- Can we confidently annotate some portion of edges with an automatic annotator? (e.g., with XL-LEXEME (Cassotti et al., 2023))

Computational Approaches for Language Change

Change is Key!
Program

XL-LEXEME: WiC Pretrained Model for Cross-Lingual LEXical sEMantic changE

Lexical Semantic Change

Lexical Semantic Change (LSC) Detection is the task of automatically identifying words that change their meaning over time.

1810-1860 *Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

1960-2010 *The President's **plane** landed at Goose Bay at 9:03 p. m.*

WSD vs WiC vs LSCD

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

plane.n.02

*The President's **plane** landed at Goose Bay at 9:03 p. m.*

airplane.n.01

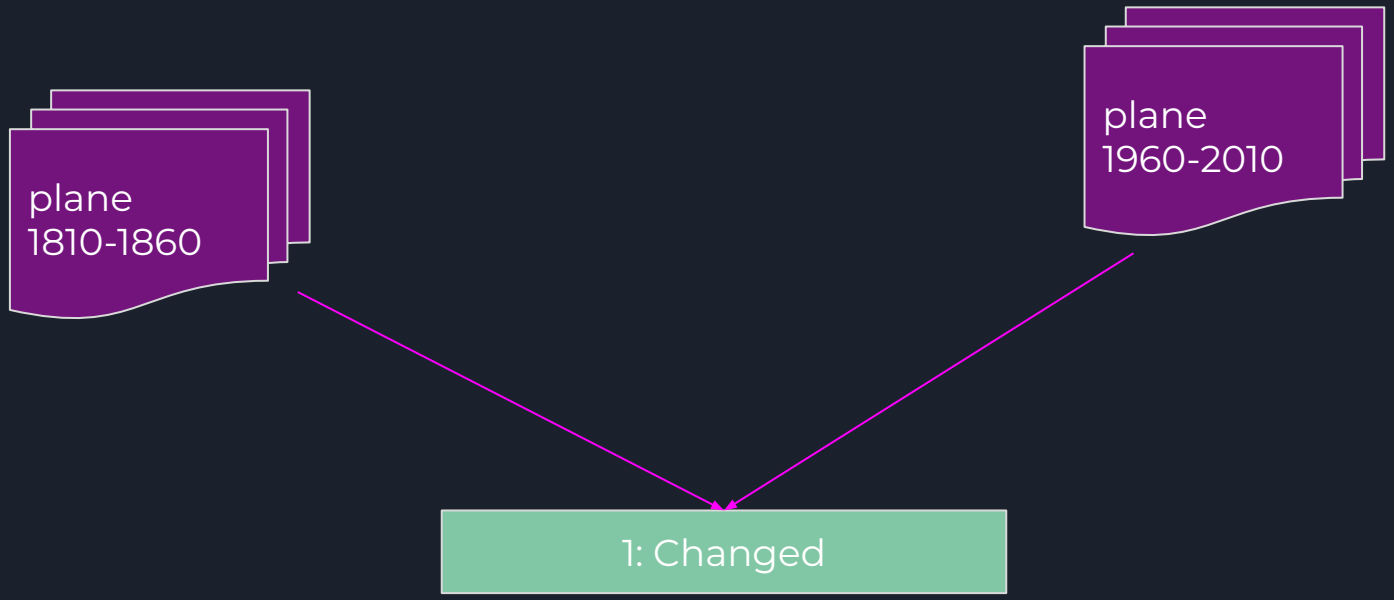
WSD vs **WiC** vs LSCD

*Provide a large table; this is a horizontal **plane**, and will represent the ground plane, viz.*

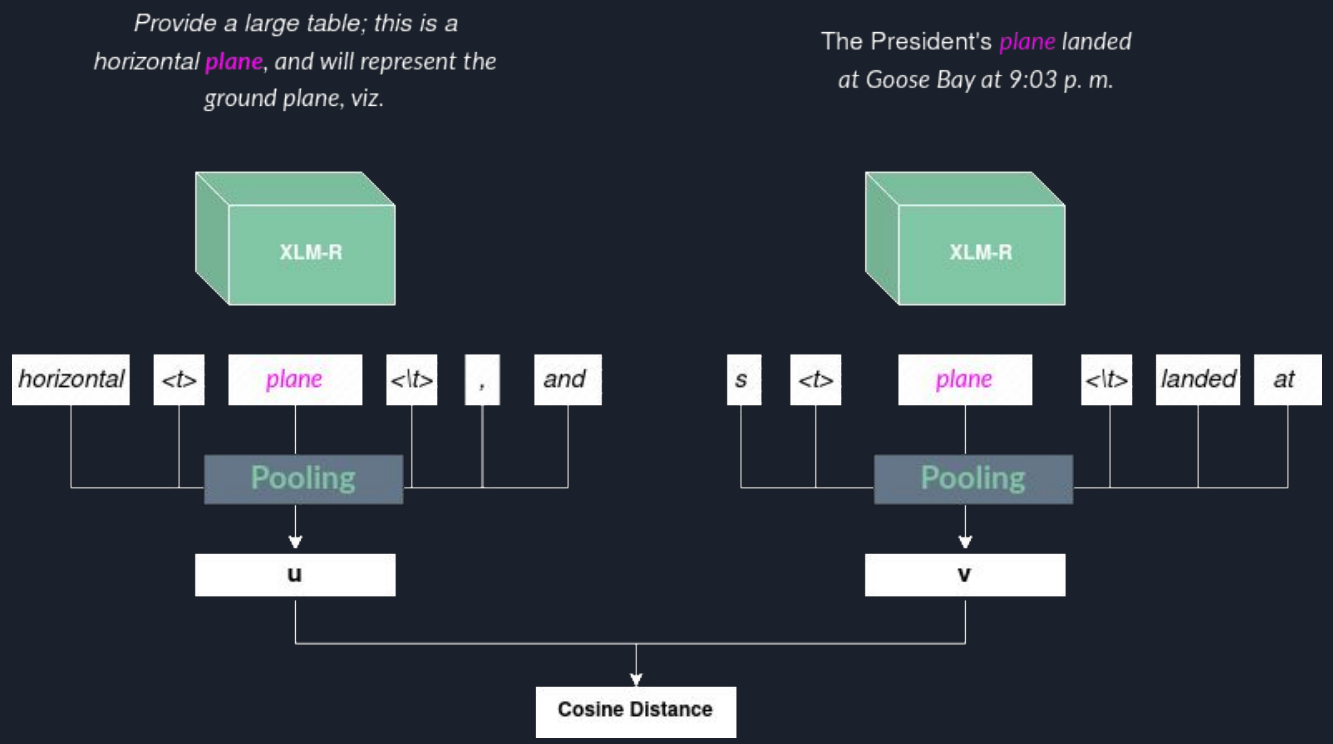
*The President's **plane** landed at Goose Bay at 9:03 p. m.*

0: Different meaning

WSD vs WiC vs **LSCD**



XL-LEXEME


Provide a large table; this is a horizontal plane, and will represent the ground plane, viz.
The President's plane landed at Goose Bay at 9:03 p. m.
XLM-R
XLM-R
horizontal
<t>
plane
<\t>
,
and
s
<t>
plane
<\t>
landed
at
Pooling
Pooling
u
v
Cosine Distance

Word-in-Context Datasets

Dataset	Languages
WiC Pilehvar et al., (2019	Monolingual EN
XL-WiC (Raganato et al., 2020)	Multilingual EN, BG, ZH, HR, DA, NL, ET, FA, FR, DE, IT, JA, KO
MCL-WiC (Martelli et al., 2021)	Multilingual **EN, AR, FR, RU, ZH**
	Crosslingual **AR, FR, RU, ZH**
AM^{2}ICO (Liu et al., 2021)	Crosslingual EN, DE, RU, JA, KO, ZH,AR, IN, FI, TR, EU, KA, UR, BN, KK

Experimental Setting

XL-LEXEME is evaluated on SemEval 2020 Task 1 Subtask 2 and RuShiftEval benchmarks.
The LSC score is computed as the ***average pairwise distances*** between pairs of sentences of different periods:

$$\mathrm{LSC}(s^{t_0}, s^{t_1}) = \frac{1}{N \cdot M} \sum_{i=0}^{N} \sum_{j=0}^{M} \delta(s_i^{t_0}, s_j^{t_1})$$

where $\boldsymbol{\delta}$ is the cosine distance and (s^{t0}, s^{t1}) are pairs of sentences sampled respectively from t_0 and t_1.

Results (SemEval 2020 Task 1 Subtask 2)

Model	**EN**	**DE**	**SV**	**LA**	Avg.
SemEval-2020 Task 1 Subtask 2 Leaderboard					
UG_Student_Intern	0.422	0.725	†0.547	0.412	0.527
Jiaxin & Jinan	0.325	0.717	†0.588	**0.440**	0.518
cs2020	0.375	0.702	†0.536	0.399	0.503
UWB	0.367	0.697	†0.604	0.254	0.481
Count baseline	0.022	0.216	-0.022	0.359	0.144
Freq. baseline	-0.217	0.014	-0.150	†0.020	-0.083
Temporal BERT					
TempoBERT	0.467	-	-	0.512	-
Temporal Attention	†0.520	†0.763	-	0.565	-
cross-encoder	†0.752	†0.837	†0.680	†0.016	0.571
XL-LEXEME	**0.757**	**0.877**	**0.754**	-0.056	**0.583**

The symbol † indicates there is no statistical difference ($p<0.05$) with the correlation obtained by XL-LEXEME.

Results (RuShiftEval)

Model	RuShiftEval1	RuShiftEval2	RuShiftEval3	Avg.
RuShiftEval Leaderboard				
GlossReader	†0.781	†0.803	**†**0.822	0.802
DeepMistake	**†**0.798	†0.773	†0.803	0.791
UWB	0.362	0.354	0.533	0.417
Baseline	0.314	0.302	0.381	0.332
cross-encoder	†0.727	†0.753	†0.748	0.743
XL-LEXEME	0.775	0.822	0.809	0.802
XL-LEXEME (Fine-tuned)	**0.799**	**0.833**	**0.842**	**0.825**

The symbol † indicates there is no statistical difference ($p<0.05$) with the correlation obtained by XL-LEXEME.

Emerging trends in gender-specific occupational titles in Italian Newspapers

Occupational titles in Italian

Alma Sabatini

OCCUPATIONAL TITLES IN ITALIAN:
CHANGING THE SEXIST USAGE

1. Introduction

The present paper is written primarily from a feminist point of view. This is a present choice of the writer, who was actively involved in linguistics long before acquiring feminist awareness. An attempt has been made to be as objective as possible in order to see things as they are, but the writer also has very strong ideas as to what they ought to be.

Feminist awareness and interest in language have been closely associated in my mind and have allowed me to see and feel to what extent the language we use misrepresents us and is directed against us.

In this paper, which is oriented towards practical usage, I shall concentrate on occupational titles, which form a most significant area of Italian sexist language, and one in particular where - contrary to accepted belief - change is possible and linguistically defendable.

Occupational titles extraction

capotreno	capotreno
sarto	sarta
predicatore	predicatrice
tessitore	tessitrice
costruttore di chitarre	costruttrice di chitarre
allenatore di cavalli	allenatrice di cavalli
segretario	segretaria
ingegnere	ingegnera
politico	donna politica
pescivendolo	pescivendola
medico scrittore	medico scrittrice
professore a contratto	professoressa a contratto
statista	donna di Stato
statista	statista
agente di polizia	poliziotta
impiegato	impiegata
direttore sportivo	direttrice sportiva
apicoltore	apicoltrice
arbitro di calcio	arbitra di calcio
imperatore	imperatrice
allenatore	allenatrice
profeta	profetessa
cartografo	cartografa
storico della Chiesa	storica della chiesa
imprenditore	imprenditrice
governatore	governatrice

Corpus

- Articles extracted by two Italian newspapers (i.e. La Stampa and L'Unità)
- Wide historical period (1948-2005)
- 3,529,820,155 tokens
- Automatically annotated with PoS tags, lemmas, morphological features and dependency relations

l'Unità

LA STAMPA

Preprocessing: PoS tags

text = E' morto Silvio Sabatelli, 92 anni Fondò la casa editrice Liguria SAVONA

1	E'	essere	AUX	VA	Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin	2	aux
2	morto	morire	VERB	V	Gender=Masc|Number=Sing|Tense=Past|VerbForm=Part	0	root
3	Silvio	Silvio	PROPN	SP	_	2	nsubj
4	Sabatelli	Sabatelli	PROPN	SP	_	3	flat:name
5	,	,	PUNCT	FF	_	3	punct
6	92	92	NUM	N	NumType=Card	7	nummod
7	anni	anno	NOUN	S	Gender=Masc|Number=Plur	3	nmod
8	Fondò	fondare	VERB	V	Mood=Ind|Number=Sing|Person=3|Tense=Past|VerbForm=Fin	2	parataxis
9	la	il	DET	RD	Definite=Def|Gender=Fem|Number=Sing|PronType=Art	10	det
10	casa	casa	NOUN	S	Gender=Fem|Number=Sing	8	obj
11	editrice	editore	ADJ	A	Gender=Fem|Number=Sing	10	amod
12	Liguria	Liguria	PROPN	SP	_	10	nmod
13	SAVONA	Savona	PROPN	SP	_	12	flat:name

Preprocessing: Morphological features

text = Presso la Scuola Convitto per infermiere professionali sono aperte le iscrizioni

1	Presso	presso	ADP	E	_	3	case
2	la	il	DET	RD	Definite=Def|Gender=Fem|Number=Sing|PronType=Art	3	det
3	Scuola	scuola	NOUN	S	Gender=Fem|Number=Sing	9	obl
4	Convitto	convitto	ADJ	A	Gender=Masc|Number=Sing	3	compound
5	per	per	ADP	E	_	6	case
6	infermiere	infermiera	NOUN	S	Gender=Fem|Number=Plur	3	nmod
7	professionali	professionale	ADJ	A	Number=Plur	6	amod
8	sono	essere	AUX	V	Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbForm=Fin	9	cop
9	aperte	aperto	ADJ	A	Gender=Fem|Number=Plur	0	root
10	le	il	DET	RD	Definite=Def|Gender=Fem|Number=Plur|PronType=Art	11	det
11	iscrizioni	iscrizione	NOUN	S	Gender=Fem|Number=Plur	9	nsubj

