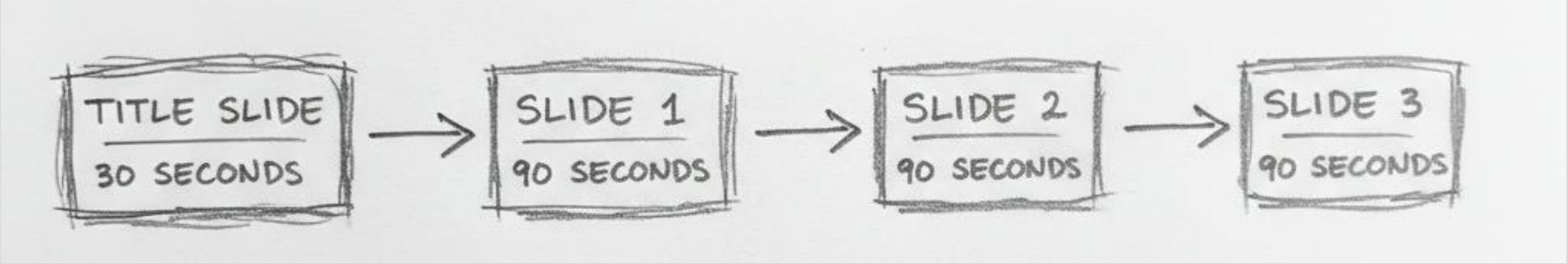
CHR 2025
Lightning Talks

Chair: Taylor Arnold

Welcome!

While all forms of papers had increased submissions this year, the lightning talks had by far the largest growth and ultimately lowest acceptance rate.

We have a fantastic set of 17 presentations. To make sure everyone has time to present, we will use the following format with automatically advancing slides:



A static PDF version of these slides is available on the CHR 2025 website.

Bringing together close reading questions and distant reading methods in the analysis of archived web

Victor Harbo Johnston, Helle Strandgaard Jensen and Sasch Berg Bogebjerg

WEB CHILD

European Research Council
Established by the European Commission

Contact Us

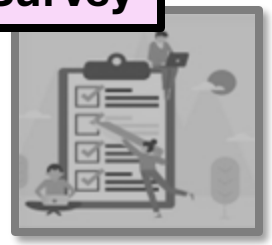
Sascha Berg **Bøgebjerg**
Helle Strandgaard **Jensen**
Victor Harbo **Johnston**
Aarhus University

Print

~50 TB Archived Web

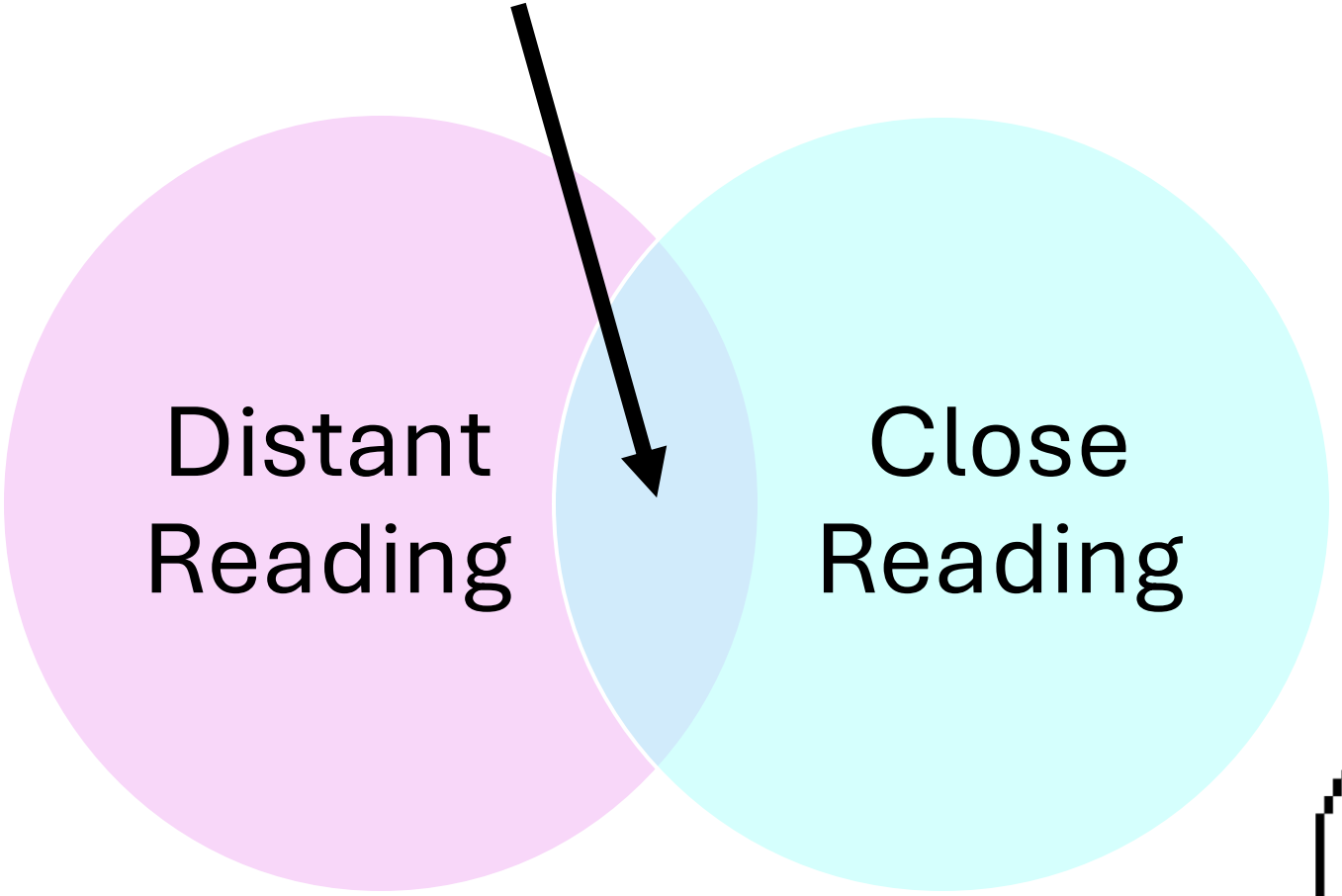
Oral History

Survey



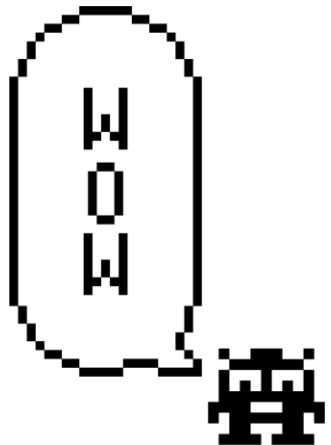
WEB CHILD

Scalable Reading

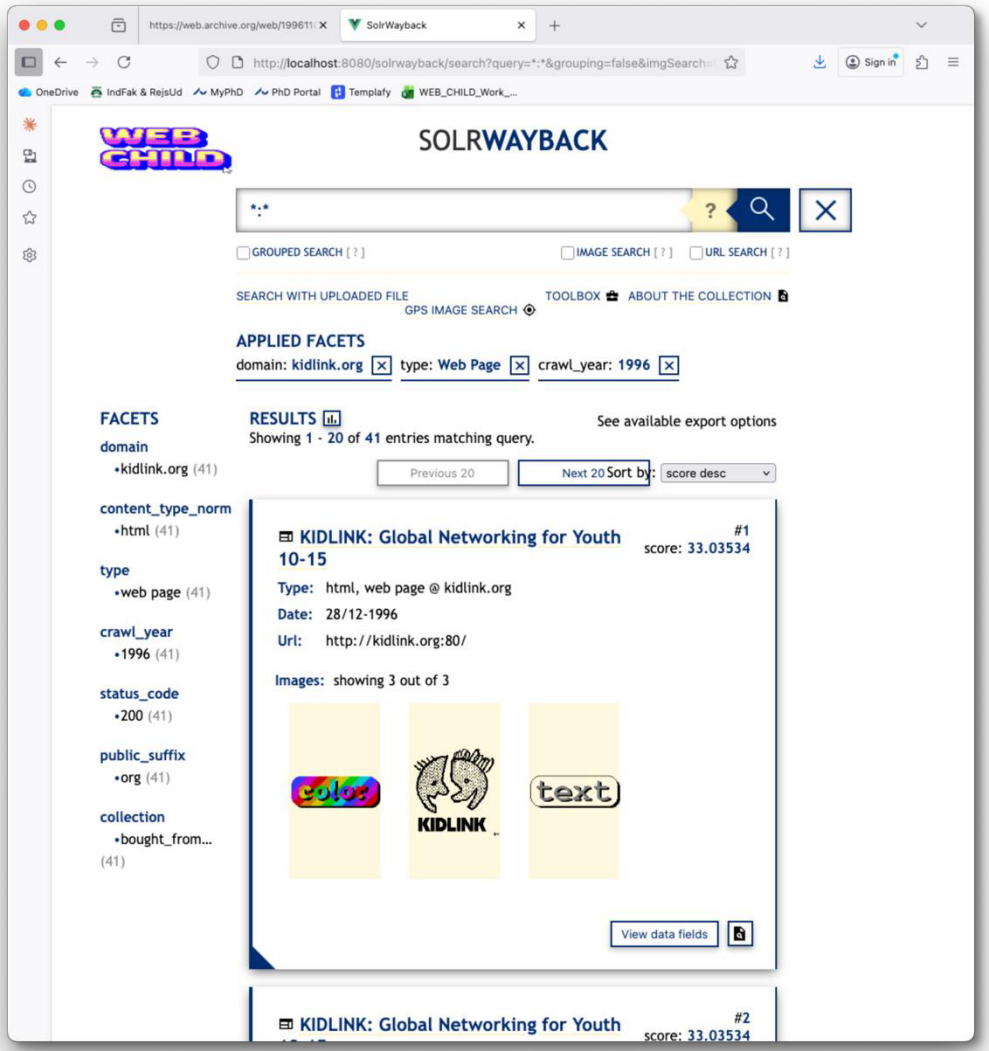


Contact Us

Sascha Berg **Bøgebjerg**
Helle Strandgaard **Jensen**
Victor Harbo **Johnston**
Aarhus University



Preliminary Solutions



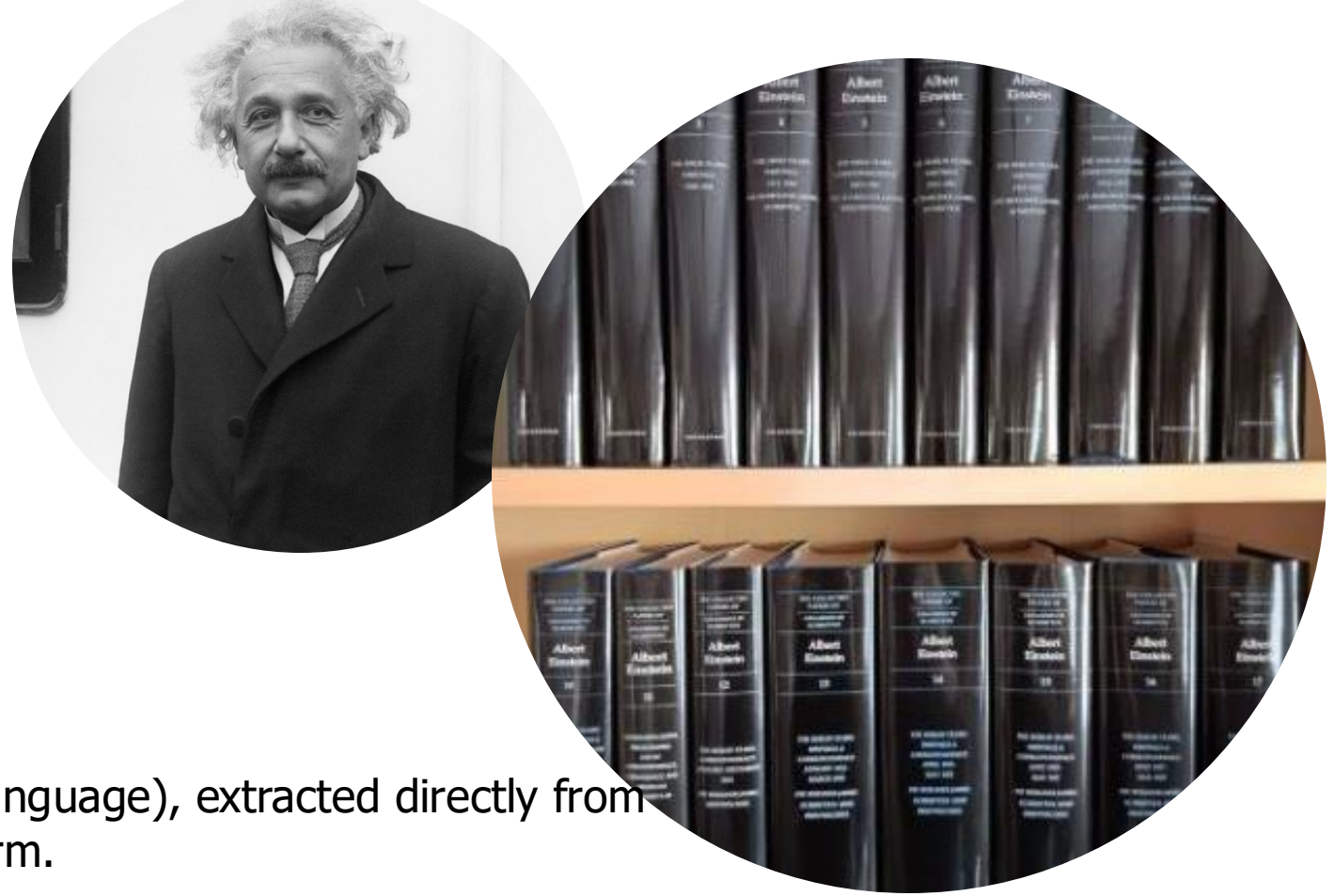
Contact Us

Sascha Berg **Bøgebjerg**
Helle Strandgaard **Jensen**
Victor Harbo **Johnston**
Aarhus University

Einstein AI: Contextual Retrieval from the Collected Papers of Albert Einstein Using RAG and GraphRAG Architectures

Florin-Stefan Morar

- *The Collected Papers of Albert Einstein* (CPAE), published by Princeton University Press. The definitive scholarly edition, established in 1977 and publishing since 1987.
- A massive ongoing project with over 30,000 unique documents total. Currently spans 17 volumes (up to 1930).
- Our dataset focuses on the "Early Berlin Years", Includes *Annus Mirabilis* (1905), the completion of General Relativity (1915), and Einstein's rise to fame (1919)



- Primary Text: The German Edition (original language), extracted directly from the open-access *Digital Einstein Papers* platform.

- Translations: English versions were integrated via custom OCR of the supplementary English volumes to ensure cross-lingual retrieval.

Research question comes from an effort to bridge history of science and LLMs

How did Einstein's personal life and social context lead to the development of his scientific ideas?

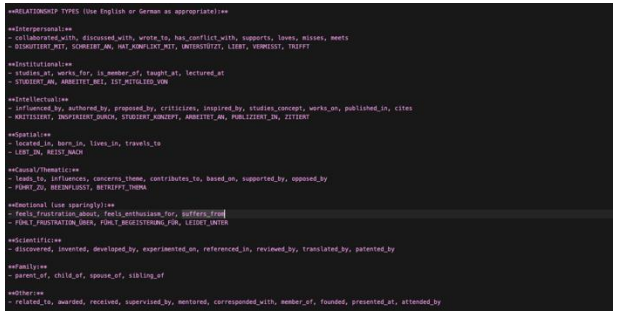
Approach

- A Custom Hybrid Pipeline for Historical Reasoning

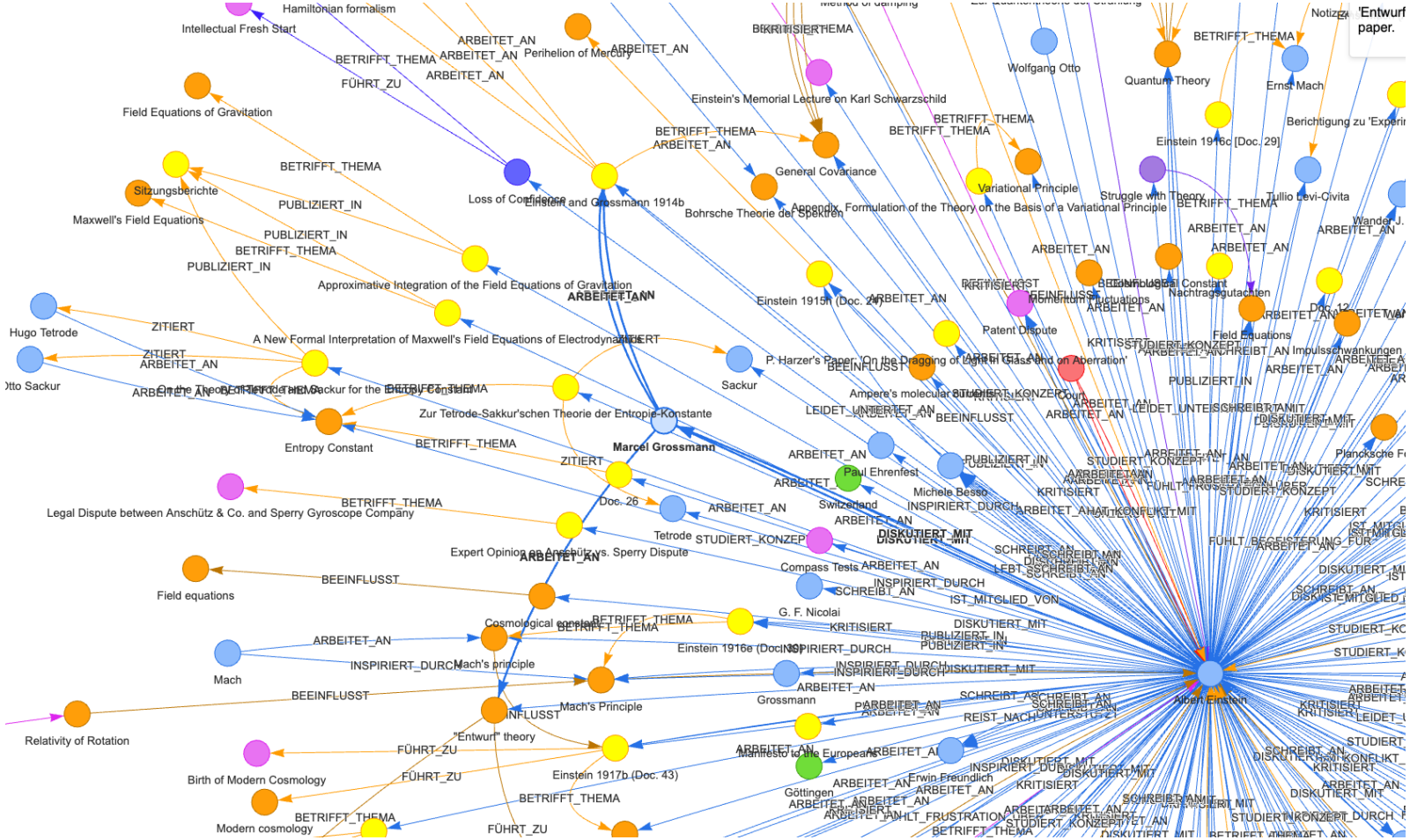
- **Foundation:** Adapted the Microsoft GraphRAG architecture (specifically "Local Search" patterns) but re-engineered for the Google ecosystem. Replaced default extractors with Google Gemini 2.5 Pro to handle the complex, multilingual schema Custom pipeline using NetworkX for graph management, FAISS for vector indexing, and SentenceTransformers (all-MiniLM-L6-v2) for embeddings.

