



CALLIDUS

Einsatz von Korpora im Erwerb historischer Sprachen

Themen

1. Motivation
2. Fragliche Prämissen
3. Ein Weg zu besserer Wortschatzkompetenz
4. Korpora
5. Forschungsfragen
6. Probleme für Latein als historische Sprache
7. Machina Callida: Softwaredesign
 - a) Korpora
 - b) Zitationsmodell
 - c) Suchfunktion
 - d) Vokabelfilter
8. Übungsdesign

1. Motivation

- Gegenwärtige **Situation** des Lateinunterrichts
 - **Selten** spezifische Wortschatzarbeit
 - sprachübergreifende **Wortgleichungen**
 - Keine einheitlichen Gütekriterien, basieren oft nur auf **rohen Frequenzen**
 - **Übersetzung** = Wortschatzkompetenz
 - **Dekontextualisierte** Einzelwörter
- res = Ding
 - res publica = öffentliches Ding ?
 - res familiares = familienartiges Ding ?
 - res frumentaria = getreidehaftes Ding ?



2. Fragliche Prämissen (I)

1. Fremdspracherwerb dient der Fähigkeit zur Kommunikation.

⚡ **Rezeption** und **Sprachreflexion** als Ziele des Lateinunterrichts

2. Im Fremdspracherwerb sollte der aktive, passive und **potentielle Wortschatz** gefördert werden, am meisten jedoch der aktive.

⚡ aktiver eher unwichtig, potentieller bei Latein besonders wichtig (Derivation, Etymologien etc.)

3. Im Lateinunterricht wird hauptsächlich Latein gelernt.

⚡ Ziele (Latein) vs. tatsächliche Effekte des LU (**Deutsch**)

2. Fragliche Prämissen (II)

4. Der Spracherwerb ist zu Beginn der Lektürephase im Wesentlichen abgeschlossen.

⚡ **Lektüre-“Schock“**, lebenslanger Spracherwerb

5. In der Lektürephase wird lateinische Literatur in umfangreicherem Maß gelesen und rezipiert.

⚡ mehr **Formenbestimmung** & Übersetzung als Leseverständnis

6. Die römische Kultur ist weit genug entfernt, um objektiv studiert zu werden.

⚡ kulturelle **Interferenzen** bei einzelnen Konzepten und bei der Textinterpretation

3. Ein Weg zu besserer Wortschatzkompetenz

- Entwicklung einer Umgebung für lateinische Wortschatzarbeit
 - **Regelmäßige** explizite Wortschatzarbeit im Unterricht (und außerhalb)
 - **Kontextualisierte** Übungen
 - Nachhaltig **vernetzt**
- Voraussetzungen
 - Systematische Analyse der **empirischen** Grundlagen
 - **Dynamische** Anpassung an die jeweiligen Gegebenheiten
 - **Verfügbarkeit:** mobil oder stationär, selbstständig oder in Gruppen
 - **Zielorientierung:** je nach zu rezipierendem Text
 - **Vorwissen:** je nach Stand des individuellen Wortschatzerwerbs

4. Korpora

- „any collection of written or spoken texts“
- **authentische** Sprache
- hohe Phänomen**frequenz**
- **Kontext**
- **Variation** & historische Linguistik
- **Distribution** von sprachlichen Phänomenen
- sprachliche **Muster**, z.B. Kollostruktionen
- **Verfügbarkeit**

Definition: Lüdeling & Kytö 2008; Authentische Sprache & Phänomenfrequenz: Campoy et al. 2010, Gilquin & Granger 2010; Kontext: Gardner 2007; Variation: Ajmer 2009; historische Linguistik: Hundt et al. 2012; Distribution: Brysbaert & New 2009, Lijffijt et al. 2011, Gries 2013; sprachliche Muster allgemein: O’Keeffe et al. 2007; Kollostruktionen: Schmid & Küchenhoff 2013; Verfügbarkeit: Römer 2009

5. Forschungsfragen

- Wie können bereits **vorhandene Ressourcen** (z.B. Korpora) für den lateinischen **Wortschatzerwerb** in Schulen eingesetzt werden?
- Lexikalische **Informationen**:
 - Wie können sie aus problematischen Texten **gewonnen** werden?
 - Wie können sie zur Unterstützung des lateinischen Wortschatzerwerbs eingesetzt werden?
- ... und einige andere mehr:
 - Softwaredesign
 - Übungsdesign
 - kognitive Progression
 - etc.