Preprocessing: Polysemy

Smoothed frequencies

$$p_w^t = \frac{f_w^t + 1}{C^t + |V^t|}$$

f_w^t Absolute frequency of the occ. title w computed on the year t

C^t Number of tokens on the year t

$|V^t|$ Vocabulary length on the year t

Odds

$$odds(w)^t = log \frac{p^t_{w_f}}{p^t_{w_m}}$$

$p^t_{w_f}$ Smoothed frequency of feminine form computed on the year t

$p^t_{w_m}$ Smoothed frequency of masculine form computed on the year t

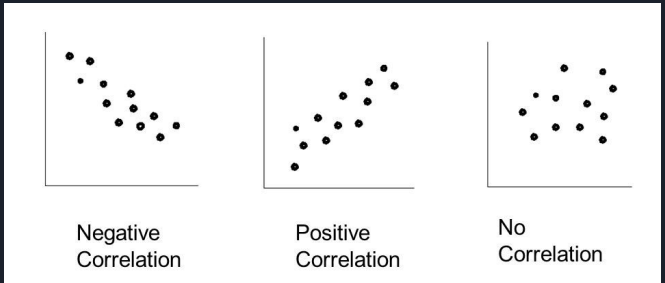
Linear Regression

$$log \frac{p^t_{w_f}}{p^t_{w_m}}$$

> 0 feminine occurrences increasing faster respect to the masculine occurrences

= 0 no correlation

< 0 masculine occurrences increasing faster respect to the feminine occurrences



Linear Regression

$$log \frac{p^t_{w_f}}{p^t_{w_m}}$$

> 0 feminine occurrences increasing faster respect to the masculine occurrences

= 0 no correlation

< 0 masculine occurrences increasing faster respect to the feminine occurrences

Linear Regression

$$log \frac{p^t_{w_f}}{p^t_{w_m}}$$

> 0 feminine occurrences increasing faster respect to the masculine occurrences

= 0 no correlation

< 0 masculine occurrences increasing faster respect to the feminine occurrences

