Temporal models for Word Sense Disambiguation in historical texts

Barbara McGillivray 12 February 2025

Turing Seminar Series University of Bristol







Available online at www.sciencedirect.com

LINEAR ALGEBRA AND ITS APPLICATIONS

ELSEVIER Linear Algebra and its Applications 418 (2006) 708–718

www.elsevier.com/locate/laa

A probabilistic algorithm for the secant defect of Grassmann varieties

Barbara McGillivray

Department of Mathematics, University of Firenze, viale Morgagni 67/a, 50134 Firenze, Italy

By Unknown - http://www-history.mcs.st-and.ac.uk/BigPictures/Grassmann.jpeg, Public Domain

Beyond disciplinary silos

"[T]he scholarly communities of Natural Language Processing, Digital Humanities, and corpus linguistics would benefit from a deeper level of interaction and awareness of each other's fields" (Jenset and McGillivray 2017:125,137)





Institute

Home +



Thanks to digitisation efforts over the past decades, humanities scholars now have access to large digital data collections which can be analysed computationally and quantitatively. Such datasets offer a huge range of opportunities to develop new approaches to answer research questions, from the practical to the high-profile.



Research areas

Social data science

Cognitive science Research methods

ds Artificial intelligence

Ğ

Supercharge research in science and \rightarrow humanities



Transforming research





How the computational sciences and the Humanities can meet

DNS

Computational humanities research: analyzing data





Develop **new** computational techniques to analyse and model humanities data

Computational research ture that aims to promote and answer new research questions

performing creative tasks with the aid of computation.



So what exactly will I cover today?



Temporal models in NLP

Simulating human-like intelligence



Assumptions about intelligence as a fixed goalpost.



The dynamic and evolving nature of intelligence.



Temporal Depth



Co-located with EMNLP 2022

EvoNLP

The First Workshop on Ever Evolving NLP

EvoNLP, the First Workshop on **Ever Evolving NLP**, a forum to discuss the challenges posed by the dynamic nature of language in the specific context of the current NLP paradigm, dominated by language models. In addition to regular research papers, at EvoNLP we will have invited speakers from both industry and academia providing insights on the challenges involved in two main areas, namely **data** and **models**. The workshop will also feature a <u>shared task on *time-aware Word-in-Context* classification</u>.

Time-sensitive NLP



Are Large Language Models Temporally Grounded?

¹Yifu Qiu ¹Zheng Zhao ¹Yftah Ziser ²Anna Korhonen ¹Edoardo M. Ponti ¹Shay B. Cohen ¹Institute for Language, Cognition and Computation, University of Edinburgh ²Language Technology Lab, University of Cambridge {yifu.qiu, zheng.zhao, yftah.ziser, eponti, scohen}@ed.ac.uk

Time-sensitive NLP



A Dataset for Answering Time-Sensitive Questions

Wenhu Chen, Xinyi Wang, William Yang Wang Department of Computer Science University of California, Santa Barbara wenhuchen@ucsb.edu, xinyi_wang@ucsb.edu, william@cs.ucsb.edu

Temporal models for historical word sense disambiguation

Word sense disambiguation



Source: https://naviglinlp.blogspot.com/2016/04/lecture-7-word-sense-disambiguation-and.html

Historical WSD, or WSD on historical texts: why?

- Lexicography
- Historical semantics
- History, and more
- Historical information retrieval in cultural heritage