- **Graph building schema**: A dual ontology, to simultaneously extract rigorous intellectual data (the "Science") and nuanced interpersonal dynamics. This allows us to map not just what Einstein discovered, but who he struggled with, what he felt about his work, and how his ideas evolved through social interaction. The Graph is built out of ENTITIES (person, organization, institution, location concept, method, theory) and RELATIONSHIPS (Interpersonal: colab with, wrote to etc.; Institutional: studied at etc.)
- Technical process: 1200 char chunking, Gemini 2.5 pro to extract entities using schema, NetworkX graph and stored as .pickle file for retrieaval

- **How the Query Works (The "Graph-Aware" Retrieval):**
  - Vector Seeding: The user's query is embedded and matched against the vector index to find the top-k "Seed Nodes" (specific relevant document chunks).
  - Graph Expansion: The system uses the graph structure to "walk" from these seed nodes to their neighbors (1-2 hops).
    - *Result:* This retrieves semantically linked content (e.g., a letter *about* a concept) that lacks shared keywords, solving the "lexical gap."
  - Community Clustering: Retrieved nodes are clustered (using DBSCAN) to group thematically related evidence before the LLM synthesizes the final answer.



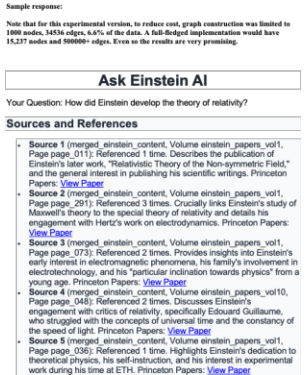
Some results



https://einstein-service-465121314425.us-central1.run.app/

^^^ simple RAG

Vs. GraphRAG [example]



Sample response:

Note that for this experimental version, to reduce cost, graph construction was limited to 1000 nodes, 34536 edges, 6.6% of the data. A full-fledged implementation would have 15,237 nodes and 500000+ edges. Even so the results are very promising.

Ask Einstein AI

Your Question: How did Einstein develop the theory of relativity?

Sources and References

- **Source 1** (merged_einstein_content, Volume einstein_papers_vol1, Page page_011): Referenced 1 time. Describes the publication of Einstein's later work, "Relativistic Theory of the Non-symmetric Field," and the general interest in publishing his scientific writings. Princeton Papers: View Paper
- **Source 2** (merged_einstein_content, Volume einstein_papers_vol1, Page page_291): Referenced 3 times. Crucially links Einstein's study of Maxwell's theory to the special theory of relativity and details his engagement with Hertz's work on electrodynamics. Princeton Papers: View Paper
- **Source 3** (merged_einstein_content, Volume einstein_papers_vol1, Page page_073): Referenced 2 times. Provides insights into Einstein's early interest in electromagnetic phenomena, his family's involvement in electrotechnology, and his "particular inclination towards physics" from a young age. Princeton Papers: View Paper
- **Source 4** (merged_einstein_content, Volume einstein_papers_vol10, Page page_048): Referenced 2 times. Discusses Einstein's engagement with critics of relativity, specifically Edouard Guillaume, who struggled with the concepts of universal time and the constancy of the speed of light. Princeton Papers: View Paper
- **Source 5** (merged_einstein_content, Volume einstein_papers_vol1, Page page_036): Referenced 1 time. Highlights Einstein's dedication to theoretical physics, his self-instruction, and his interest in experimental work during his time at ETH. Princeton Papers: View Paper

Modeling Intertextuality: An Ontological Framework for Literary Studies

Laura Untner

Intertextuality is messy, but we need structure.

Some ontologies already provide this structure:

… mainly for scholarly publishing (BIBO, CiTO, etc.)

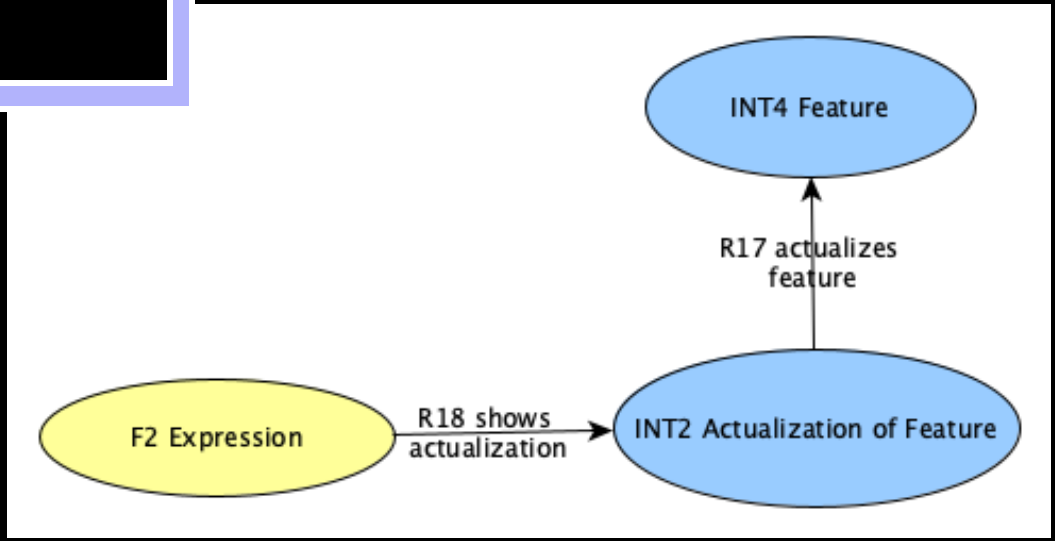
… or for specific corpora or projects (SAWS, Hypermedia Dante, OntoPoetry, GOLEM, MiMoText, etc.)

INTRO: The Intertextual, Interpictorial, and Intermedial Relations Ontology

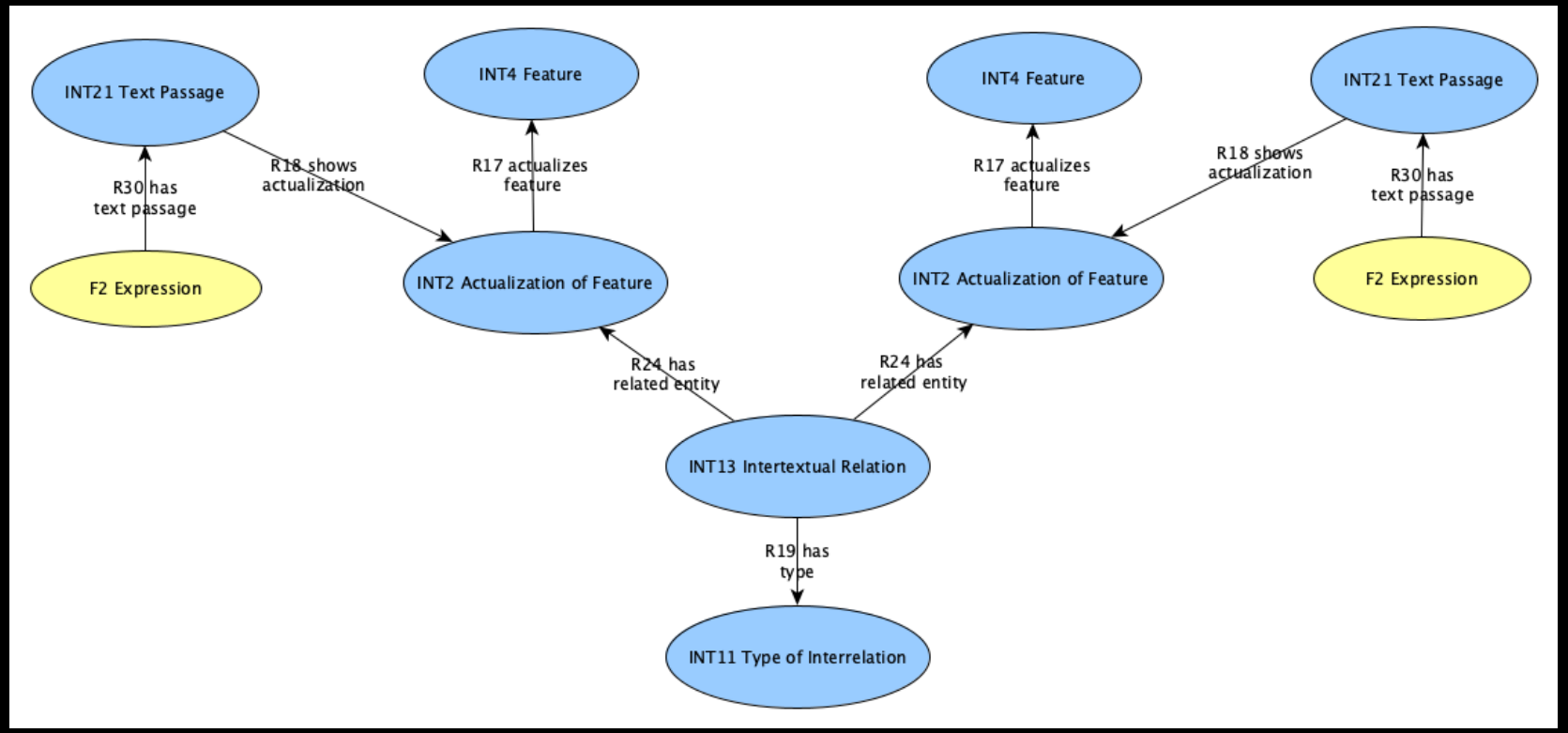
... was originally developed by Bernhard Oberreither.

... is built on CIDOC CRM and LRMoo.

... is text-centric.



>> https://boberreither.github.io/INTRO/ <<



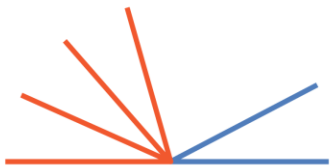
Current projects

Laura Untner (2025): From Wikidata to CIDOC CRM: A Use Scenario for Digital Comparative Literary Studies. In: Journal of Open Humanities Data [accepted].

Sappho Digital: https://sappho-digital.com

Automating the Study of Digital Literary Memory: A Multilingual LLM Pipeline for Wikipedia-Based Cultural Analysis

Botond Szemes
University of Tartu
DigiTS Research Group



DigiTS
Center for Digital
Text Scholarship

Funded by
the European Union

- Clean Wikidata query results
- Who is Polish, Hungraian etc.? Based on the language of publications
- What is literature? Before 1800: everything. After 1800: just fiction in the modern sense + drama, poetry.

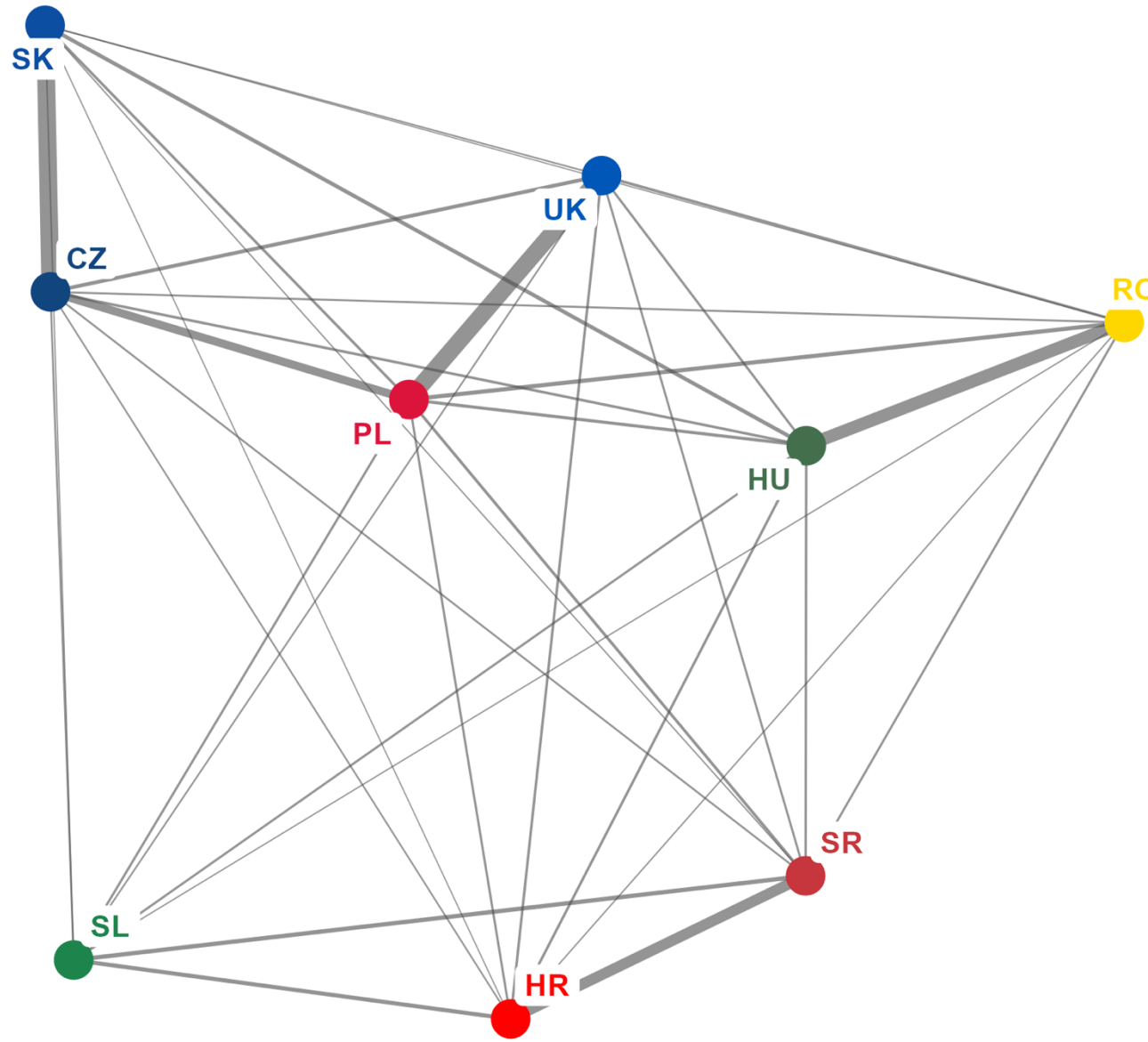
- Manual cleaned dataset for the Visegrad region (Czech, Hungarian, Polish, Slovak) as gold standard

- Test an LLM pipeline:
  - send the Wikipedia articles in a given language to a locar version of a large model with the same prompt via API
- **llama3.3:70B** overall **F1 = 0.89**
- In the extended dataset cross validation of results from different languages

LLM	Author	Wiki	F1 Score	Precision	Recall
llama3.1:8b	Cz	Hu	0,84	0,74	0,96
	Cz	Pl	0,80	0,72	0,91
	Cz	Sk	0,78	0,70	0,89
	Hu	Cz	0,81	0,73	0,92
llama3.3:70b	Cz	Hu	0,92	0,91	0,93
	Cz	Pl	0,87	0,87	0,87
	Cz	Sk	0,84	0,82	0,85
	Hu	Cz	0,90	0,92	0,89
gemma3:27b	Cz	Hu	0,78	0,71	0,86
	Cz	Pl	0,85	0,81	0,90
	Cz	Sk	0,76	0,72	0,79
	Hu	Cz	0,82	0,91	0,74
gpt-oss:120b	Cz	Hu	0,87	0,88	0,86
	Cz	Pl	0,89	0,91	0,88
	Cz	Sk	0,85	0,86	0,84
	Hu	Cz	0,92	0,98	0,87

Network of Literary Memory

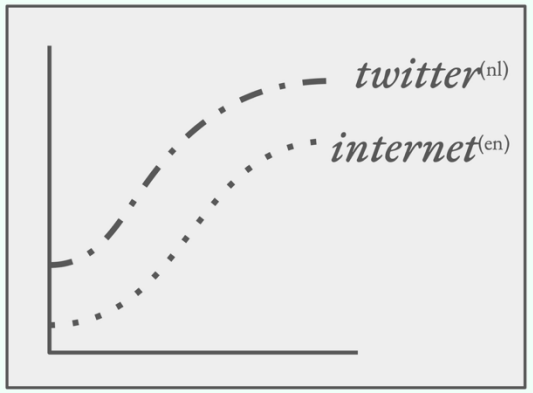
Weighted by summed memory score



Measuring the Synchronicity of Historical European Parliamentary Discourse, 1949-2018

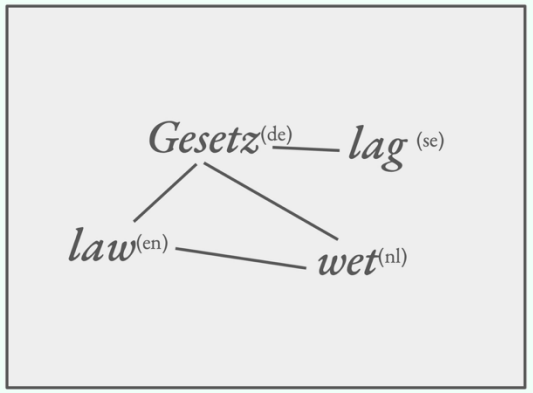
Ruben Ros and Risto Turunen

How did *British*, *Dutch*, *Swedish*, and *German* parliamentary discourse develop between 1945 and 2018?



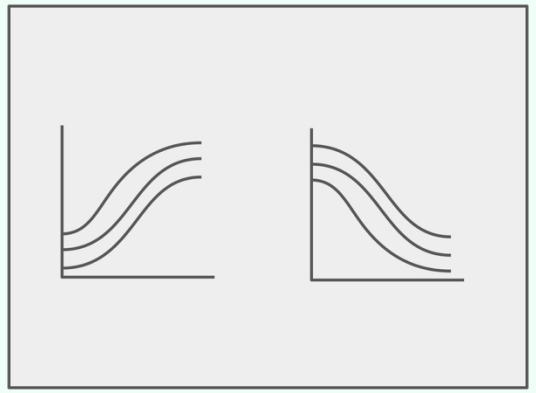
Find bilingual **word pairs** with high correlation in frequency time series.

