

LiLa: Linking Latin, June 03-04, 2019, Milano

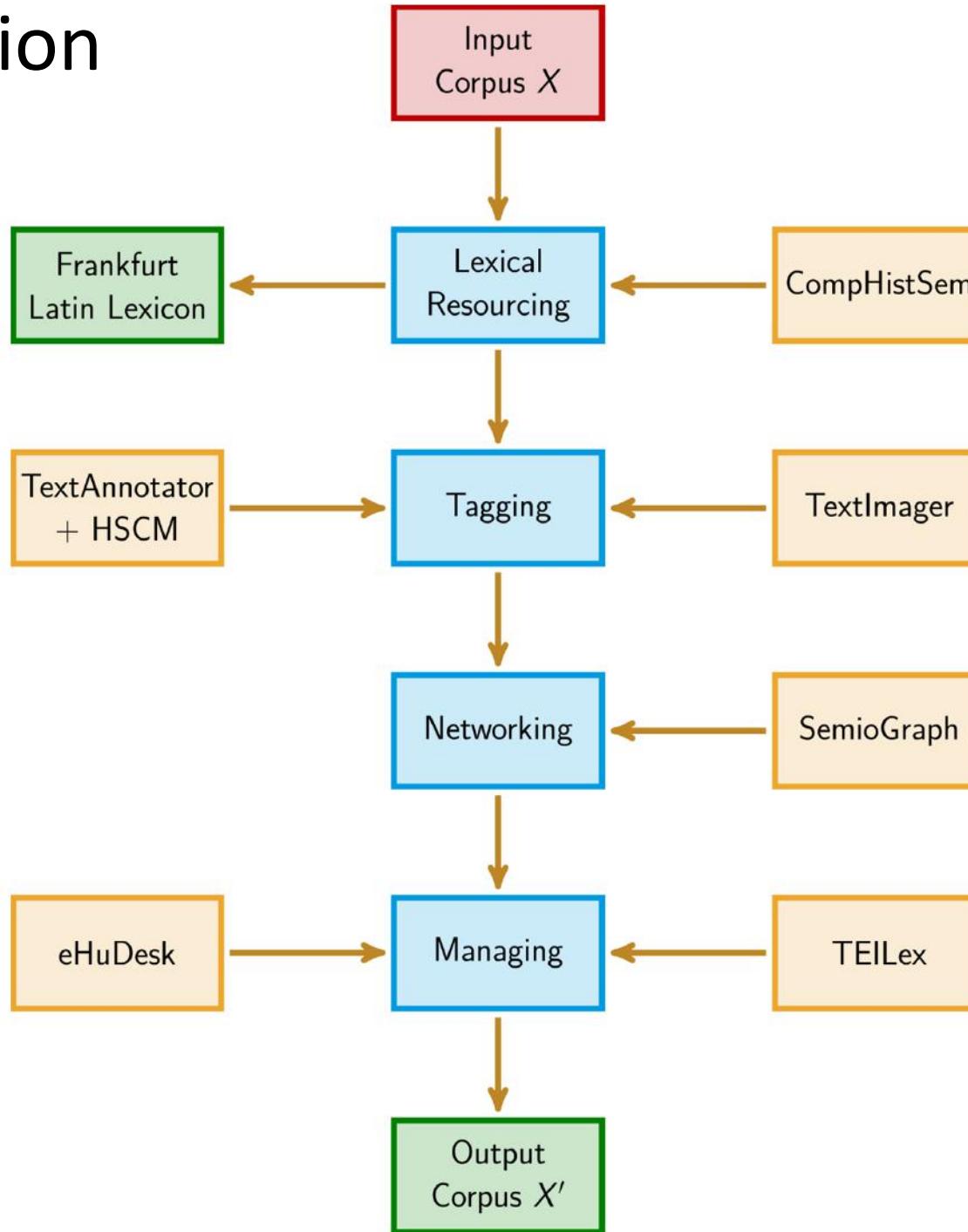
# Frankfurt Latin Lexicon: From Morphological Expansion to Latin Word Embeddings and Lexical Networks

Alexander Mehler, Tim Geelhaar  
Goethe-Universität Frankfurt



@Alexander Mehler, Tim Geelhaar, June 03, 2019, Milano

# Motivation



# Agenda

1. Resourcing

2. Tagging

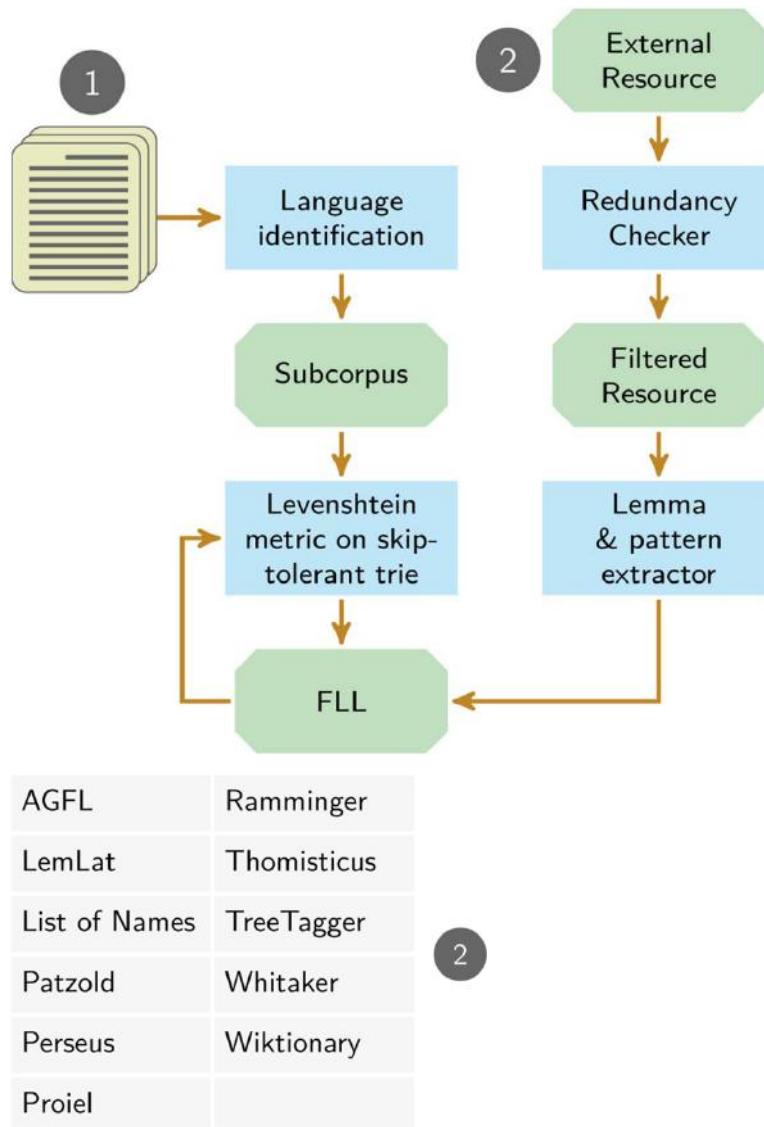
3. Networking

4. Managing

5. Conclusion

# Resourcing

## Frankfurt Latin Lexicon (FLL) (Mehler et al. 2015; v. d. Brück et al. 2016)



(by May 2019; under revision)

PoS	Superlemma	Lemma	SW	Description
ADJ	21870	26070	3337028	adjective
ADV	9682	11163	42864	adverb
AP	86	117	482	preposition
CON	101	140	519	conjunction
DIST	46	49	1321	distributive number
FM	76	109	2343	
ITJ	112	115	254	interjection
NE	5843	6649	114757	named entity
NN	35433	45383	745345	common noun
NP	26741	29657	247911	personal name
NUM	101	143	3140	number
ORD	131	194	4871	ordinal number
PRO	125	172	8041	pronoun
PTC	12	17	38	particle
V	9081	13164	5135824	verb
XY	114	117	718	
<b>Sum</b>	<b>109554</b>	<b>133259</b>	<b>9645456</b>	