ED The definitive record of the English language	Quick search:	Lost for Words? A	ivanced search Help	Dictionary Categories Timelines	⇒ Sources ⇒ Historical Thesaurus	
ck to Results Next >	Help on Dictionary Entry Print	Save Email Cite		IF TIMBIINDS		
nachine. n.		Text size: 🔥 A	This entry has	My entries (1)	-	
ew as: Outline <u>Full entry</u>	Quotations: Show all Hide all K	eywords: On <u>Off</u>	been updated (OED Third Edition March	My searches (1)	•	_
mundation (2 mile (2) malfin / 11 0 (2) (malfin /			2000; most	Jump to:		
munciation Bitt. () majin/, c.s. () majin/			version published			
orms: 1500s machyne, 1500s-1/00s machin, 1500s- machine, 1000s macheen.			online June		y	
requency (in current use): •••••••		-	2021).	Sort by: Entry D	Jate	
rigin: A borrowing from French. Etymon: French machine.			Entry history	machina, n.	1612	/
tymology: < Middle French, French machine < classical Latin māchina (compare MACIGNO n (Show More)			Entry profile	machinability, n.	1921	
				machinal, adj.	1680	
. A structure regarded as functioning as an independent body, without mechanical involvement.			Previous version:	machinament, n.	c1425	
			OED2 (1989)	machinate, v.	1537	
A material or immeterial structure, sen the fabric of the world or of the universes a construction of	r These	urus »		machinating, adj.	1751	
A material of minimaterial structure, esp. the fabric of the world of of the universe; a construction of	Categ	ories »	In this entry:	machinator, n.	1611	
nnce. Now rare.			machine-action	machine, n.	1545	
1545 in I Schäfer Farly Mod Eng Levicour (1080) II (at rited word). The hole machine of this world is divided in 2	narte. That is to save in the celes	Hall	machine aesthetic	machine, v.	?c1450	
and into the elementall regions	purce. That is to suye, in the ceres	uuu	machine age	machined, adj.	1811	
and into the elemental regions.	tabaa faa tha aaala af tha hala		machine aid	machinetul, n.	1880	
1545 in J. Schaler Edrig Mod. Eng. Lexicogr. (1989) II. (at cited word) Machine, nath many significacions, but here it is	s taken for the worke of the hole		machine-aided	machine-gun, v.	1894	
wonde.	<i>c.</i> 1 <i>v</i>		translation machine art	machine-gunner, n	1888	
c1550 Complaynt Scott. (1979) 2 The maist illustic potent prince of the maist fertil & pacebil realme, vndir the machine	of the supreme olimp.		machine-assisted	machine-gunning,	n. 1915	
1599 A. HUME Hymnes sig. B4 Be his wisedome, sa wondrouslie of nocht, This machin round, this vniuers, this vther wa	arld he wrocht.		machine helt	machineless, adj. (a 1909 1798	
1673 H. HICKMAN Hist. Quinq-articularis 518 They that asserted Universal redemption by the death of Christ destroyed	the whole Machine of the Calvini	an	machine-belting	machinery, n.	1687	
predestination.			machine bolt			
1682 N. O. tr. N. Boileau-Despréaux Lutrin 1. 239 Behind this Machine [sc. a pulpit], cover'd as with a skreen, The Snea	king Chanter scarce could then b	e	machine-boy	-	-	
seen.			machine-breaking			
1753 J. HANWAY Hist. Acct. Brit. Trade Caspian Sea II. iii. 13 Her imperial majesty is drawnin a large machine, which c conveniencesThis machine is set on a sledge, and drawn by 24 post horses.	ontains her bed, a table, and oth	er	machine candidate machine card			
1791 C. SMITH Celestina I. 129 Her new laylock bonnet. for the safety of which she was so solicitous that she would have was contained into the coach, had it not been opposed by the coachman.	taken the great machine in which	it	machine-closed machine-coated			_
a1806 J. BARRY in R. N. Wornum Lect. on Painting (1848) 196 Had the whole of this great machine of the Fontana di Tr sculptors.	evi been committed to any one o	f those	machine code machine-coded			
1820 R. HALL Wks. (1832) VI. 457. The mind casts its eve over the whole machine of society			machine-cut			
1878 R Renarran La Saisian 270 To each mortal naradvanture earth hacomes a new machine			machine-cut			
10/0 16 December 24 Contents 2/9 10 cach mortai perativenture cartin becomes a new machine.			machine-darning			
	(Hide quo	tations)	machine-divided			
b. spec. A scheme or plot. Obsolete.	These	urus »	machine drill			
1505-6 Olivery Furzavery I Let to James VI (Camden Soc) 112 In wordz of such waight as in honast dimare hit may m	ar the facon of dinelische machin	185	machine-drilling			
2355 9 Quant Essantin 1 24t, 10 Junites 11 (Calificer 50c.) 113 111 Works.or such wagnt, as, in nonest dimars, nit may m and crase the hartz of treason-mynding men.	ar the taçon of difference machin		machine-driven machine-driver			-
1601 S. DANIEL Ciuill Warres (rev. ed.) VI. viii. f. 84, in Wks. As vsually it fares with those that plot These machines of An	nbition, and high pride.		machine-electric			
1707 G. FARQUHAR Beaux Stratagem v. 65 Now unless Aimwell has made good use of his time, all our fair Machine goes s	souse into the Sea like the Edisto	ne.	machine electricity machine embroidery			
	(me da		machine finish machine-finished			.0
. A living body, esp. the human body considered in general or individually. Now chiefly financine fr	rom These	urus »	machine for living (in)			>
ense 6b (cf. sense 8b).	Categ	ories »	machine-generated			
16 a. W. Charman an Hamlet n ii 107. Thing an among most done I adu aikilat this maak's data biss			machine-ginned			