Spearman correlation coefficient + Dynamic Time Warping



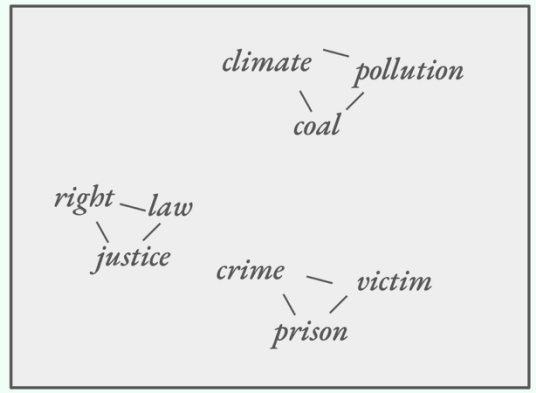
Filter **concepts**: multilingual networks of words with high internal frequency correlation.

Contextual translation + word alignment = distributions of translations



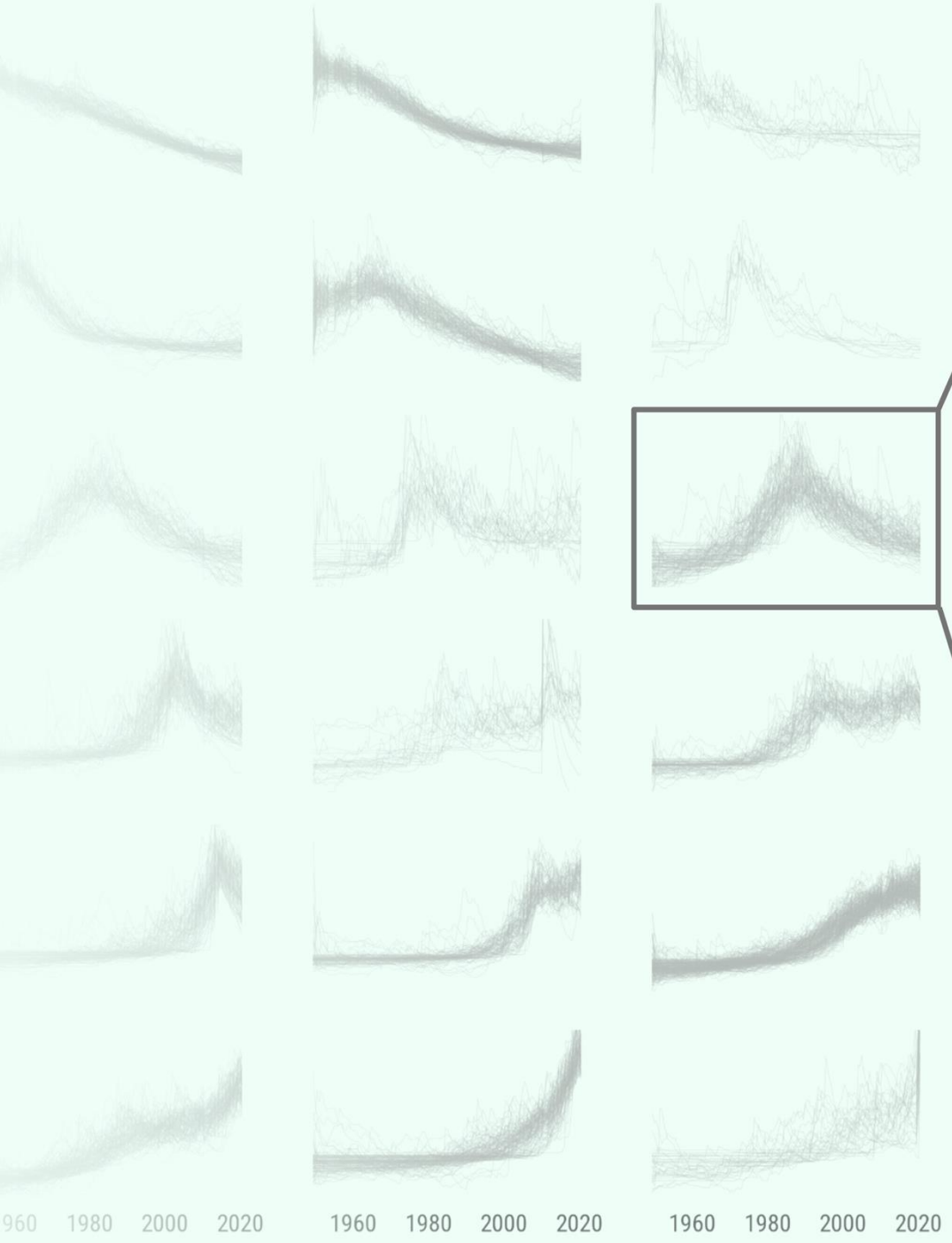
Cluster averaged concept **frequency time-series**

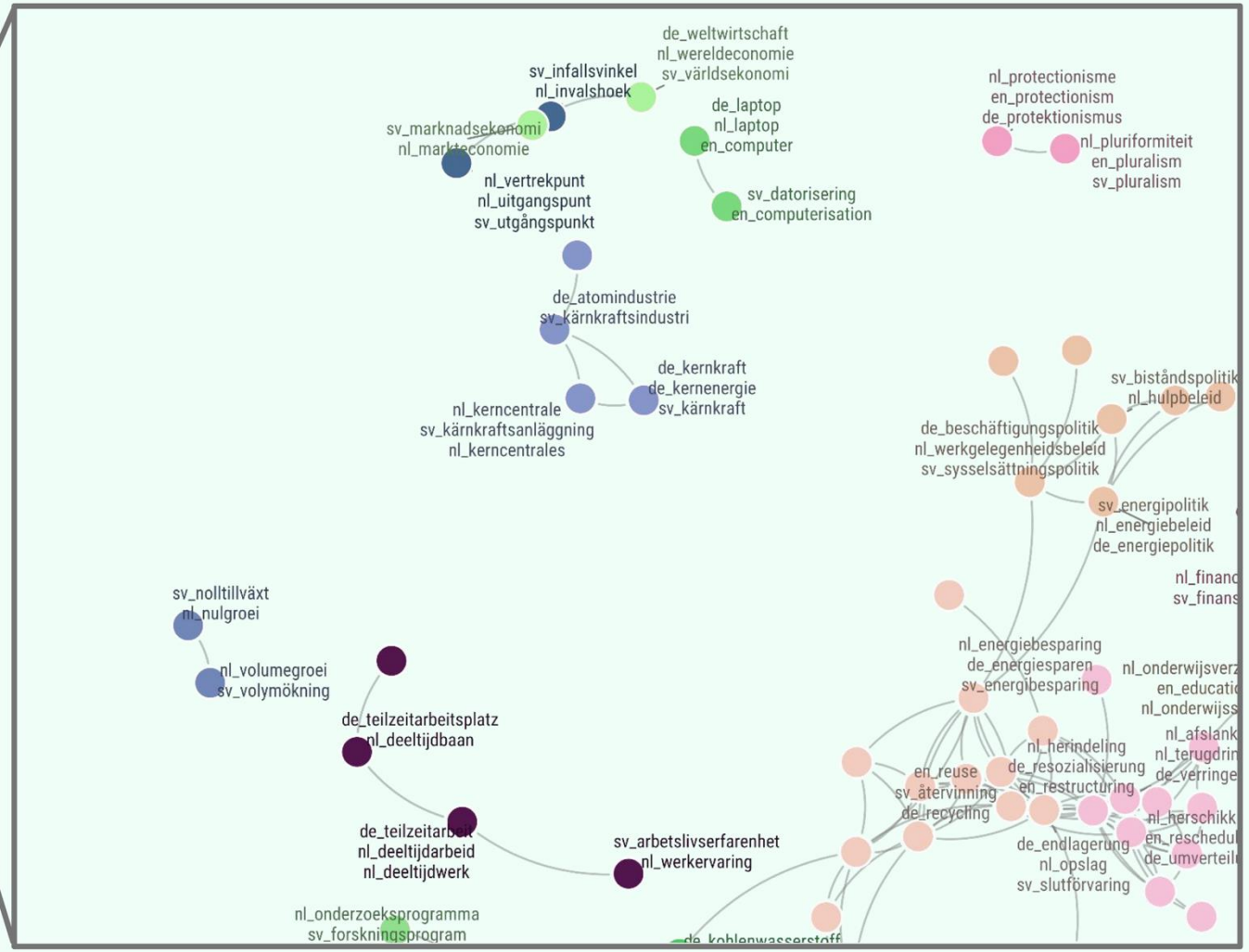
K-means clustering with normalized time series.



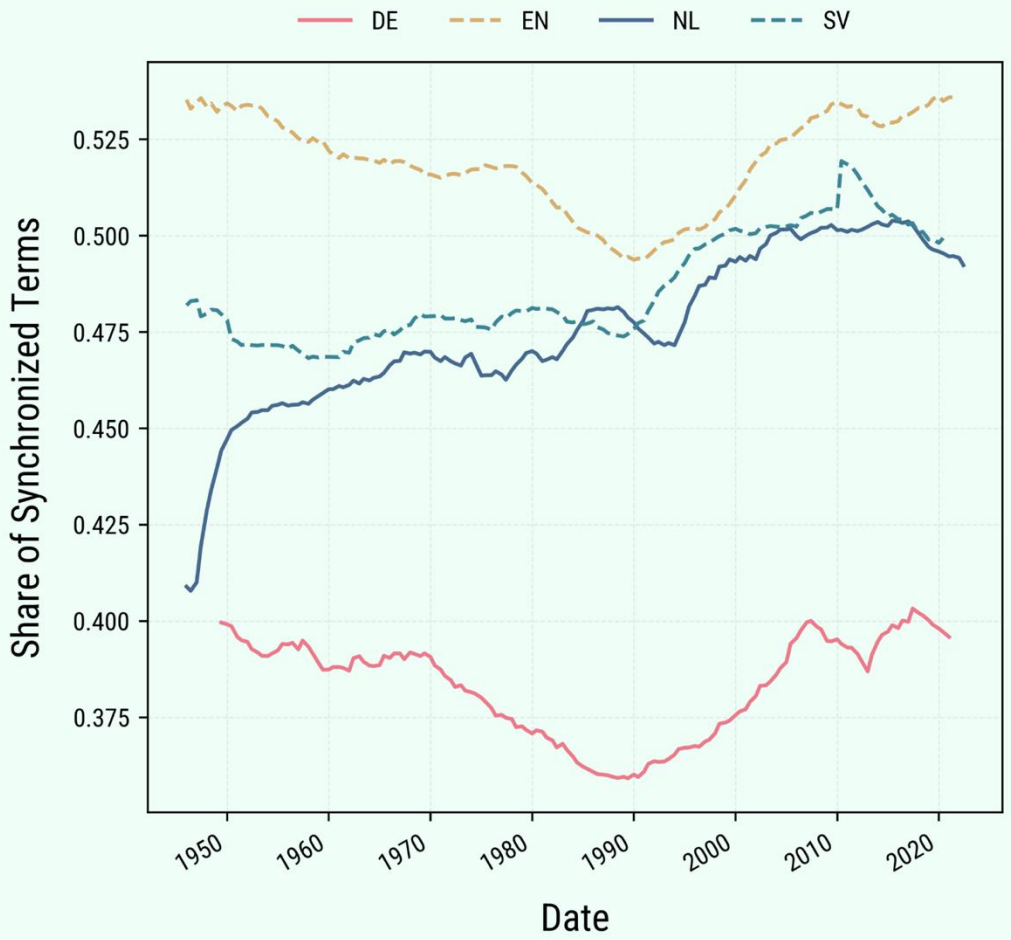
Find **semantic clusters** within time-series clusters

Type embeddings from adapted multilingual bert-base

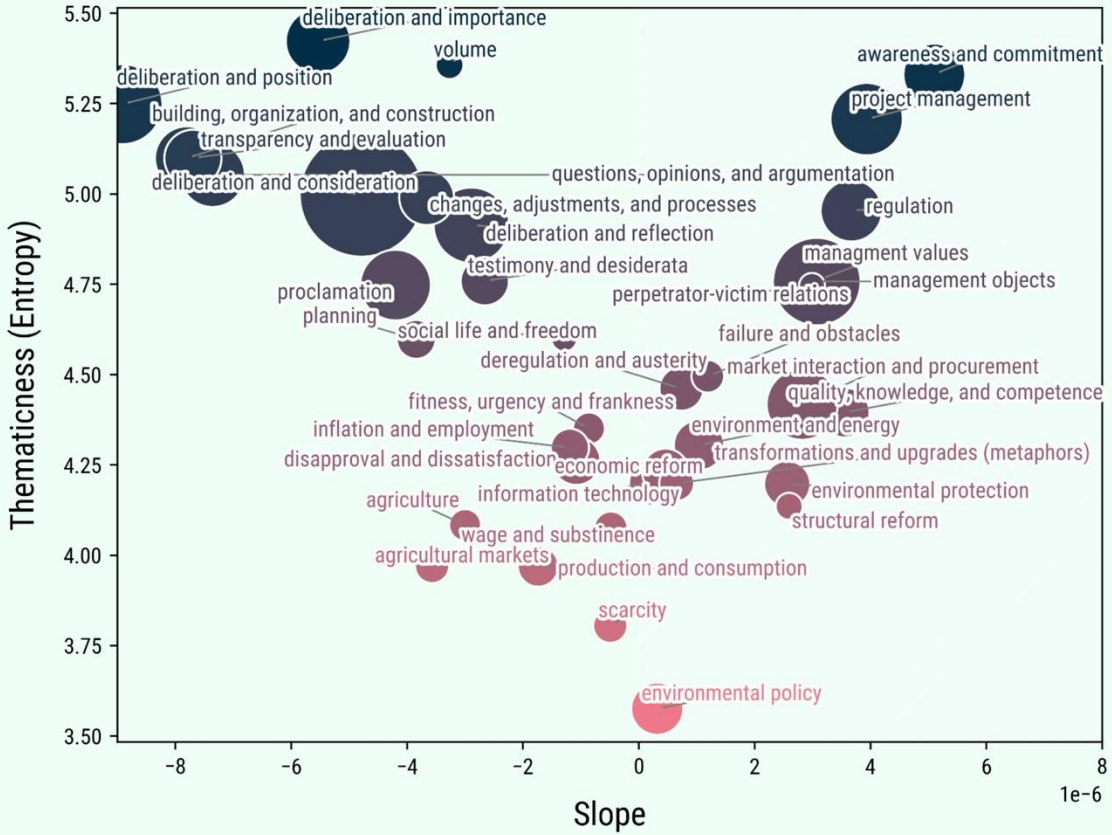




Identifying semantic clusters within temporal clusters.



The **share of synchronized terms** increases in all countries, especially from the 1990s onwards.



The slope and entropy of clusters reveals shared trends in both **procedural style** and **thematic content**

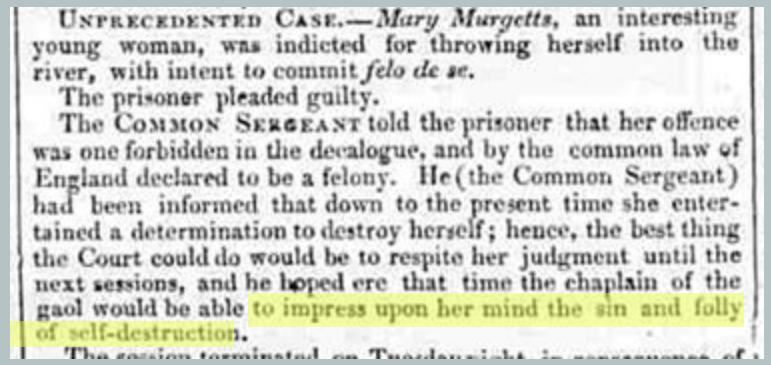
Disorder or (self-)murder? Making sense of suicide in 19th-century British newspapers

Nilo Pedrazzini and Daniel C. S. Wilson

Disorder or (self-)murder?

Making sense of suicide in 19th-century British newspapers

Nilo Pedrazzini (The Alan Turing Institute, UK)
Daniel CS Wilson (University College London, UK)



UNPRECEDENTED CASE.—*Mary Margetts*, an interesting young woman, was indicted for throwing herself into the river, with intent to commit *felo de se*.

The prisoner pleaded guilty.

The COMMON SERGEANT told the prisoner that her offence was one forbidden in the decalogue, and by the common law of England declared to be a felony. He (the Common Sergeant) had been informed that down to the present time she entertained a determination to destroy herself; hence, the best thing the Court could do would be to respite her judgment until the next sessions, and he hoped ere that time the chaplain of the gaol would be able to impress upon her mind the sin and folly of self-destruction.

23 YEARS OF SUFFERING

CLAYTON WOMAN HANGS HERSELF.

A MERCIFUL VIEW.

After suffering from rheumatism for the long period of 25 years, Margaret Heys (55), a single woman, who resided with her niece at 19, Blackburn-road, Clayton-le-Moors, put an end to her sufferings on Monday afternoon by hanging herself. Owing to the rheumatism deceased was infirm, and she attended to the house whilst her niece went to the mill. She was all right at noon on Monday, but when a next door neighbour tried the door of the house during the afternoon she found it locked. About five o'clock she suspected that something untoward had happened, and she called deceased's nephew, who, on entering the house, found his aunt hanging by a rope in the front room. The body was immedi-

1843 — sin and folly

1907 — suffering

Links to newspaper collections:

Living with Machines (LwM)

Heritage Made Digital (HMD14)

Disorder or (self-)murder?