Linear Regression

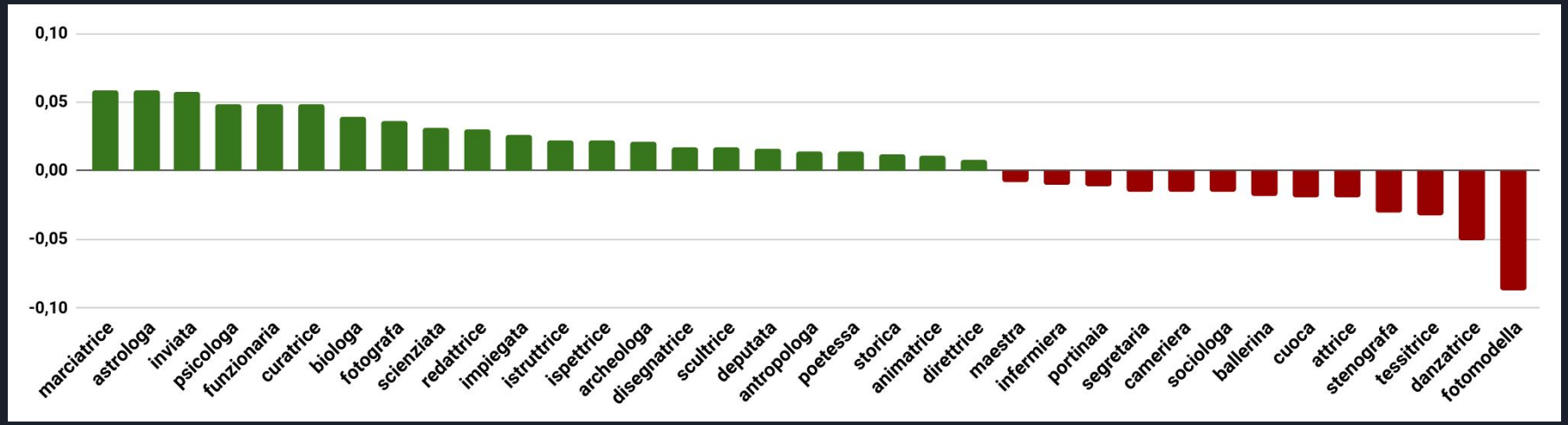
$$log \frac{p^t_{w_f}}{p^t_{w_m}}$$

> 0 feminine occurrences increasing faster respect to the masculine occurrences

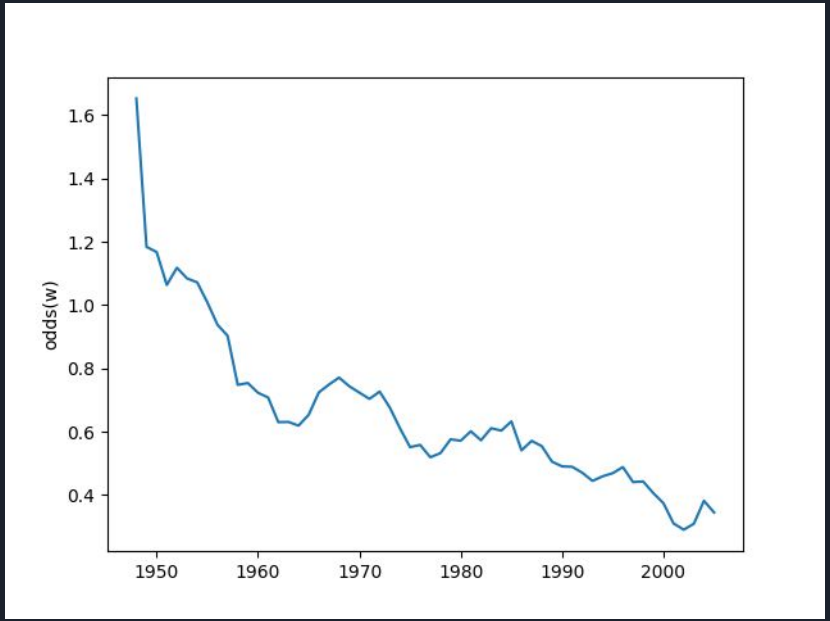
= 0 no correlation

< 0 masculine occurrences increasing faster respect to the feminine occurrences

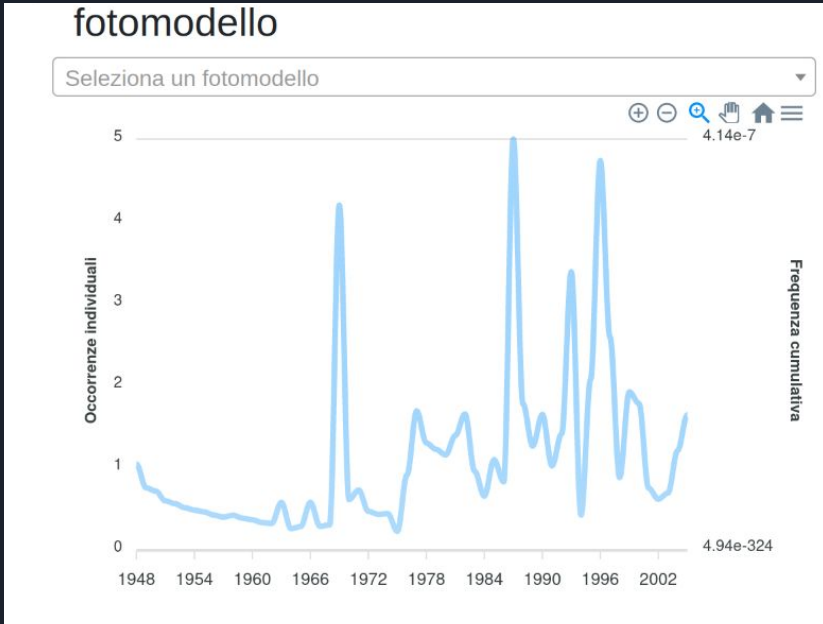
Slope of the odds

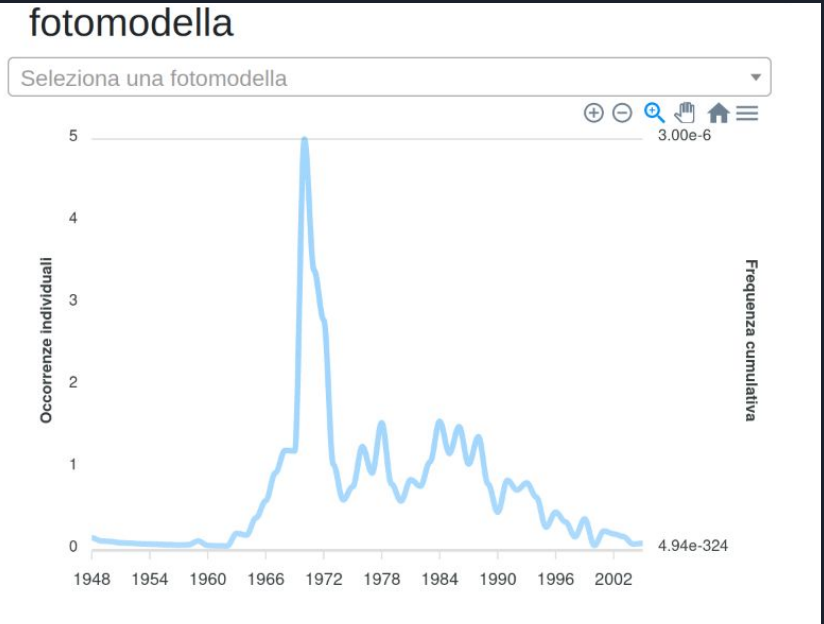


Decreasing odds: *infermiere* example

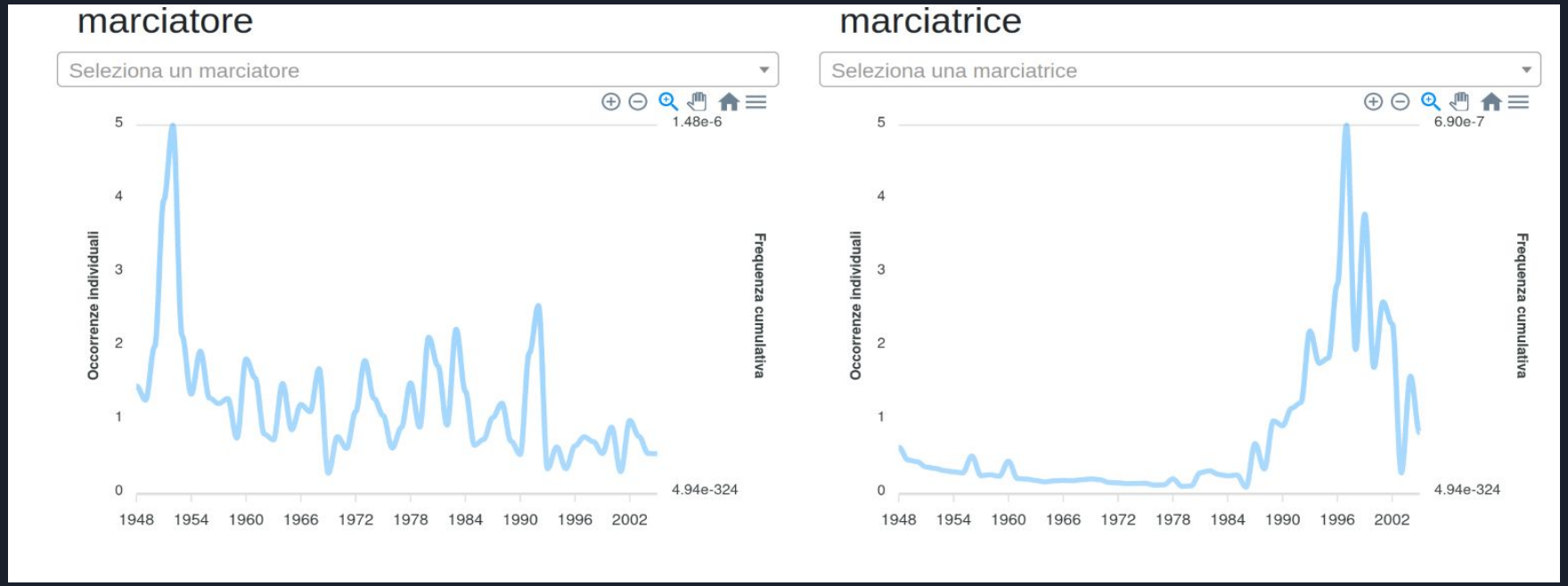


Frequencies of fotomodello/fotomodella

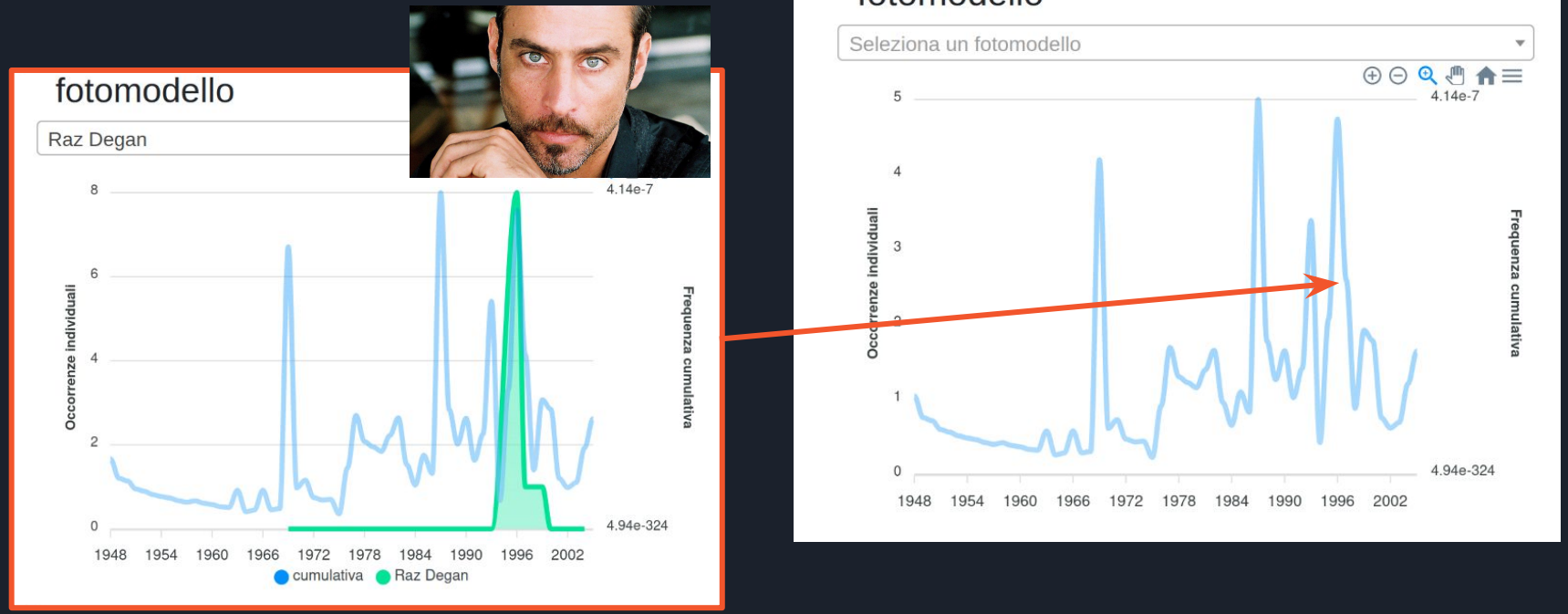




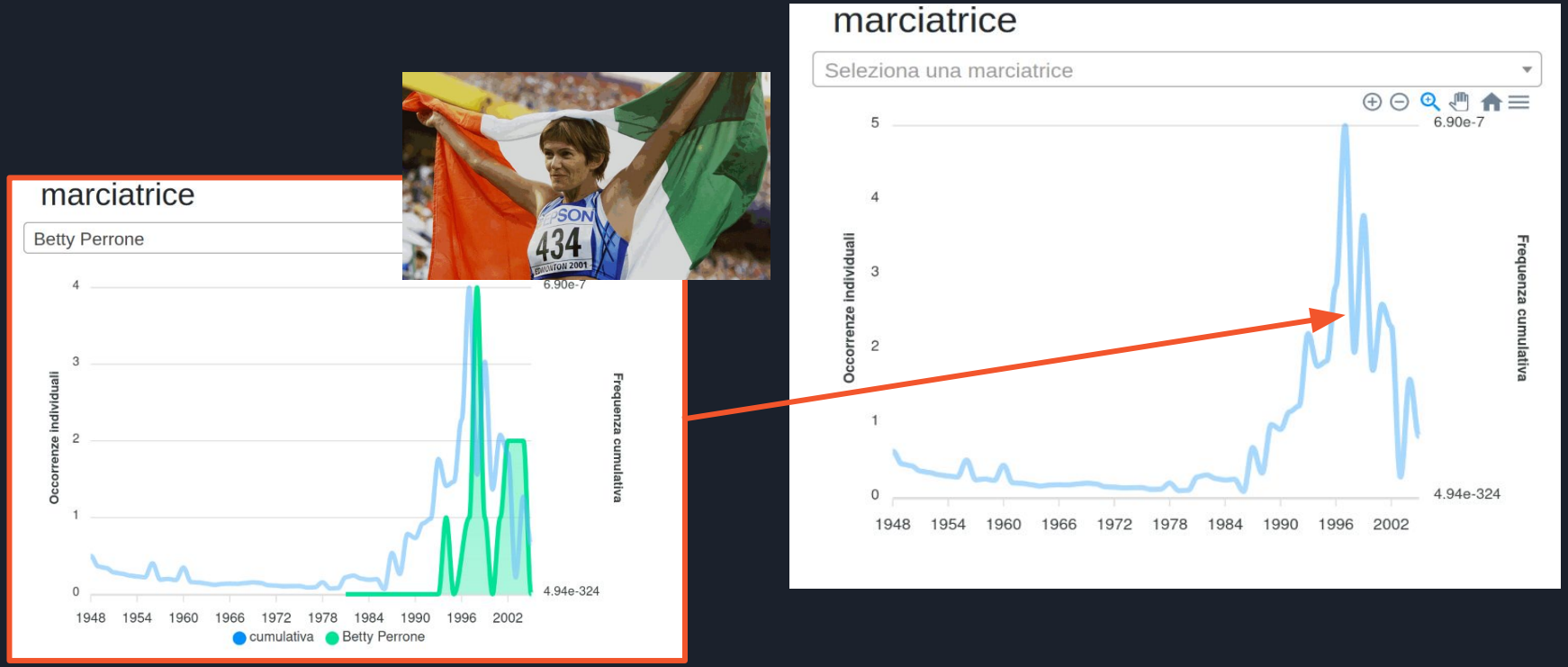
Frequencies of marciatore/marciatrice



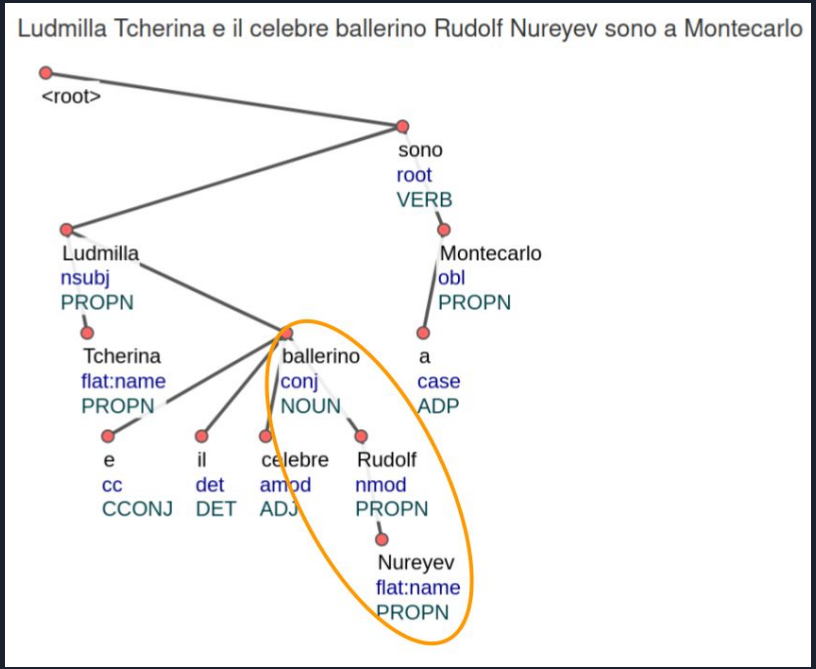
Frequencies of fotomodello



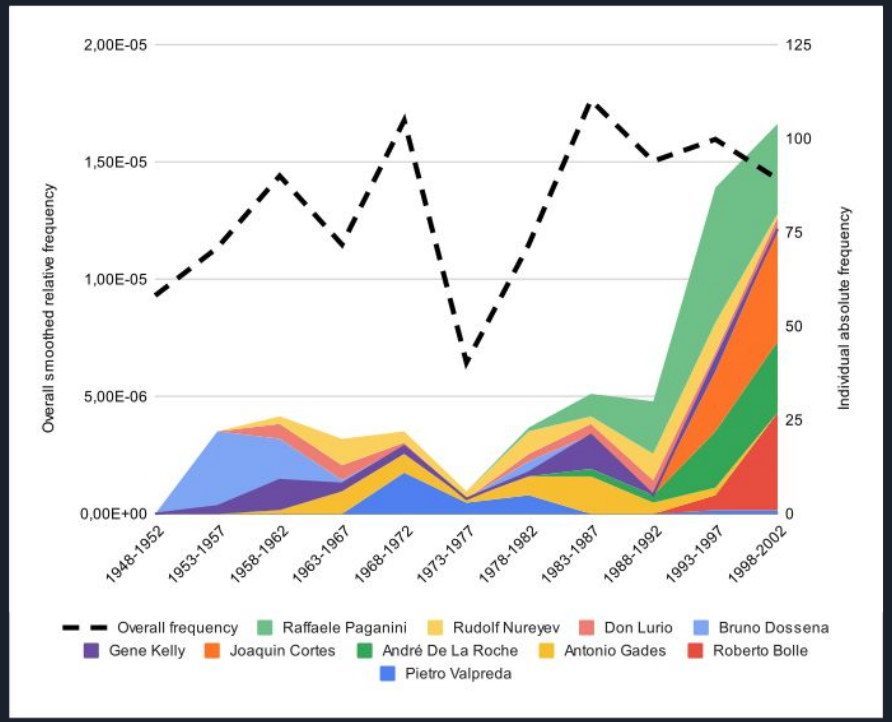
Frequencies of marciatrice



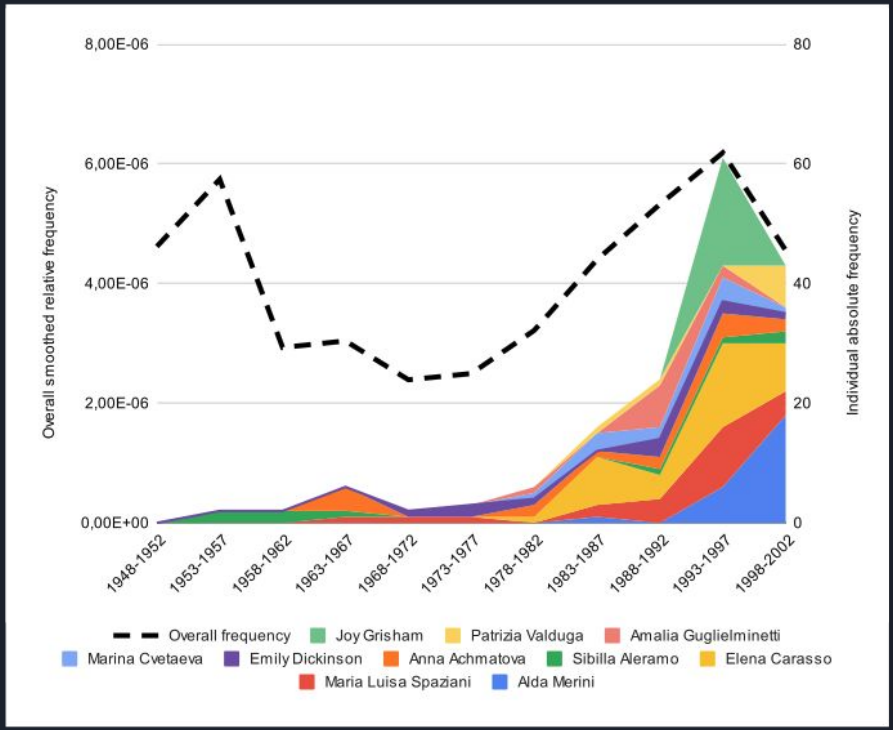
Entities extraction



Entities: *ballerino* example

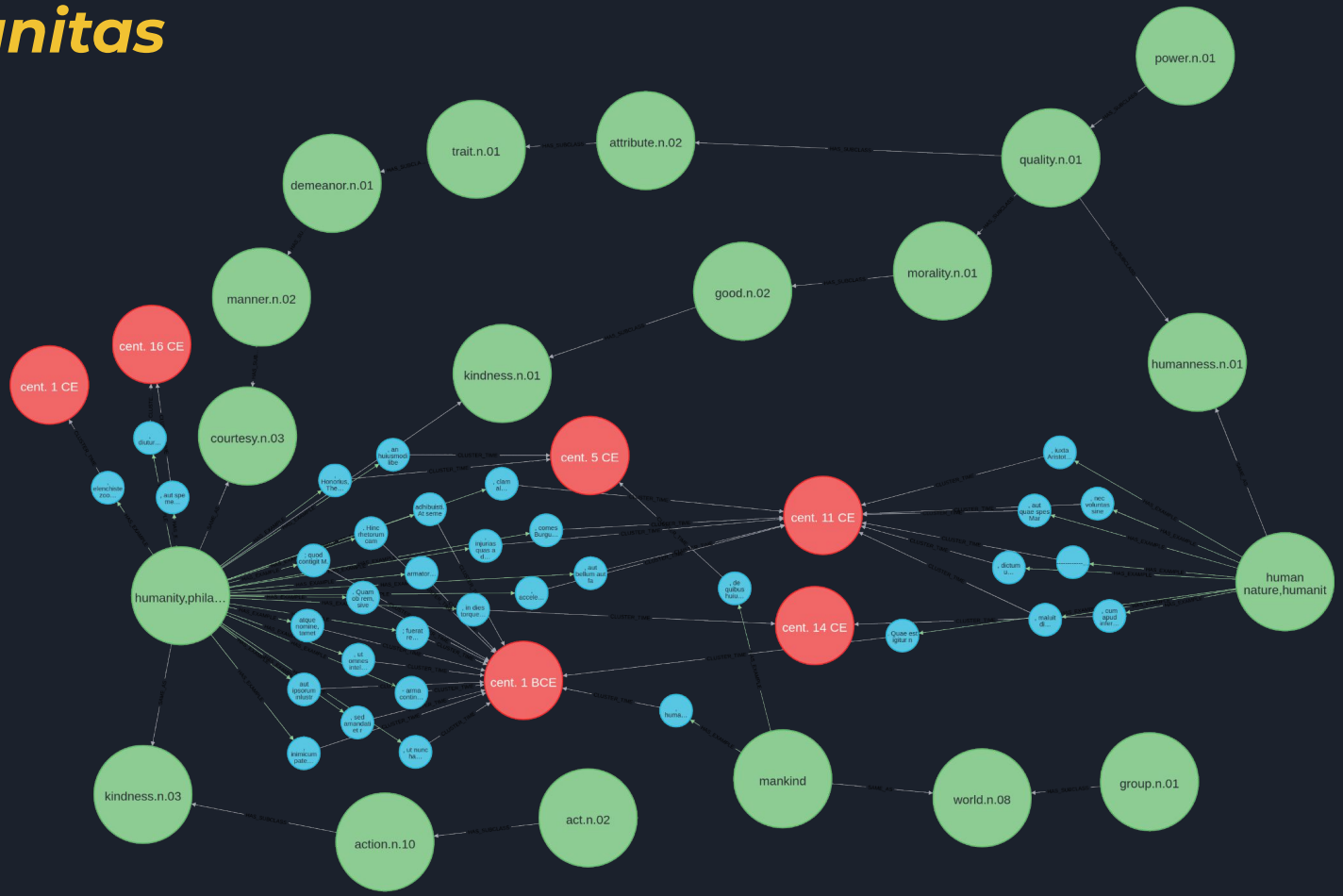


Entities: *poetessa* example

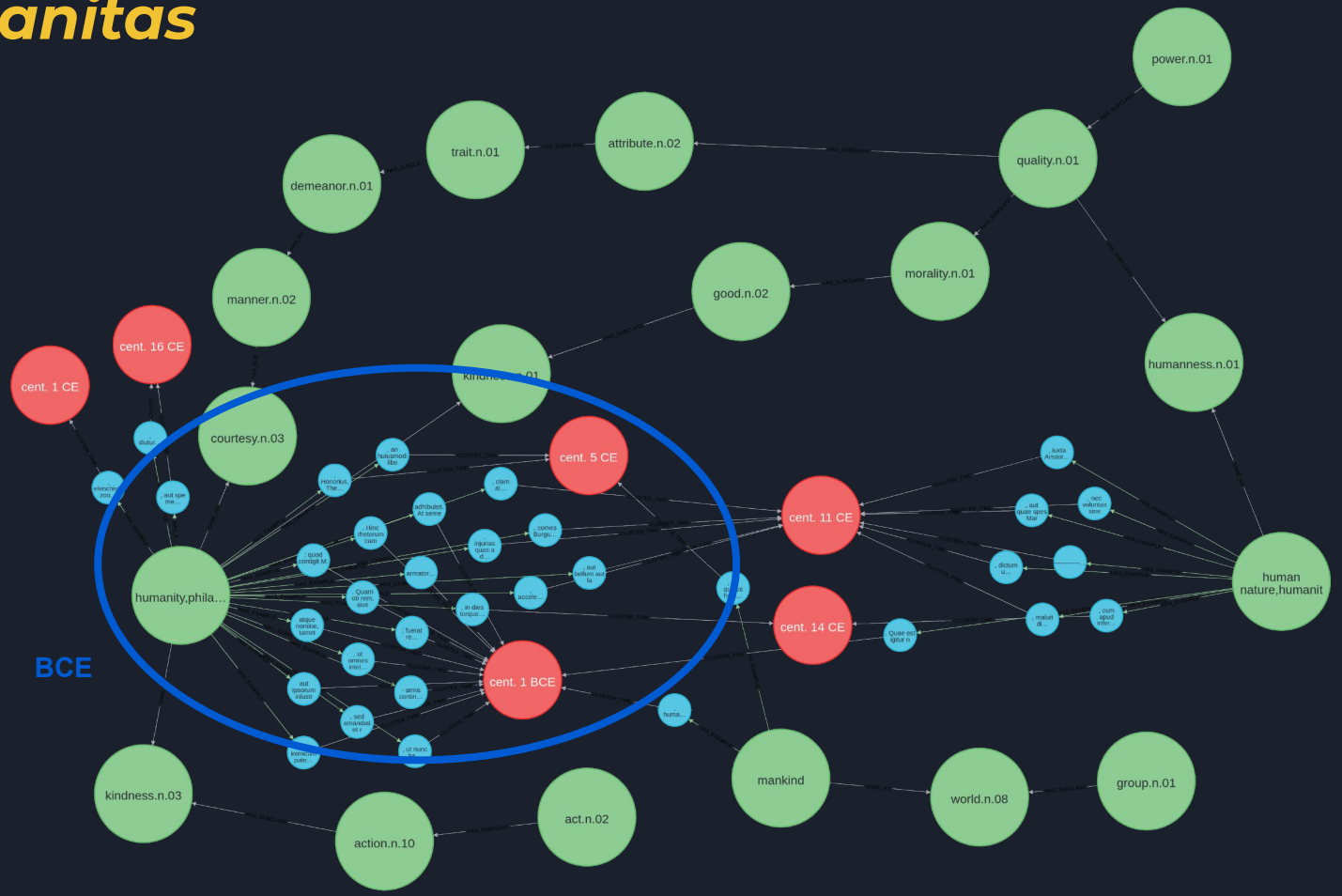


Graph Databases for Diachronic Language Data Modelling

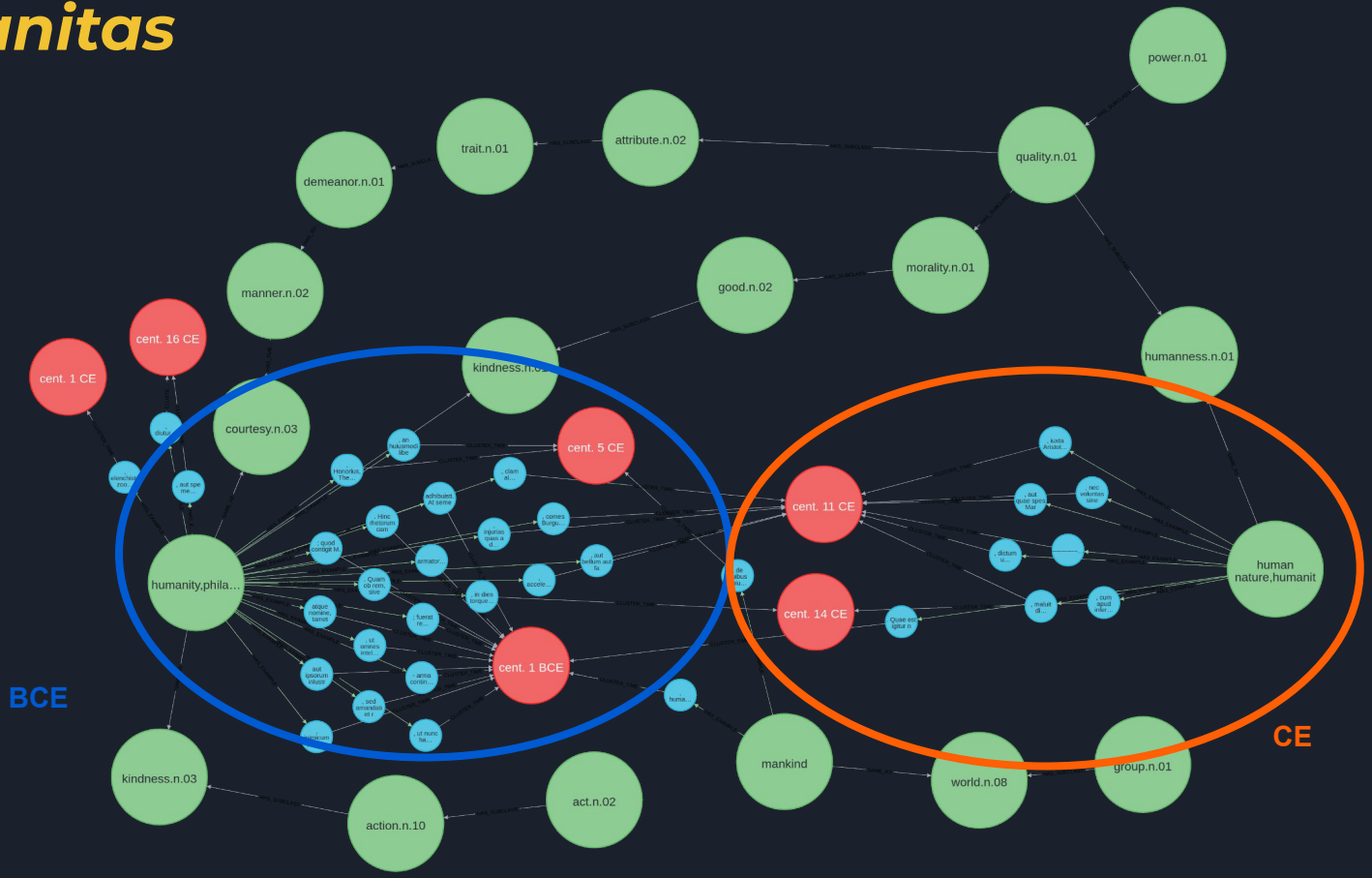
Exploiting the WordNet Hierarchy: the case of *humanitas*