Sources lexica: referenced per lemma + syntactic word

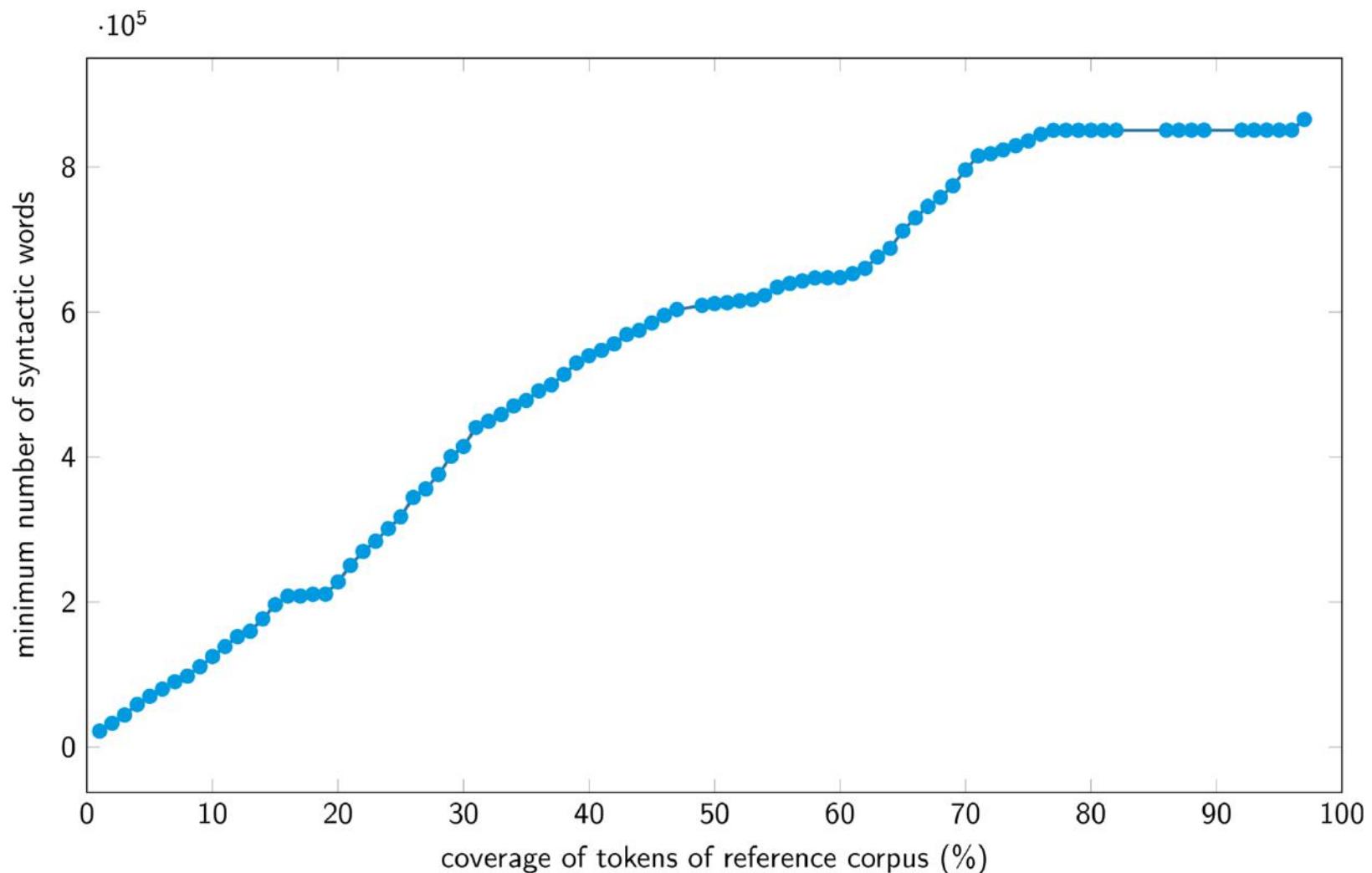
# Resourcing

Coverage:  
Overgeneration  
& Under-  
representation

Attribute	Value
All texts	111 515
All tokens	185 808 777
Tokens mapped to lexicon	180 535 369 97.16%
Tokens unassigned	5 273 408
All superlemmas	109 554
Superlemmas used	83 780 76.47%
All lemmas	133 259
Lemmas used	102 728 77.09%
All syntactic words	9 645 456
Syntactic words used	871 452 9.04%

# Resourcing

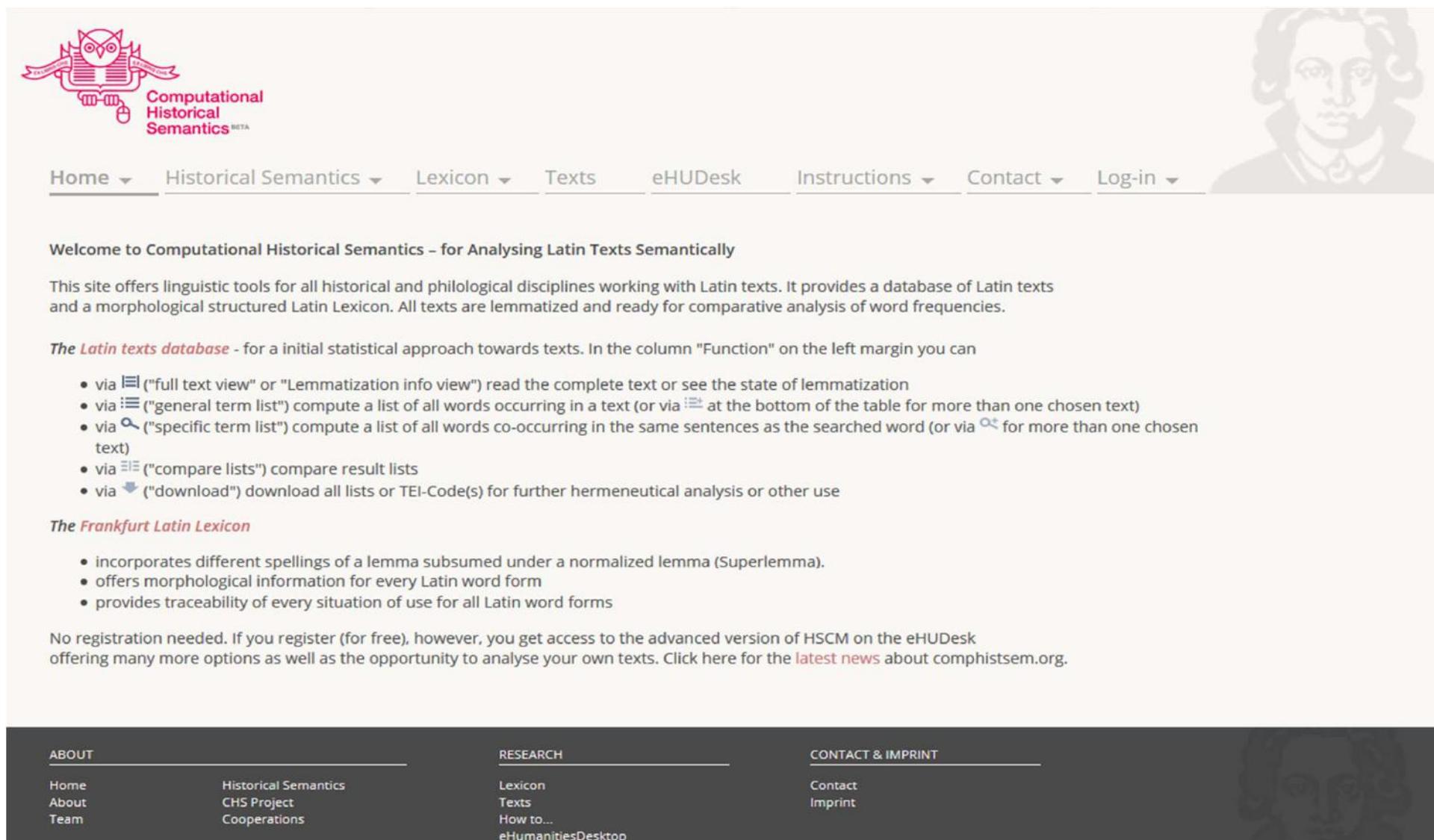
## Coverage: Overtgeneration & Underrepresentation



# Resourcing

## Interface: CompHistSem

<http://www.comphistsem.org/home.html>



The screenshot shows the homepage of the Computational Historical Semantics (CHS) website. At the top left is the CHS logo featuring a red owl with a ribbon. The top right features a large, faint watermark of a classical bust. The navigation bar includes links for Home, Historical Semantics, Lexicon, Texts, eHUDesk, Instructions, Contact, and Log-in. Below the navigation, a welcome message reads: "Welcome to Computational Historical Semantics – for Analysing Latin Texts Semantically". A text block explains the site's purpose: "This site offers linguistic tools for all historical and philological disciplines working with Latin texts. It provides a database of Latin texts and a morphologically structured Latin Lexicon. All texts are lemmatized and ready for comparative analysis of word frequencies." A section titled "The Latin texts database" provides instructions for using the text analysis features. Another section, "The Frankfurt Latin Lexicon", highlights its strengths. A note at the bottom states: "No registration needed. If you register (for free), however, you get access to the advanced version of HSCM on the eHUDesk offering many more options as well as the opportunity to analyse your own texts. Click here for the latest news about comphistsem.org." The footer contains links for About, Research, and Contact & Imprint.

**ABOUT**

[Home](#)  
[About](#)  
[Team](#)

**RESEARCH**

[Historical Semantics](#)  
[CHS Project](#)  
[Cooperations](#)

[Lexicon](#)  
[Texts](#)  
[How to...](#)  
[eHumanitiesDesktop](#)

**CONTACT & IMPRINT**

[Contact](#)  
[Imprint](#)

# Resourcing

## Interface: Latin Text Archive

<http://www.comphistsem.org/home.html>

IBETA VERSION. TESTING ENVIRONMENT. NOT FOR PRODUCTION USE]

Start About Documentation Providers Contact Imprint

L T A      Texts Lexicon Join in

MONUMENTA GERMANIAE HISTORICA  
berlin-brandenburgische AKADEMIE DER WISSENSCHAFTEN

## The Latin Text Archive

- Latin texts provided by leading institutions
- Lemmatized and ready for semantic analysis
- Quick and easy semantic analysis
- Use reference corpora or build your own corpora
- Enter texts via *Frankfurt Latin Lexicon*

search within LTA find

### Texts

Find texts and build your own corpus.

- List by author

### Lexicon

*Frankfurt Latin Lexicon*

- incorporates different spellings of a lemma subsumed under a

### Providers

Institutions and Series

MONUMENTA GERMANIAE HISTORICA

# Resourcing

## Interface: CompHistSem

<http://www.comphistsem.org/home.html>

The screenshot shows the CompHistSem interface with a navigation bar at the top featuring an owl logo, the text "Computational Historical Semantics BETA", and links for Home, Historical Semantics, Lexicon (selected), Texts, eHUDesk, Instructions, Contact, and Log-in. A portrait of a historical figure is visible on the right.

The main area displays a table titled "Query Results" with the following columns: Lemma, Description, Pos, Super lemma, Corpus freq., and Report Error. The table lists various forms of the word "ecclesia" along with their corresponding details. The "Report Error" column contains small red circular icons with white numbers.