Making sense of suicide in 19th-century British newspapers

Nilo Pedrazzini (The Alan Turing Institute, UK)
Daniel CS Wilson (University College London, UK)

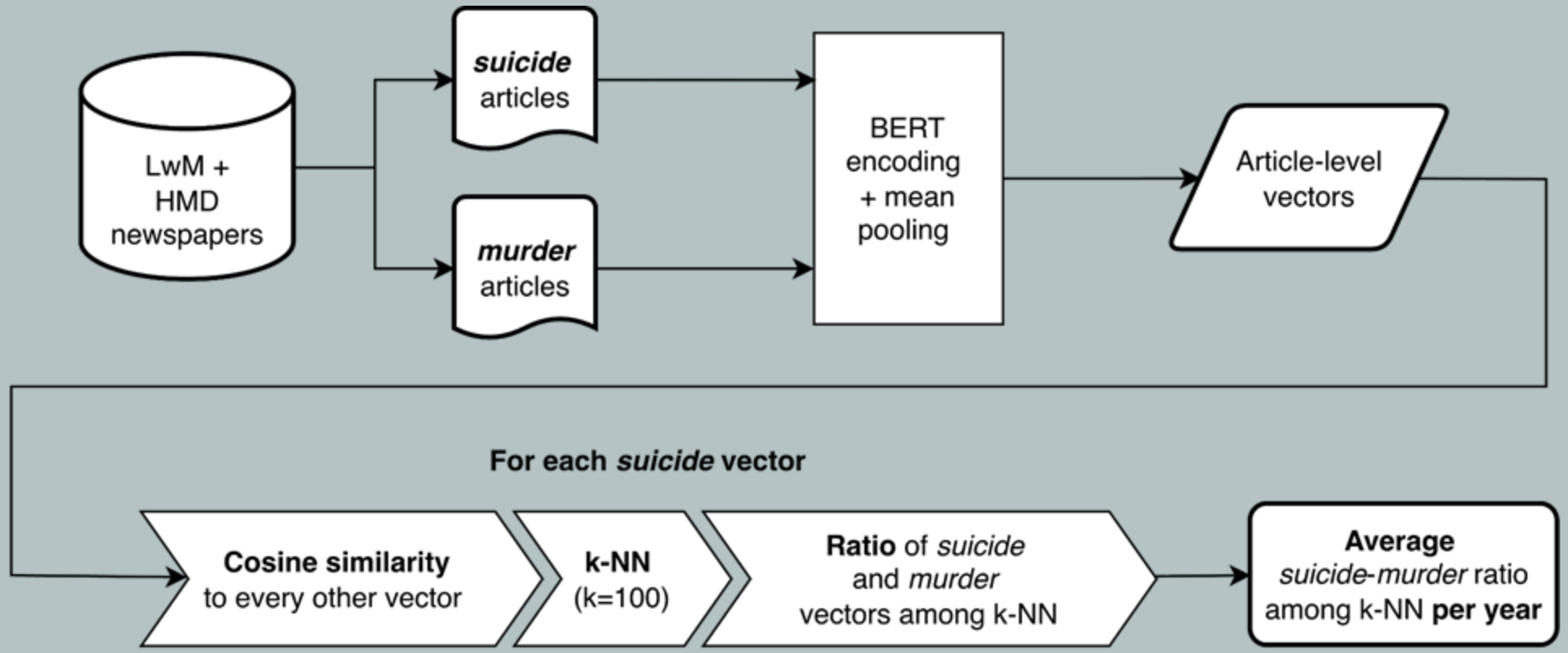


Figure 2. Flowchart of the method. Newspaper articles on *suicide* and *murder* are encoded with BERT into article-level vectors. For each suicide vector we compute cosine k-NN over all articles, derive the suicide-murder neighbour ratio, and average this ratio by year.

Disorder or (self-)murder?

Making sense of suicide in 19th-century British newspapers

Nilo Pedrazzini (The Alan Turing Institute, UK)
Daniel CS Wilson (University College London, UK)

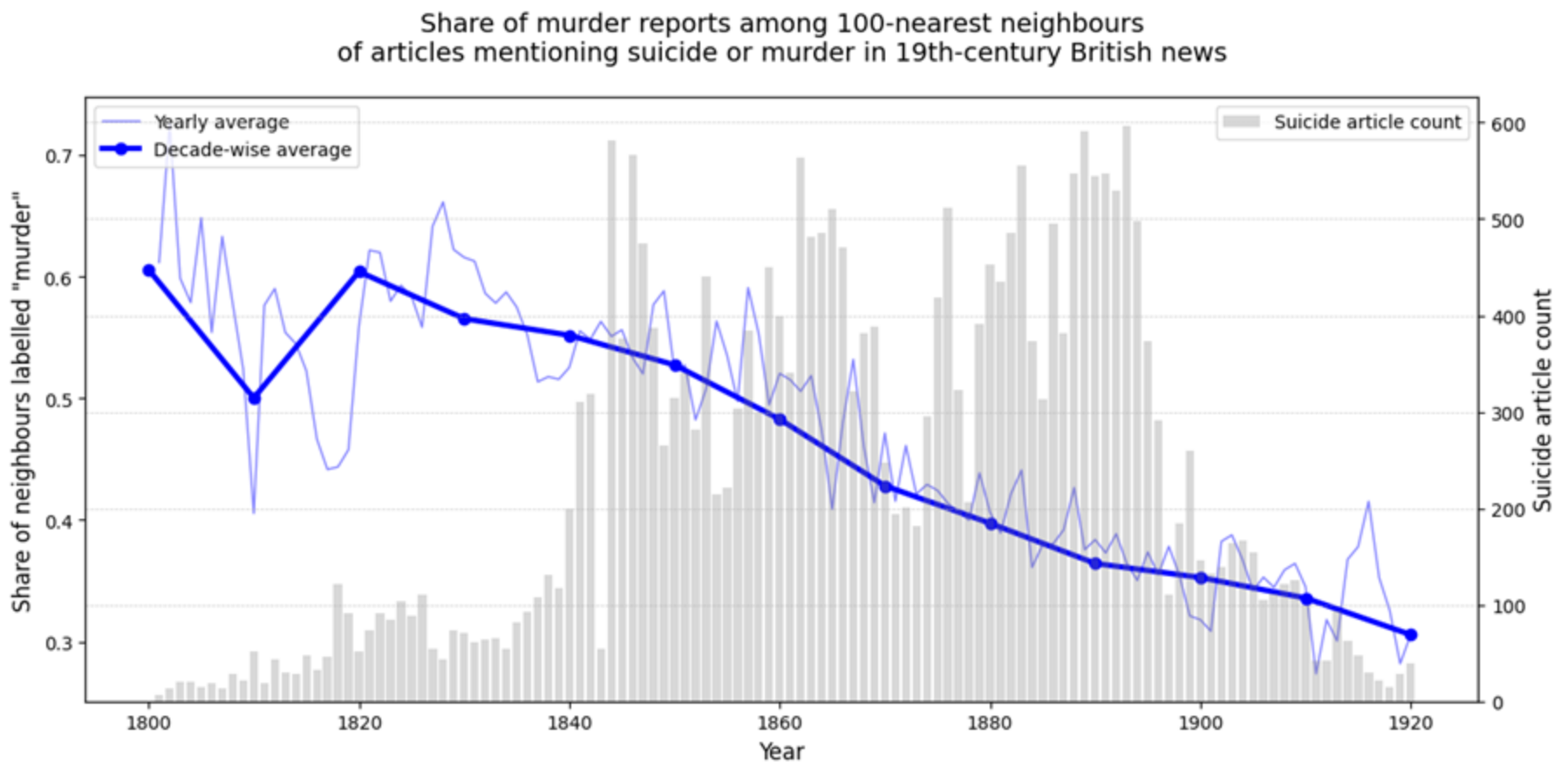
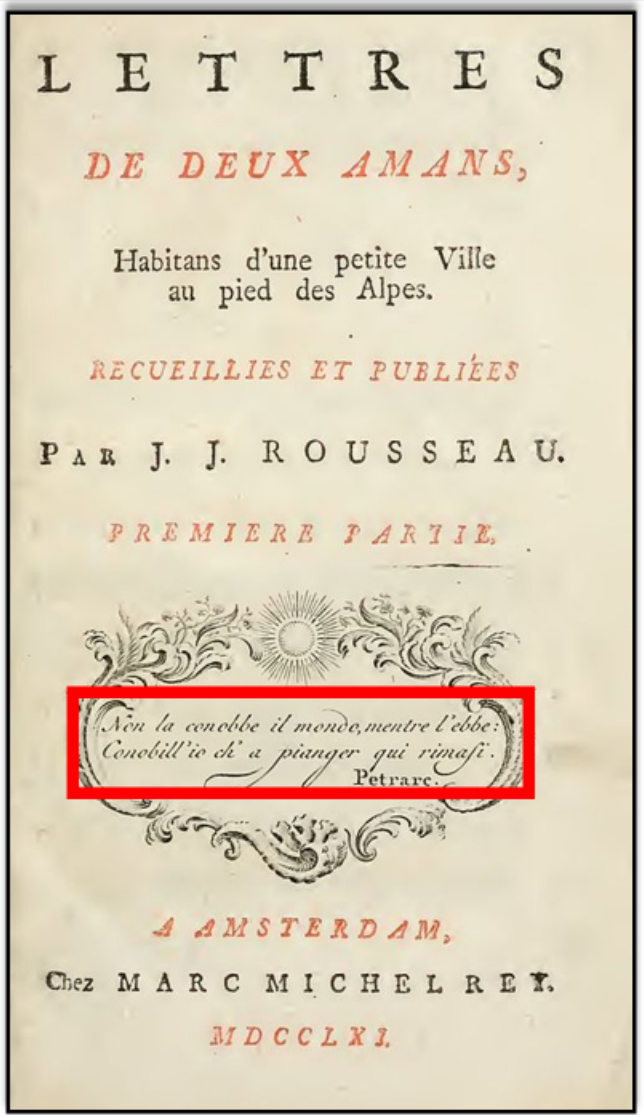


Figure 3. Average ratio of neighbours from *murder* articles among the 100 nearest neighbours of suicide articles (yearly and decade-wise averages, 1800–1920). Grey bars show the yearly count of suicide articles.

Mapping Literary Networks through Epigraphs

Tomás Espino Barrera

EPIMAPS


LETTRES

DE DEUX AMANS,

Habitans d'une petite Ville
au pied des Alpes.

RECUEILLIES ET PUBLIÉES

PAR J. J. ROUSSEAU.

PREMIERE PARTIE.

Nèn la conobbe il mondo, mentre l'ebbe:
Conobill'io ch' a pianger qui rimasi.
Petrarc.

A AMSTERDAM,

Chez MARC MICHEL REY.

MDCCLXI.

EPIMAPS - Mapping Epigraphical Networks

1. ELTeC corpus, aprox. 2000 novels
2. Epigraph extraction and enrichment (incl. sources)
   – 1st open-access multilingual epigraph database
3. Interactive mapping (Nodegoat)
4. Distant Reading
5. Case studies

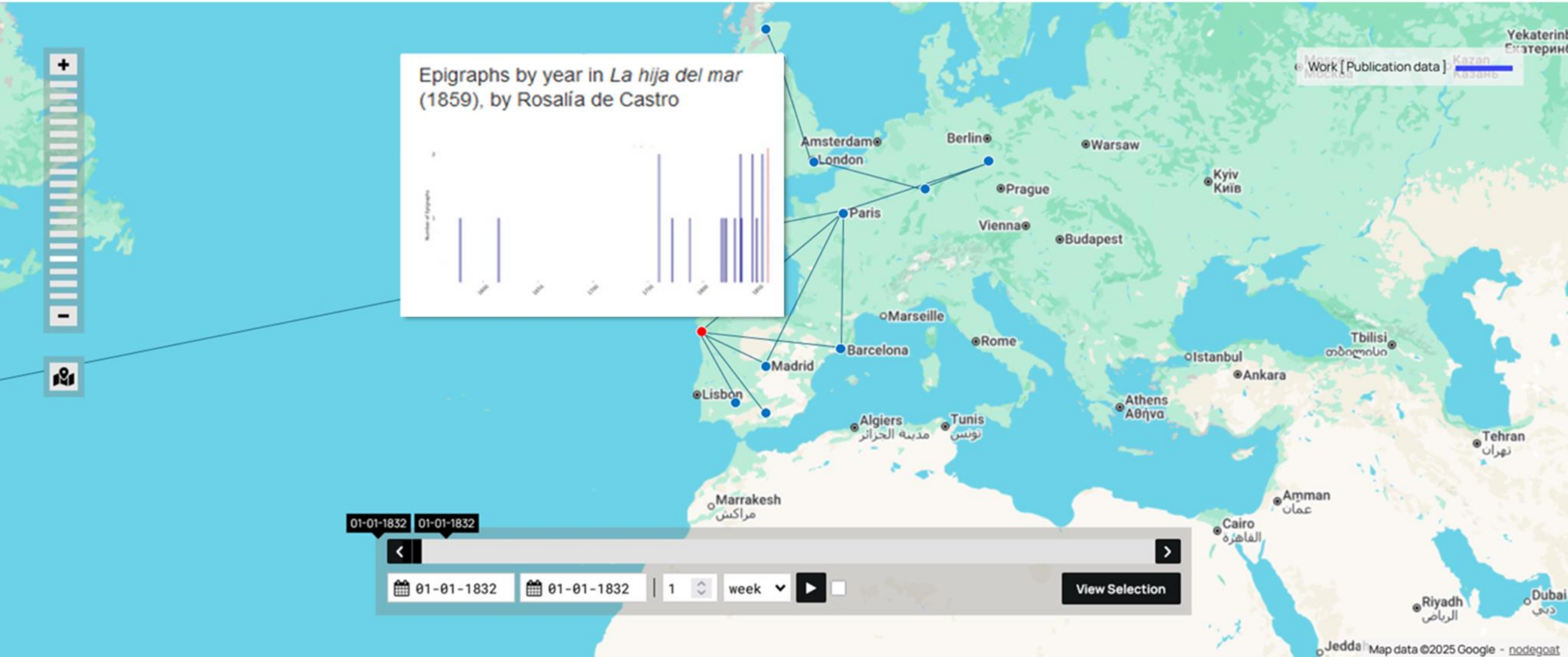
EPIMAPS

subgenre	narrative-p	size-categ	reprint-co	time-slot	Epigraphs	Feuilleton	title	chapters	epigraph	e-author-r	e-author-r	e-author-id	e-title	e-title-id	e-publicati	e-original	e-translati	e-publicat
NA	NA	long	unspecifie	1750	Y	N	Y	N	"O quantu	Persius	"Pers."	wikidata:Q332785	*Saturae* 1.	wikidata:Q87143921	Rome?			1st centur
NA	NA	short	unspecifie	1750	Y	N	Y	N	"Haec legit	Voltaire	"Ovid."	wikidata:Q9068	*Mérope*	wikidata:Q3333324	Paris	"Hoc legite	austeri, cri	1744
NA	NA	short	unspecifie	1750	Y	N	Y	N	"Non sem	Phaedrus	"Phaedr."	wikidata:Q52166	*Fabulae Ae*	wikidata:	Rome?			1st centur
NA	NA	long	unspecifie	1760	Y	N	Y	N	"successer	Tacitus	"Tacit."	wikidata:Q2161	*Histories* 2	wikidata:Q1247073	Rome?			2nd centur
NA	NA	short	unspecifie	1760	Y	N	Y	N	"Nulla Viro	Catullus	"Cat."	wikidata:Q163079	LXIV "Argo	wikidata:				1st centur
epistolary	NA	long	unspecifie	1760	Y	N	Y	N	"Non la co	Petrarch	"Petrarc"	wikidata:Q1401	*Canzonere*	wikidata:Q777574	Avignon / Italy			14th C.
epistolary	NA	long	unspecifie	1761	Y	N	Y	Y (2.12)	"O qual fia	Metastasic	N/A	wikidata:Q29473	*Attilio Reg*	wikidata:	Several locations (opera)			1750
NA	NA	short	unspecifie	1760	Y	N	Y	N	"Des jeune	Gresset, Je	"Gresset"	wikidata:Q942163	*Le Méchan*	wikidata:Q3224937	Paris	selected fragments of		1747
didactic no	NA	long	unspecifie	1760	Y	N	Y	N	"Sanabilibu	Seneca	"Sen"	wikidata:Q2054	"de ira. L. I	wikidata:Q3704115	Rome?			1st centur
NA	NA	short	unspecifie	1760	Y	N	Y	N	"J'admire l	N/A	N/A	N/A	N/A	wikidata:	N/A			N/A
libertine no	NA	long	unspecifie	1760	Y	N	Y	N	"Priape, so	Piron, Alex	"Piron"	wikidata:Q983437	*Ode à Priaj*	wikidata:	N/A			ca. 1710
NA	NA	medium	unspecifie	1760	Y	N	Y	N	"..... Ut nec	Horace	"Hor."	wikidata:Q6197	*Ars poetica*	wikidata:Q677997	Rome?			1st centur
NA	NA	long	unspecifie	1760	Y	N	Y	N	"Tout ce q	Self-epigra	N/A	N/A	"Tome II. F	N/A	London?	"tout ce qui	est hors d	1766
NA	NA	short	unspecifie	1760	Y	N	Y	N	"Ecce spec	Seneca	Senec.	wikidata:Q2054	*De provide*	wikidata:Q2264770	Rome?			1st centur
NA	NA	short	unspecifie	1760	Y	N	Y	N	"Non miro	Seneca	Senec.	wikidata:Q2054	*De provide*	wikidata:Q2264770	Rome?	"Ego uero	non miror, s	1st centur
NA	NA	short	unspecifie	1760	Y	N	Y	N	"La Vertu c	N/A	N/A	N/A	N/A	N/A	N/A			N/A
NA	NA	short	unspecifie	1760	Y	N	Y	N	"Virtue car	Richardsor	"Row."	wikidata:Q295941	*Clarissa*	wikidata:Q980534	London			1748
NA	NA	short	unspecifie	1760	Y	N	Y	N	"Si vuole à	Guarini, Gi	"Guarini"	wikidata:Q542039	*Il pastor fic*	wikidata:Q477990	Venice	"Si vuole à	punto / Far	1590
NA	NA	short	unspecifie	1760	Y	N	Y	N	"On peut t	Boileau, Ni	"Boil."	wikidata:Q188857	Satire X [Sa	wikidata:	Paris			1694
NA	NA	medium	unspecifie	1770	Y	N	Y	N	"Le Temps	Leibniz, Go	"Leibnitz"	wikidata:Q9047	Different p	N/A	Different p	"Le présent	est gros d	ca. 1710-1
NA	NA	long	unspecifie	1770	Y	N	Y	N	"One Almig	Milton, Jol	"Milton"	wikidata:Q79759	*Paradise L*	wikidata:Q28754	London		"Il est un s	1667
NA	NA	short	unspecifie	1770	Y	N	Y	N	"Nunc scio	Virgil	"Virg."	wikidata:Q1398	*Ecglogues*	wikidata:Q546203	Rome?			1st centur
libertine no	NA	medium	unspecifie	1770	Y	N	Y	N	"La faute e	N/A	N/A	N/A	N/A	N/A	N/A			N/A
epistolary	NA	long	unspecifie	1770	Y	N	N	Y (2.34)	"O serpent	N/A	N/A	N/A	N/A	N/A	N/A			N/A

EPIMAPS

You know my method. It is founded upon the observation of trifles.