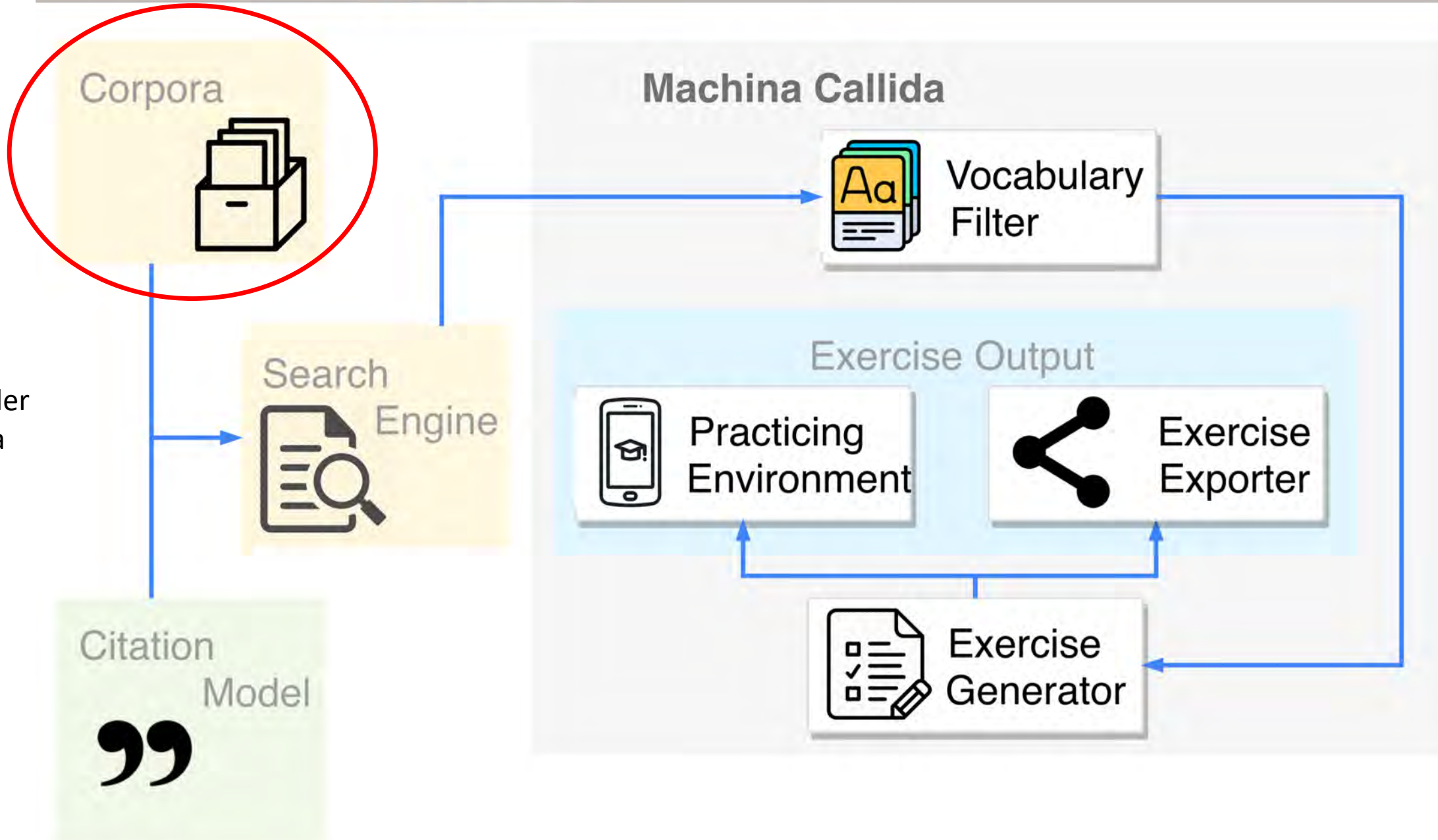
6. Probleme für Latein als historische Sprache

Probleme	Lösungsansätze
Keine L1-Lernenden → fehlende Goldstandards für Korpora	Annotation durch Forschende und Studierende; Crowdsourcing / Cyberscience
Geringe Datenbasis → statistisch generalisierende Modelle schwer anwendbar	Vereinheitlichung der Annotationsformate
Starke Diachronie → höhere Variation → schwerer verarbeitbar für maschinelle Sprachmodelle	Normalisierung, Klassizismus

A faint, light gray watermark of a fox's head is centered in the background of the slide. The fox has large, pointed ears and a smiling mouth.