Lemma	Description	Pos	Super lemma	Corpus freq.	Report Error
aaeclesia	aaeclesia	Noun	ecclesia@NN	3	0
acclesia	acclesia	Noun	ecclesia@NN	66	0
aclesia	aclesia	Noun	ecclesia@NN	4	0
aeclæsia	aeclæsia	Noun	ecclesia@NN	37	0
aecclesia	aecclesia	Noun	ecclesia@NN	12192	0
aeclæsia	aeclæsia	Noun	ecclesia@NN	8	0
aeclæsia	aeclæsia	Noun	ecclesia@NN	4	0
aeclesia	aeclesia	Noun	ecclesia@NN	899	0
aeclisia	aeclisia	Noun	ecclesia@NN	11	0
aëcclesia	aëcclesia	Noun	ecclesia@NN	6	0
ecclesia	ecclesia	Noun	ecclesia@NN	22	0
ecclæsia	ecclæsia	Noun	ecclesia@NN	88	0
ecclecia	ecclecia	Noun	ecclesia@NN	6	0
ecclesia	ecclesia	Noun	ecclesia@NN	540091	0
ecclesiae	ecclesiae	Noun	ecclesia@NN	1	0
ecclessia	ecclessia	Noun	ecclesia@NN	11	0
ecclisia	ecclisia	Noun	ecclesia@NN	16	0
ecclesia	ecclesia	Noun	ecclesia@NN	1	0

# Resourcing

## Interface: CompHistSem

<http://www.comphistsem.org/home.html>

The screenshot shows the CompHistSem interface with the following elements:

- Logo:** An owl icon with a ribbon.
- Header:** Computational Historical Semantics (BETA).
- Navigation:** Home, Historical Semantics, Lexicon (selected), Texts, eHUDesk, Instructions, Contact, Log-in.
- Search Fields:** SuperLemma (ecclesia), Lemma (empty), Word Form (empty).
- Search Buttons:** Search for SuperLemma, Lemma, and Word Form.
- Query Results:** A table titled "Query Results" showing results for ecclesia\*(SuperLemma). The table has columns: Super lemma, Lemma, Word form, Pos, Corpus freq., Casus, Genus, Numerus, Person, Tense, Mood, Voice, Is MWU, Lemma sot, Superle, Report Err.
- Data:** The table lists numerous entries for various forms of "ecclesia" (e.g., ecclesia@NN, aecclesia, aecclesiae, aecclesiaeque) across different grammatical categories (Noun, Verb) and cases (Nom, Gen, Dat, Acc, Voc).

# Agenda

1. Resourcing

2. Tagging

3. Networking

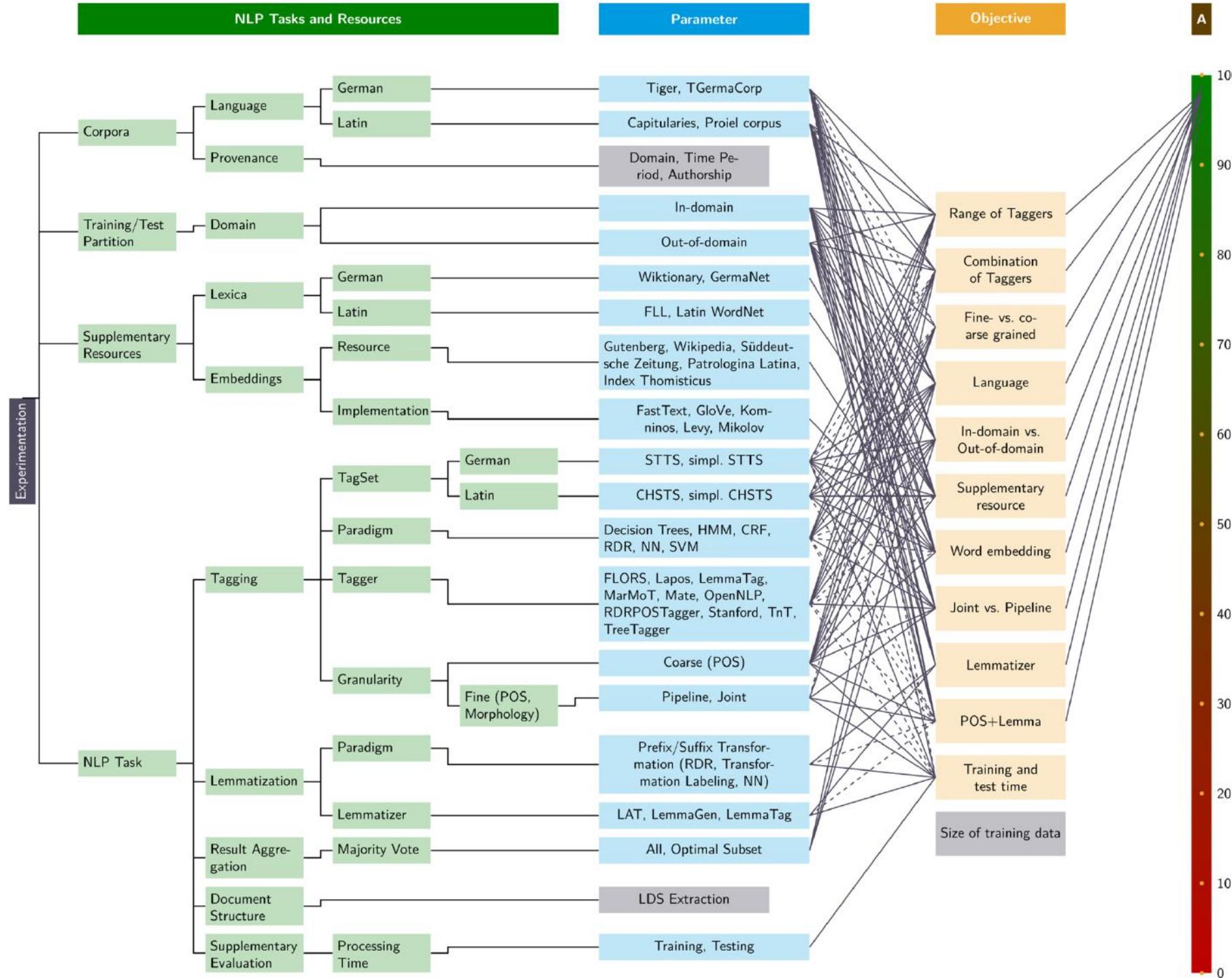
4. Managing

5. Conclusion

# Tagging

Task (Eger et al. 2016; Gleim et al. 2019)

Sub.-cats.	[gender=f case=abl number=pl]	[gender=f case=gen number=sg]	[gender=n case=nom number=sg]	[gender=n case=nom number=sg]	[number=sg person=3 mood=ind voice=pass tense=pres]	
PoS	AP	NN	NN	NN	PRO	V
Lemma	ex	frux	terra	corpus	noster	sustento
Form	Ex	frugibus	terrae	corpus	nostrum	susten- tatur
(1) Coarse-grained tagging						
(2) Fine-grained tagging						



# Experimentation

## POS Accuracies (%) of Using MarMoT, Lexica & Embeddings (Gleim et al. 2019)

corpora	lexicon	tagset	no em.	PL skip	PL cbow
Capit	none	CHSTS	96,10	96,25	95,05
Capit	FLL	CHSTS	96,52	<b>96,57</b>	95,52
Capit	none	sCHSTS	96,35	96,46	95,42
Capit	FLL	sCHSTS	96,67	<b>96,72</b>	95,77
Capit	La-WN	sCHSTS	96,37	96,47	95,40
Capit	all	sCHSTS	96,68	96,72	95,73
Cp→Pro	none	CHSTS	87,58	88,71	85,00
Cp→Pro	FLL	CHSTS	88,59	<b>88,90</b>	85,92
Pro→Cp	none	CHSTS	63,05	68,81	67,86
Pro→Cp	FLL	CHSTS	69,88	<b>72,26</b>	70,34

- Including lexical resources to train MarMoT consistently improves accuracies in ID as well as OD experiments.
- Particularly, for OD scenarios including embeddings has positive results.

# Tagging

Availability: TextImager

<https://textimager.hucompute.org/>

 TextImager

Legal Notice Cite Info Register Login

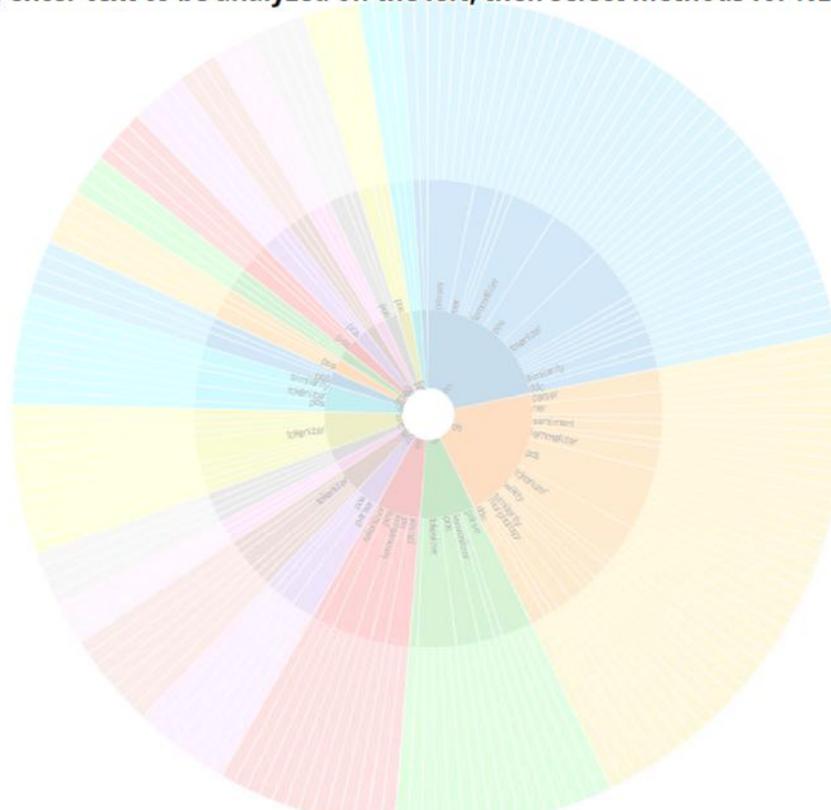
**Input**

Enter text to be classified. You can also up load a file (.txt, .tei) directly to classify it. This file is only stored temporarily for the purpose of classification- it will not be archived.

+ Unnamed Text X

**NLP Selector**

First, enter text to be analyzed on the left, then select methods for NLP here.