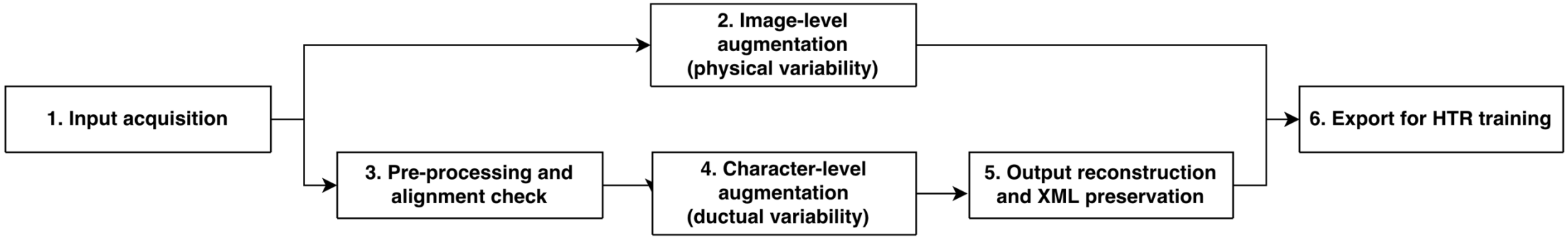
—Arthur Conan Doyle



Low-Cost Synthetic Data Generation for HTR Training: Evaluating a Multimodal Strategy for Historical Manuscript Processing

Serena Carlamaria Crespi and Carlos-Emiliano González-Gallardo

A Modular Approach for Data Augmentation

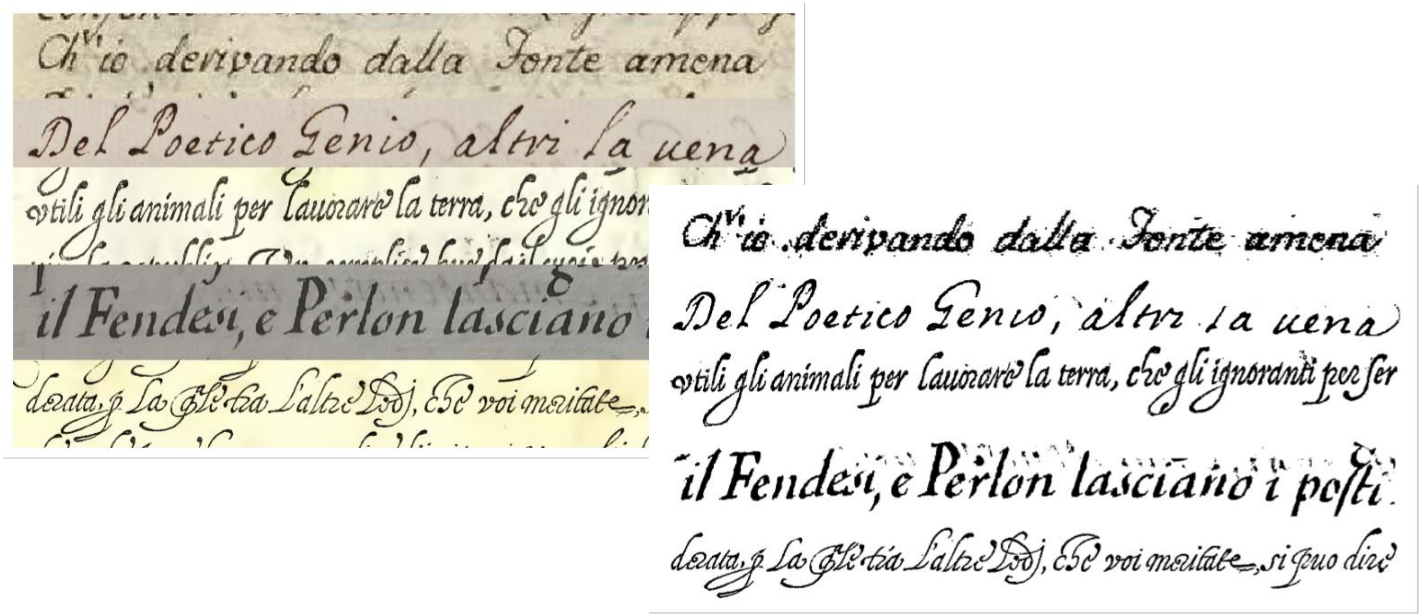


Training corpus

- 446 manuscript pages (images & ALTO transcriptions)
- Mostly poetic Italian codices (17th and 18th centuries)

SCAN ME

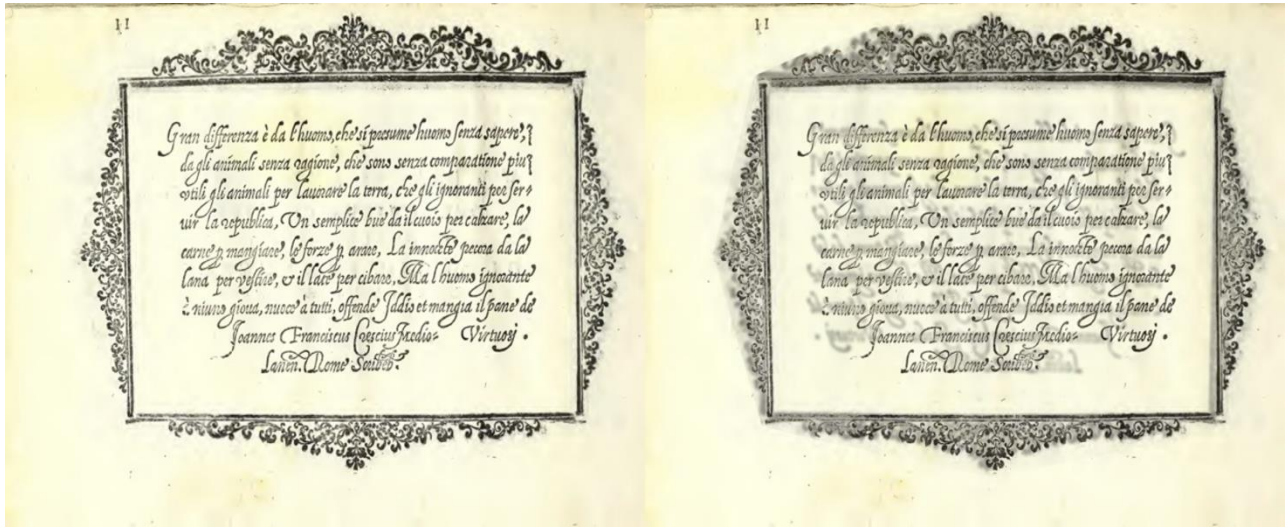
GitLab



Corpus & Augmentations

I. Image-level augmented corpus

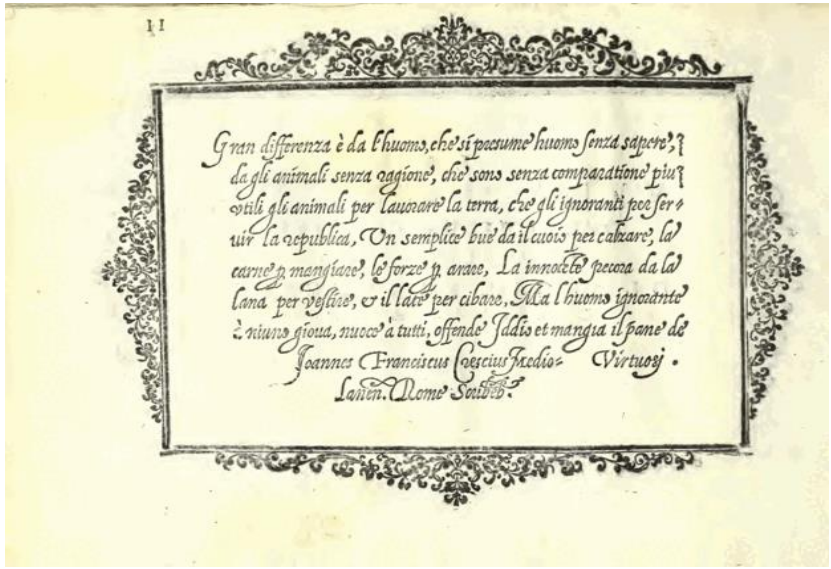
- **1,338 manuscript pages (images & ALTO transcriptions)**



Ink-bleeding simulation. Cresci, Essemplare, Presso Altobello Salicato, alla Libraria della Fortezza, Venice, fol. 2r. ***Left:*** *original line before augmentation ;* ***right:*** *result after applying the ink-bleeding simulation.*

II. Character-level augmented corpus

- **1,074 manuscript pages (images & ALTO transcriptions)**

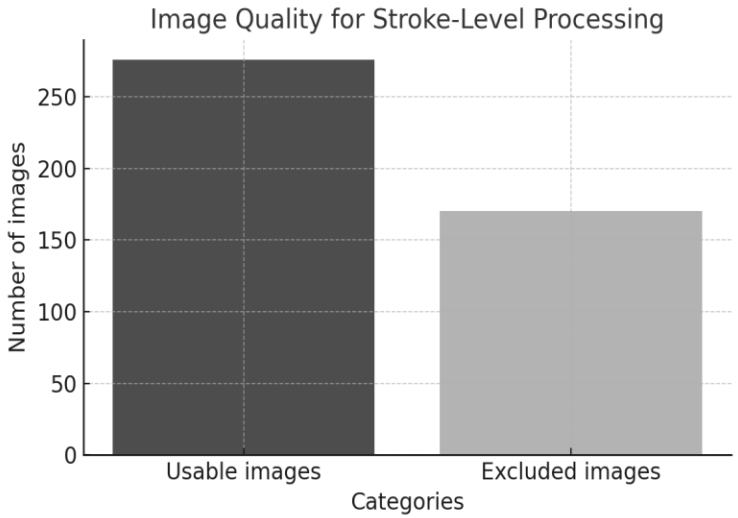


Example of Bézier-based ductual augmentation on the same Cresci Essemplare

Preliminary Results

III. Image+character-level augmented corpus

- **2,806 manuscript pages (images & ALTO transcriptions)**



IV. Test corpus

- **720 in-domain lines**

Model	Lines	Characters	WER	CER
mcCATMuS *(generic, unadapted)*	720	3,975	47.08%	11.17%
CatPrima *(fine-tuned on raw corpus)*			19.72%	3.97%
CatPrima_imgAugmented *(fine-tuned on image-augmented corpus)*			7.08%	1.41%
CatPrimaDouble *(fine-tuned on image & graphic augmented corpus)*			0.97%	0.23%

Beyond the Statistics: Migration to a Kyiv Suburb through the Lens of the 1897 Census

Konstantin Mogarichev, Tetiana Shyshkina and Maria Volkova

Rethinking Census 1897: Demiivka Individual Census Sheets

ПЕРВАЯ ВСЕОБЩАЯ ПЕРЕПИСЬ НАСЕЛЕНІЯ
РОССІЙСКОЙ ИМПЕРІИ, 1897 Г.

ИЗДАНІЕ ЦЕНТРАЛЬНАГО СТАТИСТИЧЕСКАГО КОМИТЕТА МИНИСТЕРСТВА ВНУТРЕННИХЪ ДѢЛЪ

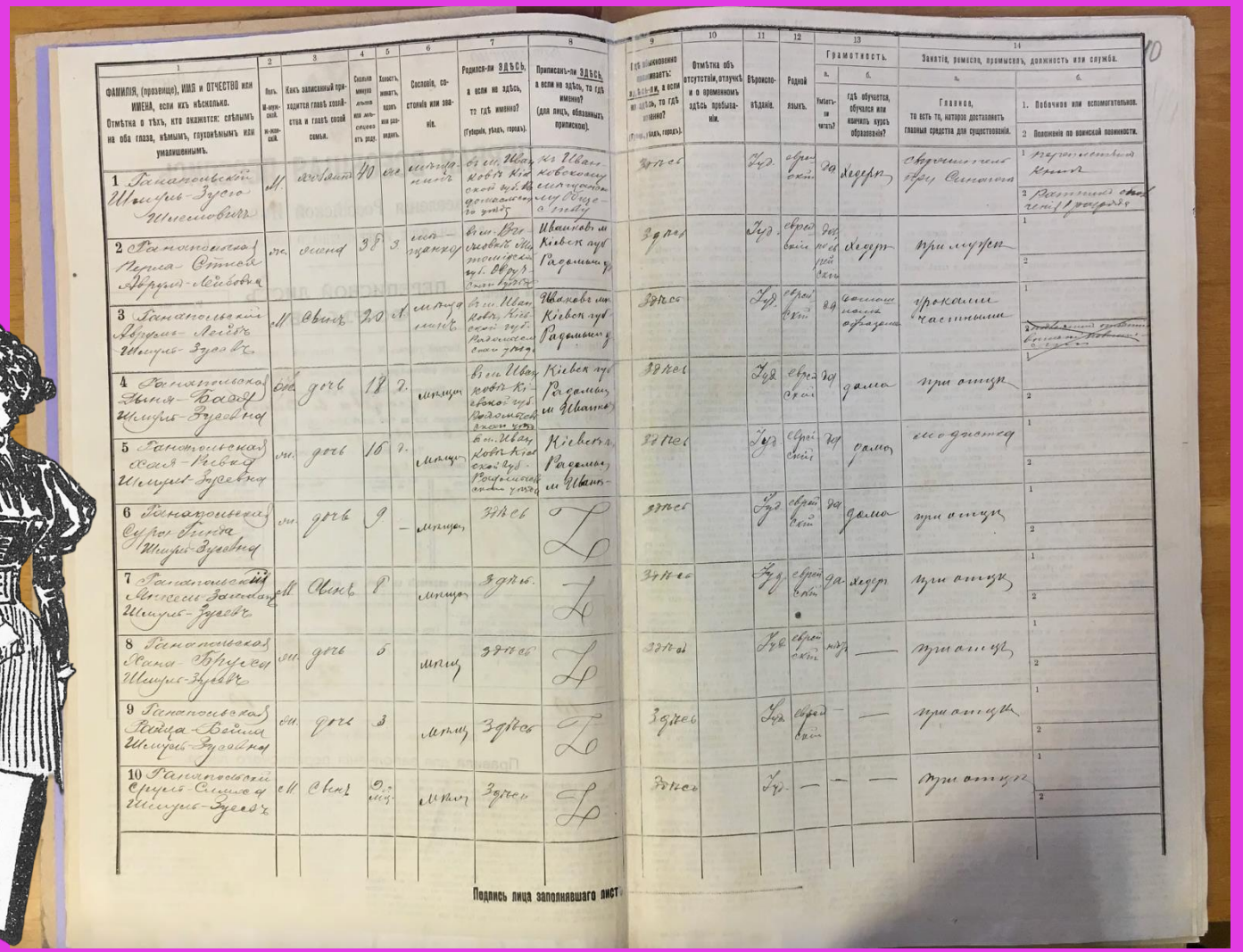
ПОДЪ РЕДАКЦІЕЮ Н. А. ТРОЙНИЦКАГО.

XVI. КІЕВСКАЯ ГУБЕРНІЯ.

1904.

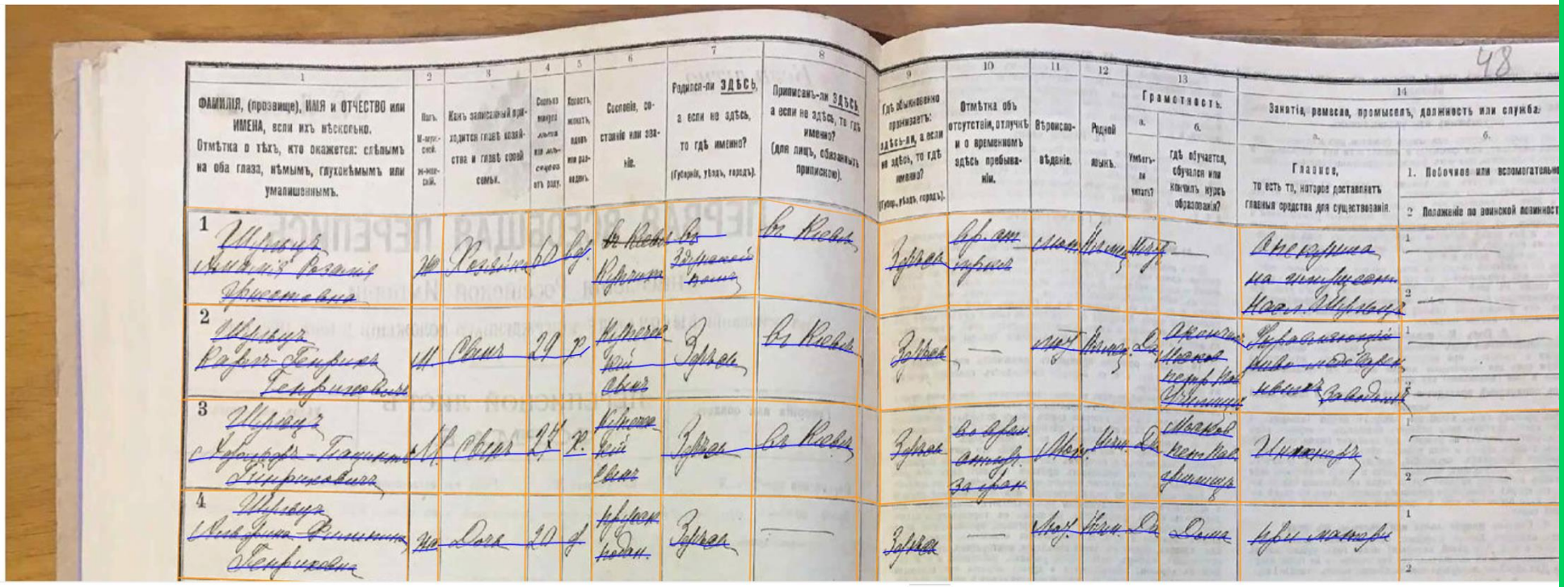
— What is your occupation, madam?
— How indiscreet you are, really.