7. Machina Callida

Resource Workflow

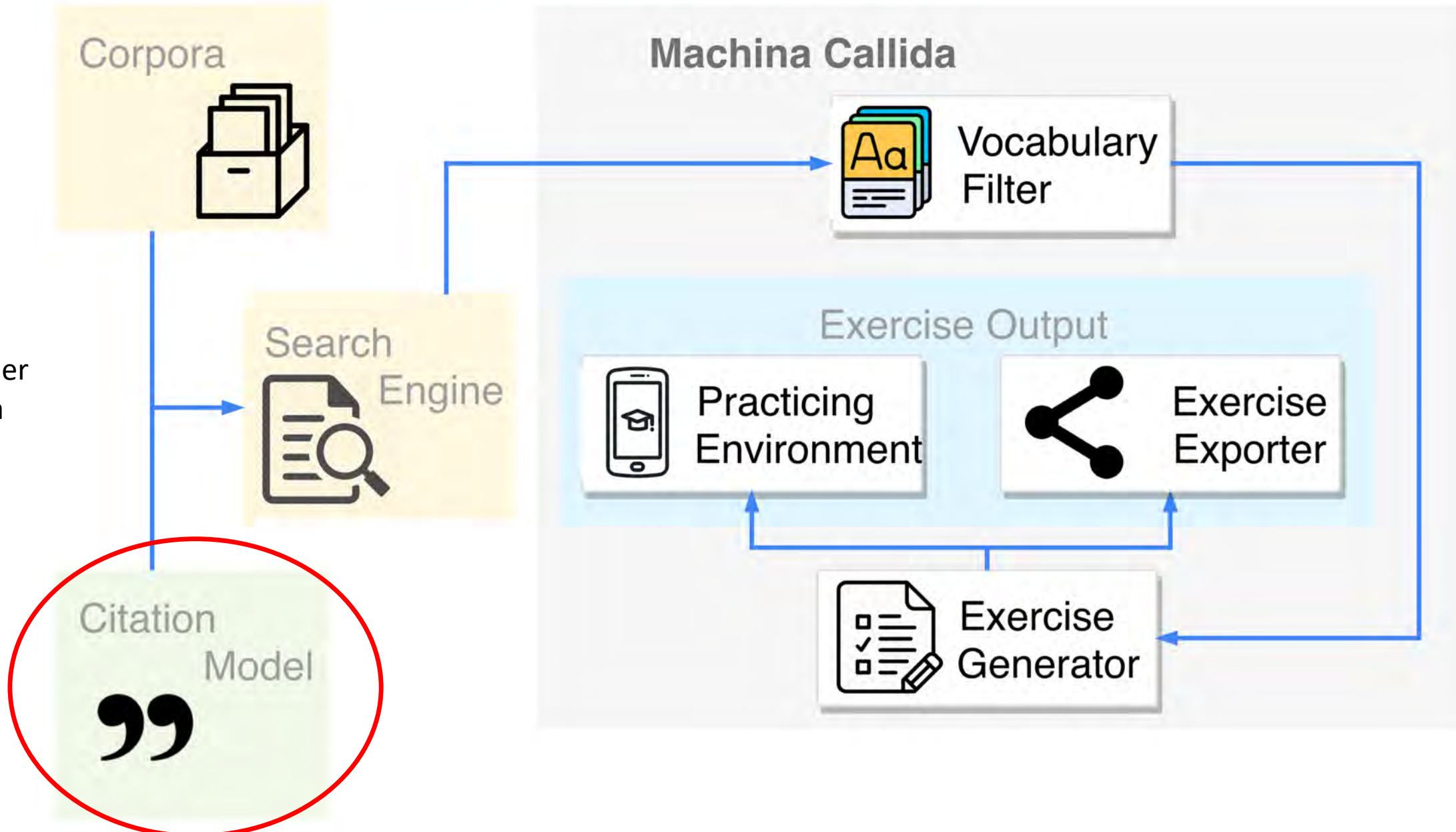


7.1. Softwaredesign: Korpora

- a) Treebanks (ca. 0,5 Mio. *tokens*)
 - Annotationen & Vokabular: PROIEL
 - Trainingsdaten für automatische Annotationen: ITTB
 - Vokabular: AGLDT
- b) Rohtexte (> 15 Mio. *tokens*)
 - Perseids CTS
 - Lehrbuch VIVA
- c) experimentell: maschinelle Generierung autorenspezifischer pseudo-literarischer Texte