Pedrazzini, Nilo, and Barbara McGillivray. (2022) Let Viacin Proceedings of the second of the secon

So, why historical WSD? You guessed it!

- Manual annotation is very time-consuming
- Semantic change detection algorithms still too rough and don't capture the nuances needed in humanities research

BUT

- Humans disagree on word senses, let alone historical ones!
- New, under-developed research area, still



Historical English

Beelen, K., Nanni, F., Coll Ardanuy, M., Hosseini, K., Tolfo, G., McGillivray, B. (2021) "When time makes sense: A historically-aware approach to targeted sense disambiguation." *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*.

The industrial revolution and the English language

- The British Industrial Revolution was a time of profound changes and mechanization was its primary driver.
- New machines and new technology heavily affected the English lexicon: new words are coined and existing words get new meanings.
- Impact of mechanisation on everyday life by exploring large historical collections (19th century books, newspapers etc.)



Sense annotation of historical texts

6 V II

The records of the American associations show hscores of remarkable captures. A few years ago, +a salesman for a Maiden-lane firm chanced to be travelling with fifty thousand dollars worth of +unset diamonds in his pocket, when s man sitting beside him quietly rose, and, thrusting a revolver before his face, demanded the gems. The diamonds were produced, the robber next pulled the bell rope and, still covering the passengers with his revolver, waited until the car stopped and jumped off.

The detective machinery was at once put is operation, but it was two years before the thief was arrested.

The detectives employed by the unions also watch where stolen goods are likely to find their way, and keep in touch in every possible way with the "under world." It has happened several times that the detectives have notified jewellers of thefts from their stock which they had not discovered themselves. Several important plots have also been discovered in advance in the same way, and the proposed victims placed upon their guard. TASK

TUTORIAL

How did the word 'car' change over time and place?

Read the text carefully. Select the option that best matches the use of 'car'.

You can also note irrelevant or hard to read images.