Illustration published in Budilnik, No. 5, 1897



L'DOR V'DOR FOUNDATION · PURDUE UNIVERSITY · GCSC International Graduate Centre for the Study of Culture · JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Training the Model



Model: PI Russian v.10e
Versions: 10
Attempts: 40

Error rate v.1: **32.92%**
Error rate v.10: **9.75%**

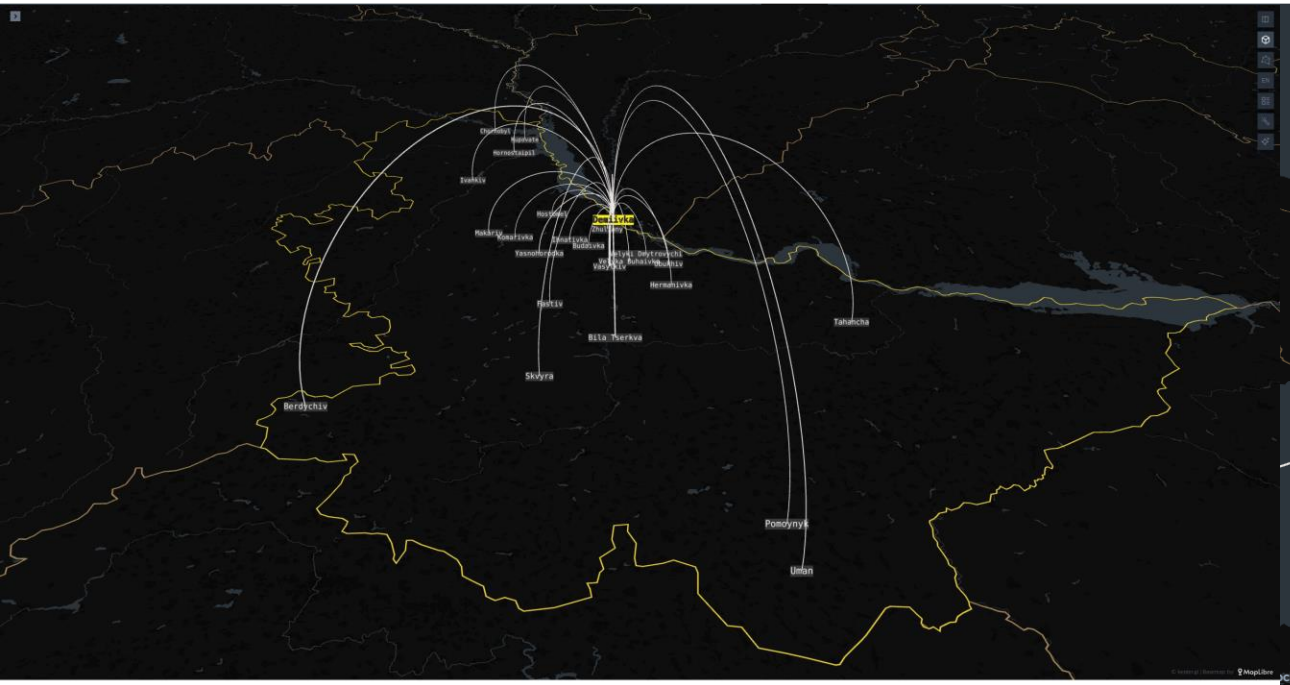
Tr. Words v.10: 97.567
Val. Words v.10: 7.496

Credits:
Jan Gronski
Artemii Plekhanov

Region1

1 Шульцъ 2 Амалия Розалия 3 Эрнестовна	1 Ж	1 Хозяйка	1 60	1 вд.	1 Купчиха	1 в 2 Здунской 3 Воле	1 въ Киеве		1 Здесь	1 вр. от. 2 лучке	1 Лют	1 Немец	1 Нетъ		1 Опекунша 2 на имущест 3 Насл. Шульцъ
1 Шульцъ 2 Карлъ-Генрихъ 3 Генриховичъ	1 М	1 Сынъ	1 29	1 Х	1 Купечес 2 кий 3 сынъ	1 Здесь	1 въ Киеве		1 Здесь		1 Лют	1 Немец	1 Да	1 Окончилъ 2 Москов 3 Петр. Пав 4 училище	1 Управляющий 2 Пиво - Медоварен 3 нымъ

Migration to Demiivka from Beyond Kyiv Gubernia (male + female)



Migration to Demiivka within Kyiv Gubernia (male+female)

Speculative Reconstruction and the Ethics of the Fragment: Early Experiments with Generative AI in Art History

Katarina Mohar

Fragments, AI, and Speculative Reconstruction

- Generative AI used to explore **visual hypotheses**, not restorations
- Two micro-datasets: **Selo fresco fragments** and **four paintings by Almenak**
- Early Selo tests exposed the **base model's generic "medievalness"**
- Reveals a core issue: AI fills **absence** with the *probable*, not the *historically specific*

original, Selo

DreamBooth inpainting tests

What AI Learns — and What It Doesn’t

- Fine-tuning captured **surface style** from minimal data
- But models struggled with **composition** and **narrative coherence**
- Outputs work as **multiple speculative possibilities**, not reconstructions
- Key insight: AI exposes the **range of interpretations** a fragment can support

Almenak: *The Peddler*

(late 17th century, oil/canvas, National Gallery of Slovenia)

LoRa tests

Prompt: An elderly peddler showing trinkets to a peasant woman outside a cottage, depicted in the style of a 17th-century oil painting, with expressive gestures, ochre tones, and soft painterly edge in style of almnk

The Ethics of Completing Fragments

- Fragments carry meaning through **incompleteness**
- AI “completions” risk producing **false wholeness** or stylistic flattening
- Our guidelines for responsible use:
  - Label AI outputs as **speculative visualizations**
  - Always show them **alongside the fragment**
  - Present **multiple options**, never a single authoritative fill
- Aim: use AI to **clarify absence**, not erase it

original, Selo

Hypothetical AI-generated fill

Speculative visualisation 1

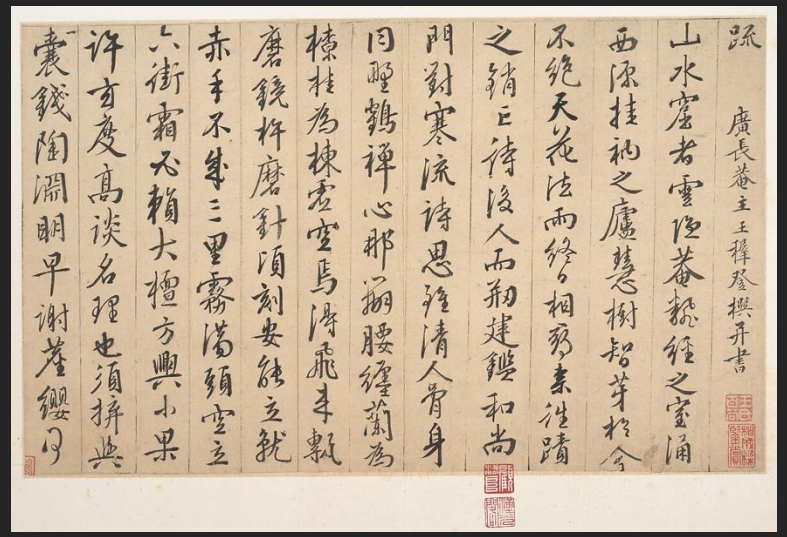
Speculative visualisation 2

Speculative visualisation 3

When Larger LLMs Aren’t Enough: Word Segmentation in Historical Chinese Texts

Hao Tan

• “Word Segmentation” - What is the Problem Here?



“removespacesbetweens
entencesthisisasentencet
hisisanothersentence”

“東京大学院生”

“東京大学 + 院生”

A graduate student at
the University of Tokyo

“東京 + 大学院生”

A graduate student
in Tokyo

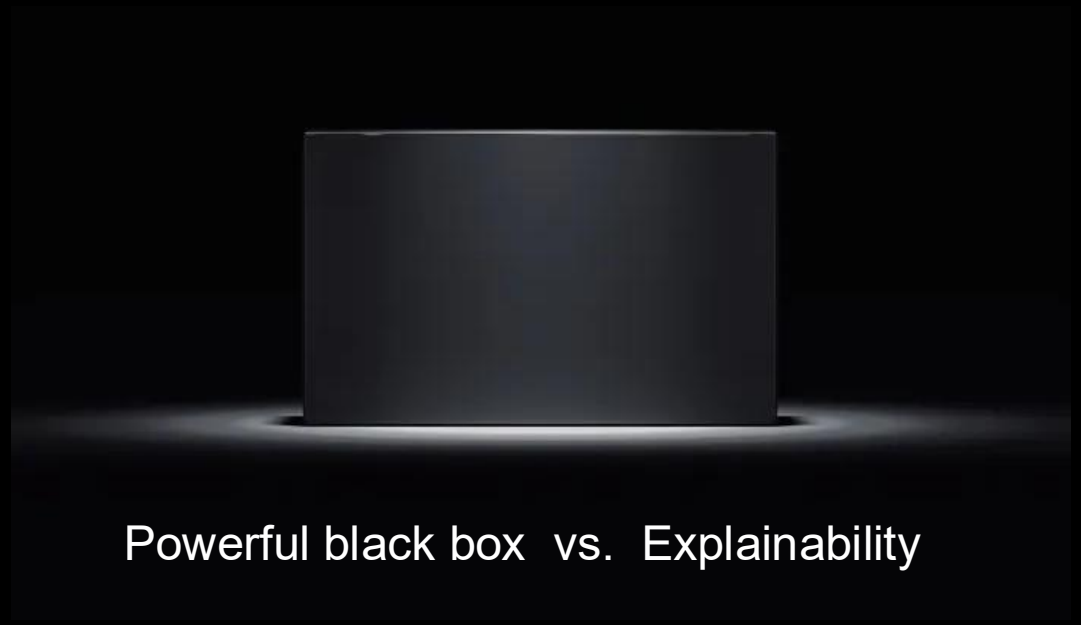
• Why LLMs Aren’t Enough for This Problem?

-- Language evolves: “One Size Doesn’t Fit All Eras”

	‘yellow’	‘river’	‘enter’	‘sea’	‘mouth’
	黄	河	入	海	口
Modern Chinese	the Yellow River		estuary		
	黄	河	入	海	口
Medieval Chinese	the Yellow River		enter	estuary	
	黄	河	入	海	口
Ancient Chinese	yellow	river	enter	sea	mouth
	黄	河	入	海	口

- **LLMs in Humanities:**
  “What are the Hidden Assumptions Here?”

Digital humanities ←→ Interpretive scholarship



Powerful black box vs. Explainability

Towards animal-centric affective analysis in poetry

Thomas Haider

Framework for Affective Analysis of Animals

Animals in NH German Poetry:
over 20k mentions (in 65k poems)

Taxon	Frequency	Taxon	Frequency	Taxon	Frequency
Nachtigall	1036	Hund	822	Adler	776
Pferd	653	Esel	521	Vieh	493
Wurm	462	Lamm	406	Hahn	378
Fisch	339	Schlangen	338	Bienen	333
Raben	315	Schlange	314	Tauben	308
Fuchs	299	Löwen	293	Hirsch	291
Brut	286	Wolf	265	Löwe	264
Taube	263	Reh	251	Schwan	250
Drachen	249	Schmetterling	236	Schimmel	233
Grillen	224	Aar	211	Schafe	208
Ungeheuer	198	Schwalbe	195	Kukuk	186
Kuh	184	Schaf	175	Bär	173

Table 1: Most frequent animal taxa in German poetry corpus.

Dimensions of Interest:

- Diegetic Function
- Agency
- Power
- Target Sentiment
- Representation/Connotation

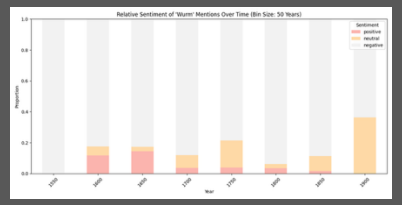
Target Sentiment Classification (of Animals)

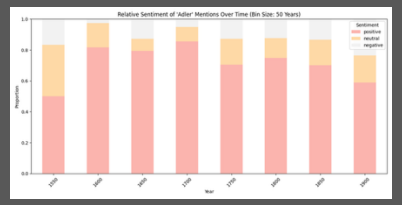
- LLMs in zero shot -> .7 F1 macro
- Few shot hurt
- Animal name as generic ‘Tier’ -> minus 7 points (sentiment encoded in symbol only)

Metric / Class	Qwen3	Mistral7B	DeepSeekV3	LLaMA3.3
Macro F1	0.659	0.337	0.692	0.696
Negative F1	0.797	0.226	0.762	0.778
Neutral F1	0.415	0.444	0.573	0.573
Positive F1	0.767	0.340	0.742	0.736
Negative Precision	0.724	0.875	0.752	0.690
Neutral Precision	0.595	0.292	0.526	0.623
Positive Precision	0.737	0.735	0.814	0.812
Negative Recall	0.885	0.130	0.773	0.891
Neutral Recall	0.319	0.924	0.630	0.531
Positive Recall	0.798	0.221	0.681	0.672

Table 2: Comparison of Model Performance on Target Sentiment Task

Worm





Eagle

- Why is an animal evaluated a certain way?
- Prompt DeepSeekV3 to interpret ‘meaning’ of animal in context -> connotation labels (representation)
- Calc. association measures sentiment & connotation
- Counterintuitive connotation -> non-canonical?

Top Keywords for Sentiment: POSITIVE (Adler)

	keyword	observed	expected	pmi	z_score	fisher_p
2	strength	351	302.76	0.2133	2.8888	0.0
3	transcendence	160	130.4	0.2952	2.6376	0.0
5	vision	40	29.98	0.4162	1.838	9e-06
15	victory	37	28.48	0.3777	1.603	0.000229
23	divine	60	49.46	0.2787	1.5083	0.000891

Top Keywords for Sentiment: NEGATIVE (Adler)

	keyword	observed	expected	pmi	z_score	fisher_p
0	vulnerability	21	3.18	2.7236	10.0219	0.0
1	destruction	10	1.48	2.7601	7.0247	0.0
4	retreat	7	1.02	2.7761	5.9188	7e-06
6	suffering	5	0.57	3.1386	5.8852	1.9e-05
7	predator	7	1.14	2.6241	5.509	2.1e-05

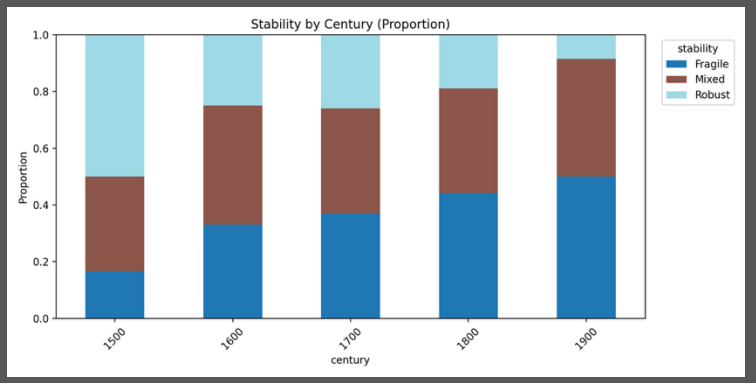
How do we evaluate connotation/representation?

- Which meaning comes from from the symbol (animal name), the context, what's encoded in the model?
- What does a reader need to know about the symbol (to understand the poem)?

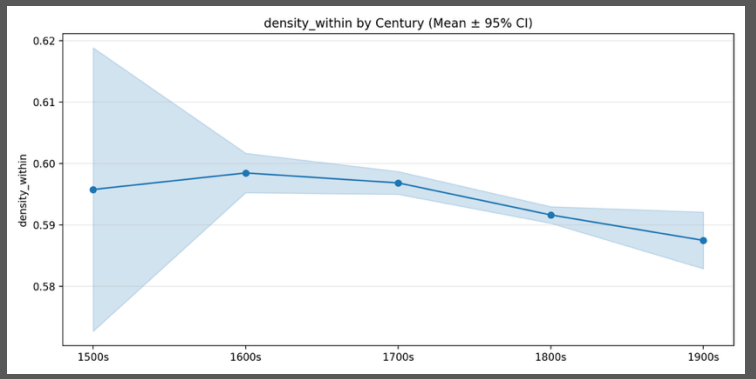
- What's the signal in the context?
- Create synthetic ground truth?
- Context stays constant -> Inject various Animals
- -> DeepSeek 'interprets'

- Here stood once a cow and thought
- Here stood once a dragon and thought
  - contemplation; simplicity; existence; reflection; ordinary
  - mythical; transience; memory; decay; contemplation

Jaccard coef. Keyword overlap (thresholded) -> context matters less and less



Topic diversity increases (more contexts) (density ↓)