Resource Workflow

Verarbeitung bestehender Ressourcen in der Machina Callida



7.2. Softwaredesign: Zitationsmodell (I)

- CITE

- urn:cts:CTS_NAMESPACE:WORK:PASSAGE

- WORK = TEXTGROUP.WORK.VERSION.EXEMPLAR

- PASSAGE = CITABLENODE@SUBSECTION

- Problem: keine Subsections beim Perseids CTS-Repository möglich

- Lösung: Segmentierung nach Satzzeichen und dann Verweis auf Satz-IDs

- Neues Problem: kein genauere Kontext als Satzebene möglich

- Lösungsansatz: Verweis auf Token-IDs (über graphANNIS)

- Beispiel urn:cts:latinLit:phi0448.phi001.perseus-lat2:1.1.1-1.2.3

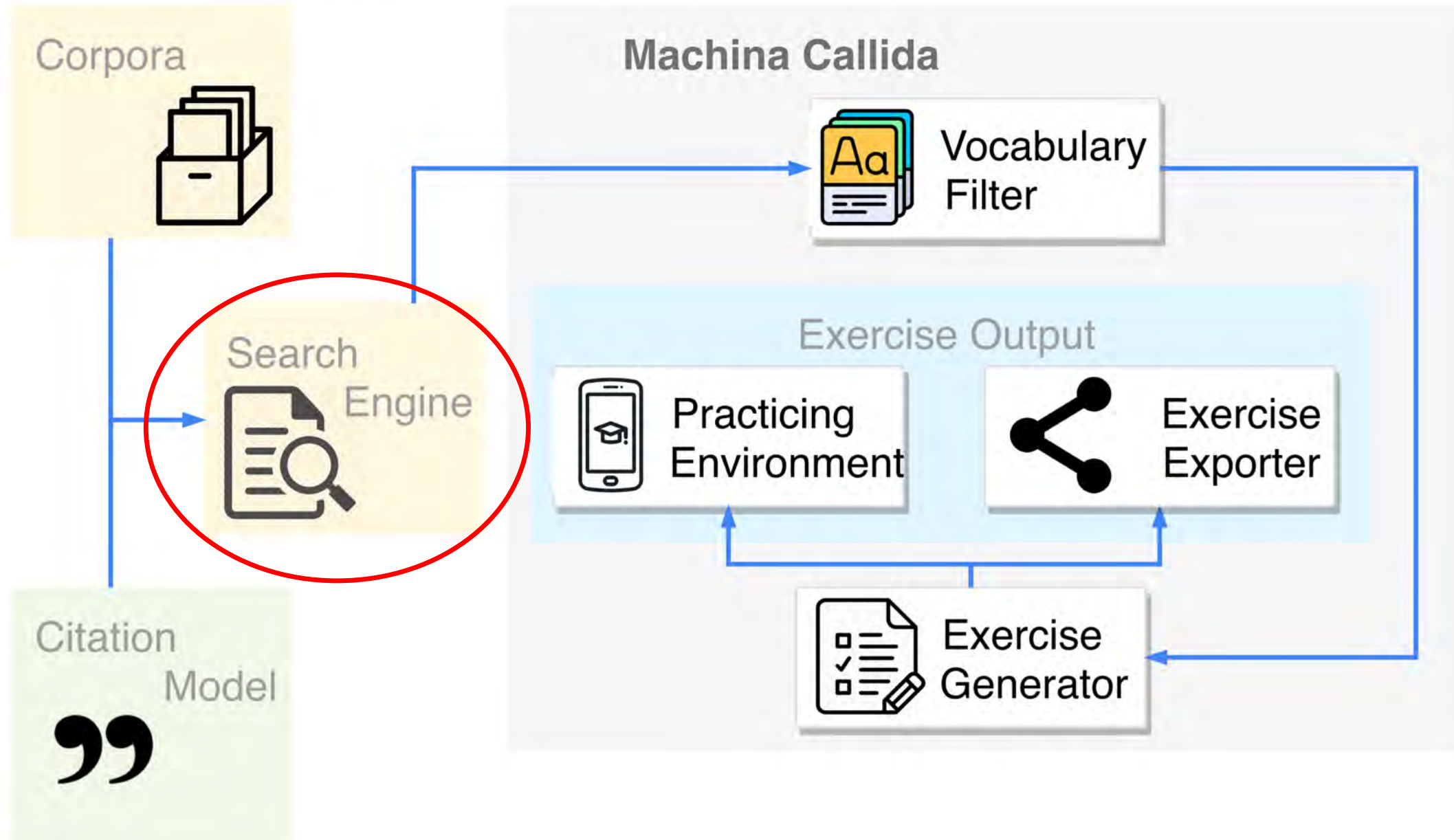
- Sprache Autor Werk Textedition Textpassage

- = Latein, Caesar, Bellum Gallicum, Oxford 1901, Buch 1, Kapitel 1, Abschnitt 1 ...