A wheeled, usually horse-drawn conveyance

A railway carriage or wagon, usually powered by steam

An electric tramcar (also called streetcar)

A road vehicle powered by a combustion engine, usually for a driver and a small number of passengers 19th-century (1800-1920) British newspaper corpus (4.5B tokens) Voluntary crowdsourcing *Trolley, train, coach, bike*

Sense distributions in historical texts









Sense distributions in historical texts





Targeted Sense Disambiguation

Given a target sense, determine whether a token in a given context is relevant to the sense

Example: disambiguate tokens in text as being relevant to "machine" as "a vehicle or conveyance"

 This can include "machine" but also "car" referring to vehicle

DUT COMMUNITY BLOG	
DED Oxford English Dictionary The definitive record of the English language	Quick search:
Back to Results Next »	Help on Dictionary Entry Print Save E
machine, <i>n</i> .	Text siz
View as: Outline <u>Full entry</u>	Quotations: Show all <u>Hide all</u> Keywords:
Yronunciation: Brit. ▷/mə'ʃi:n/, U.S. ▷/mə'ʃin/ Forms: 1500s machyne, 1500s-1700s machin, 1500s- machine, 1600s macheen. Frequency (in current use): ●●●●●● Origin: A borrowing from French. Etymon: French machine. Etymology: < Middle French, French machine < classical Latin māchina (compare MACIGNO	17 (Show More)
I. A structure regarded as functioning as an independent body, without me 1.	chanical involvement.
a. A material or immaterial structure, <i>esp</i> . the fabric of the world or of the wedifice. Now <i>rare</i> .	universe; a construction or Thesaurus » Categories »
1545 in J. Schäfer <u>Early Mod. Eng. Lexicogr</u> . (1989) II. (at cited word) The hole <u>machyn</u> and into the elementall regions.	ie of this world is divided in .2. parte. That is to saye, in the celestiall
1545 in J. Schäfer <i>Early Mod. Eng. Lexicogr.</i> (1989) II. (at cited word) <u>Machine</u> , hath m worlde.	any significacions, but here it is taken for the worke of the hole
c1550 <u>Complaynt Scotl.</u> (1979) 2 The maist illustir potent prince of the maist fertil & pac	sebil realme, vndir the machine of the supreme olimp.
 1999 A. HOME <u>riginites</u> sig. 64 Be his wisedome,sa wondrousne of nocht, This <u>machin</u> r 1673 H. HICKMAN <u>Hist. Quinq-articularis</u> 518 They that asserted Universal redemption b predestination. 	y the death of Christ destroyed the whole <u>Machine</u> of the Calvinian
1682 N. O. tr. N. Boileau-Despréaux <u>Lutrin</u> I. 239 Behind this <u>Machine</u> [sc. a pulpit], cov seen.	ver'd as with a skreen, The Sneaking Chanter scarce could then be
1753 J. HANWAY <i>Hist. Acct. Brit. Trade Caspian Sea</i> II. iii. 13 Her imperial majesty is dra conveniencesThis machine is set on a sledge, and drawn by 24 post horses.	wnin a large $\underline{\text{machine}}$, which contains her bed, a table, and other
1791 C. SMITH <u>Celestina</u> I. 129 Her new laylock bonnetfor the safety of which she was so was contained into the coach, had it not been opposed by the coachman.	solicitous that she would have taken the great <u>machine</u> in which it
a1806 J. BARRY in R. N. Wornum Lect. on Painting (1848) 196 Had the whole of this great sculptors.	at <u>machine</u> of the Fontana di Trevi been committed to any one of those
beuptorbi	

Step 1

Select target sense: "machine_nn01-ABC" as "complex device, consisting of a number of interrelated parts [...]"

Include sense synonyms (e.g. "plant_nn01-XYZ")

Headword: machine, n. **Quotation:** *Windmills as hitherto made are very costly* machines. Sense ID: machine_nn01-38475286 **Definition:** "A complex device, consisting of a number of interrelated parts, each having a definite function [...]" Text daterange: 1659-Keyword: machines Offset: 43 Year: 1881 Source: Nature, by W. Thomson.