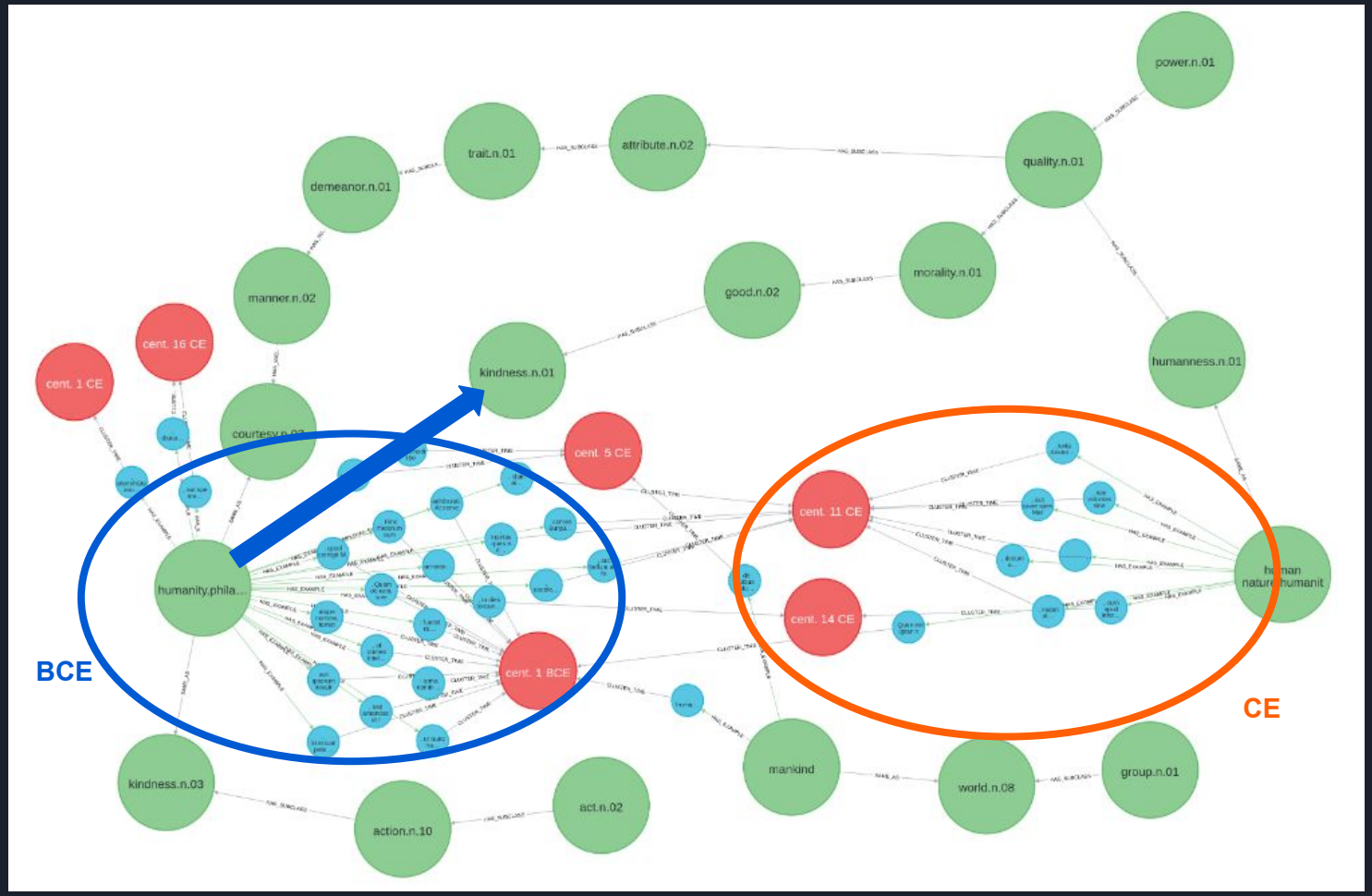
Exploiting the WordNet Hierarchy: the case of *humanitas*



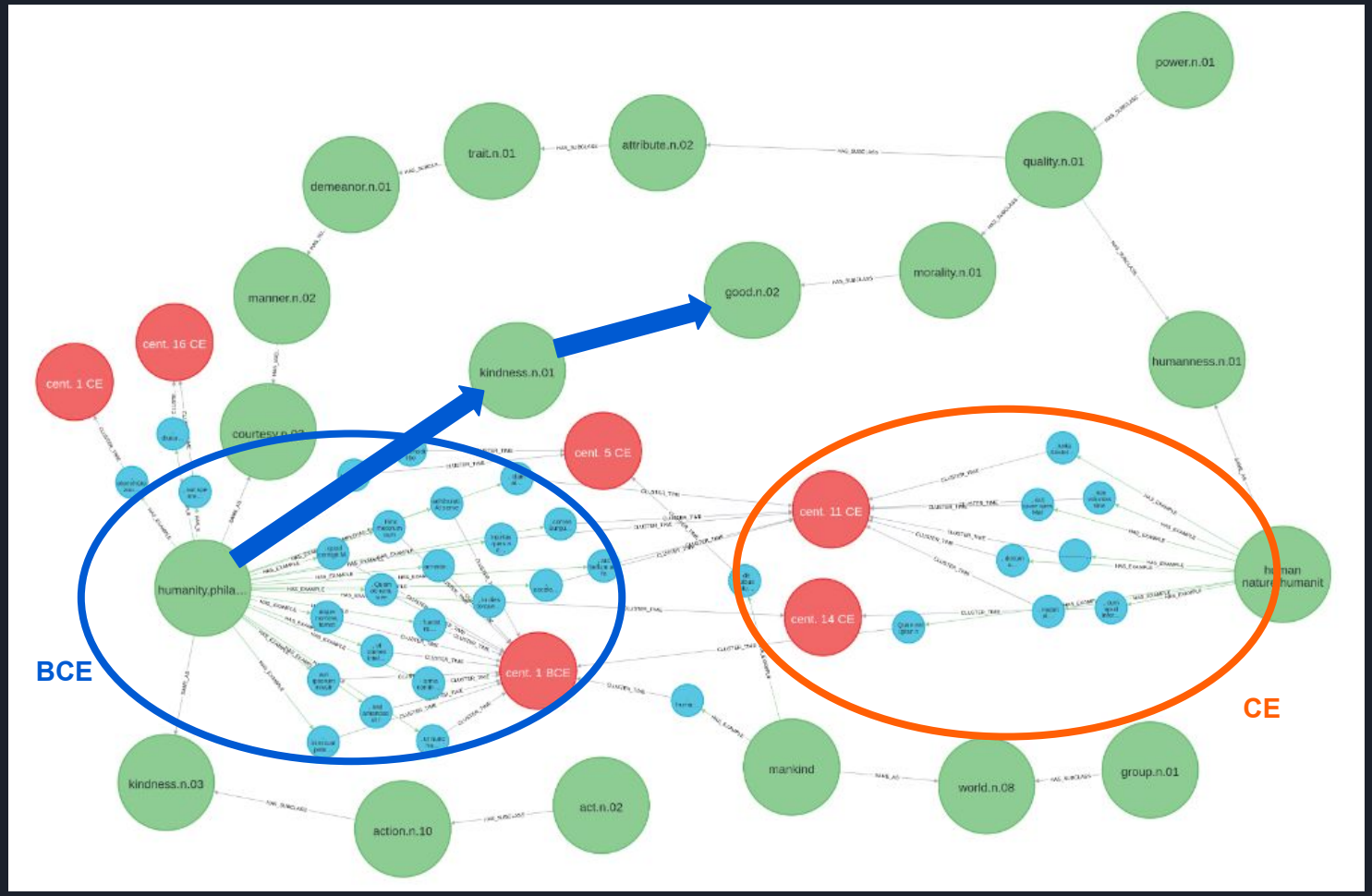
Exploiting the WordNet Hierarchy: the case of *humanitas*



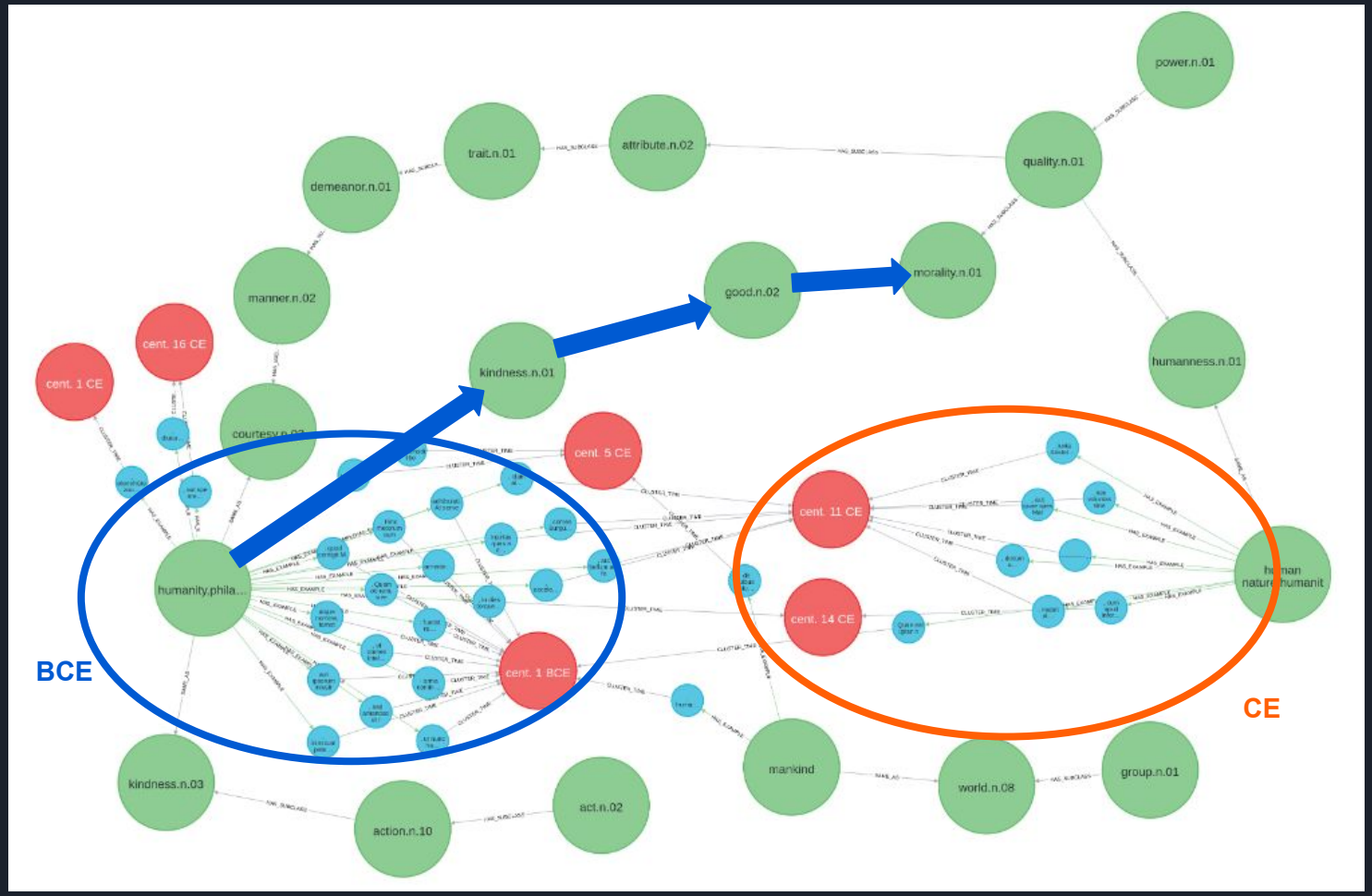
Analysis: Exploiting the WordNet Hierarchy



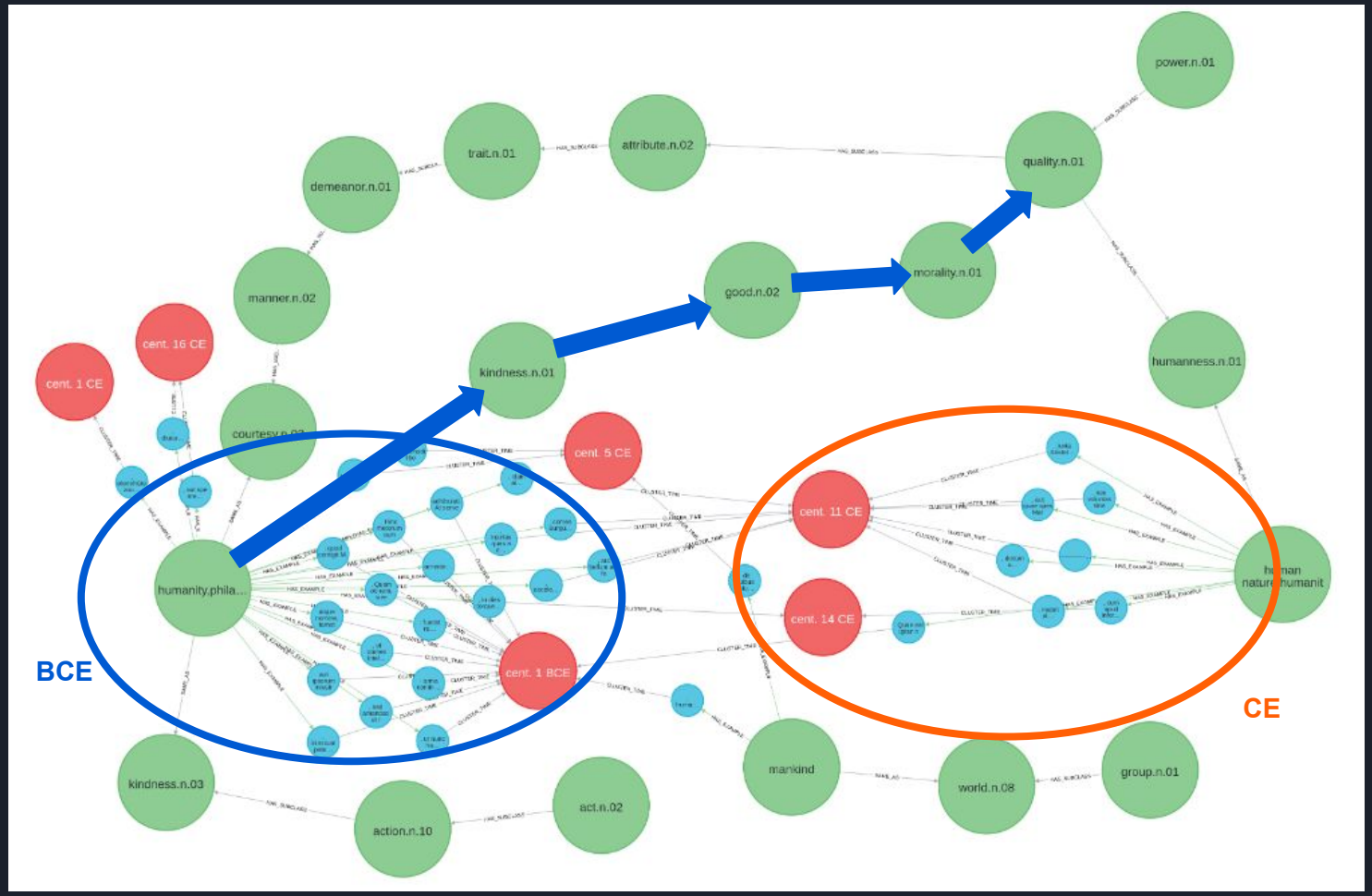
Analysis: Exploiting the WordNet Hierarchy



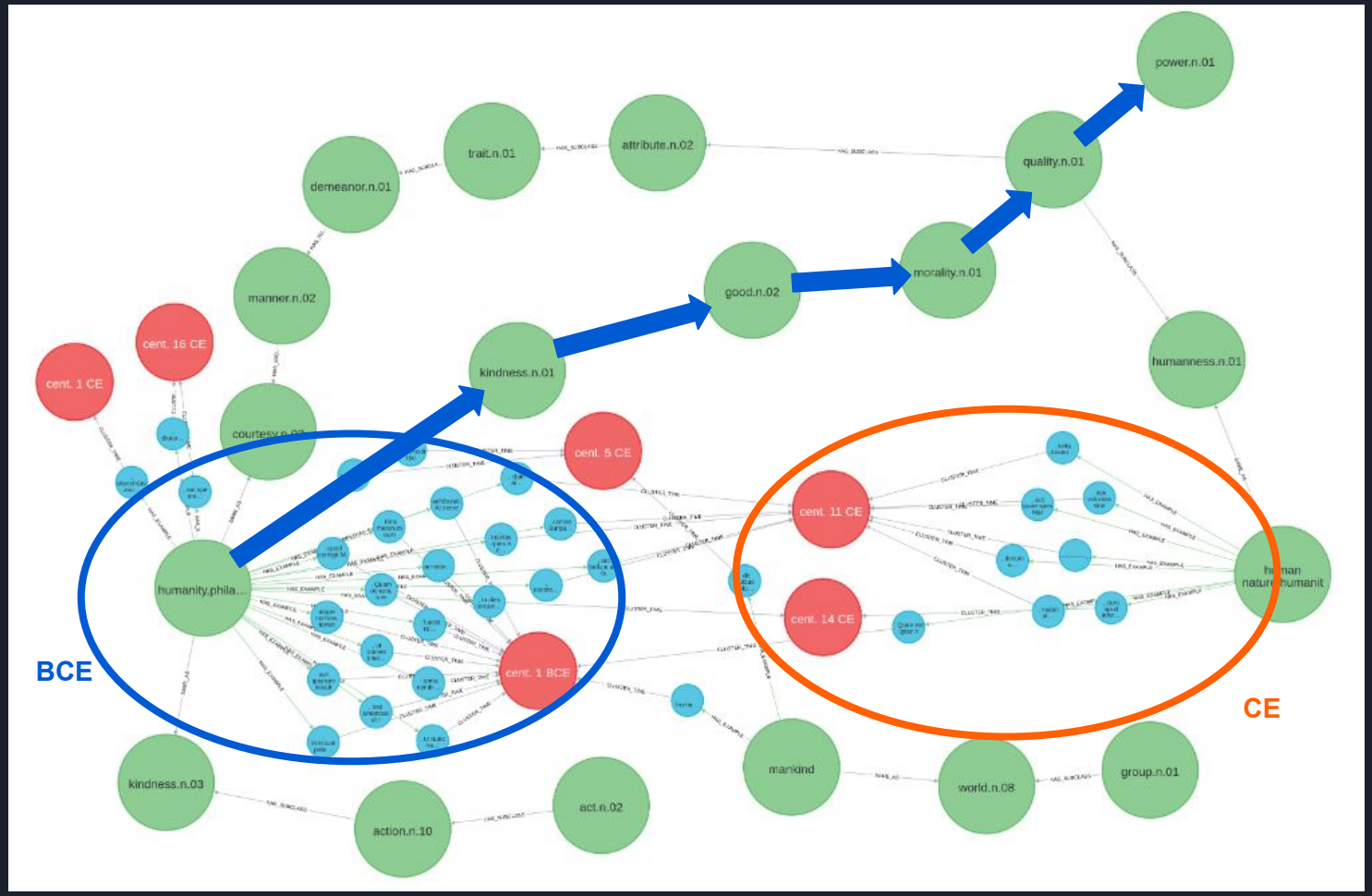
Analysis: Exploiting the WordNet Hierarchy