# Tagging

## Creating Training Data: Requirements Analysis (Abrami et al. 2019)

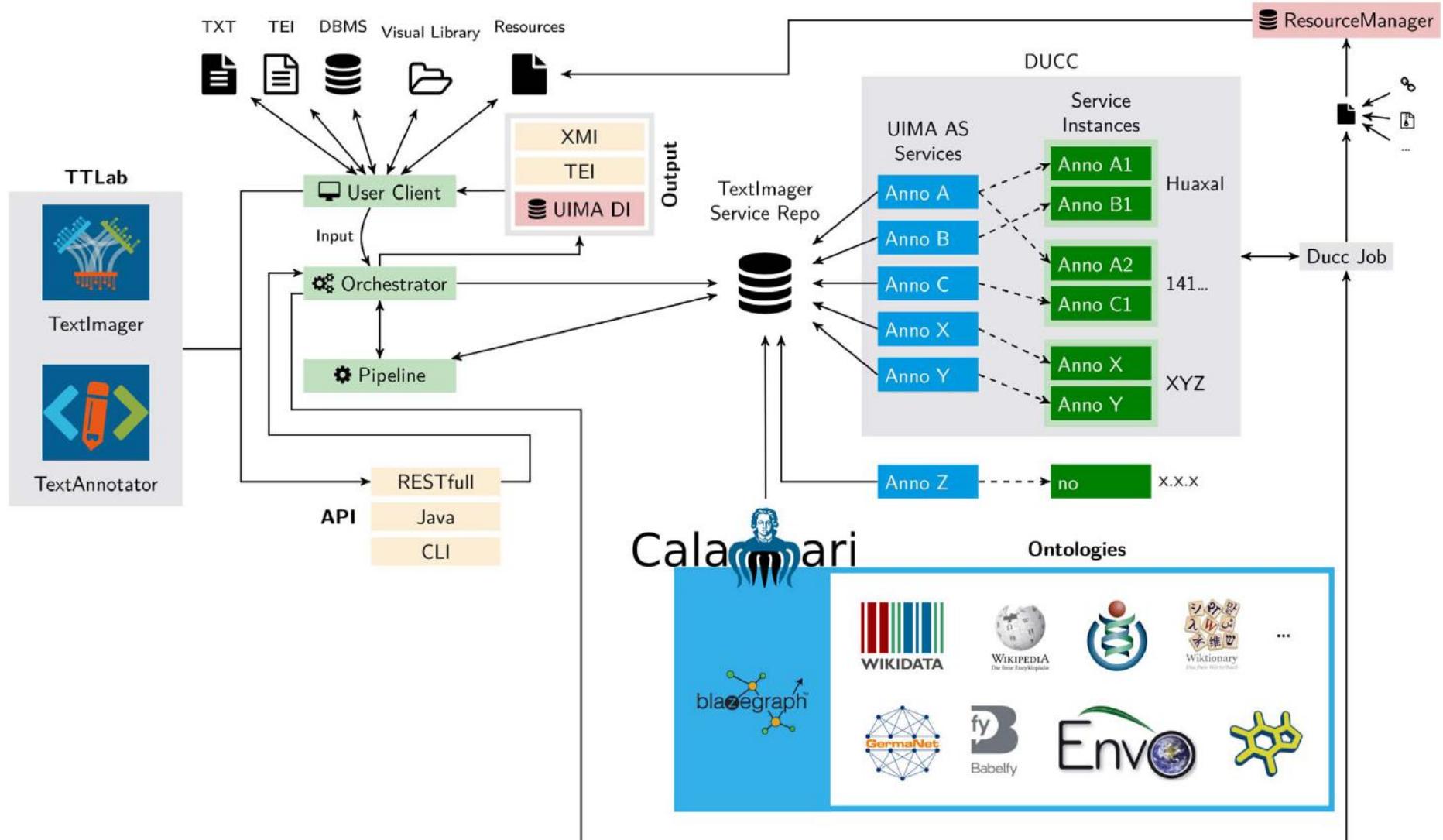
- 1 Addressing a variety of annotation tasks within the same framework
- 2 Employing automatic processing as much as possible
- 3 Allowing for collaborative annotation using a browser-based interface

→ [TextAnnotator](#)

# Tagging

<http://www.textannotator.hucompute.org/>

## TextAnnotator: Architecture (Abrami et al. 2019)



# Tagging

<http://www.textannotator.hucompute.org/>

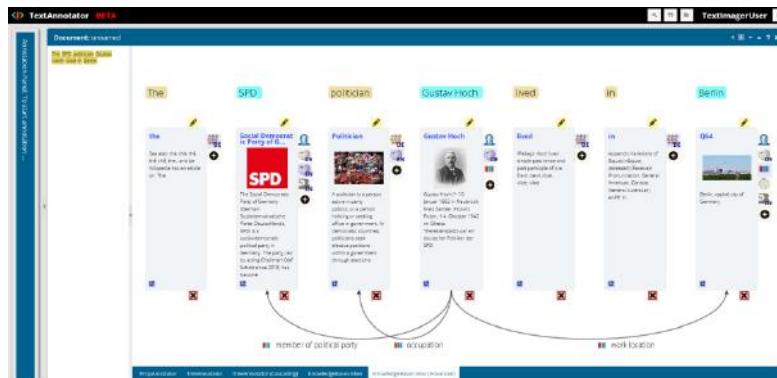
## TextAnnotator: Modules

(Abrami et al. 2018, 2019; Helfrich et al. 2018)

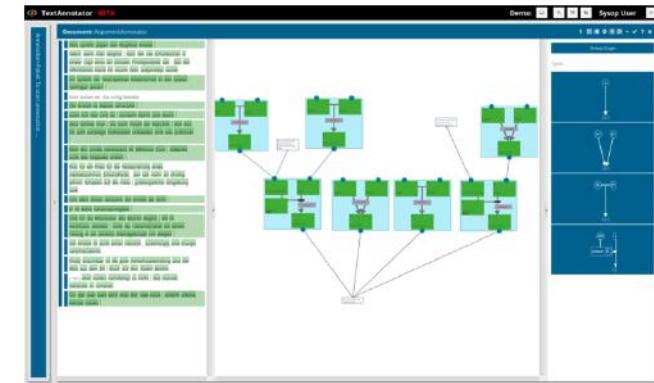
### QuickAnnotator



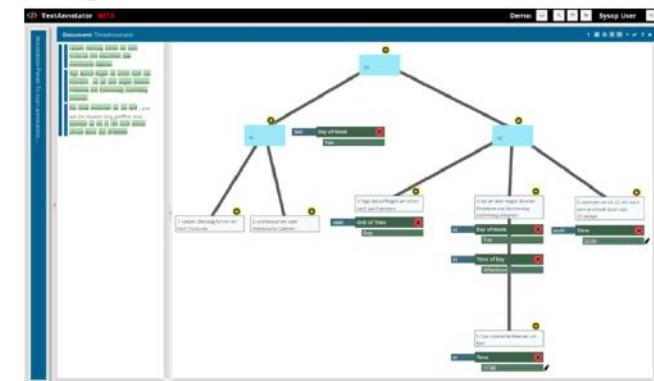
### KnowledgeBaseLinker



### TimeAnnotator



### ArgumentAnnotator



### Additional Annotators:

TreeAnnotator: RST

PropositionAnnotator

# Agenda

1. Resourcing

2. Tagging

3. Networking

4. Managing

5. Conclusion

# Networking

## Confusion Matrix

1. word2vec:  
[Mikolov \(2013\)](#)
2. Levy &  
Goldberg  
(2014)
3. Komninos &  
Manandhar  
(2016)
4. wang2vec:  
[Ling et al.  
\(2015\)](#)
5. ELMo:  
[Peters et al.  
\(2018\)](#)
6. BERT:  
[Devlin et al.  
\(2018\)](#)
7. ...

	Mikolovian Embeddings				#Texts	#Tokens
	WF	L	L+P	L+D		
Overall Corpus	✗	✗	✗	✗	111 515	185 808 777
Training Corpus	✓	✓	✓	✗	33 791	61 451 677
Epistolographical Texts	✓	✓	✓	✗	844	16 406 556
Legal Texts	✓	✓	✓	✗	31 461	12 097 990
Liturgical Texts	✓	✓	✓	✗	252	2 667 784
Narrative Texts	✓	✓	✓	✗	663	7 635 906
Philosophical Texts	✓	✓	✓	✗	10	243 386
Poetical Texts	✓	✓	✓	✗	1	4 018
Political Texts	✓	✓	✓	✗	31	3 197 879
Theological Texts	✓	✓	✓	✗	494	18 305 475
👤 Bernard of Clairvaux	✓	✓	✓	✗	56	773 136
👤 John of Salisbury	✓	✓	✓	✗	4	380 534
👤 William of Ockham	✓	✓	✓	✗	1	437 242
👤 ...	...	...	...	...	...	...