Historical Thesaurus of English (HTE)

Step 2: data expansion

Classify head words in quotations of the dictionary as related to the target (seed) sense

"The calculating <i>machine</i> now constructing under the superintendence of the inventor."			
"The Church was excellent as a national refrigerating machine."	0		
"Examples of mobile earth moving <i>plant</i> are bulldozers, graders and scrapers."	1		
"I could lift the <i>plant</i> and be far away before daylight."	0		

Lemmas and senses

	Seeds	Expanded senses	Quotations
anger	6	17/121	103/564
apple	19	6/61	36/300
art	19	5/47	33/212
democracy	7	6/61	37/301
happiness	5	9/46	57/186
labour	18	4/30	29/148
machine	25	8/83	42/361
man	48	9/78	55/380
nation	15	8/85	53/430
power	39	5/49	34/244
slave	10	20/158	103/670
woman	17	10/81	64/379

Table 1: Headwords used in the experiments, with the number of seed senses, their expanded senses (synonym and unrelated) averaged per sense, and number of quotations (positive and negative examples) averaged per sense.

Language models

Search ou Our web Catalogues & Collections + Discover & Learn + What's On + Visit + Business Support + Sh

Collection guides

Digitised printed books (18th-19th century)



Concise Guide to London, with map, etc. 1885

Printed books digitised in partnership with Microsoft from the 18th & 19th century

Subjects

Digital scholarsh Undertake innovative collections and data

Printed books Printed material in a r current researchers a

Neural Language Models for Nineteenth-Century English

Kasra Hosseini^a*, Kaspar Beelen^{a,b}, Giovanni Colavizza^c, Mariona Coll Ardanuy^{a,b}

^aThe Alan Turing Institute, London, UK ^bQueen Mary University of London, London, UK

^cInstitute for Logic, Language and Computation, University of Amsterdam, NL

Language models

base



BERT 1900

Token Embeddings from Transformers



Extract BERT embeddings for each quotation keyword

- Concatenate last four layers
- If a word has multiple subtokens*, we averaged their vector representations

* in case the BERT tokenizer splits a token into two sub ##tokens

Sense Centroids



Does time matter?

Impact of time-sensitive disambiguation

- Quotation and time-stamp is input for classifier
- Influences the construction of centroid vector
 - Filtering: only use the temporally closest quotations (in training data)
 - Weighting: a weighted average of quotations with weight determined by proximity to input time-stamp

Does time make sense?

	1760–1850		1760–1920			1760-2000			
	Prec	Recall	F_1Score	Prec	Recall	F_1Score	Prec	Recall	F_1Score
random	0.102	0.511	0.170	0.087	0.483	0.148	0.087	0.503	0.148
Lesk: token overlap	0.234	0.266	0.249	0.245	0.278	0.261	0.248	0.277	0.261
Lesk: sentence embedding	0.269	0.196	0.227	0.266	0.198	0.227	0.280	0.215	0.243
Lesk: w2v	0.323	0.291	0.306	0.288	0.270	0.279	0.286	0.257	0.271
SVM classifier	0.500	0.091	0.155	0.495	0.083	0.143	0.509	0.077	0.133
BERT_base binary centroid	0.254	0.699	0.373	0.238	0.702	0.356	0.236	0.716	0.355
BERT_base sense centroid	0.756	0.464	0.575	0.665	0.471	0.552	0.618	0.493	0.548
BERT_base perceptron	0.578	0.425	0.490	0.575	0.448	0.504	0.580	0.456	0.510
BERT_1900 binary centroid	0.234	0.698	0.351	0.221	0.715	0.338	0.222	0.728	0.340
BERT_1900 sense centroid	0.766	0.498	0.604	0.702	0.512	0.592	0.630	0.497	0.556
BERT_1900 perceptron	0.575	0.429	0.492	0.588	0.453	0.511	0.586	0.463	0.517
BERT_1850 binary centroid	0.229	0.678	0.343	0.224	0.713	0.340	0.222	0.722	0.339
BERT_1850 sense centroid	0.789	0.486	0.602	0.688	0.500	0.579	0.613	0.495	0.548
BERT_1850 perceptron	0.587	0.424	0.492	0.568	0.437	0.494	0.570	0.456	0.506

Table 2: Precision, recall and macro F_1 scores of the positive class over all senses computed for different time periods. The table highlights the top performing methods for each experiment.