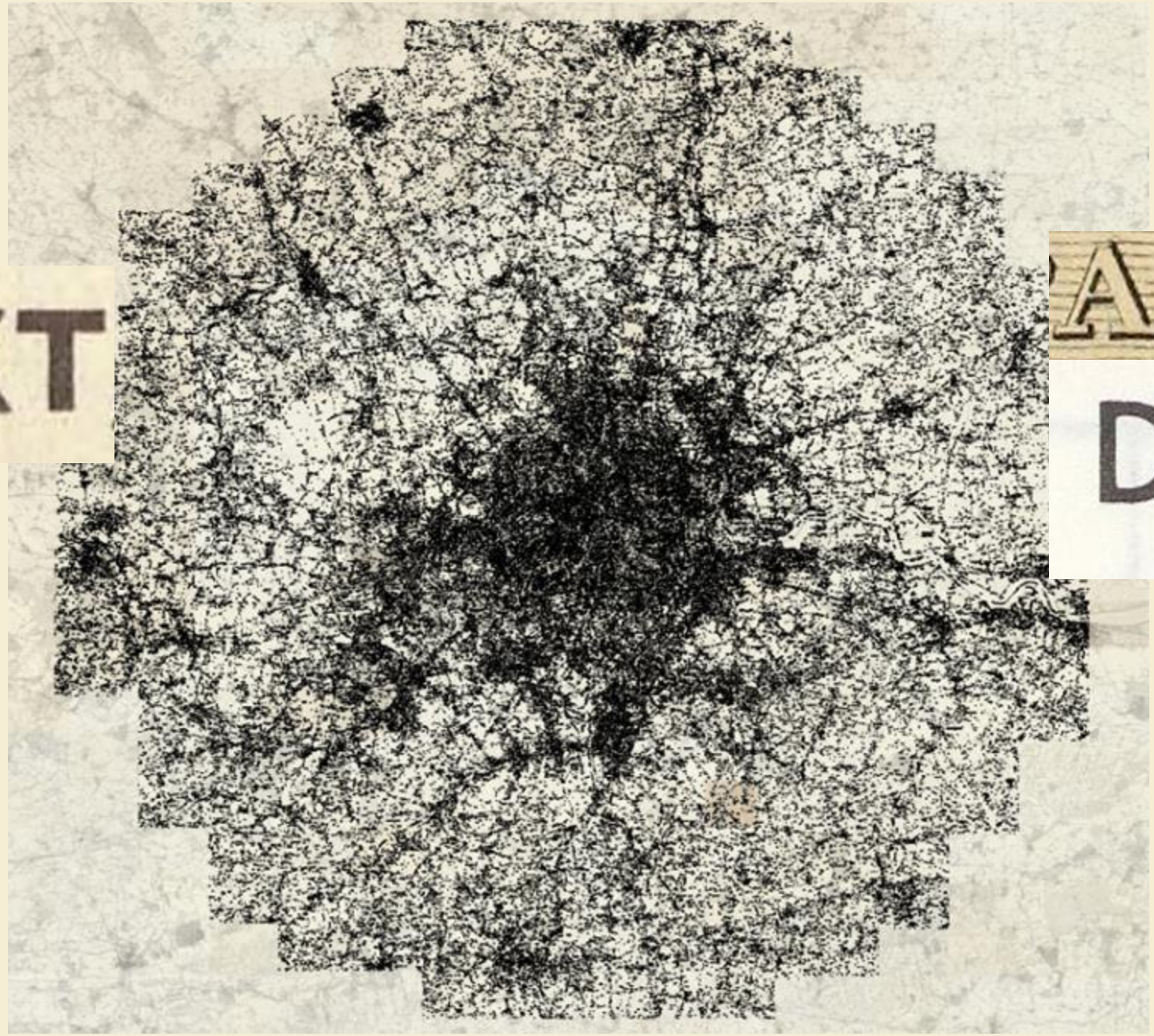


Neighbourhood Walks: A New Semantic Topology for Historical Map Text

Katherine McDonough, Kaspar Beelen and Daniel C. S. Wilson

MAP

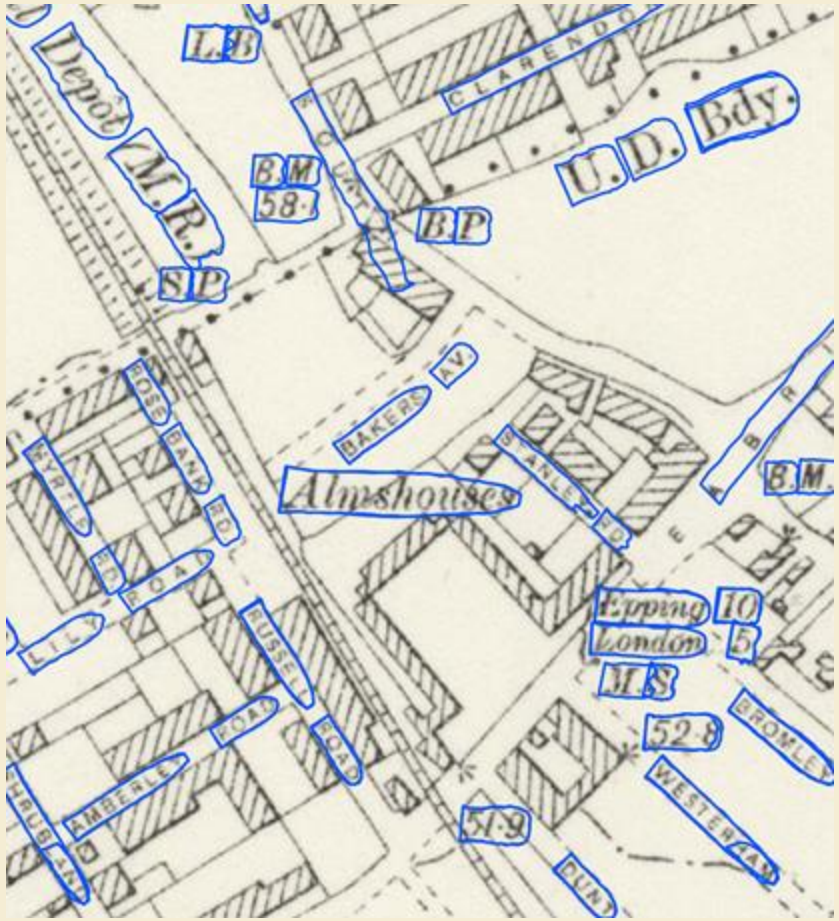
TEXT

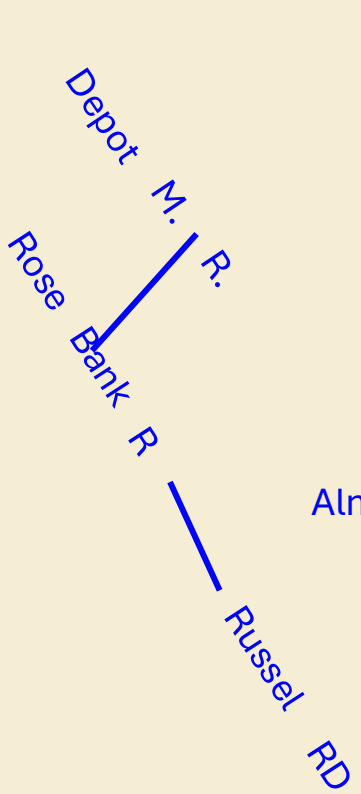


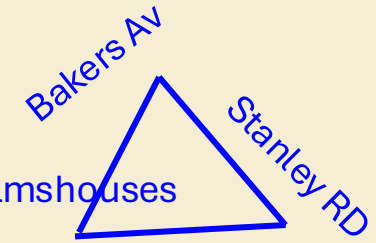
AS

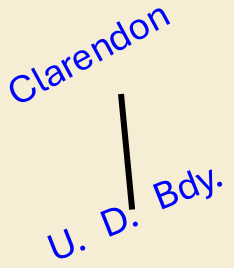
DATA

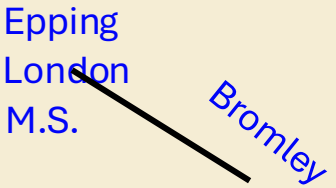
Maps 2 Labels 2 Networks



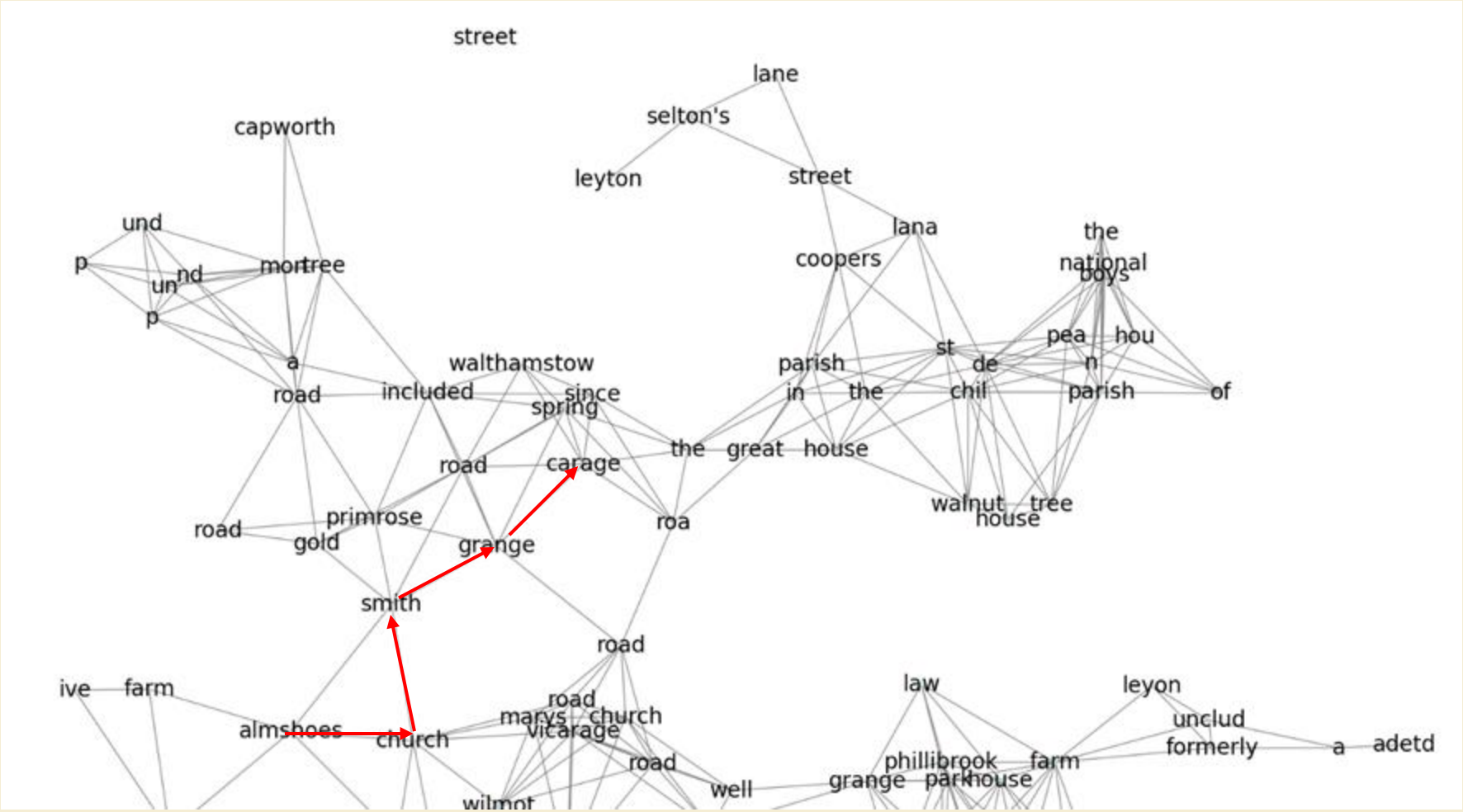








Networks 2 Walks



walk: almsho[us]es ⇨ smith ⇨ orange ⇨ [g]arage ⇨ ...

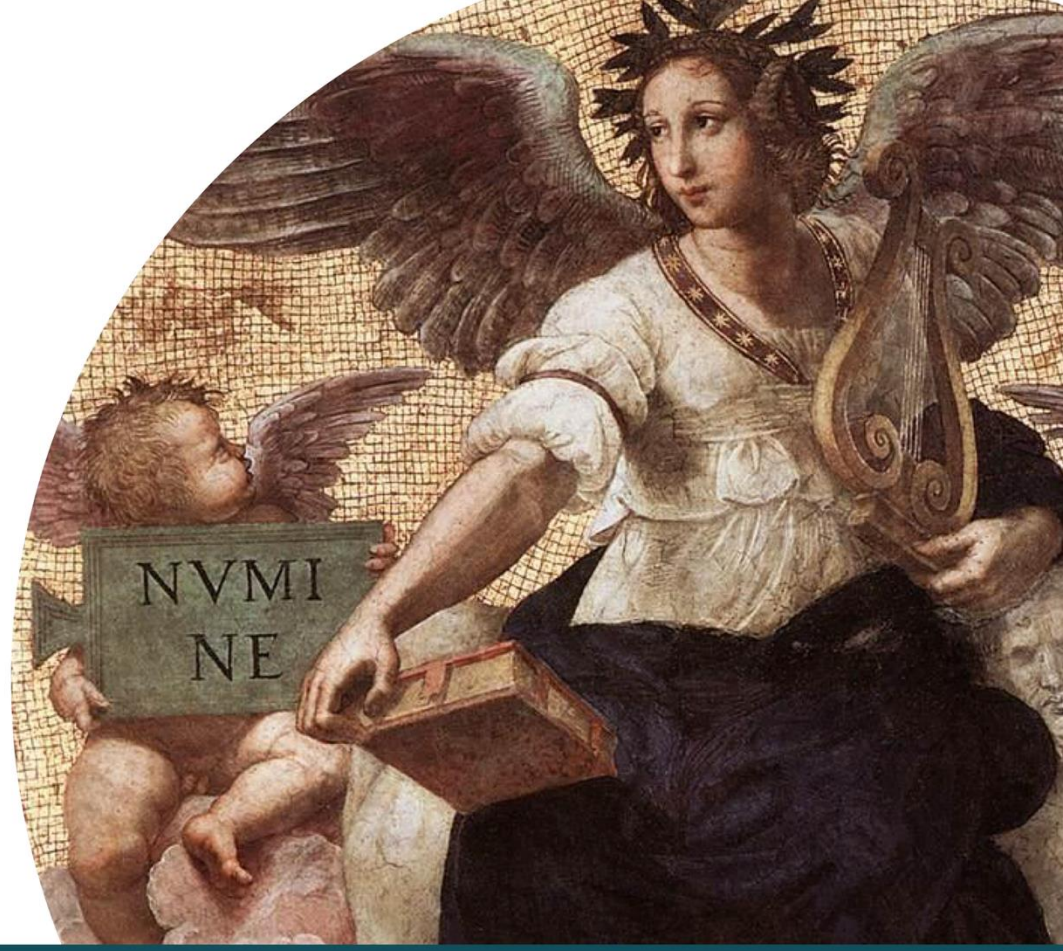
Where Empires End: Tracing the Geography of a "Soaring Spirit" in Poetry

Antonina Martynenko, Artjoms Šeļa and Petr Plecháč



Ústav pro českou literaturu AV ČR
Institute of Czech Literature of the CAS

From A to B:
Soaring view in poetry in six languages



Poet's gaze:

'*From Malta's temples to the gates of Rome*'

Six European poetic corpora (17–20th c.)
Czech, German, English, French, Russian, Slovenian

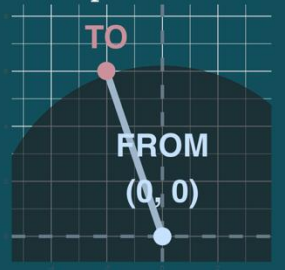
1086 poetic formulas

Method

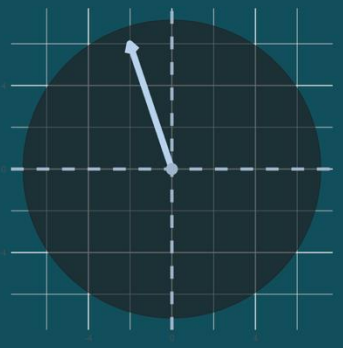
1. Extract
from A — to B
formula from a poem

2. Plot & calculate
Haversine distance

3. Transpose to FROM(0,0)



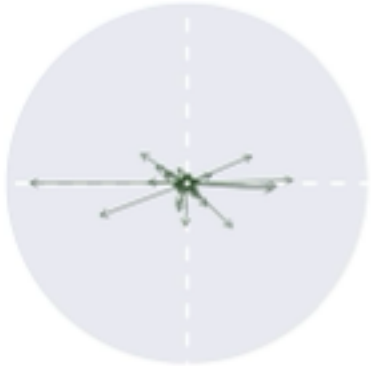
4.



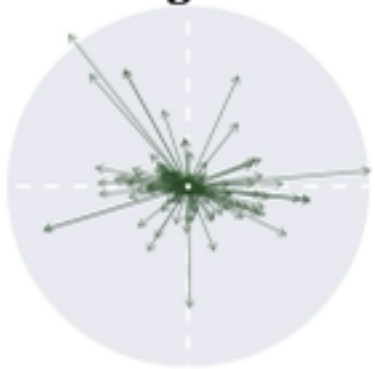
East <—> West gaze of empire

Czech

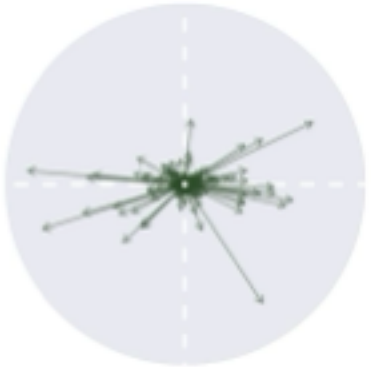
German



English



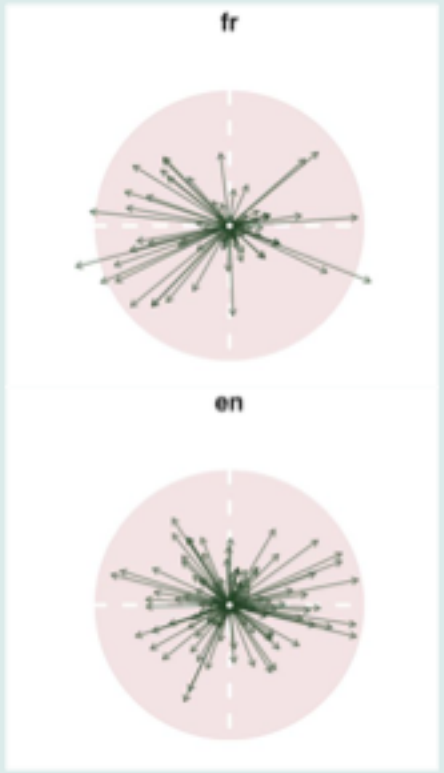
French



Russian

Slovenian

Longer distances (>1000 km)
East <-> West direction (except English!)
"Younger" national literatures explore more **local** spaces



Shorter distances
do not have
particular directions

Shared symbolic **borders and centers** of European poetry

Locations appeared in 4, 5, or all 6 corpora and their *from* directions (normalised distances)

Rapid Cultural Analytics Using LLMs: A Case of Dreams

Andres Karjus
Tallinn University
Estonian Business School
University of Tartu

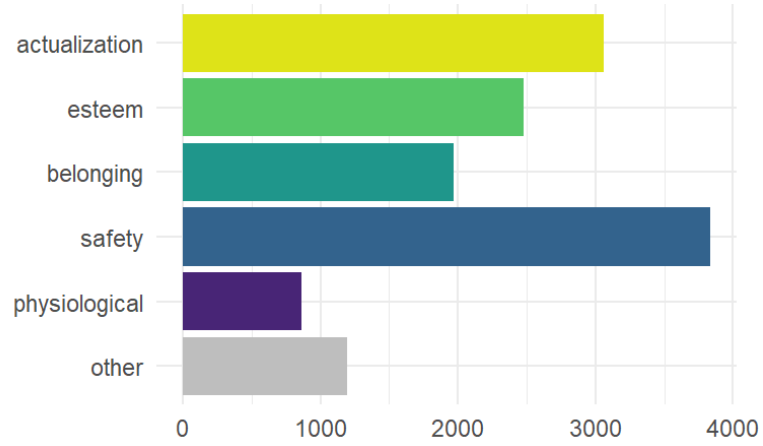
@andreskarjus on LinkedIn/Bluesky/X

- The Estonian “It’s coming together!” NGO (Hakkab looma) collected dreams and aspirations at various events across the country
- 873 groups of people (~5503 people total; btw Estonia is 1.3M)
- Not very standardized data collection though… lots of text files

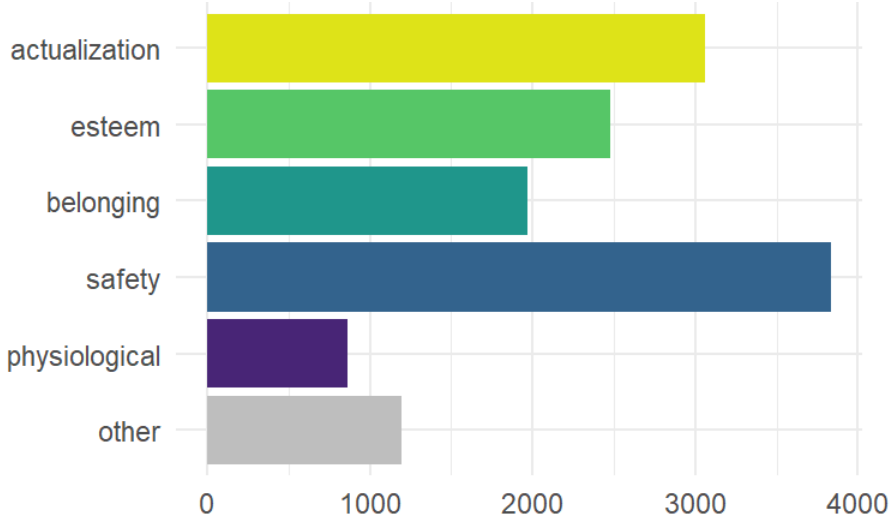
```
**Ages:**  16 16 17

**Gender:*
Boys: 1 man, 1 man, and another man
Girls: no girls!

**Place of residence:** First man – Milky Way, solar system, Earth. Other men: Narva

**Question 1:** Answers:  Right now I dream of living abroad. As a kid I used to dream of playing video games all day. Dominus hat in roblox. To live in America.

**Question 2:** In 2050 there’s new tech like super fast cars, and Minecraft is round.
```

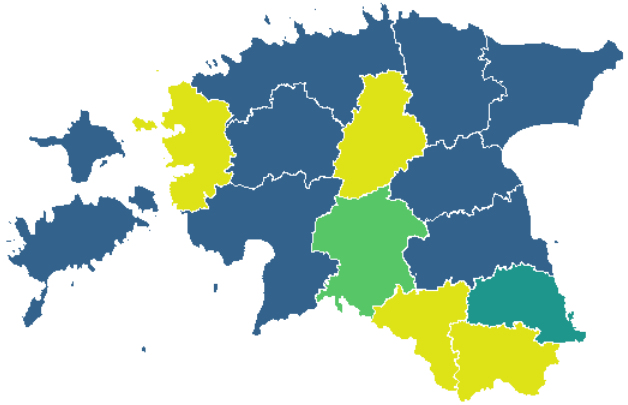
- The Estonian “It’s coming together!” (Hakkab looma) NGO collected dreams and aspirations at various events across the country
- 873 groups of people (~5503 people total; btw Estonia is 1.3M)
- No standardized data collection procedures though, lots of text files