# Networking

## Multi-codal Graphs

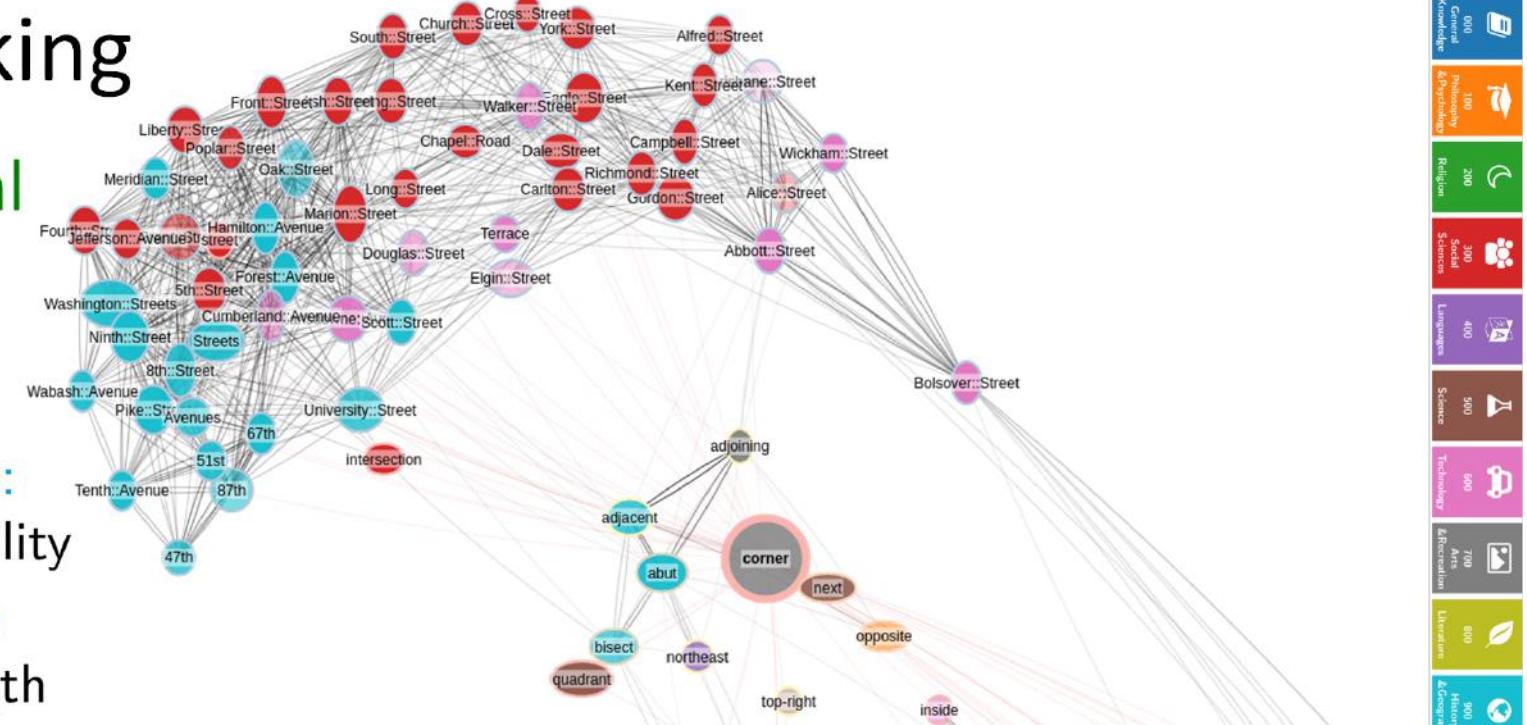
### 1. Node size:

#### (a) height:

centrality

#### (b) width:

strength



### 2. Node transparency:

class membership value

### 3. Node color:

class

### 4. Border color (size):

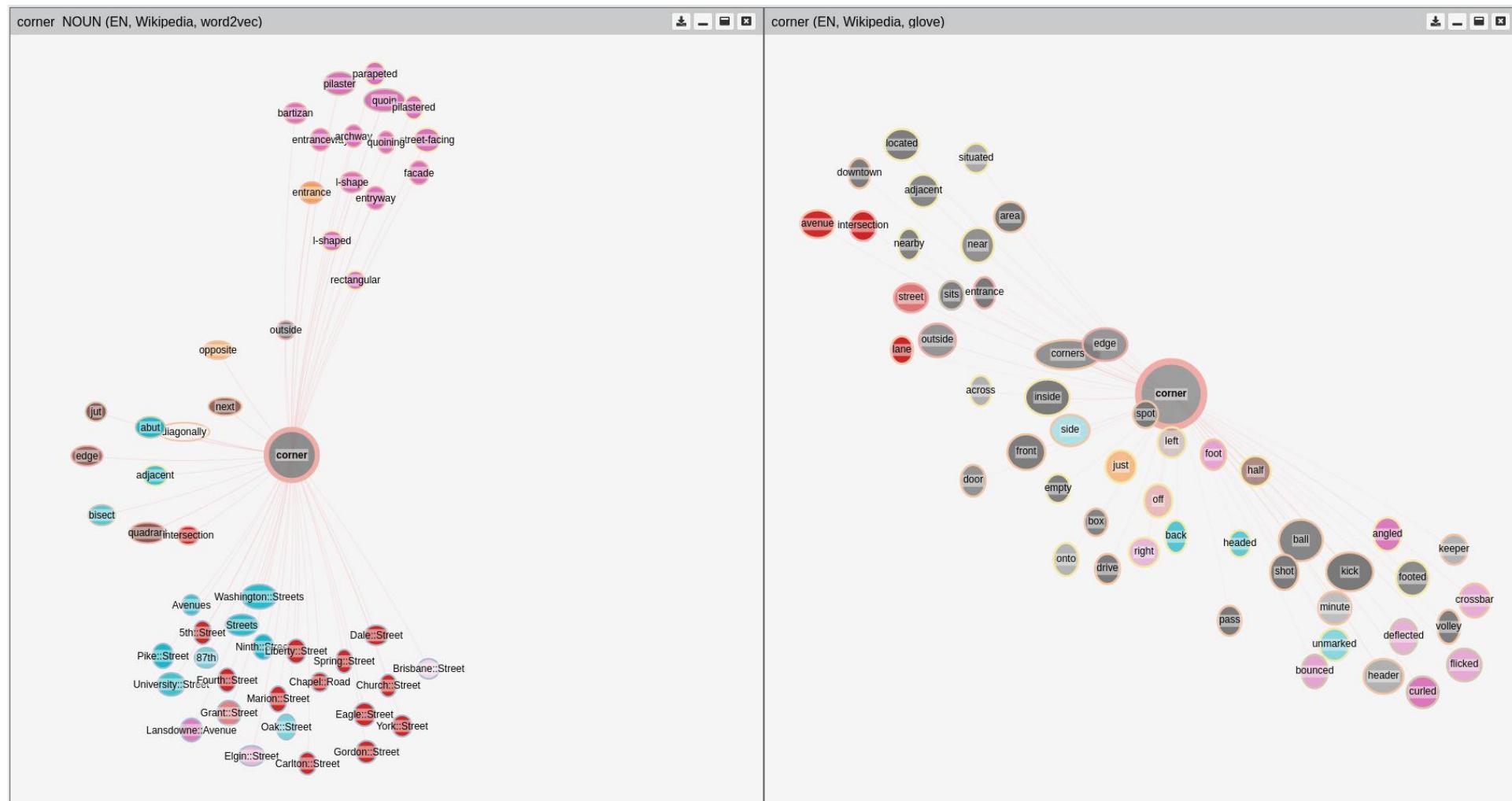
POS (seed node)