Does time make sense?



Figure 1: Optimal date range for each language model as measured by the F_1 score of the positive class, using the *sense centroid* method: the x-axis represents the average points of rolling 100-year quotation date ranges.

Time matters

1850	1920
0.575	0.552
0.458	0.433
0.593	0.556
0.604	0.592
0.505	0.464
0.627	0.584
0.602	0.579
0.489	0.441
0.609	0.562
	1850 0.575 0.458 0.593 0.604 0.505 0.627 0.602 0.489 0.609

Table 3: Macro Fscores for time (in)sensitive models sense embeddings.

Latin

Iacopo Ghinassi, Simone Tedeschi, Paola Marongiu, Roberto Navigli, Barbara McGillivray (2024). Language Pivoting from Parallel Corpora for Word Sense Disambiguation of Historical Languages: a Case Study on Latin. In *Proceedings of LREC-COLING 2024*.

Latin WSD

- Historical language data (e.g. Latin) vs historical data (e.g. historical English)
- Low-resource setting
- Challenge: improve state of the art in Latin WSD with automatically generated training data

Compos.: concrete.
Compos.: concrete.
Civitās, -ātis f. a civis (cf. VARRO ling. 10, 39 GRAMM. suppl. 68, 14 ISID. orig. 15, 2, 1 l. 77). scribitur: ceivitas: LEX repetund. (Corp. I 198) 12 al. LEX Tarent. (Ephem. epigr. IX p. 1 sqq.) 39 CORP. V 7231, 2(aet. Aug.) al. IX 5834. FRONTIN. aq. 1? (celintatis cod.; sed cf. 89. 94 al.). civitas: LEX de flam. Narbon. (Corp. XII 6038) 17 al. CORP. V 4954. XII 1855. III 388 et saepius. cibitas: CORP. VI 34676. XIV 2080. IX 648 (bis). X 478. 6226. III 218 (Cypr., a. 198 p. Chr.), 11 EDICT. imp.

Previous research

- Dictionary quotations
- Data sparsity
- Skewed distribution of quotations over time
- Bamman & Burns 2020

ab-dĭco , āvi, ātum, 1, v. a. (prop.

I.to indicate, announce something as not belonging to one; hence),

I. In gen., to deny, disown, refuse, reject.—With acc. und *inf*.: mortem ostentant, regno expellunt, consanguineam esse abdicant, *deny her to be*, Pac. ap. Non. 450, 30 (Trag. Rel. p. 84 Rib.): "abdicat enim voluptati inesse bonitatem," *Pseudo Apul. de Dogm. Plat. 3 init.*—With acc. (so very freq. in the elder Pliny): naturam abdico, Pac. ap. Non. 306, 32 (Trag. p. 120 Rib.): "ubi plus mali quam boni reperio, id totum abdico atque eicio," **Cic. de Or. 2, 24, 102**: "legem agrariam," **Plin. 7, 30, 31, § 116**: "corticem," <u>id. 13, 22, 43, § 124</u>: "ea (signa) in totum," id. 10, 4, 5, § 16; cf.: "utinam posset e vita in totum abdicari (aurum)," *be got rid of*, id. 33, 1, 3, § 6: "omni venere abdicata," id. 5, 17, 15, § 73 al.