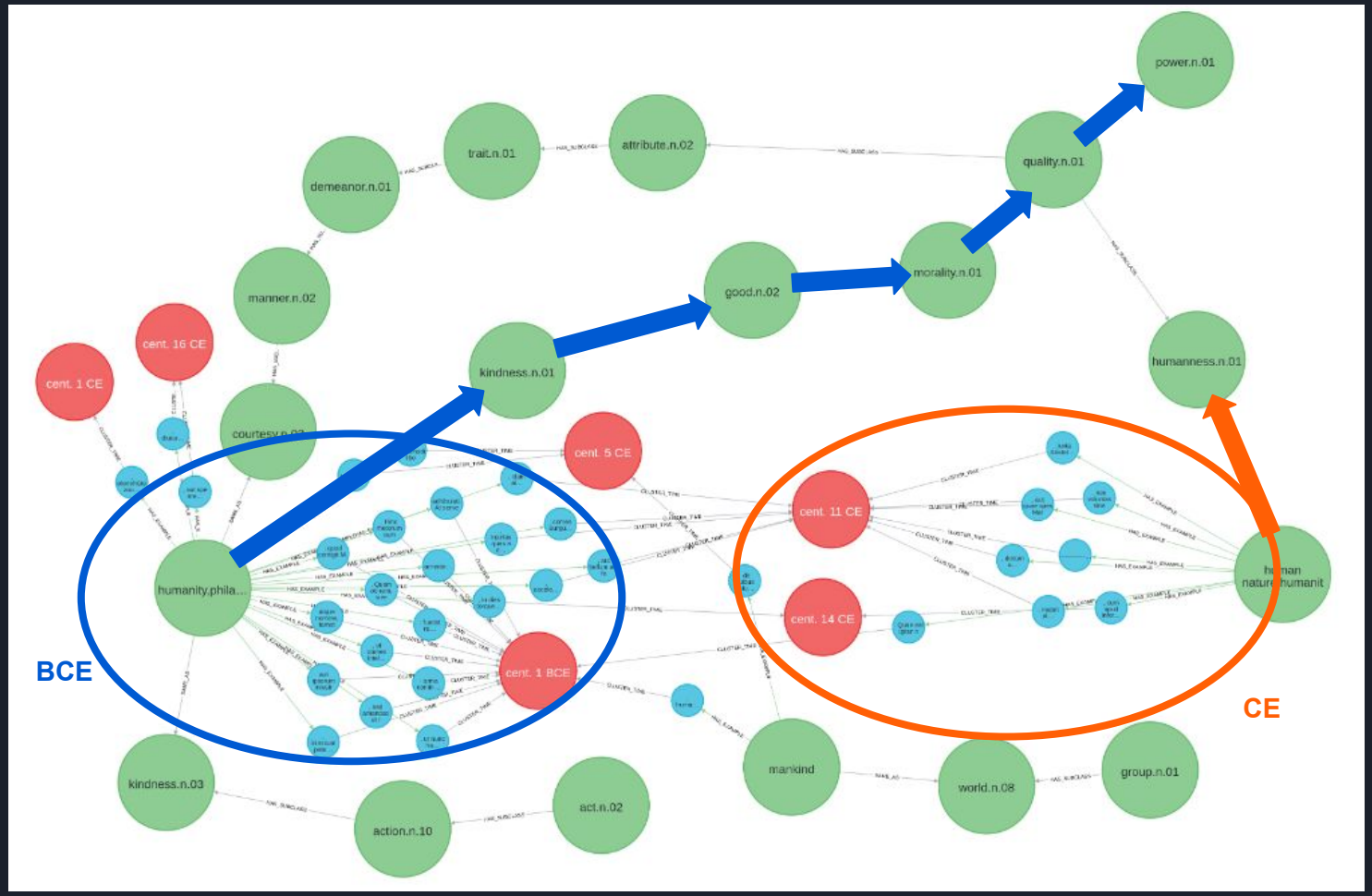
Analysis: Exploiting the WordNet Hierarchy



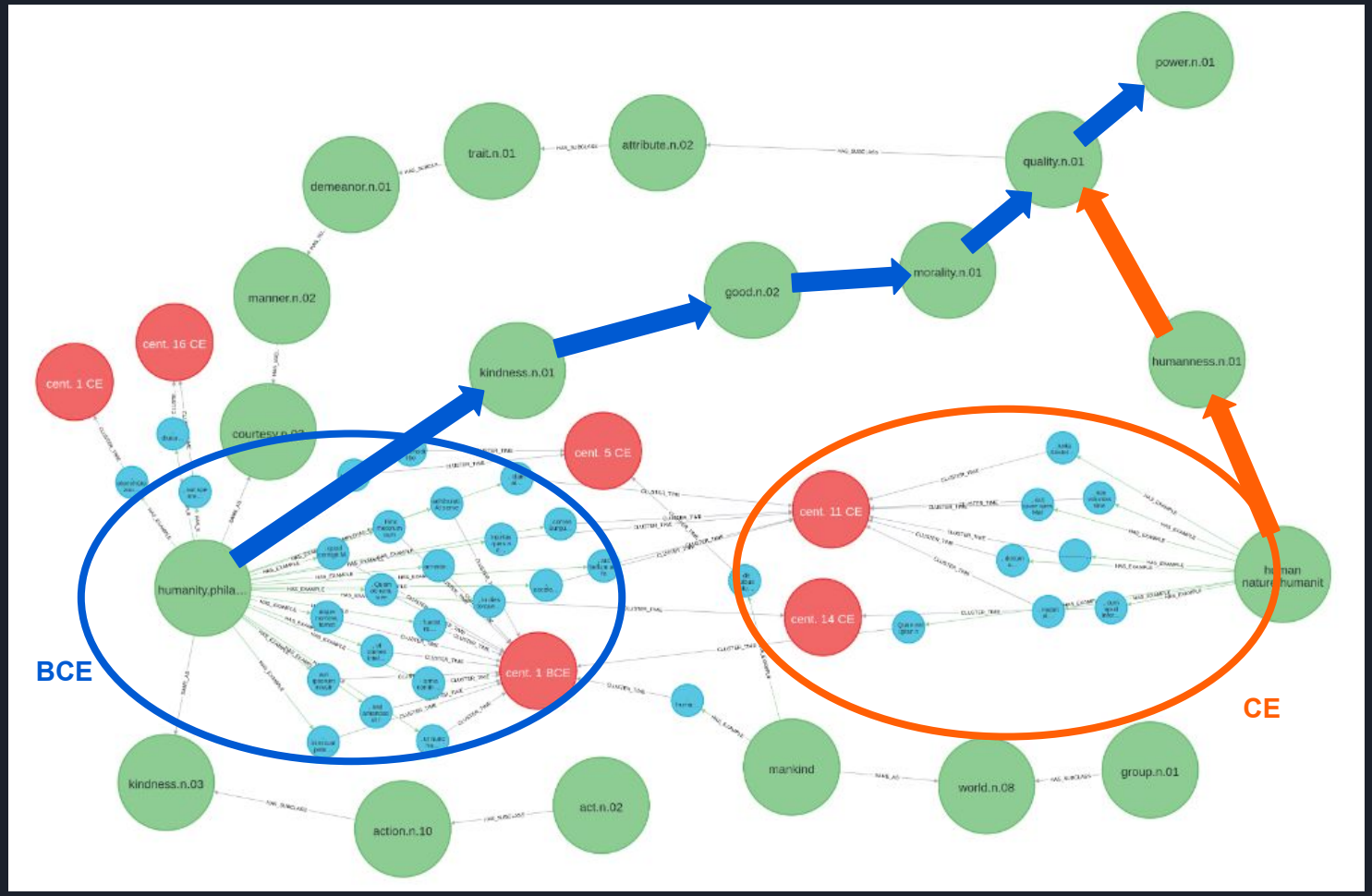
Analysis: Exploiting the WordNet Hierarchy



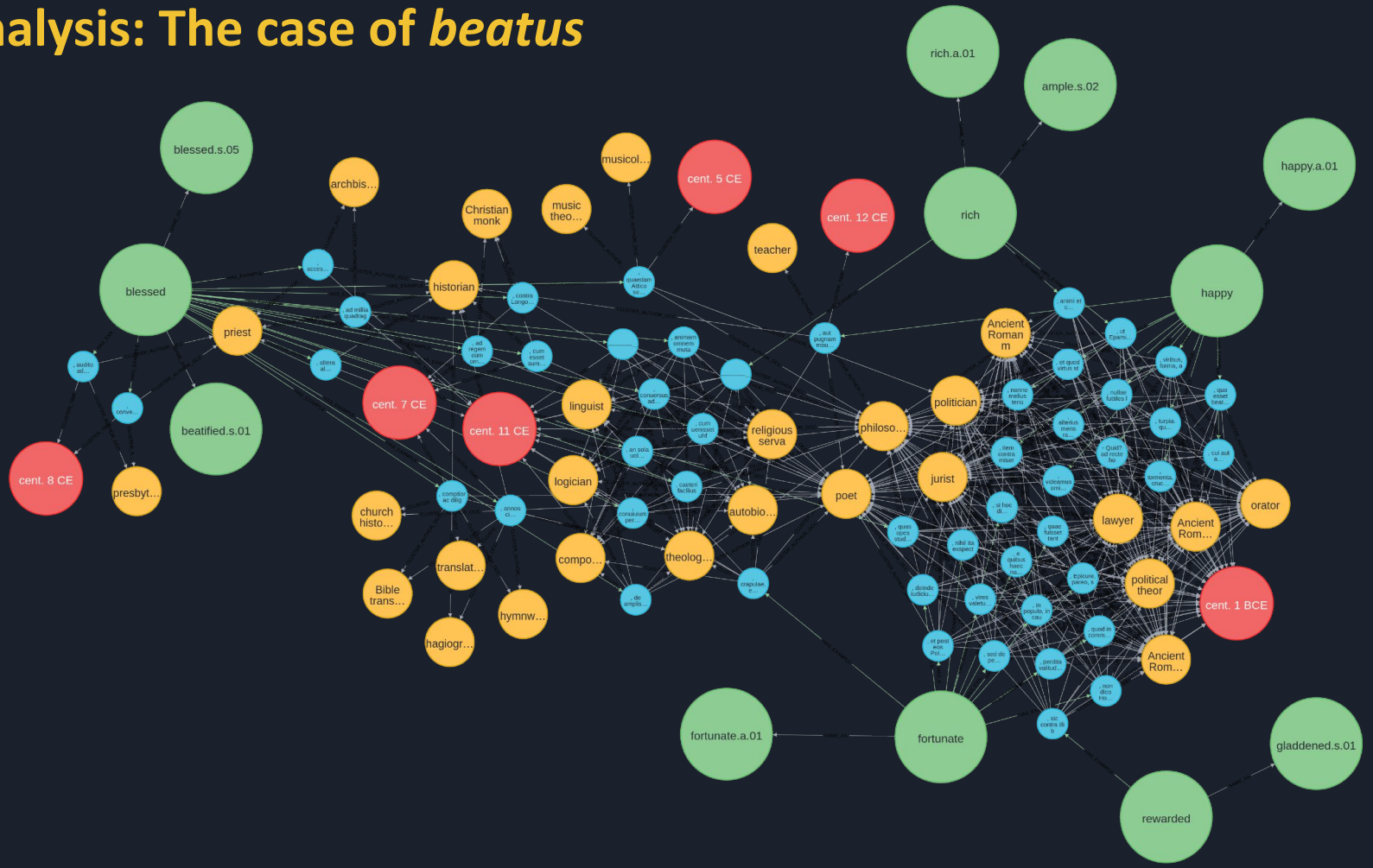
Analysis: Exploiting the WordNet Hierarchy



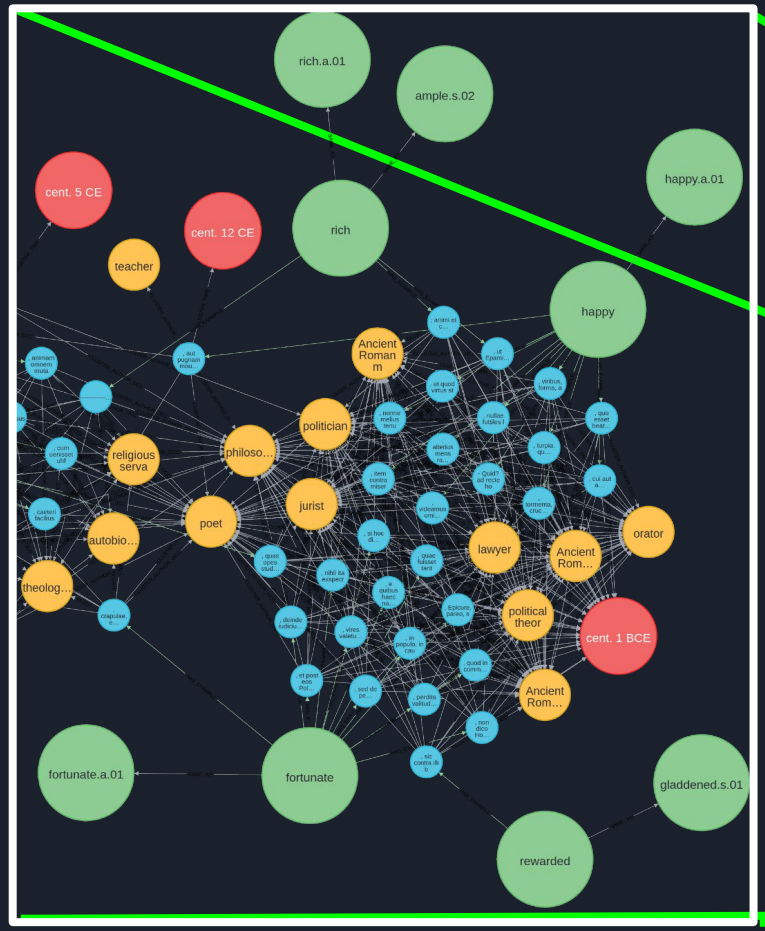
Analysis: Exploiting the WordNet Hierarchy

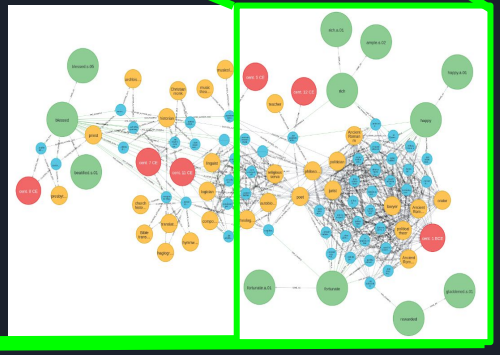


Analysis: The case of *beatus*



Analysis: The case of *beatus*





Thank you for your attention!

e-mail: pierluigi.cassotti@uniba.it

Revealing Semantic Variation In Swedish Using Computational Models Of Semantic Proximity

(A Case Study)

Dominik Schlechtweg
Shafqat Mumtaz Virk
Emma Sköldberg

November 1, 2023

Background: Aims of the current case study

- Lexicography is one the application areas that we promised to focus on in the *Change Is Key* program.
- In summary, we promised to develop methods/tools to assist lexicographers in their work to identify and record semantic changes in the vocabulary of a language (Swedish); collaborative work
- The current work is a first step towards fulfilling that promise.
  - Make the available computational resources more usable for lexicographers.

DURel

Background: The Contemporary Dictionary of the Swedish Academy' (SO, 2021)

65 000 headwords

SOME QUESTIONS WITHIN THE DICTIONARY PROJECT:

Are the semantic descriptions of the headwords up to date?