7.2. Softwaredesign: Zitationsmodell (II)

- Übertragung des CITE URN-Modells auf nicht-CTS-Texte:
 - urn:cts:latinLit:phi0448.phi001.perseus-lat2:1.1.1-1.2.3
 - urn:custom:latinLit:proiel.caes-gal.lat:1.1.1-1.2.3
 - Problem: schwammige Definition von Autor, Werk und Edition
 - urn:custom:latinLit:viva.lat:1.3-1.5 (= Lehrbuch VIVA, Buch 1, Lektion 3-4Z)
 - Problem: Analogie „Lehrbuchlektion=Kapitel“ funktioniert schlecht, wegen Zusatztexten
 - Lösungsansatz: weitere Zitationsebene definieren (Level 3), für Zusatztexte

Resource Workflow

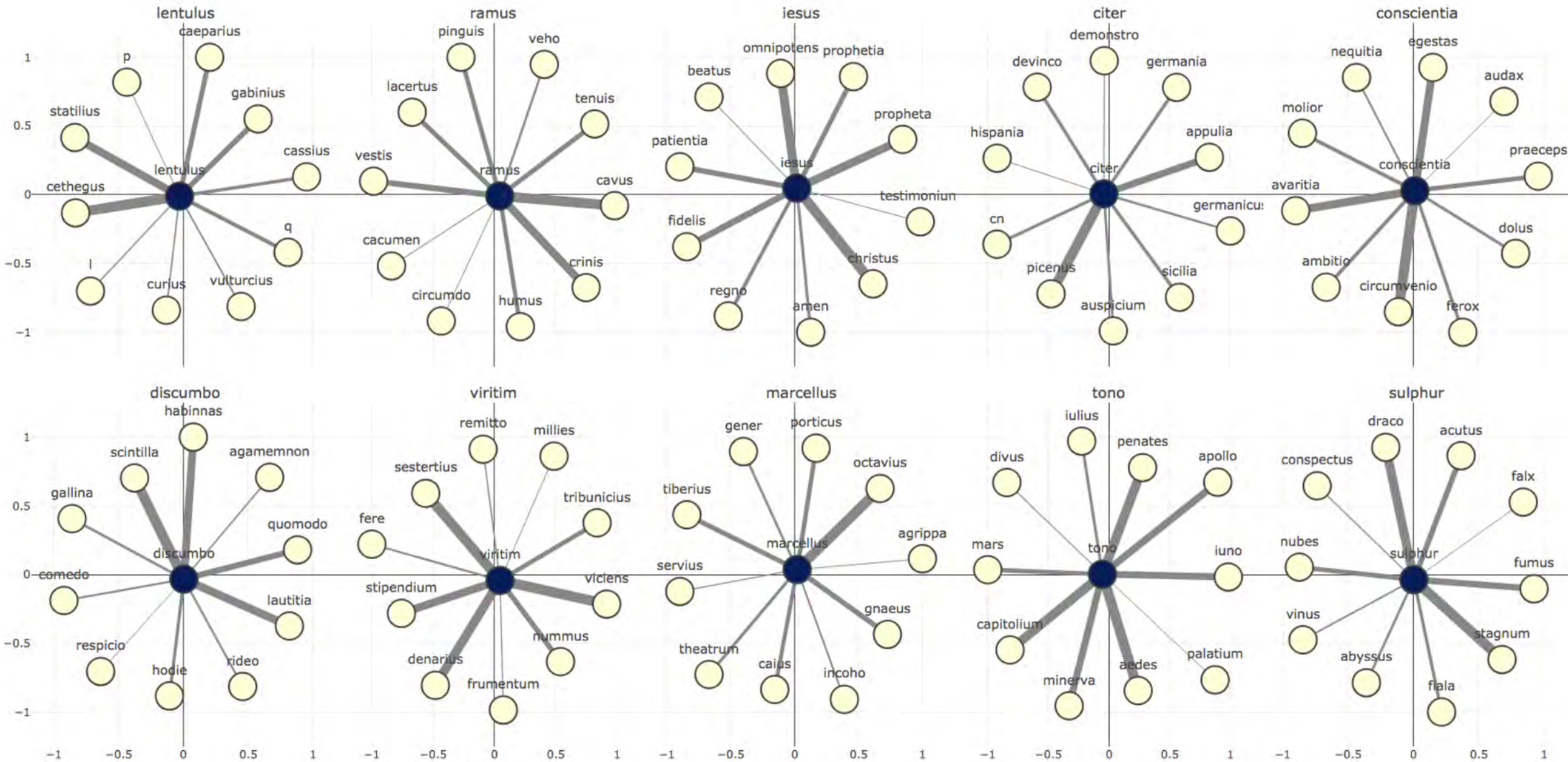


Verarbeitung bestehender Ressourcen in der Machina Callida

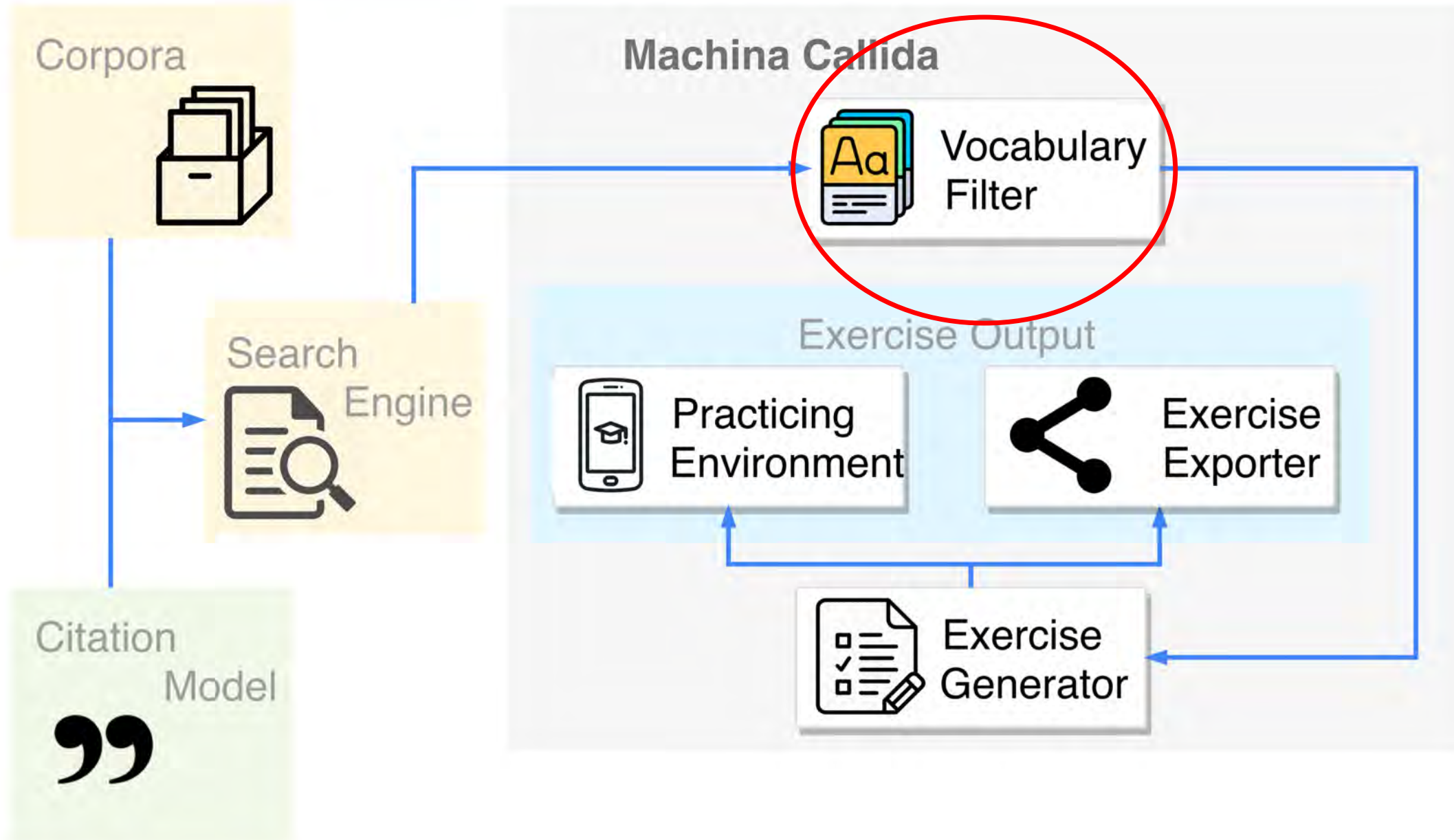
7.3. Softwaredesign: Suchfunktion ...fueled by