II. In partic.

A. Jurid. t. t., to renounce one, partic. a son, to disinherit (post-Aug.): "qui ex duobus legitlmis alterum in adoptionem dederat, alterum abdicaverat," Quint. 3, 6, 97; cf.: "minus dicto audientem fllium," id. 7, 1, 14: "ex meretrice natum," id. 11, 1, 82 al.: "quae in scholis abdicatorum, haee in foro exheredatorum a parcntibus ratio cst," id. 7, 4, 11.—*Absol.*: "pater abdicans," Quint. 11, 1, 59; cf.: "filius abdicantis," id. 4, 2, 95; and: "abdicandi jus," id. 3, 6, 77.—Hence, patrem, to disoun, Curt. 4, 10, 3. B. Polit. t. t.: abdicare se magistratu, or *absol.* (prop. to detach one's self from an office, hence), to renounce an office, to resign, abdicate (syn.: "deponere magistratum): consules magistratu se abdicaverunt," Cic. Div. 2, 35, 74; so, so magistrutu, id. Leg. 2, 12, 31; Liv. 4, 15, 4 al.: "se dictatu. rā," Caes. B. C. 3, 2; Liv. 2, 31, 10; 9, 26, 18 al.: "sc consulatu," id. 2, 2, 10; Vell. 2, 22, 2: "se praeturā," Cic. Cat. 3, 6, 14: "se aedilitate," Liv. 39, 39, 9 etc. Likewise: "se tutelā," Cic. Att. 6, 1, 4; and fig.: se scriptu, Piso ap. Gell. 6, 9, 4; cf.: "eo die (Antonius) se non modo consulatu, sed etiam libertate abdicavit," Cic. Phil. 3, 5, 12. — Absol.: augures rem ad senatum; "senatus, ut abdicarent consules: abdicaverunt," Cic. N. D. 2, 4, 11.—

b. With acc. a few times in the historians: "(patres) abdicare consulatum jubentes et deponere imperium," **Liv. 2**, **28** *fin.*: "abdicando dictaturam," **id. 6**, **18**, **4**.—In *pass*.: "abdicato magistratu," **Sall. C. 47**, **3**; cf.: "inter priorem dictaturam abdicatam novamque a Manlio initam," **Liv. 6**, **39**: "causa non abdicandae dictaturae," **id. 5**, **49** *fin.*

Language pivoting

We pivot on a high-resource language (contemporary English)





Credits: lacopo Ghinassi

Latin-English pivoting



Latin-English alignment

- Perseus project
- Automatic alignment with Giza++ (Och & Ney 2003)
- 123,000 sentence pairs
- 2.33 million Latin words



Manually annotated dataset

- 40 lemmas annotated in SemEval 2020
- LatinISE corpus (McGillivray & Kilgarriff 2013)
- lemmatization and PoS tagging of Perseus corpus with Cracovia lemmatizer and PoS tagger (Wróbel & Nowak 2022)
- Fine-tuned Latin BERT (Bamman & Burns 2020)

left context	target word	right context	the condition or pri vileges of a (Roman) citiz en, citizenship	the citizens united in a community	a city	Rome
 ex quo intellegi necesse est aut neminem ex sociis civem fieri posse aut etiam posse ex foederatis, tum vero ius omne noster iste magister mutandae civitatis ignorat, quod est, iudices, non solum in legibus publicis positum, sed etiam in privatorum voluntate. Iure enim nostro neque mutare civitatem quisquam invitus potest, neque si velit mutare non potest, modo adsciscatur ab ea civitate cuius esse se civitatis velit: ut, si Gaditani sciverint nominatim de aliquo cive Romano ut sit is civis Gaditanus, magna potestas sit nostro civi mutandae 	civitatis	, nec foedere impediatur quo minus ex civi Romano civis Gaditanus possit esse. Duarum civitatum civis noster esse iure civili nemo potest: non esse huius civitatis qui se alii civitati dicarit potest. Neque solum dicatione, quod in calamitate clarissimis viris Q. Maximo, C. Laenati, Q. Philippo Nuceriae, C. Catoni Tarracone, Q. Caepioni, P. Rutilio Zmyrnae vidimus accidisse, ut earum civitatum fierent cives,; cum; hanc ante amittere non potuissent quam hoc solum civitatis mutatione vertissent, sed etiam postliminio potest civitatis fieri mutatio. Neque enim sine causa	3: Closely Related	2: Distantly Related	4: Identica I	1: Unrela ted
 obseditque illam civitatem anno illo integro. Sed, Deo misericorditer favente et civibus viriliter se defendentibus, munitionem irrumpere non potuit. Eodem anno Aelfred, Angulsaxonum rex, post incendia urbium stragesque populorum, Lundoniam civitatem honorifice restauravit et habitabilem fecit; quam Aetheredo, Merciorum comiti, commendavit servandam. Ad quem regem omnes Angli et Saxones, qui prius ubique dispersi fuerant aut cum paganis sub captivitate erant, voluntarie converterunt, et suo dominio se subdiderunt. Anno Dominicae Incarnationis DCCCLXXXVII, nativitatis autem Aelfredi regis trigesimo 	civitatem	derelinquens incolumem, eo quod aliter proficere sibimet non poterat, classem suam sub illo ponte sursum contra Signe longe remigando, tam diu direxit, donec ad ostium fluminis, quod Materne nominatur, pervenisset; tunc Sigonam deserentes in ostium Materne divertunt, contra quod diu ac longe navigantes, demum non sine labore usque ad locum, qui dicitur Caziei, id est villa regia, pervenerunt. In quo loco hiemaverunt integro anno. Sequenti anno in ostium fluminis, quod dicitur lona, intraverunt, non sine magno regionis damno, et illic remorati sunt	1: Unrelated	3: Closely Related	4: Identica I	1: Unrela ted