Have the meanings of the headwords developed in some way since the 2nd edition (2021)?

Background: The Contemporary Dictionary of the Swedish Academy' (SO, 2021)

65 000 headwords

SOME QUESTIONS WITHIN THE DICTIONARY PROJECT:

Are the semantic descriptions of the headwords up to date?

Have the meanings of the headwords developed in some way since the 2nd edition (2021)?

The SO-lexicographers currently do not use any formal, computational methods for discovering semantic changes.

Data: 50 polysemous SO headwords in focus

Some examples

lemma	part of speech	meanings in SO (2021)	English (rough translation)
bagage	noun	1 main sense, 1 subsense (fig.)	luggage, baggage
baksida	noun	1 main sense, 2 subsenses (ext., fig.)	back, downside, drawback
enkelspårig	adjective	1 main sense, 1 subsense (fig.)	one-track, simplistic
fasad	noun	1 main sense, 1 subsense (fig.)	front, facade
fotavtryck	noun	1 main sense, 1 subsense (fig.)	footprint

Data: Corpora

- The SVT corpora (including news from the Swedish public service television company, 2004-2021) in Språkbanken Text/Korp.
  - 21 corpora
  - 240,393,329 tokens
  - 15,991,049 sentences

Data Preparation

- Selection of 20 polysemous words with at least two senses represented in the data.
- Usage extraction from the SVT corpora. 50 random uses (a sentence in our case) per word.
- Filtering to exclude duplicates (5 tokens on either side of the candidate word) among the uses.

Semantic Proximity and Word-Usage-Graphs

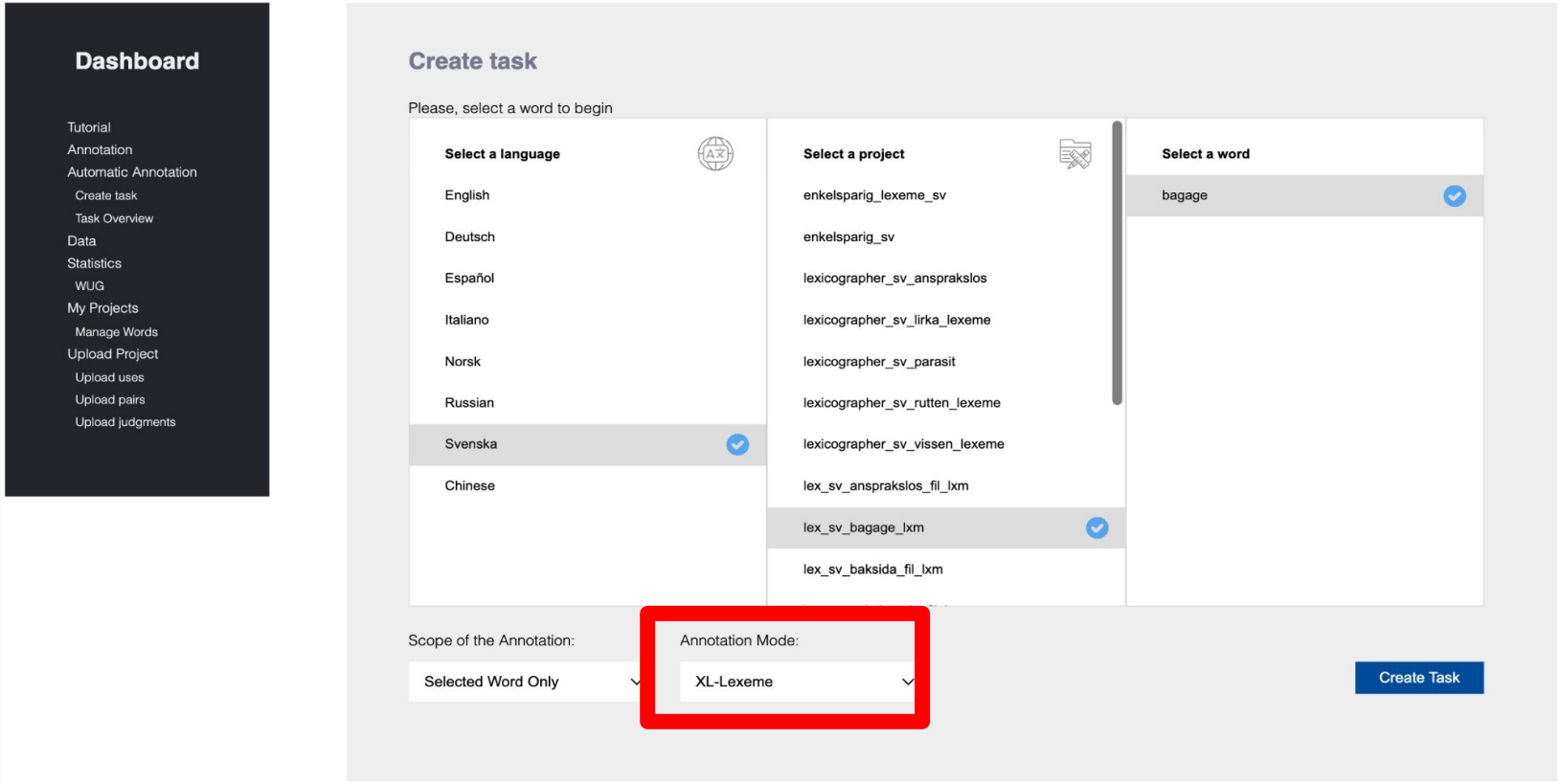
DURel

Uses

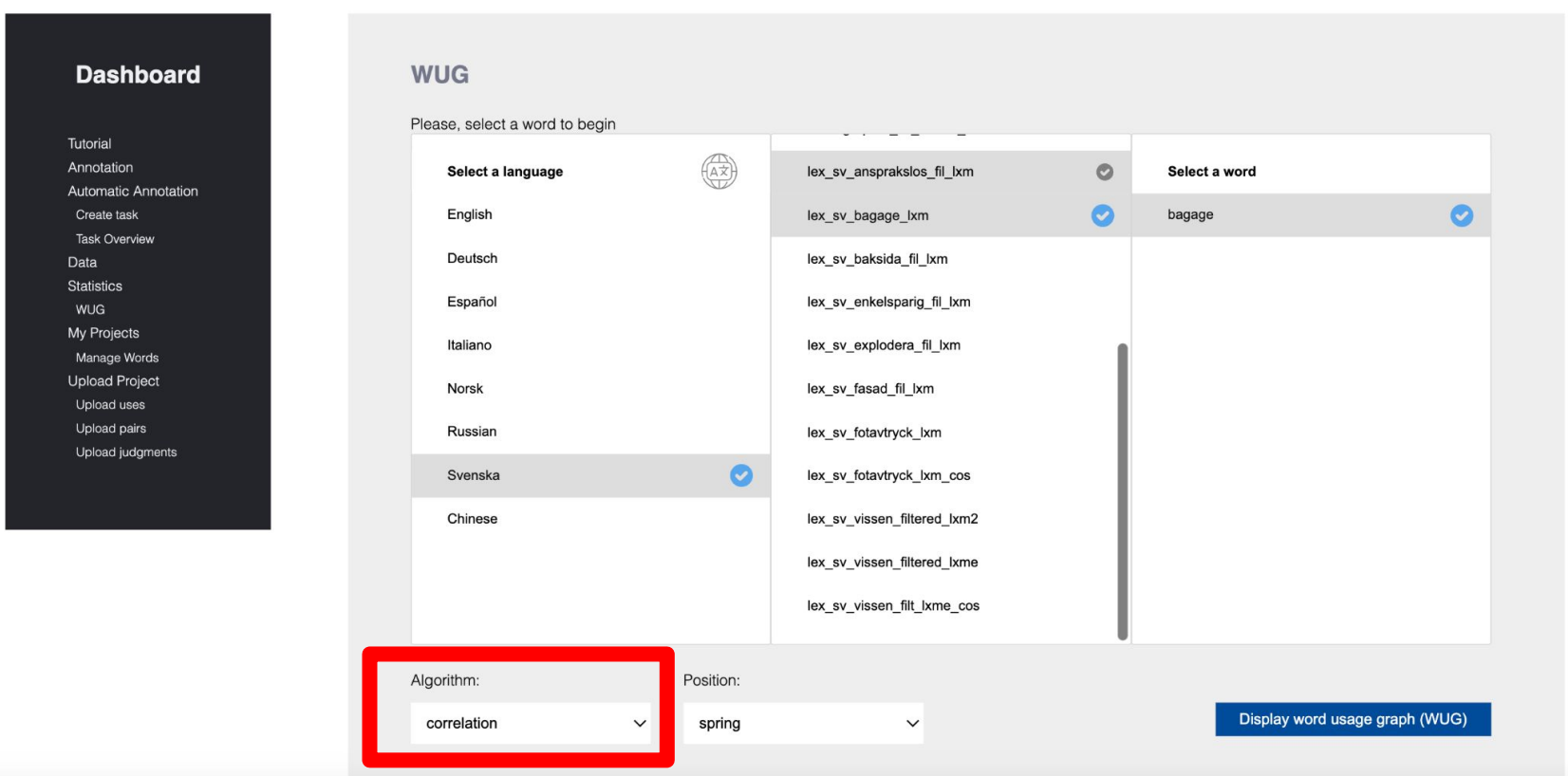
Back

ID	Lemma	POS	Date	Left Context	Target	Right Context
1283960	bagage	NN	2019	Varken piloten eller framsätespassageraren använde axelremsbälte, och	**bagaget**	i lastutrymmet var inte fastsurrat.
1283961	bagage	NN	2019	Med i	**bagaget**	har S ytterligare en bottennotering från riksdagsvalet.
1283962	bagage	NN	2007	Maria Norrfalk har inte företrädarens kulturprofil, men andra kunskaper och erfarenheter i	**bagaget**	som också har alla utsikter att bli en tillgång i det nya uppdraget. ”
1283963	bagage	NN	2010	I	**bagaget**	hade de avancerad sk skimmingsutrustning.
1283964	bagage	NN	2019	Hon har 160 landskamper i	**bagaget**	och var med och tog SM-guld 2012.
1283965	bagage	NN	2013	– Man får billigare pris, men man får ta slitet och betala för	**bagage**	och kolla vikt och sådär, säger Ylva Bailey.
1283966	bagage	NN	2013	Inför sista heatet i går hade Vetlanda chansen att gå segrande med tio poäng i	**bagaget**	till i dag.
1283967	bagage	NN	2013	Försvunnet	**bagage**	kommer tillbaka
1283968	bagage	NN	2013	Frida Hansdotter från Norberg, som hittills har fyra andraplatser i	**bagaget**	den här säsongen, knep VM-bronset.
1283969	bagage	NN	2016	Inget	**bagage**	– planet blev för tungt
1283970	bagage	NN	2013	En av de mest namnkunniga är Maksim Vylegzjanin, en 31-åring som tillhört världseliten i snart tio år med två VM-silver från femmilen och ett från skiathlon i	**bagaget**	– och en av få som har slagit norrmannen Petter Northug i en spurtduell.
1283971	bagage	NN	2020	Innebandyklubben Iksu, med sju SM-guld i	**bagaget**	, kommer inte spela i SSL i nästa säsong.
1283972	bagage	NN	2013	Ingemar Isaksson är kriminalkommissarie med 25 års erfarenhet av kvalificerat mordutredande i	**bagaget**	.

Semantic Proximity and Word-Usage-Graphs



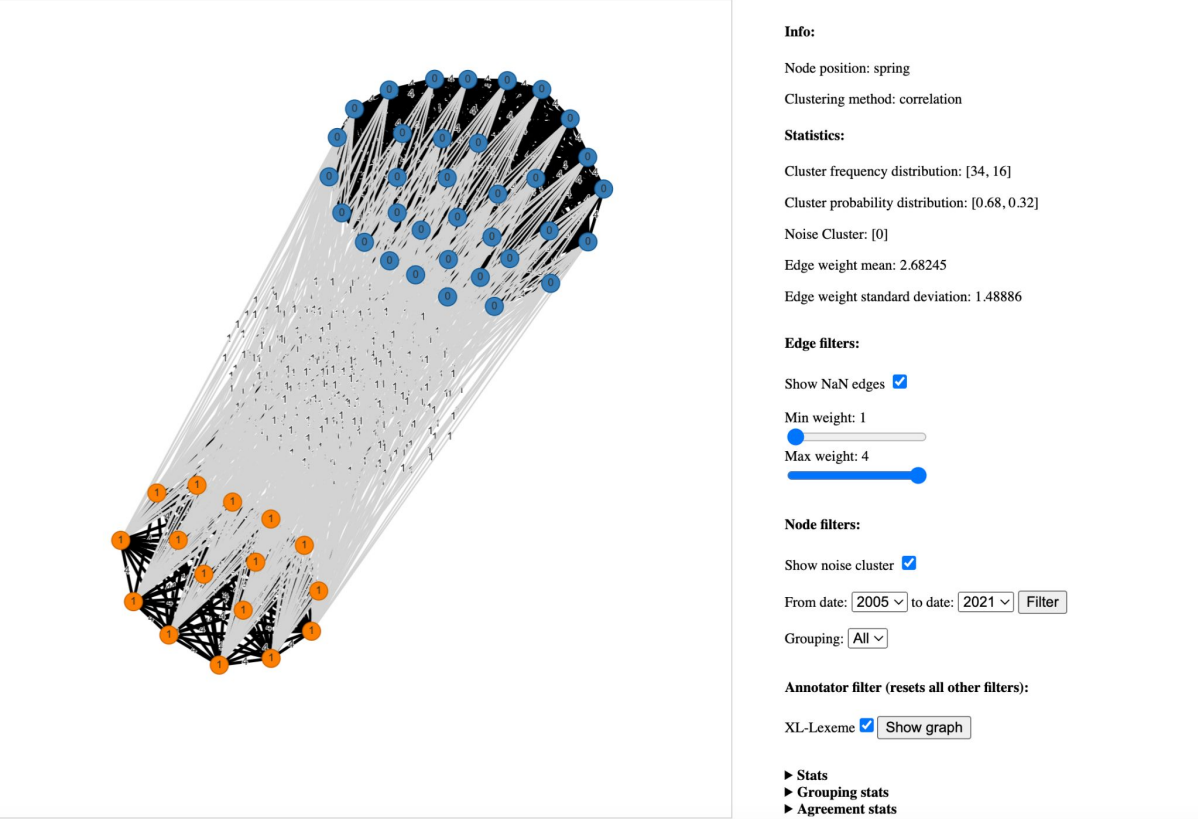
Semantic Proximity and Word-Usage-Graphs



Semantic Proximity and Word-Usage-Graphs: *enkelspårig*

BLUE: Den **enkelspåriga** järnvägen mellan Motala och Hallsberg är idag en flaskhals (...). (‘The single-track railway between Motala and Hallsberg is today a bottleneck’)

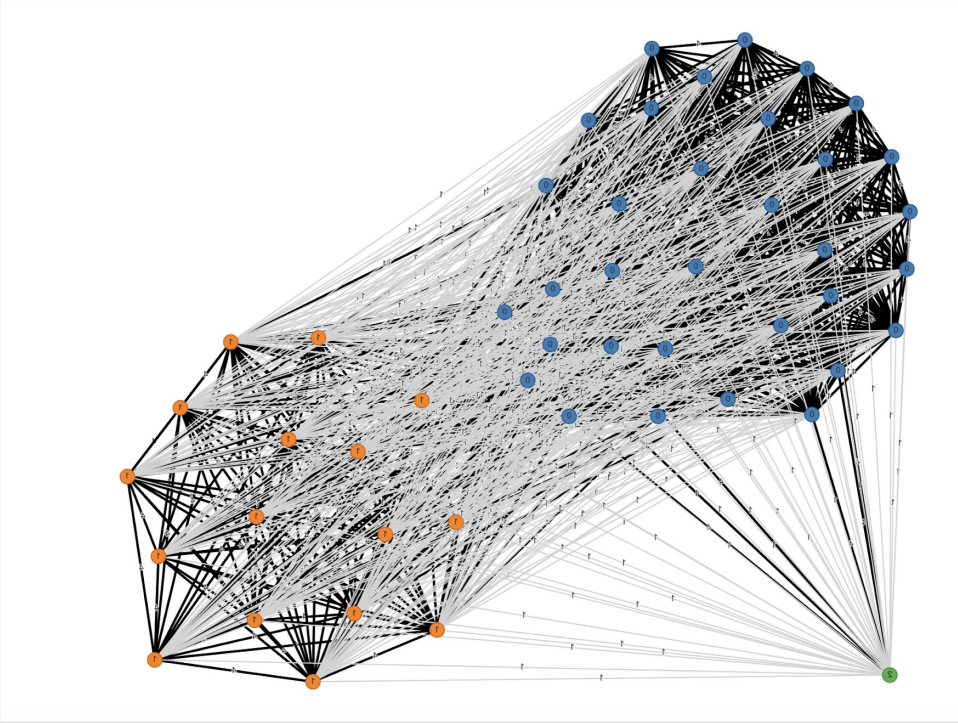
ORANGE: De tror att vi är **enkelspåriga** lantisar, de tror att vi är trångsynta, att vi är rasister och homofober. (They think we're narrow-minded peasants, they think we're bigoted, that we're racists and homophobes’)



Semantic Proximity and Word-Usage-Graphs: *bagage*

ORANGE: I **bagaget** hade de avancerad s.k. skimmingsutrustning. (‘In the luggage they had advanced so-called skimming equipment.)