- Verschiedene linguistische Phänomene: Flexion (→ Morphologie), Lemmata (→ Lexik), Dependenzgrammatik (→ Syntax)
- Multi-Layer-Annotationen
- Verschiedene Modelle: Bäume, Graphen...
- Domänenspezifische Sprache zur Anwendung der ANNIS Query Language
- Unicode für entlegene Zeichen
- Unterstützung für Von-rechts-nach-links gelesene Texte & Annotationen
- Experimentell: maschinell generierte Wortfelder (→ Semantik)

Top 10 Kookkurrenzfelder (nach Ähnlichkeit)



Resource Workflow



Verarbeitung bestehender Ressourcen in der Machina Callida

7.4. Softwaredesign: Vokabelfilter (I)

- Status quo: Bamberger Wortschatz
 - Lemmaliste, als Basis für die meisten Lehrbücher
 - Erstellungsprozedur fragwürdig: Korpus nicht zugänglich, rohe Frequenzen, willkürliche Schwellwerte
 - Einteilung in Top 500 und vollständig (1.248)
 - Ergänzungen / Alternativen:
 - Treebanks:
 - AGLDT: 7.182 Lemmata
 - PROIEL: 16.402 Lemmata
 - Lehrbuch: VIVA → 1.164 Lemmata (bisher ohne Frequenzen)
- ➔ Granulare Hierarchie (nach Frequenz sortiert), größerer Umfang, korpuspezifisch

7.4. Softwaredesign: Vokabelfilter (II)

1. Korpus auswählen (z.B. Caesar, Bellum Gallicum, Buch 1)
2. Referenzvokabular auswählen (z.B. BWS Top 500 oder VIVA oder PROIEL Top 287)
3. Lemmatisierung
 - a) Treebank: Annotationen abfragen mittels ANNIS
 - b) Rohtext: automatisch annotieren mittels UDPipe, dann Übertragung in ANNIS
4. Lemma-Abgleich für jedes Token (Korpus vs. Referenzvokabular)
5. Ranking nach prozentualer Abdeckung
6. Auswahl eines Subkorpus
7. Weitere Operationen
 - a) Unbekannte Vokabeln markieren (z.B. um sie dann zu erklären)
 - b) Übungen nur zu bekannten Vokabeln erstellen

7.4. Softwaredesign: Vokabelfilter (III)

Probleme	Lösungsansätze
Lemmaliste als Repräsentation der Wortschatzkompetenz → alles gleich gewichtet, nur Gesamtpunktzahl zählt → fehlende adäquate Diagnosemittel	<ul style="list-style-type: none">• Übertragung des Toolbox Picture Vocabulary Test auf den Schriftbereich• Anwendung der Item Response Theory<ul style="list-style-type: none">→ Individuelle Anpassung der Item-Schwierigkeit→ Unterschiedliche Gewichtungen
Fehlende Alternative zu Einzelwortlisten → überprüft nur dekontextualisiert Form/Bedeutung	<ul style="list-style-type: none">• strategisches Leseverständnis prüfen<ul style="list-style-type: none">• Focus on Form• Form/Bedeutung, Derivation, Polysemie, Kollokationen• Textbasiert (?)
Produktion/Kommunikation als Antrieb für Spracherwerb fehlt	<ul style="list-style-type: none">• Intensivere Auseinandersetzung mit den jeweiligen Vokabeln<ul style="list-style-type: none">• Spreading Activation nutzen• Stemming, Word Sense Induction, Kollokationsanalysen

Toolbox Picture Vocabulary Test & Item Response Theory: Gershon et al. 2013; Leseverständnis statt Form-Bedeutung: Schmitt 2014; Focus-on-form: Ellis 2009; Dimensionen der Wortschatzkompetenz: González-Fernández & Schmitt 2019; Kommunikation im Spracherwerb: Kersten 2010; Spreading Activation : Bruza et al. 2009