Propagation

 w/inter : intersection of synsets for target Latin lemma and all synsets from the aligned English sentence.

2) w /align : use word alignment from Perseus



Credits: Iacopo Ghinassi

Training set

Dataset	SemEval	Pers
total data	2400	6020
total lemmas	40	33
total word senses	125	84
$A \setminus B$	52	11

Table 1: Statistics of the datasets used in our experiments. The last row shows the relative complement of the set of word senses of the dataset from the current column (A) and the set of word senses from the other dataset (B).



Results

Results obtained by fine-tuning LatinBERT on:

- 1) Expert annotated dataset (SemEval).
- 2) Pivoted annotations from Perseus using the intersection method.
- 3) Pivoted annotations from Perseus using the alignment method.
- 4) SemEval + Pivoted annotations from Perseus using the w/inter method.
- 5) SemEval + Pivoted annotations from Perseus using the w/align method.
- 6) SemEval + Pivoted annotations from Perseus using the w/inter method but limiting to senses which occur rarely in SemEval training test.

Training Data	F1	Precision	Recall
SemEval	60.30	55.63	67.15
Persinter	36.93	36.16	45.38
Persalign	36.93	36.21	45.78
SemEval+Persinter	61.45*	57.83	68.01
SemEval+Persalign	61.48*	57.46	68.00
SemEval+Persnere	61.83*	58.17	67.93

Conclusions

- As language is historically situated, making computational approaches more sensitive to the past should improve performance on semantic tasks relevant to cultural analysis and history
- Method applicable to other languages
- Not confined to historical dictionary data
- Lightweight but efficient tool for exploring the fine-grained semantics of historical texts

COALA:

COmputational Corpus Annotation for Quantitative Analysis of LAtin Lexical Semantics



Project's aims



Large-scale annotation of word meaning in context Quantitative analysis of meaning over two millennia

TRANSFORM HISTORICAL SEMANTICS

COALA: Computational Corpus Annotation for Quantitative Analysis of Latin Lexical Semantics

Methodology: team



Methodology: overview



COALA: Computational Corpus Annotation for Quantitative Analysis of Latin Lexical Semantics

Methodology: annotation



Outcome



COMPUTATIONAL Corpus2Sense



COALA: Computational Corpus Annotation for Quantitative Analysis of Latin Lexical Semantics

Impact and summary





Thank you!

Barbara McGillivray

Barbara.mcgillivray@kcl.ac.uk

© 2020 King's College London. All rights reserved