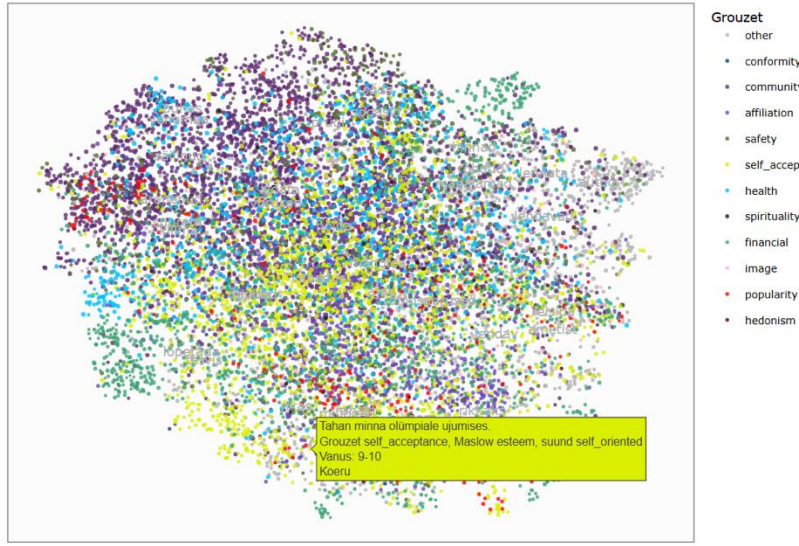
- A little inter-sector collaboration (with Innar Liiv at Taltech)
- Challenge: how to structure and analyze?
- Solution: LLM, structured outputs (GPT-4.1 + Pydantic)
- Parse, unitize and classify - all in one pass
- Result: 13407 dream units (1 to 145 per group)
- Classified into several standard psychological taxonomies + topics



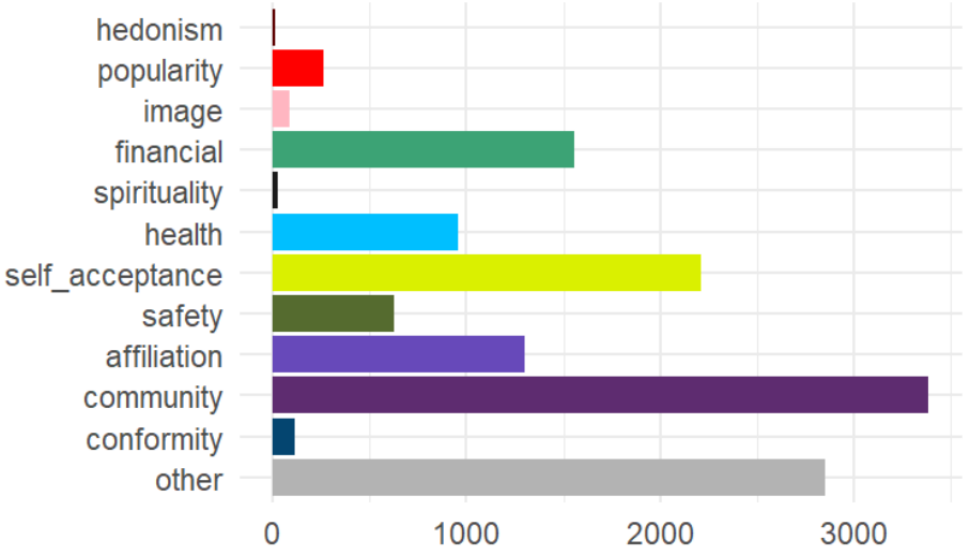
Maslow's needs:



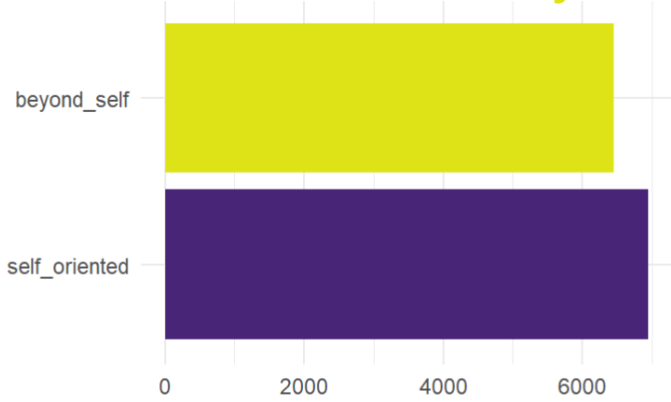


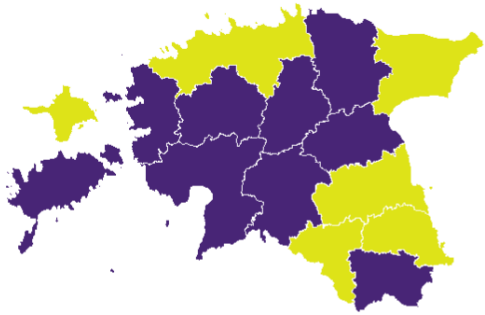


Grouzet's goals; top: community feeling, to improve the world



Damon's goals: self-oriented or beyond self



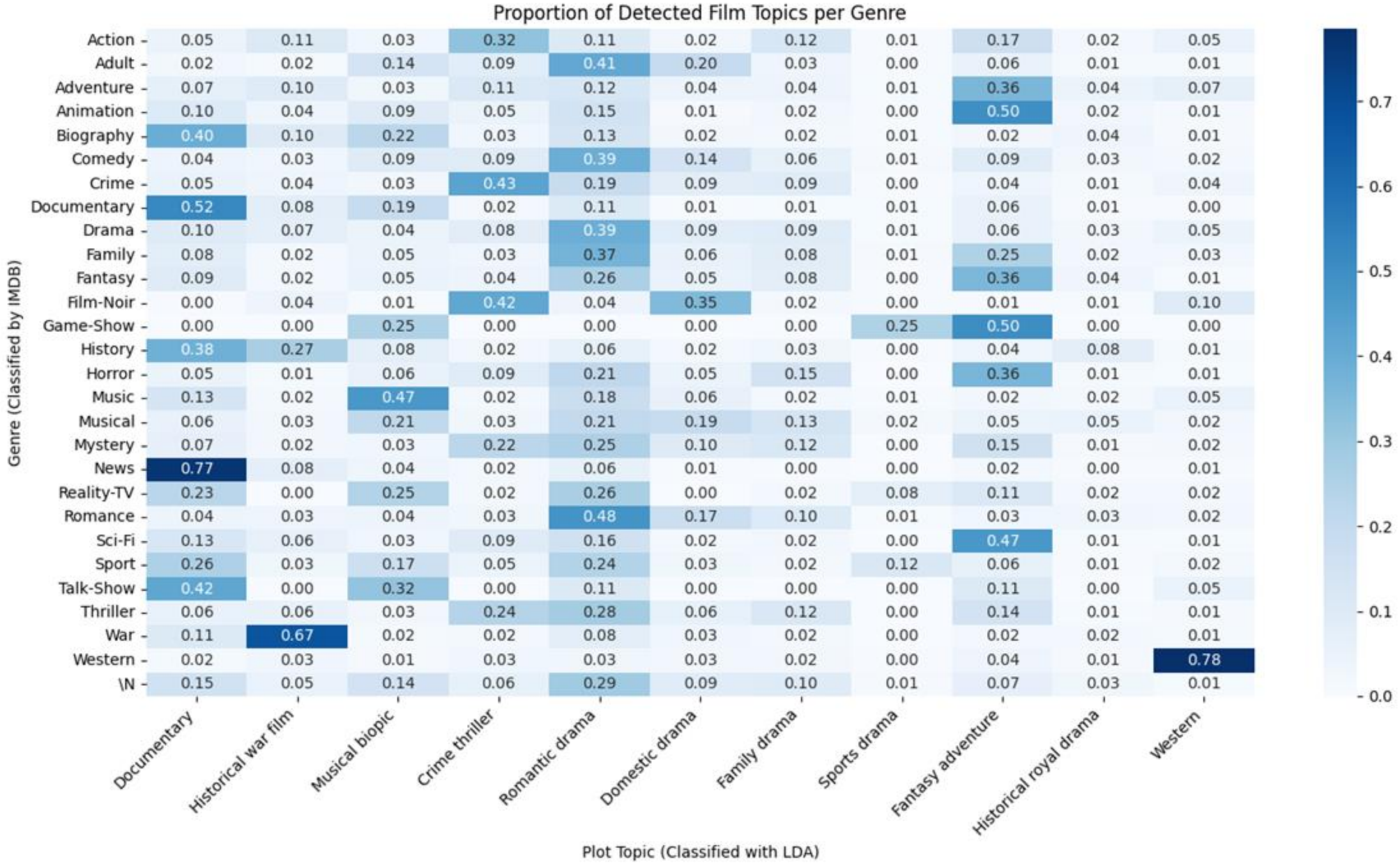


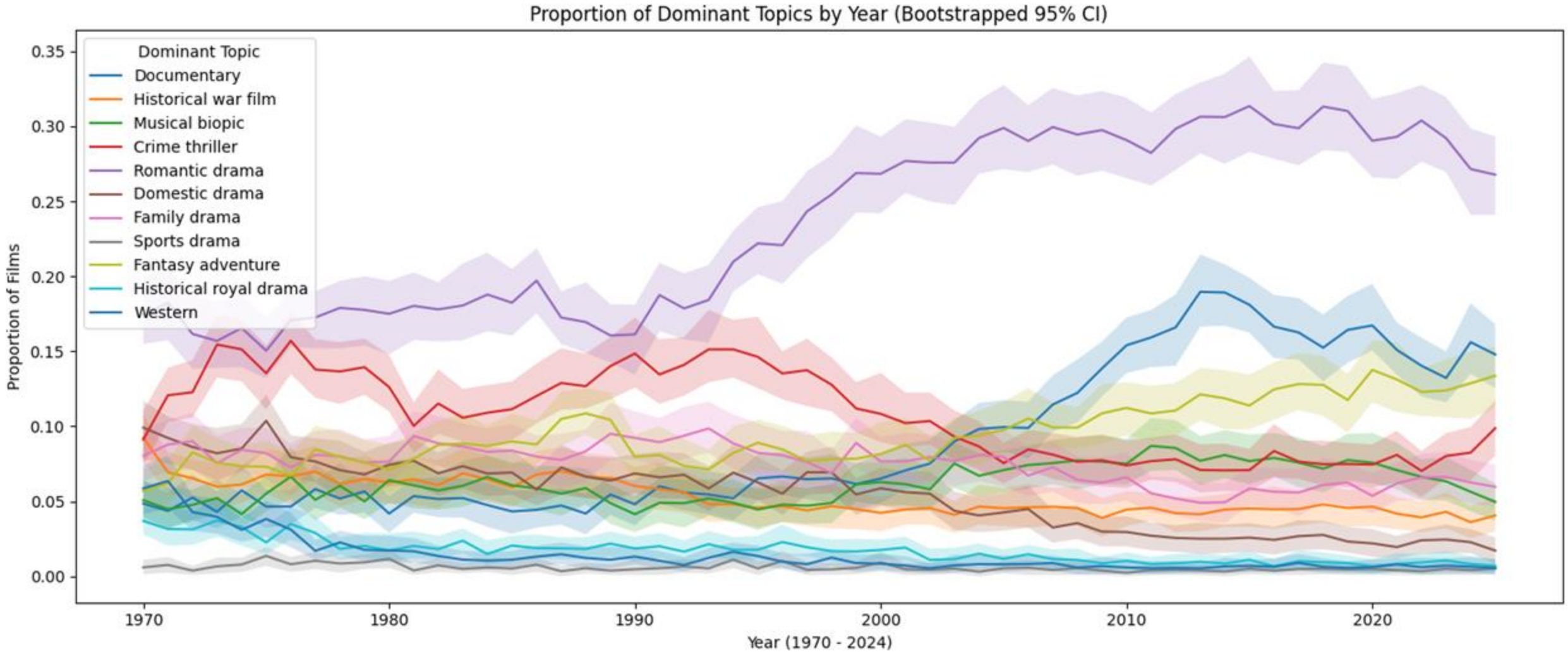
Also, we're hiring postdocs, ERC level salary, remote possible

A Diachronic Analysis of Cinematic Trends and Their Reception

Jan Tvrz


Topic 0 - Documentary
life
world
story
documentary
people
time
journey
year
history
follow
Topic 1 - Historical war film
world
war
american
army
force
military
soldier
german
country
Topic 2 - Musical biopic
star
music
movie
show
new
make
story
Topic 3 - Crime thriller
police
crime
find
get
kill
murder
criminal
drug
gang
Topic 4 - Romantic drama
friend
life
love
young
find
family
Topic 5 - Domestic drama
wife
love
husband
get
marry
Topic 6 - Family drama
find
love
get
family
father
son
Topic 7 - Sports drama
football
team
game
player
coach
mary
school
Topic 8 - Fantasy adventure
find
world
must
take
friend
group
save
evil
Topic 9 - Historical royal drama
love
king
family
father
prince
young
Topic 10 - Western
town
adam
gang
john
kill
take


Proportion of Detected Film Topics per Genre
Genre (Classified by IMDB)
Plot Topic (Classified with LDA)
Documentary
Historical war film
Musical biopic
Crime thriller
Romantic drama
Domestic drama
Family drama
Sports drama
Fantasy adventure
Historical royal drama
Western
Action 0.05 0.11 0.03 0.32 0.11 0.02 0.12 0.01 0.17 0.02 0.05
Adult 0.02 0.02 0.14 0.09 0.41 0.20 0.03 0.00 0.06 0.01 0.01
Adventure 0.07 0.10 0.03 0.11 0.12 0.04 0.04 0.01 0.36 0.04 0.07
Animation 0.10 0.04 0.09 0.05 0.15 0.01 0.02 0.00 0.50 0.02 0.01
Biography 0.40 0.10 0.22 0.03 0.13 0.02 0.02 0.01 0.02 0.04 0.01
Comedy 0.04 0.03 0.09 0.09 0.39 0.14 0.06 0.01 0.09 0.03 0.02
Crime 0.05 0.04 0.03 0.43 0.19 0.09 0.09 0.00 0.04 0.01 0.04
Documentary 0.52 0.08 0.19 0.02 0.11 0.01 0.01 0.01 0.06 0.01 0.00
Drama 0.10 0.07 0.04 0.08 0.39 0.09 0.09 0.01 0.06 0.03 0.05
Family 0.08 0.02 0.05 0.03 0.37 0.06 0.08 0.01 0.25 0.02 0.03
Fantasy 0.09 0.02 0.05 0.04 0.26 0.05 0.08 0.00 0.36 0.04 0.01
Film-Noir 0.00 0.04 0.01 0.42 0.04 0.35 0.02 0.00 0.01 0.01 0.10
Game-Show 0.00 0.00 0.25 0.00 0.00 0.00 0.00 0.25 0.50 0.00 0.00
History 0.38 0.27 0.08 0.02 0.06 0.02 0.03 0.00 0.04 0.08 0.01
Horror 0.05 0.01 0.06 0.09 0.21 0.05 0.15 0.00 0.36 0.01 0.01
Music 0.13 0.02 0.47 0.02 0.18 0.06 0.02 0.01 0.02 0.02 0.05
Musical 0.06 0.03 0.21 0.03 0.21 0.19 0.13 0.02 0.05 0.05 0.02
Mystery 0.07 0.02 0.03 0.22 0.25 0.10 0.12 0.00 0.15 0.01 0.02
News 0.77 0.08 0.04 0.02 0.06 0.01 0.00 0.00 0.02 0.00 0.01
Reality-TV 0.23 0.00 0.25 0.02 0.26 0.00 0.02 0.08 0.11 0.02 0.02
Romance 0.04 0.03 0.04 0.03 0.48 0.17 0.10 0.01 0.03 0.03 0.02
Sci-Fi 0.13 0.06 0.03 0.09 0.16 0.02 0.02 0.00 0.47 0.01 0.01
Sport 0.26 0.03 0.17 0.05 0.24 0.03 0.02 0.12 0.06 0.01 0.02
Talk-Show 0.42 0.00 0.32 0.00 0.11 0.00 0.00 0.00 0.11 0.00 0.05
Thriller 0.06 0.06 0.03 0.24 0.28 0.06 0.12 0.00 0.14 0.01 0.01
War 0.11 0.67 0.02 0.02 0.08 0.03 0.02 0.00 0.02 0.02 0.01
Western 0.02 0.03 0.01 0.03 0.03 0.03 0.02 0.00 0.04 0.01 0.78
\N 0.15 0.05 0.14 0.06 0.29 0.09 0.10 0.01 0.07 0.03 0.01
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

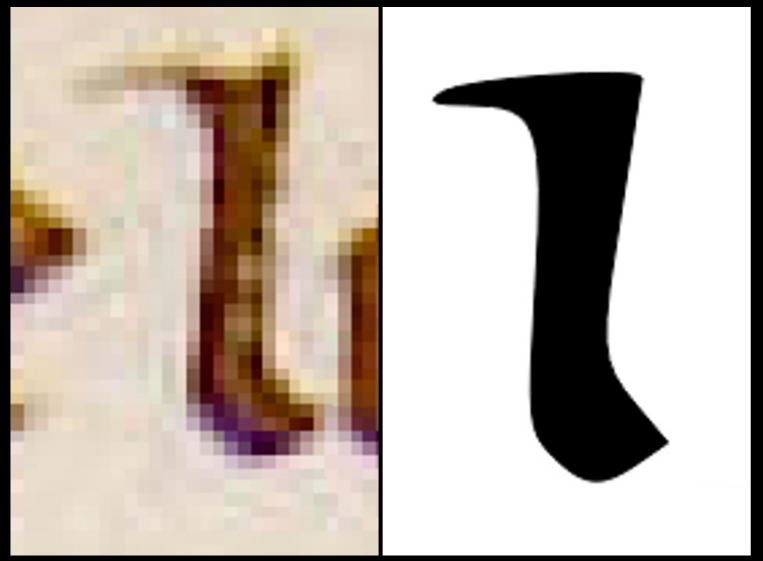

Proportion of Dominant Topics by Year (Bootstrapped 95% CI)
Dominant Topic
Documentary
Historical war film
Musical biopic
Crime thriller
Romantic drama
Domestic drama
Family drama
Sports drama
Fantasy adventure
Historical royal drama
Western
Proportion of Films
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1970
1980
1990
2000
2010
2020
Year (1970 - 2024)

Quill2Vec: A Tool for Vector Manipulation of Medieval Latin Script

Mart Herman Gerrit Makkink

WHEN TO USE QUIL2VEC?

- New and innovative pipeline
  - Based on SVGs
  - For palaeography
- Convert individual
  - Characters
  - Pen strokes
  - Part of strokes
- Scalable Vector Graphics (SVG)
  - Mathematically expressed shapes
  - XML-like
- GUI-based
  - No programming required



TECHNICAL

- Front end is built in Qt
- All processes are done through Python
- CV2 for image preprocessing
- Potrace for image tracing