### 5. Node tiling (pie chart): coding of different class memberships

# Networking



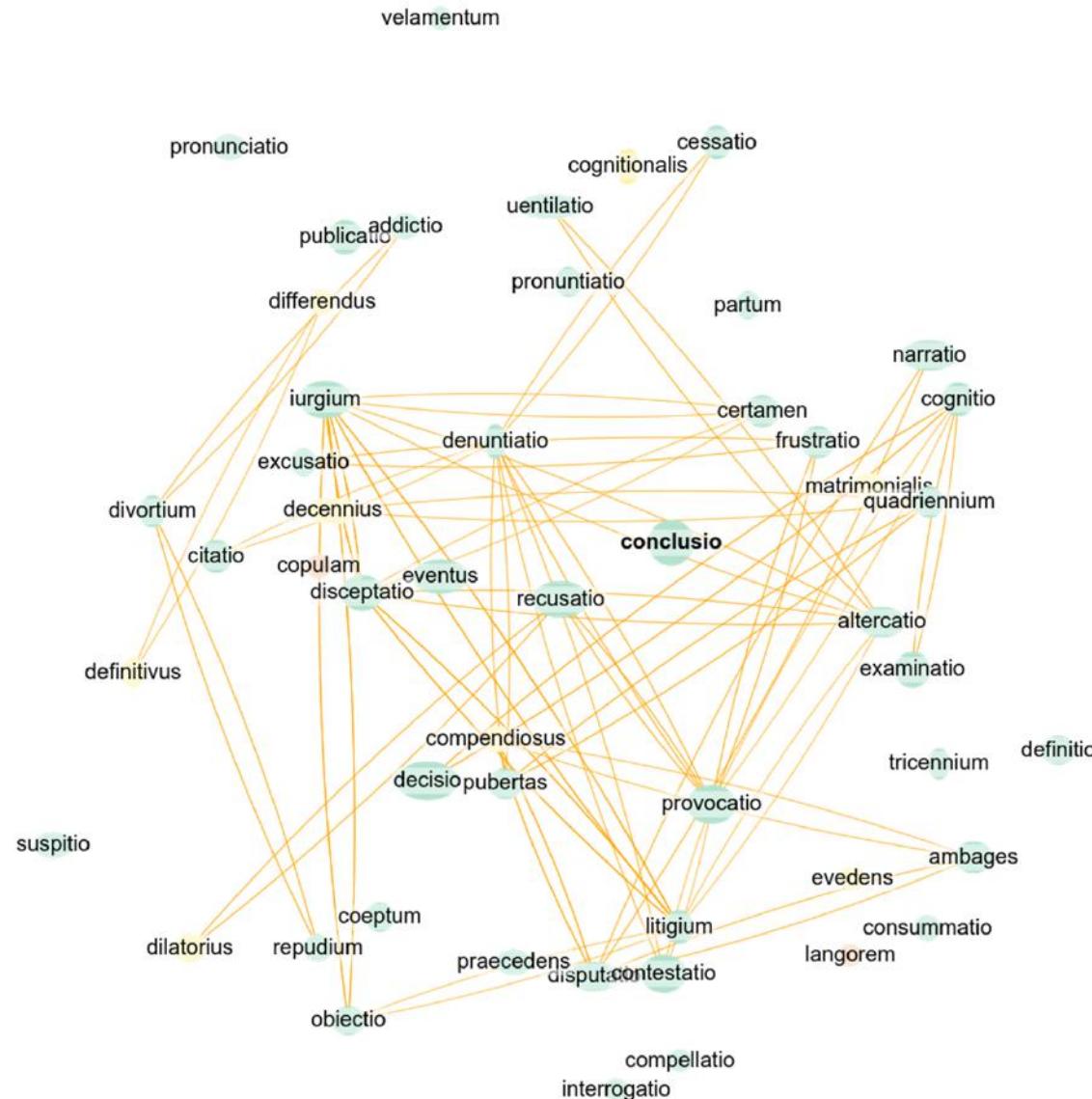
## Multi-codal Graphs



# Networking

<http://semiograph.texttechnologylab.org/>

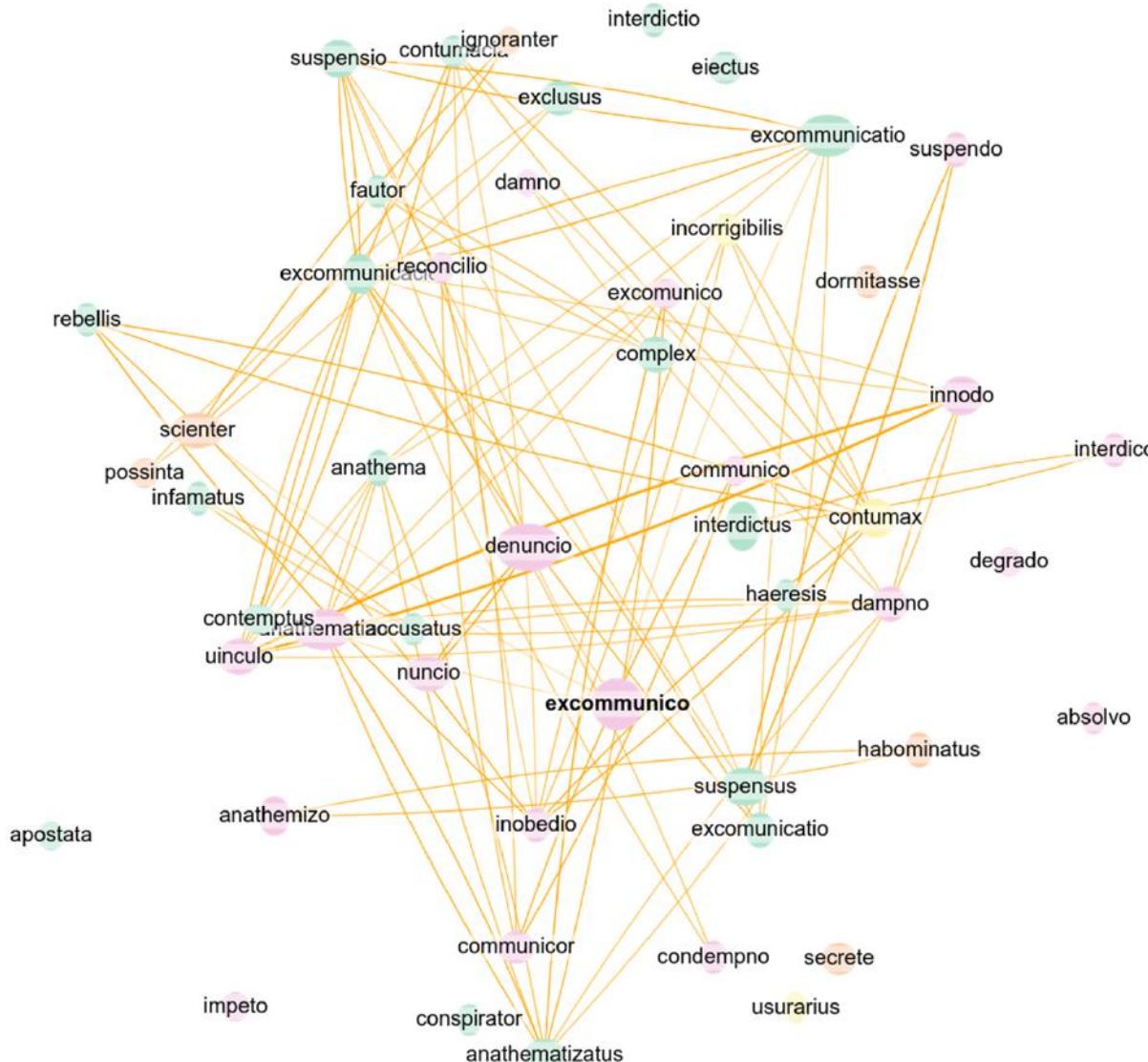
## SemioGraph: Example 1: *conclusio* in Legal Texts



# Networking

<http://semiograph.texttechnologylab.org/>

## SemioGraph: Example 1: *excommunico* in Legal Texts



# Agenda

1. Resourcing
2. Tagging
3. Networking
4. Managing
5. Conclusion

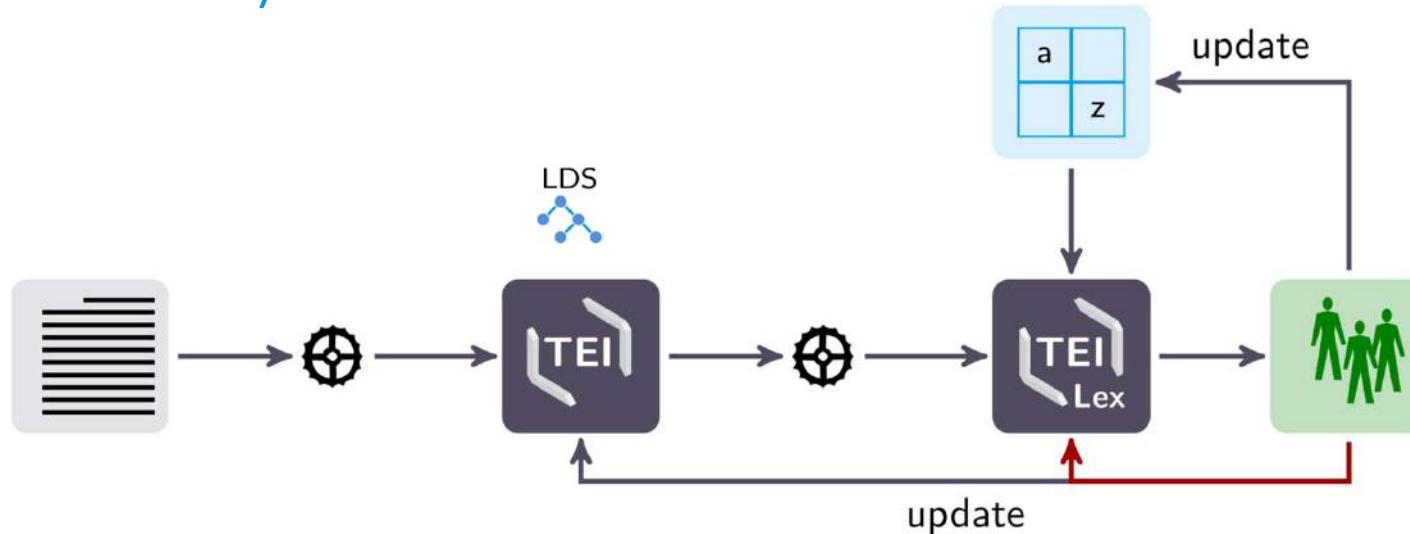


TEILex

# Managing

Used by: <https://hudesktop.hucompute.org/>

## Dynamic Corpus & Lexicon Update with TEILex (Gleim et al. 2012)



1. Automatic text processing (lemmatization, tagging, ...)
2. Automatic indexing (TEILex)
3. Human computation (lexicon update)
  - Corpus-related changes are automatically carried out (adjustable)
  - TEILex index is automatically revised (adjustable)



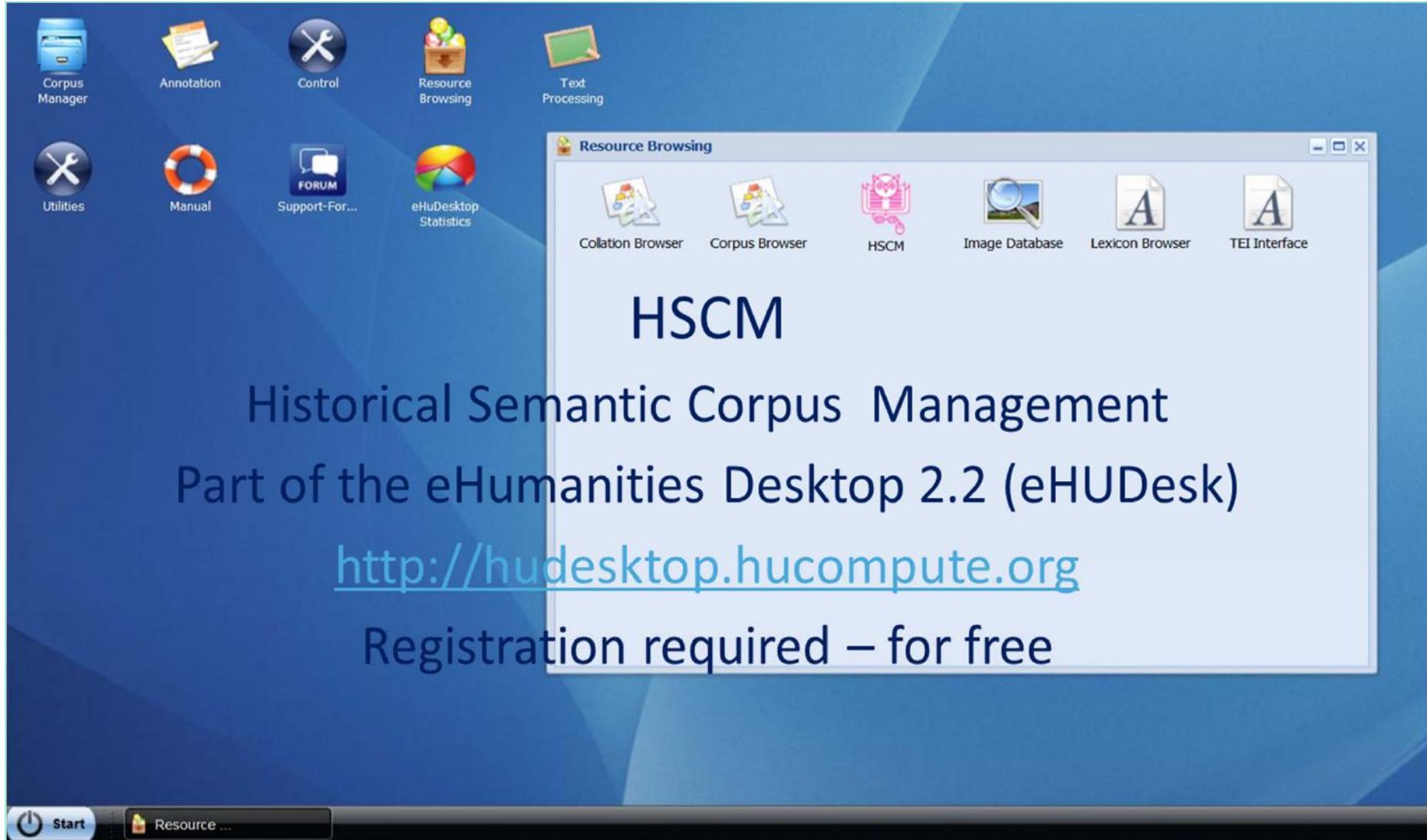
All lemmas	User-created
133 037	18 564
100%	13,93%