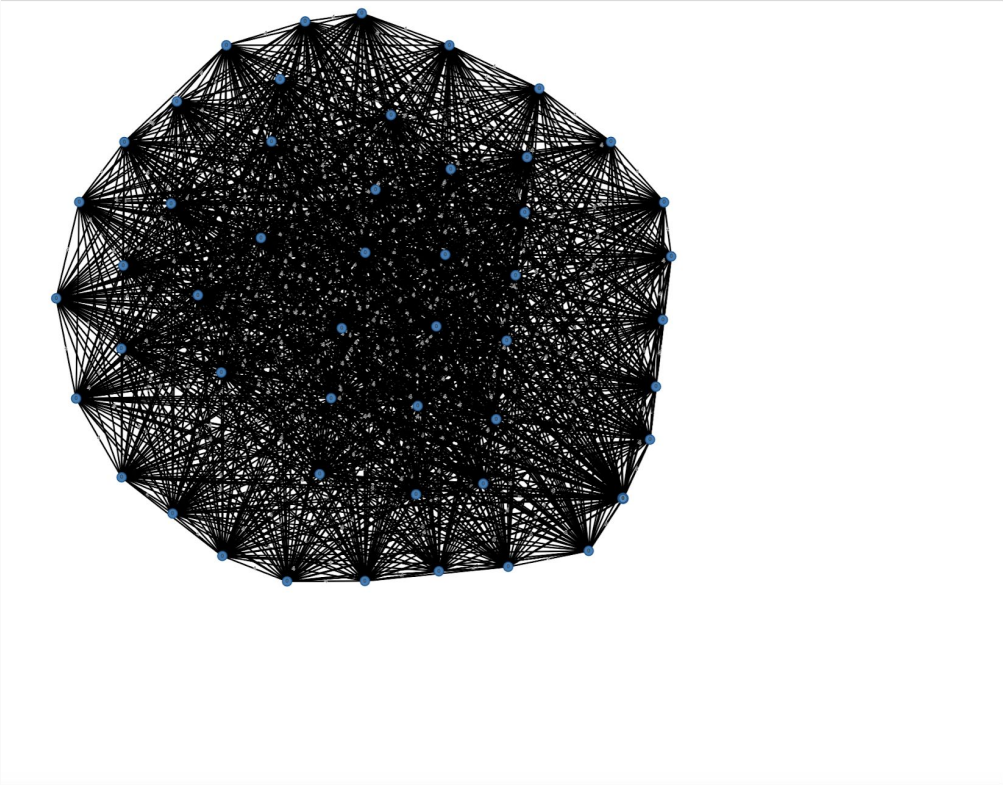
BLUE: Många hade uppslitande händelser i **bagaget**, som dödsfall och skilsmässor. (‘Many had upsetting events in their baggage, such as deaths and divorces.)



Semantic Proximity and Word-Usage-Graphs: *fotavtryck*

BLUE: Arkeologer fann **fotavtrycket** i lera (...) när de höll på att undersöka en antik plats i Siwa. (‘Archaeologists found the footprint in clay (...) while investigating an ancient site in Siwa.’)

BLUE: Det ekologiska **fotavtrycket** från maten är alldeles för stort och köttet är det viktigaste att ta itu med (...). ‘The ecological footprint of food is far too large and the meat is the most important thing to deal with (...)’



Evaluation (manual and limited to 5 polysemous words)

Word/Cluster	Orange			Blue			Green			
	C	I	U	C	I	U	C	I	U	Accuracy
bagage	16	1	1	31				1		47/50
enkelspårig	16			34						50/50
baksida	12	1		37						49/50
fasad	1			47		1		1		48/50
fotavtryck				29	21					29/50

C: Correct **I**: Incorrect: **U**: Unclear

Relevant CL Tasks

- Assign word usages to different clusters (Word Sense Induction)
- Detect different word senses in a usage sample (Semantic Variation Detection)
- Detect non-recorded word senses (Non-Recorded Word Sense Detection)

Future Work

- More polysemous Swedish words
- More usages per word
- Lexicographic error analysis
- More fine-ingrained computational predictions
- Clustering on cosine similarity scores

Thanks for listening!

Questions/Comments/Suggestions?

Change is Key!

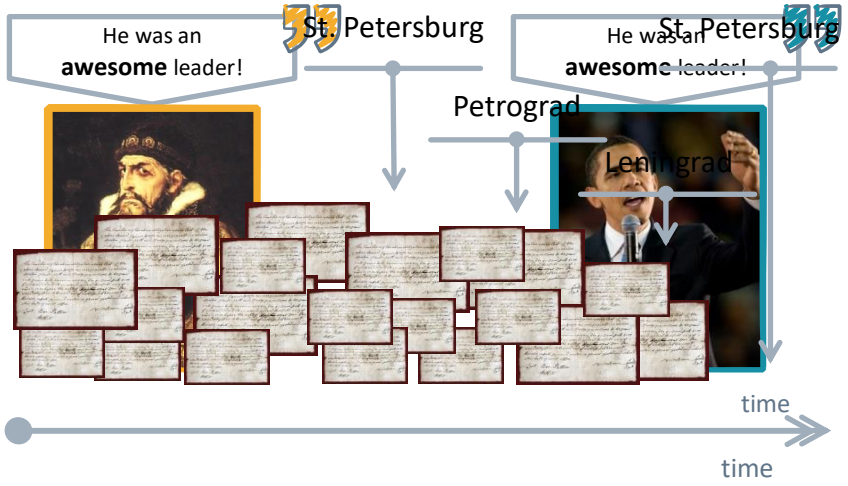
The study of contemporary and historical societies

Change is Key! | GU Seminar | November 1st, 2023

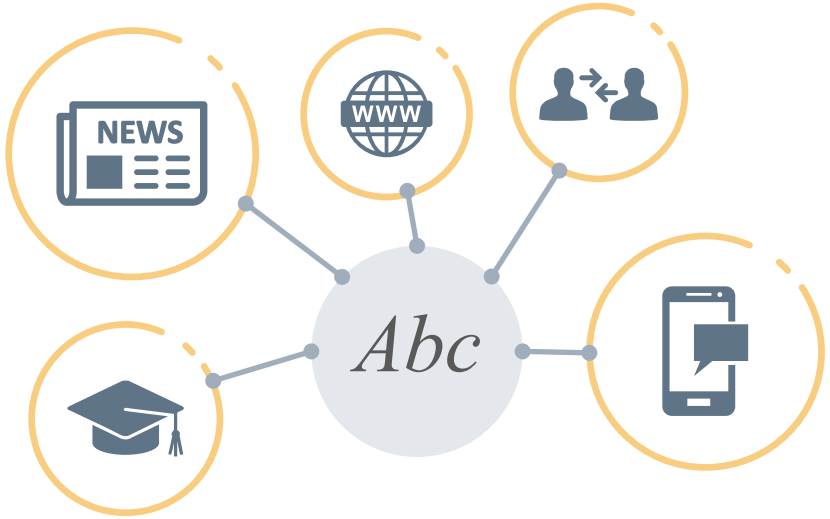

he asked, glancing at
at drop parritch?
was his own supper.
I can do fine wanting it
ckens my cough.' He drank the cup
an eye upon me as he drank, and
nd. 'Let's see the letter,' said he.
letter was for Mr Balfour; not for him
think I am?' says he. 'Give me

Word meaning change

Over time



In different contexts (at the same time)

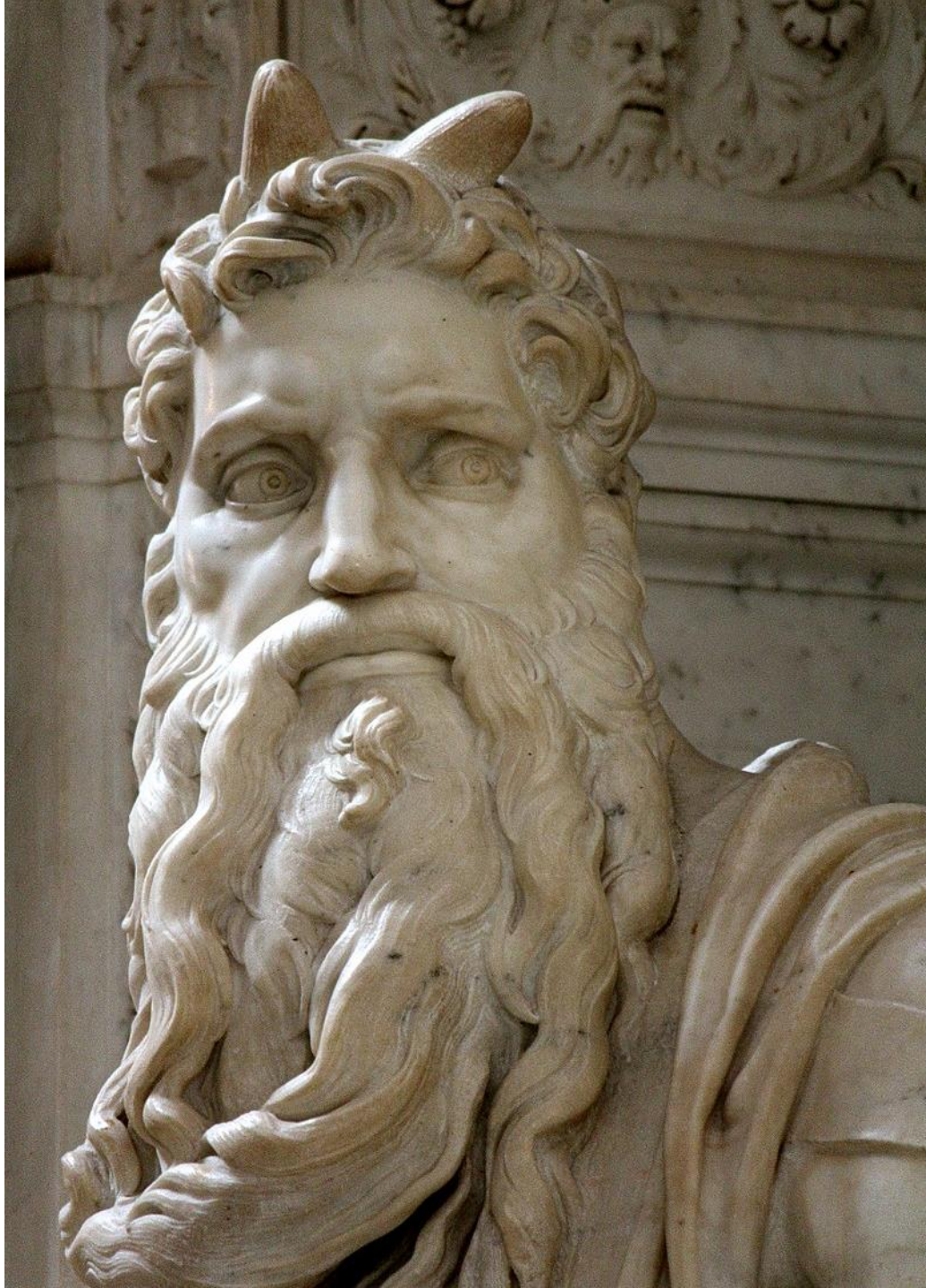


Michelangelo’s Moses,

San Pietro in Vincoli in Rome

1513-1515

קָרַן (qāran)



main CHALLENGES for
computational models of meaning and change

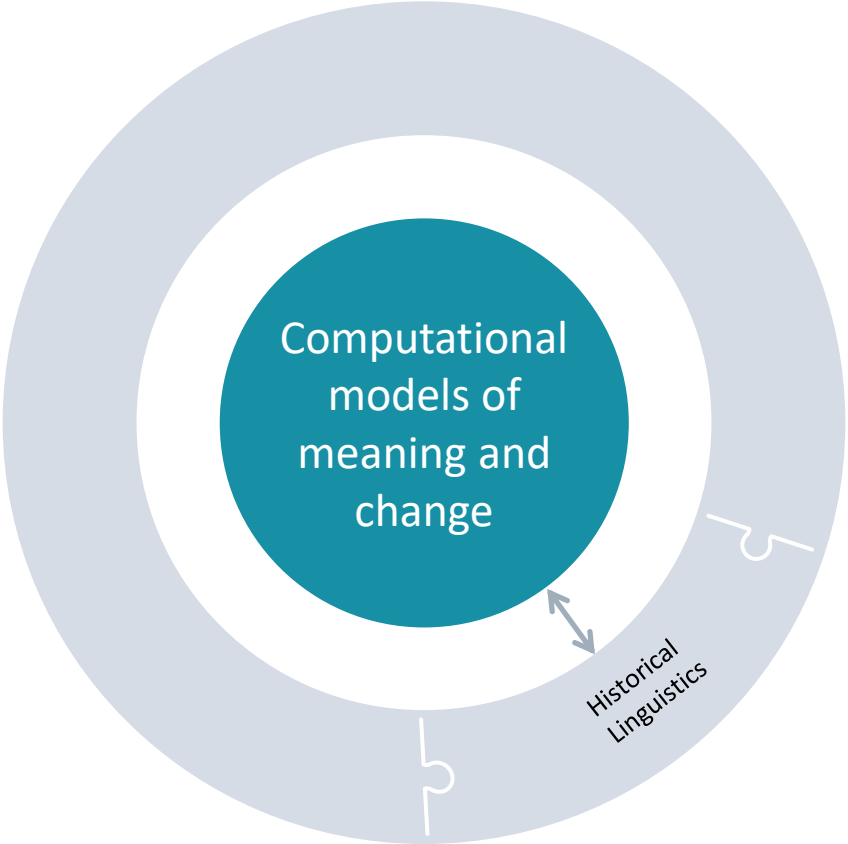
Computational models of meaning and change

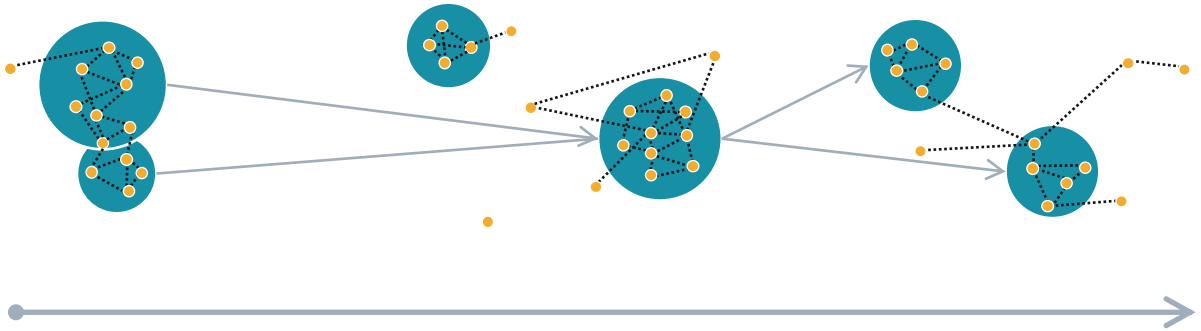
- Handle languages with smaller amounts of data
- Generalize to multiple languages
- Sense-aware models
- Find out WHAT changed, HOW and WHEN