(May 2019)

# Managing

<https://hudesktop.hucompute.org/>

## eHumanities Desktop (eHuDesk) & HSCM



# Managing

<https://hudesktop.hucompute.org/>

## Lexicon Browser

The screenshot shows the Lexicon Browser interface with the following components:

- Query Bar:** Active Lexicon: Latin, Token: obtineo, Token Type: Lemma.
- Query Result (Lemma) Tab:** Shows a table of word forms and their descriptions. One row is highlighted with a red box, showing "obtineo@V".
- Lemmas Editor Tab:** Shows a table of lemmas with columns: Lemma, Description, and Pos. The first row is highlighted with a red box, showing "obtineo".
- Forms Editor Tab:** Shows a table of word forms with detailed grammatical features like Casus, Genus, Mood, Tense, Voice, etc. The first row is highlighted with a red box, showing "obtineo".

Red boxes highlight the rows in the main table, the lemma table, and the forms table, while arrows point from the highlighted rows in the main table to the lemma table, and from the lemma table to the forms table, illustrating the hierarchical nature of the data.

# Managing

<https://hudesktop.hucompute.org/>

## HSCM – Main page

HSCM - Historical Semantics Corpus Management

Corpus Selection/Search Results Result Comparison Document Statistics Term-Document Matrix NGrams Co-occurrence Graph

Step 1 - Available Repositories

Selected: 64

Step 2 - Subcorpus Selection

Page 1 of 7 Max Page Size: 50 Set for... Save Col. Config. Clear filters Selected: 1 Displaying 1 - 50 of 348

Document	LT...	Author (Latin)	Author (Germ.)	C1	C2	C Stat	Title	Time	Year of...	Text Type	Text Type Class	Edition	Volume	Column	Bibl...	Lic...	Provided by	URL	Status	Num. of words
4294969...	true	Einhardus	Einhard	09	08	09	Vita Karoli Imperatoris	814-830	0830	Biography; Historiography	Biographical; Historiography	MGH Scriptores rerum Germanicarum 25	SS. rer. Germ. 25	1	<a href="#">Link</a>	CC BY 4.0	Monumenta Germaniae Historica	<a href="#">Link</a>	c	7174
Antapodis...	true	Liutprandus Cremonensis	Liutprand von Cremona	10	10	10	Antapodis...	Antapodis...	Antapodis...	Full Text View	Lemmatization Editor	MGH Scriptores rerum Germanicarum 41	SS. rer. Germ. 41	1	<a href="#">Link</a>	CC BY 4.0	Monumenta Germaniae Historica	<a href="#">Link</a>	a	35879
bsb000007...	false	Rimbertus archiepiscopus Hammaburgensis et Bremensis	Rimbert von Hamburg-Bremen	09	09	09	Vita Rimberti	886 post	0888	Hagiography	Biographical; Historiography	MGH Scriptores rerum Germanicarum 55	SS. rer. Germ. 55	13	<a href="#">Link</a>	CC BY 4.0	Monumenta Germaniae Historica	<a href="#">Link</a>	a	1024
bsb000007...	false	Rimbertus archiepiscopus Hammaburgensis et Bremensis	Rimbert von Hamburg-Bremen	09	09	09	Vita Rimberti	886 post	0888	Hagiography	Biographical; Historiography	MGH Scriptores rerum Germanicarum 55	SS. rer. Germ. 55	18	<a href="#">Link</a>	CC BY 4.0	Monumenta Germaniae Historica	<a href="#">Link</a>	a	18963
bsb000007...	true	Anonymous	Anonymous	10	09	10	Vita Rimberti	886 post	0888	Hagiography	Biographical; Historiography	MGH Scriptores rerum Germanicarum 55	SS. rer. Germ. 55	80	<a href="#">Link</a>	CC BY 4.0	Monumenta Germaniae Historica	<a href="#">Link</a>	a	6420
bsb000007...	true	Rimbertus archiepiscopus Hammaburgensis et Bremensis	Rimbert von Hamburg-Bremen	09	09	09	Vita Anscharii	865-876	0876	Hagiography	Biographical; Historiography	MGH Scriptores rerum Germanicarum 55	SS. rer. Germ. 55	13	<a href="#">Link</a>	CC BY 4.0	Monumenta Germaniae Historica	<a href="#">Link</a>	a	19987
bsb000007...	true	Norbertus abbas Iburgensis	Norbert von Iburg	11	12	11	Vita Bonnensis II. episcopi Osnabrugensis	1090-1100	1100	Hagiography	Biographical; Historiography	MGH Scriptores rerum Germanicarum 56	SS. rer. Germ. 56	1	<a href="#">Link</a>	CC BY 4.0	Monumenta Germaniae Historica	<a href="#">Link</a>	a	10776
bsb000007...	true	Willibaldus presbyter Moguntinus	Willibald von Mainz	08	08	08	Vita Bonifatii	754-769	0769	Hagiography	Biographical; Historiography	MGH Scriptores rerum Germanicarum 57	SS. rer. Germ. 57	1	<a href="#">Link</a>	CC BY 4.0	Monumenta Germaniae Historica	<a href="#">Link</a>	a	9230
bsb000007...	true	Auctor incertus	Unbekannter Autor	10	10	10	E Martyrologio Fuldense	900(?)	0900	Hagiography; Martyrologia	Biographical; Historiography	MGH Scriptores rerum Germanicarum 57	SS. rer. Germ. 57	59	<a href="#">Link</a>	CC BY 4.0	Monumenta Germaniae Historica	<a href="#">Link</a>	a	321
bsb000007...	true	Radbodus	Radbod	10	09	10	Vita altera	899	0917	Hagiography	Biographical	MGH Scriptores rerum Germanicarum 62	SS. rer. Germ. 62	62	<a href="#">Link</a>	CC BY	Monumenta Germaniae Historica	<a href="#">Link</a>	a	3867

Reset Reload Select Subcorpus Selection... Download TEI... Corpus... Compute... Go to query

Step 3 - Query

Process History

# Managing

<https://hudesktop.hucompute.org/>

## Lemmatization Editor

### Lemmatization Statistics

Unique Wordform	14134	34,29%
Unique Wordform. Tagger based result. Multiple Lemmata possible	283	0,69%
Unique Wordform. Tagger based result. Multiple Super lemmata possible	7276	17,65%
Multiple Wordforms under a unique Lemma	8635	20,95%
Multiple Wordforms. Tagger based result. Multiple Lemmata possible	214	0,52%
Multiple Wordforms. Tagger based result. Multiple Super Lemmata possible	7933	19,25%
Multiple Lemmata under a unique Super lemma	63	0,15%
Multiple Super lemmata	1434	3,48%
Not yet linked or unknown Wordform	1248	3,03%
Word count	41220.0	

### Table of Contents.

#### ► Details

- [1.1.0. De summa trinitate et de fide...](#) 1.1.0. De summa trinitate et de fide catholica et ut nemo de ea publice contendere audeat.  
1.1.1  
1.1.2 Imperatores Gratianus, Valentinianus, Theodosius .  
1.1.3 Cunctos populos, quos clementiae nostrae regit temperamentum, in tali volumus religione versari, quam divinum Petrum  
1.1.4 apostulum tradidisse Romanis religio usque ad nunc ab ipso insinuata declarat quamque pontificem Damasum sequi claret et  
1.1.5 Petrum Alexandriae episcopum virum apostolicae sanctitatis, hoc est ut secundum apostolicam disciplinam evangelicamque  
1.1.6 doctrinam patris et filii et spiritus sancti unam deitatem sub pari maiestate et sub pia trinitate credamus.  
1.1.7 1. Hanc legem sequentes christianorum catholicorum nomen iubemus amplecti, reliquos vero dementes vesanosque iudicantes  
1.1.8 haeretici dogmatis infamiam sustinere, divina primum vindicta, post etiam motus nostri, quem ex caelesti arbitrio  
1.2.0. De sacrosanctis ecclesiis et... sumpserimus, ultione plectendos. \*
- 1.2.1  
1.2.2  
1.2.3 GRAT. VALENTIN. ET THEODOS. AAA. AD POP. URB. CONSTANTINOP. \*(A 380 D. III K. MART. THESSALONICA  
1.2.4 GRATIANO V ET THEODOSIO AA. CONSS.)  
1.2.5 1.1.2  
1.2.6 Imperatores Gratianus, Valentinianus, Theodosius .

# Managing

<https://hudesktop.hucompute.org/>

## Lemmatization Editor and W-Link Editor

**Lemmatization Statistics**

Unique Wordform | 39292 | 32.20%

**W-Link Editor**

Link Editor for *populus* | Link Editor for *populus* | **Link Editor for *populus***

Page 1 of 1 | Max. Page Size: 50 | Clear filters | Save Col. Config. | Selected: 0 | Displaying Count of Possible Connections 1 - 4 of 4

Word f...	Pos	Casus	Genus	Numerus	Co...	Co...	Declen...	Mood	Pe...	Pr...	Tense	Verb ...	Voice	Lemma	Super lem...	Dis...	Lemma descn...
<input type="checkbox"/> <i>populus</i>	Noun	Ablative	Feminine	Plural										populus	populus@NN	red	populus, poplar
<input type="checkbox"/> <i>populus</i>	Noun	Dative	Feminine	Plural										populus	populus@NN	red	populus, poplar
<input type="checkbox"/> <i>populus</i>	Noun	Ablative	Masculi...	Plural										populus	populus@NN	red	populus, people
<input type="checkbox"/> <i>populus</i>	Noun	Dative	Masculi...	Plural										populus	populus@NN	red	populus, people

Sentence:  
Decedente atque immo potius pereunte ab urbibus Gallicanis liberalium cultura litterarum, cum nonnullae res gererentur vel rectae vel inprobae, ac feretas gentium desaeviret, regum furor acueretur, eclesiae inpugnarentur ab hereticis, a catholicis tegerentur, ferveret Christi fides in plurimis, tepisceret in nonnullis, ipsae quoque eclesiae vel ditarentur a devotis vel nudarentur a perfides, nec repperire possit quisquam peritus dialectica in arte grammaticus, qui haec aut stilo prosaico aut metrico depingeret versu: ingemescebant saepius plerique, dicentes: 'Vae diebus nostris, quia perit studium litterarum a nobis, nec reperitur rethor in *populis*, qui gesta praesentia promulgare possit in paginis'.

(4 Items)

Selection... | Confirm... | Delete... | ReLink...

Next blue | Next red | Next purple | Next yellow | Next yellow-purple | Next yellow-red | Next green-purple | Next green-red | Close all Tabs

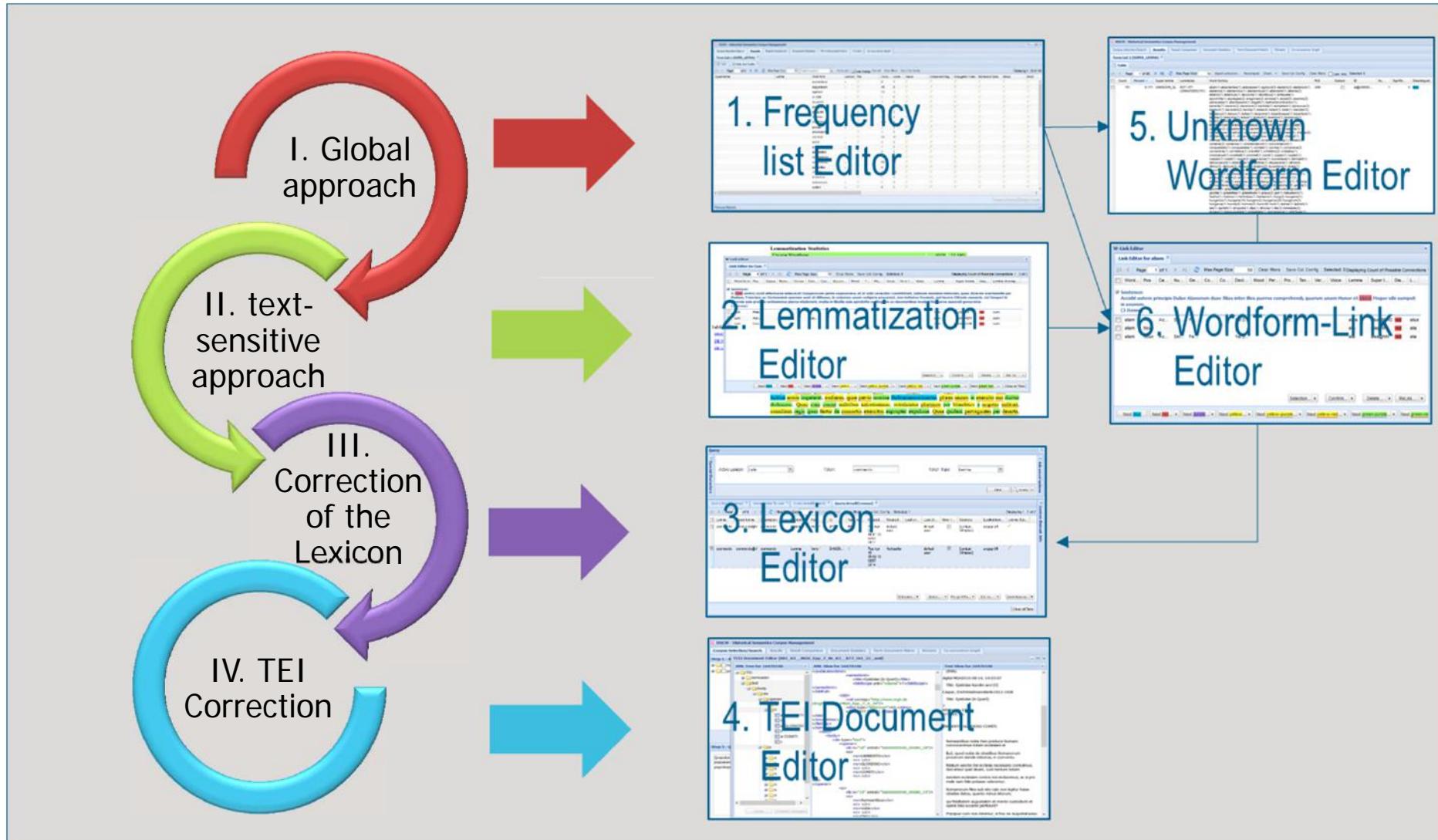
2. EXHIC CONCUM genus in facinus ...  
3. Decedente atque immo potius pereunte ab urbibus Gallicanis liberalium cultura  
litterarum, cum nonnullae res gererentur vel rectae vel inprobae, ac feretas gentium  
desaeviret, regum furor acueretur, eclesiae inpugnarentur ab hereticis, a catholicis  
tegerentur, ferveret Christi fides in plurimis, tepisceret in nonnullis, ipsae quoque  
eclesiae vel ditarentur a devotis vel nudarentur a perfides, nec repperire possit  
quisquam peritus dialectica in arte grammaticus, qui haec aut stilo prosaico aut metrico  
depingeret versu: ingemescebant saepius plerique, dicentes: 'Vae diebus nostris, quia  
perit studium litterarum a nobis, nec reperitur rethor in *populis*, qui gesta praesentia  
promulgare possit in paginis'.

4. Dominus ergo commotus contra  
iniquitates...  
5. Habebat ergo Noe post diluvium tres...  
6. Cumque multiplicati hominis  
dispergerentur...  
7. Primus autem filius Noe Sem; de ...  
8. His cum centum esset annorum,

# Managing

<https://hudesktop.hucompute.org/>

## Post-Lemmatization process



# Agenda

1. Resourcing
2. Tagging
3. Networking
4. Managing
5. Conclusion

# Conclusion

- 1 Frankfurt Latin Lexicon
- 2 Frankfurt Latin Tagger
- 3 SemioGraph: Latin Word Embeddings
- 4 Tools: TextImager, eHuDesk, CompHistSem

# Thank you!



# References

- Giuseppe Abrami and Alexander Mehler. A UIMA database interface for managing NLP-related text annotations. In Proceedings of the 11th edition of the Language Resources and Evaluation Conference, May 7 - 12, LREC 2018, Miyazaki, Japan, 2018. accepted.
- Giuseppe Abrami, Alexander Mehler, Andy Lücking, Elias Rieb, and Philipp Helfrich. TextAnnotator: A flexible framework for semantic annotations. In Proceedings of the Fifteenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation, (ISA-15), ISA-15, May 2019. accepted.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- Steffen Eger, Rüdiger Gleim, and Alexander Mehler. Lemmatization and morphological tagging in German and Latin: A comparison and a survey of the state-of-the-art. In Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016, 2016.
- Rüdiger Gleim, Steffen Eger, Tolga Uslu, Wahed Hemati, Andy Lücking, Alexander Henlein, Sven Kahlsdorf, and Armin Hoenen. Practitioner's view: A comparison and a survey of lemmatization and morphological tagging in German and Latin. Journal of Language Modelling, 2019. accepted.
- Rüdiger Gleim, Alexander Mehler, and Alexandra Ernst. SOA implementation of the eHumanities Desktop. In Proceedings of the Workshop on Service-oriented Architectures (SOAs) for the Humanities: Solutions and Impacts, Digital Humanities, Hamburg, 2012.
- Philipp Helfrich, Elias Rieb, Giuseppe Abrami, Andy Lücking, and Alexander Mehler. TreeAnnotator: Versatile visual annotation of hierarchical text relations. In Proceedings of the 11th edition of the Language Resources and Evaluation Conference, May 7 - 12, LREC 2018, Miyazaki, Japan, 2018.
- Alexandros Komninos and Suresh Manandhar. Dependency based embeddings for sentence classification tasks. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1490–1500, 2016.
- Omer Levy and Yoav Goldberg. Dependency-based word embeddings. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), volume 2, pages 302–308, 2014.
- Wang Ling, Chris Dyer, Alan Black, and Isabel Trancoso. Two/Too Simple Adaptations of word2vec for Syntax Problems. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2015.
- Alexander Mehler, Tim vor der Brück, Rüdiger Gleim, and Tim Geelhaar. Towards a network model of the coreness of texts: An experiment in classifying Latin texts using the TTLab Latin Tagger. In Chris Biemann and Alexander Mehler, editors, Text Mining: From Ontology Learning to Automated Text Processing Applications, Theory and Applications of Natural Language Processing, pages 87–112. Springer, Berlin/New York, 2015.
- Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In Proceedings of NAACL 2013, pages 746–751, 2013.
- Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In Proceedings of NAACL-HLT 2018, pages 2227–2237, 2018.
- Tim vor der Brück and Alexander Mehler. TLT-CRF: A lexicon-supported morphological tagger for Latin based on conditional random fields. In Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016, 2016.