Our Research Questions

Language level change

Historical Linguistics

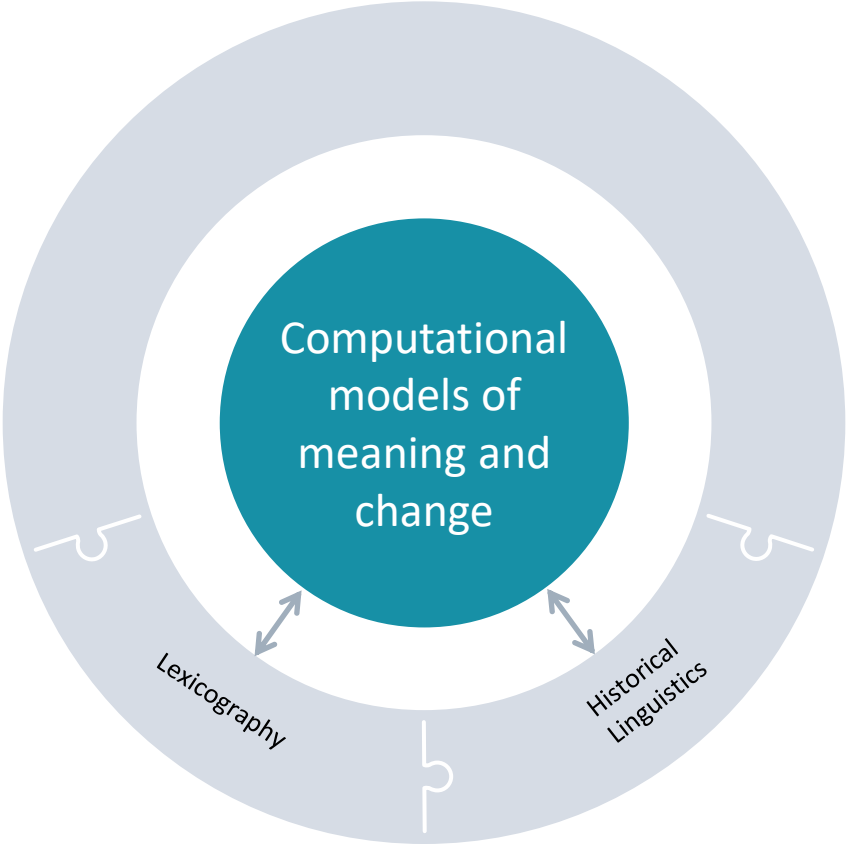




Our Research Questions

Language level change

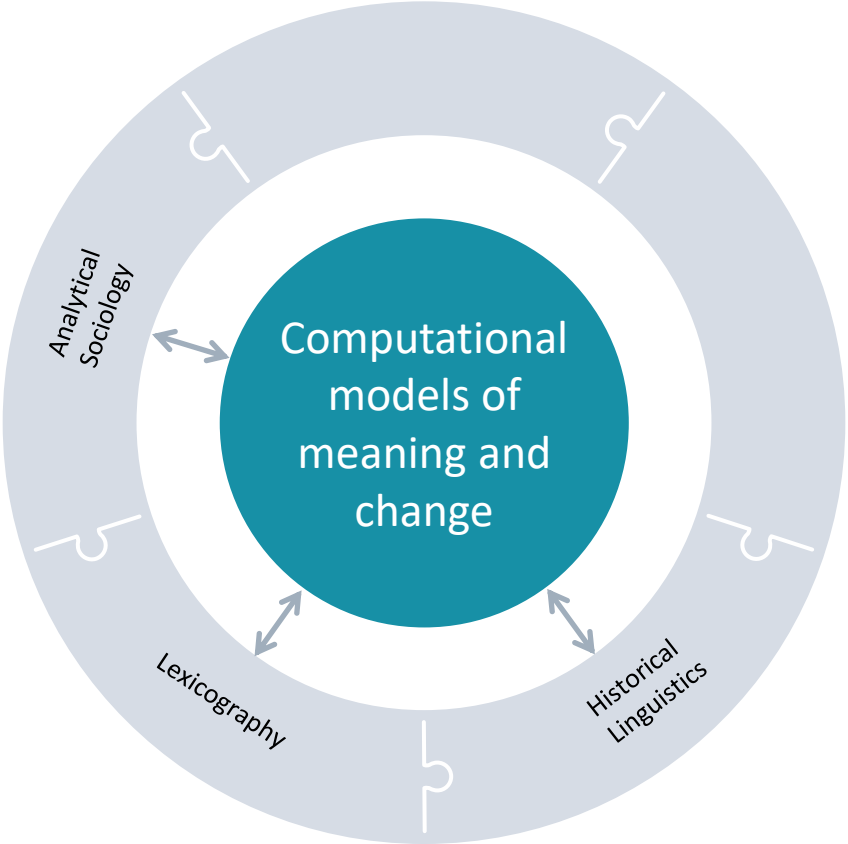
Lexicography



Our Research Questions

Societal level change

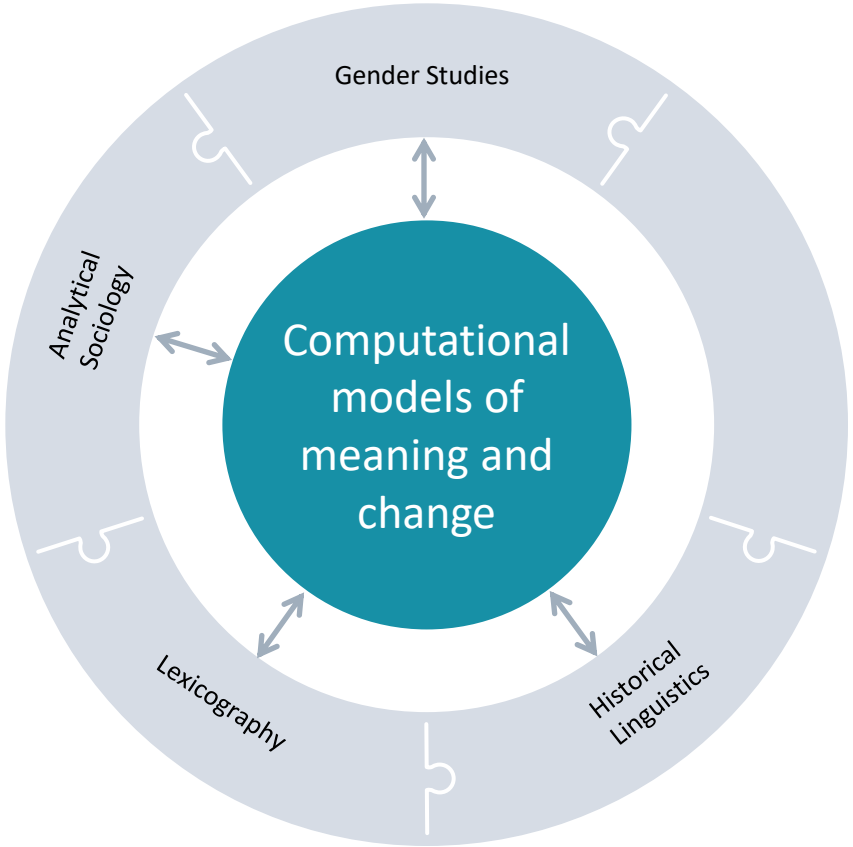
Analytical Sociology

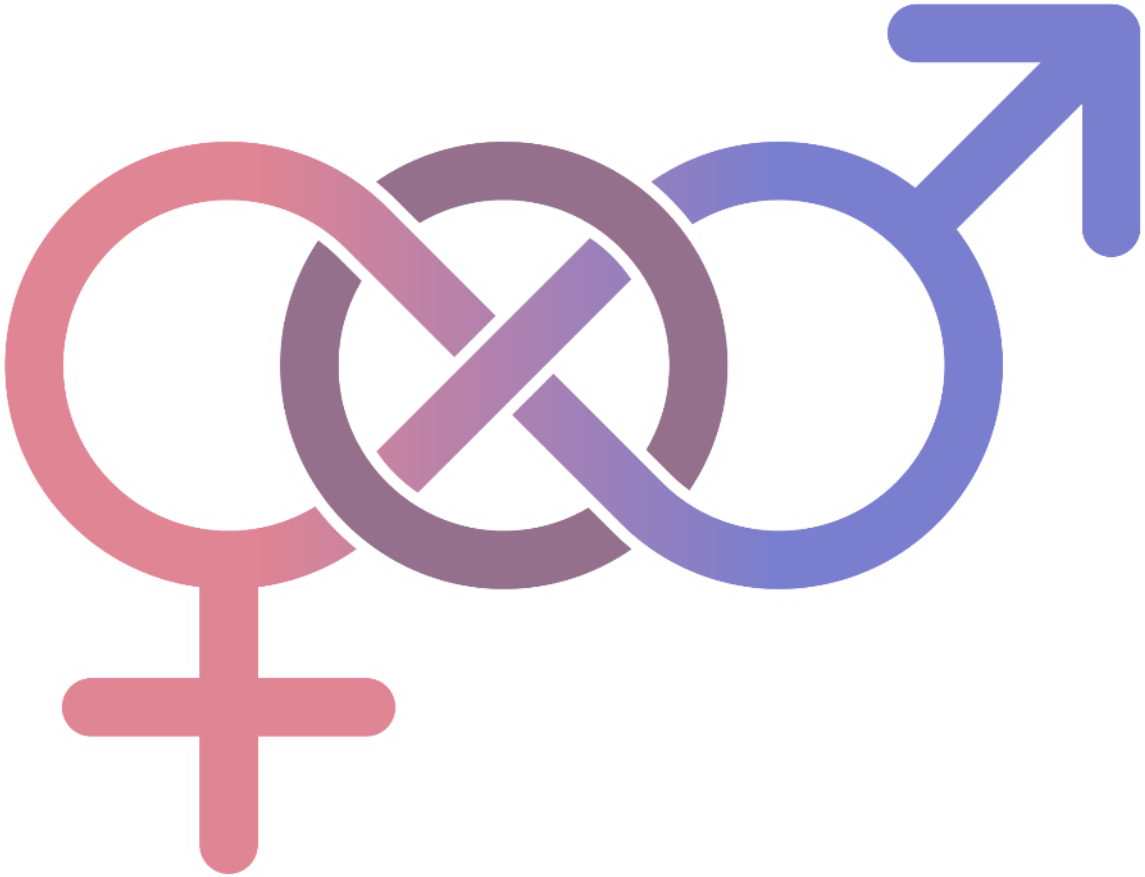


Our Research Questions

Societal level change

Gender Studies





The Market Language

Marknadens språk: **Studier i talet om marknader från medeltid till nutid**

Problem formulation: How did the market language change over time?




Funded by MAW (2022-2025)

https://www.gu.se/forskning/marknadens-sprak-studier-i-talet-om-marknader-fran-medeltid-till-nutid

The Market Language

The **productive** market

an ever-expanding core concept, a *diversification* took place and the market became immensely productive as a concept

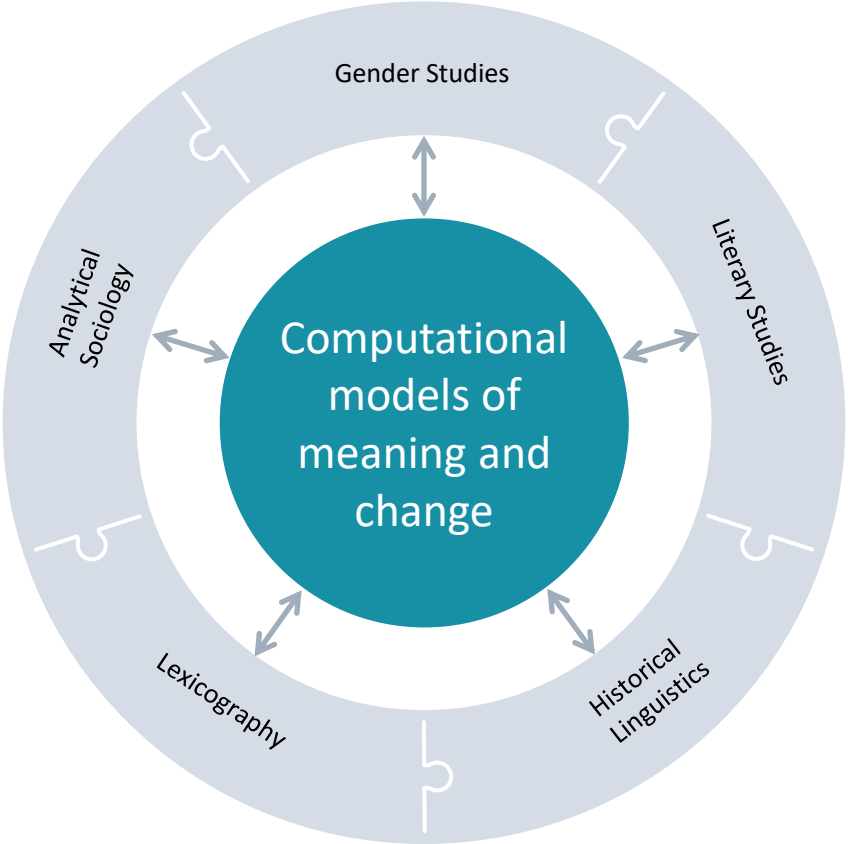
The **problematic** market

the market as a phenomenon was debated and an *area for conflicts* in an ever-changing society

Our Research Questions

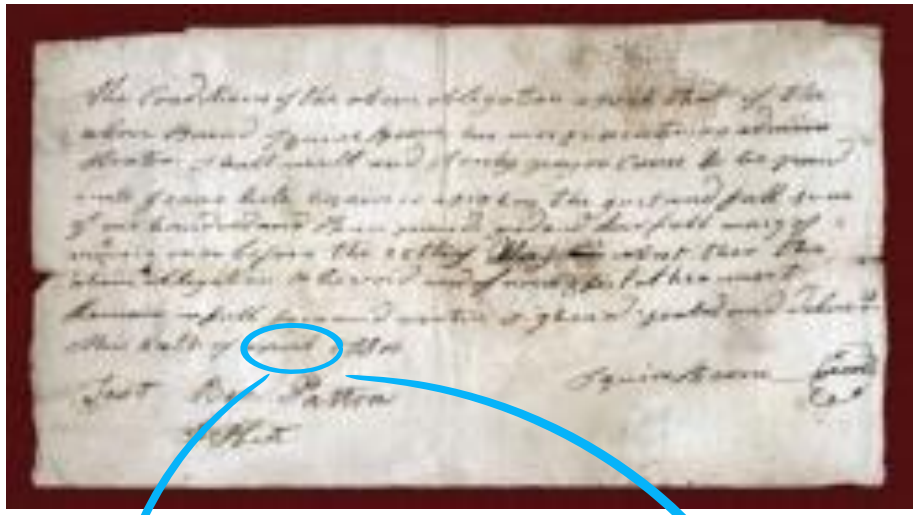
Societal level change

Literary Studies



Our societal contribution

Meaning for everyone



'gay adjective \gā\

Definition of GAY

1 a : happily excited : MERRY <in a gay mood>

b : keenly alive and exuberant : having or inducing high spirits <a bird's gay spring song>

'gay adjective \gā\

4 a : HOMOSEXUAL <gay men>

b : of, relating to, or used by homosexuals <the gay rights movement> <a gay bar>

Idag, 12:49

[redacted]

Medlem

Reg: Mar 2004

Inlägg: 1 790

Väldigt, väldigt vanligt att musslor öker holk. Är väldigt säker på bland unga idag. Jag brukar höra att alkoholkonsumtionen bland

Med det sagt har jag inget emot musslor. Angående alkoholen så 20+ som öppet dricker bira på stan, det tycker jag såklart är skoj

Bosnier har länge haft stark ölkultur trots islam.

clams

muslim

Our Two Research Aims

Computational linguistics

Humanities and social sciences

Understand and create computational methods for lexical semantic change and variation

Answer research questions in different text-based HSS

Generate methods, methodology and proper evaluation

Some facts

- **6** years
- **6** partner universities
- **4** Members from 4 countries
- **6** Countries, with advisors
- **17** People including a SE & PM
- **33.5** MSek from RJ + 10.5MSek from the University, Faculty, Dept

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung

Queen Mary University of London

KU LEUVEN

LINKÖPING UNIVERSITY

LUNDS UNIVERSITET

GÖTEBORGS UNIVERSITET


Belgium (KUL)

Steering committee

Germany (USTUTT)

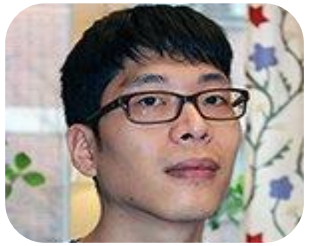
UK (QMUL)

Belgium

Local team (GU)





Lund university

IAS

GU

GU

Program management I

Principal investigator: Nina Tahmasebi

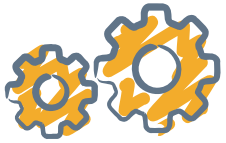
Program manager: Netta Huebscher

Steering group: Nina, Haim, Dominik, Simon

Advisors: Maria Koptjevskaja Tamm, Claire Bowern, Adam Jatowt, Dirk Geeraerts

Open source

Data



Methods

(code, pipelines, test data, tutorials on how to use the code)

Models

(Topic models, Swedish word embeddings)

Results

https://zenodo.org/record/3928474

2017-00626) and its 10 partner institutions, to NT. The Swedish list of potential change words were provided by the research group at the Department of Swedish, University of Gothenburg that work with the Contemporary Dictionary of the Swedish Academy. This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1, to BMcG. Additional thanks go to the annotators of our datasets, and an anonymous donor.

Preview

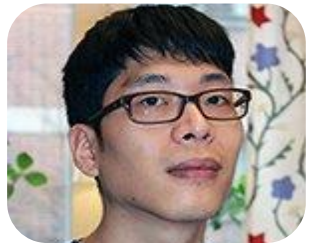
submissions_results.csv	64.2 kB
scoring_program	
evaluation.py	5.0 kB
metadata	120 Bytes
starting_kit	
README.html	3.3 kB
README.md	2.7 kB
code	
cd.py	2.2 kB
ci.py	1.9 kB
class.py	1.9 kB
cnt.py	2.5 kB
diff.py	2.4 kB
freq.py	2.1 kB
utils_.py	1.9 kB
requirements.txt	51 Bytes
run.sh	3.5 kB
test_data_public.zip	4.6 kB
test_data_truth	

Files (4.2 MB)

Name	Size	
semeval2020_ulscd_posteval.zip	4.2 MB	Preview Download

md5:1a97bf696c4c56e8ed0071c51be1e9fb











Belgium

UK

Germany





GU

Lund university

IAS

GU

GU