Kollokationsanalyse: rohe Frequenzen

- AGLDT, dependenzbasiert, Substantiv + Adjektiv

127: **res publicus**

42: populus Romanus

18: **vox magnus**

16: nomen meus

14: pars tertius

13: **res novus**

13: **eques Romanus**

11: res omnis

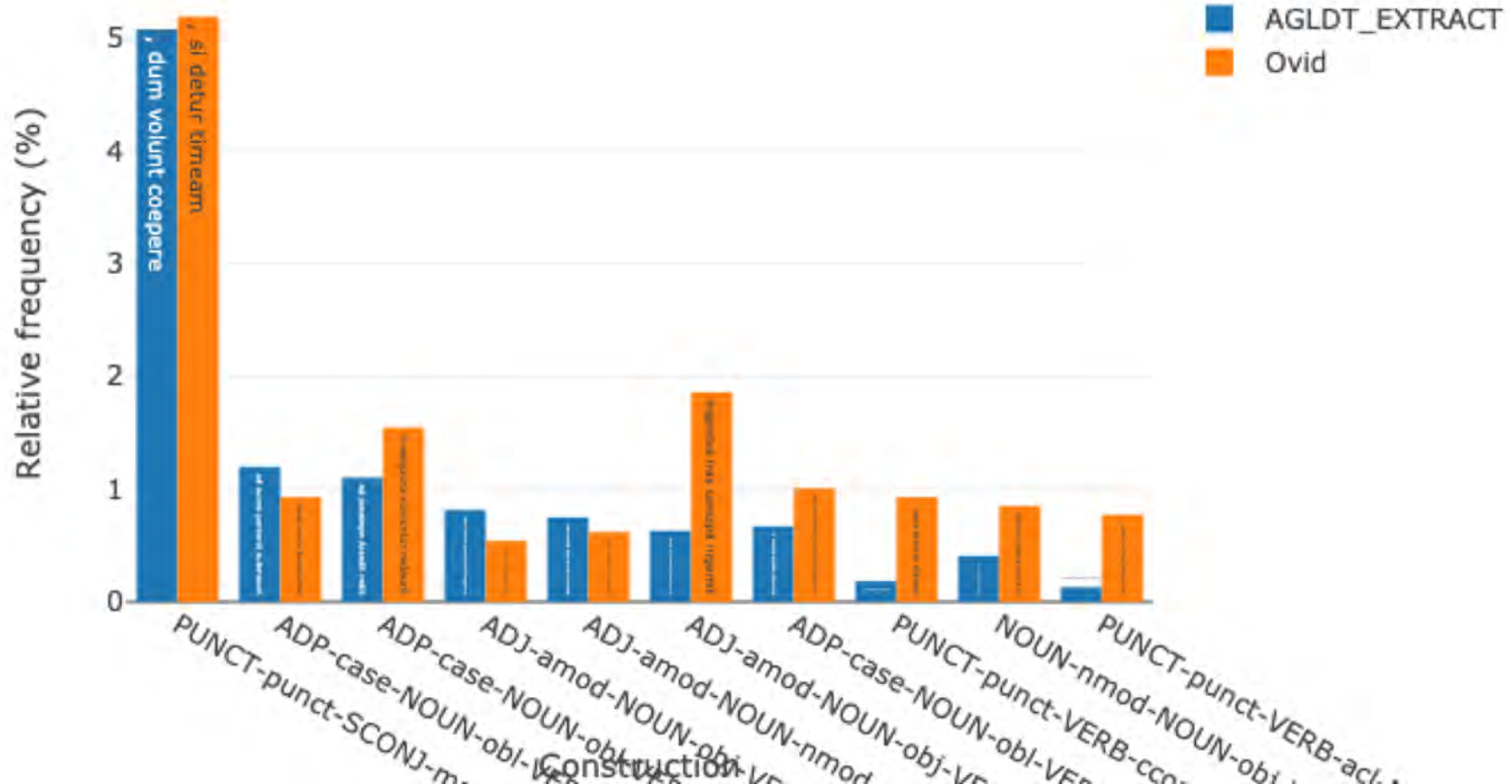
10: deus immortalis

10: **aes alienus**

- ➔ ca. 50% Kollokationen nach der lexikografischen Definition: „word combination whose semantic and/or syntactic properties cannot be fully predicted from those of its components“
- ➔ Genauigkeit könnte möglicherweise durch Pointwise Mutual Information noch weiter erhöht werden

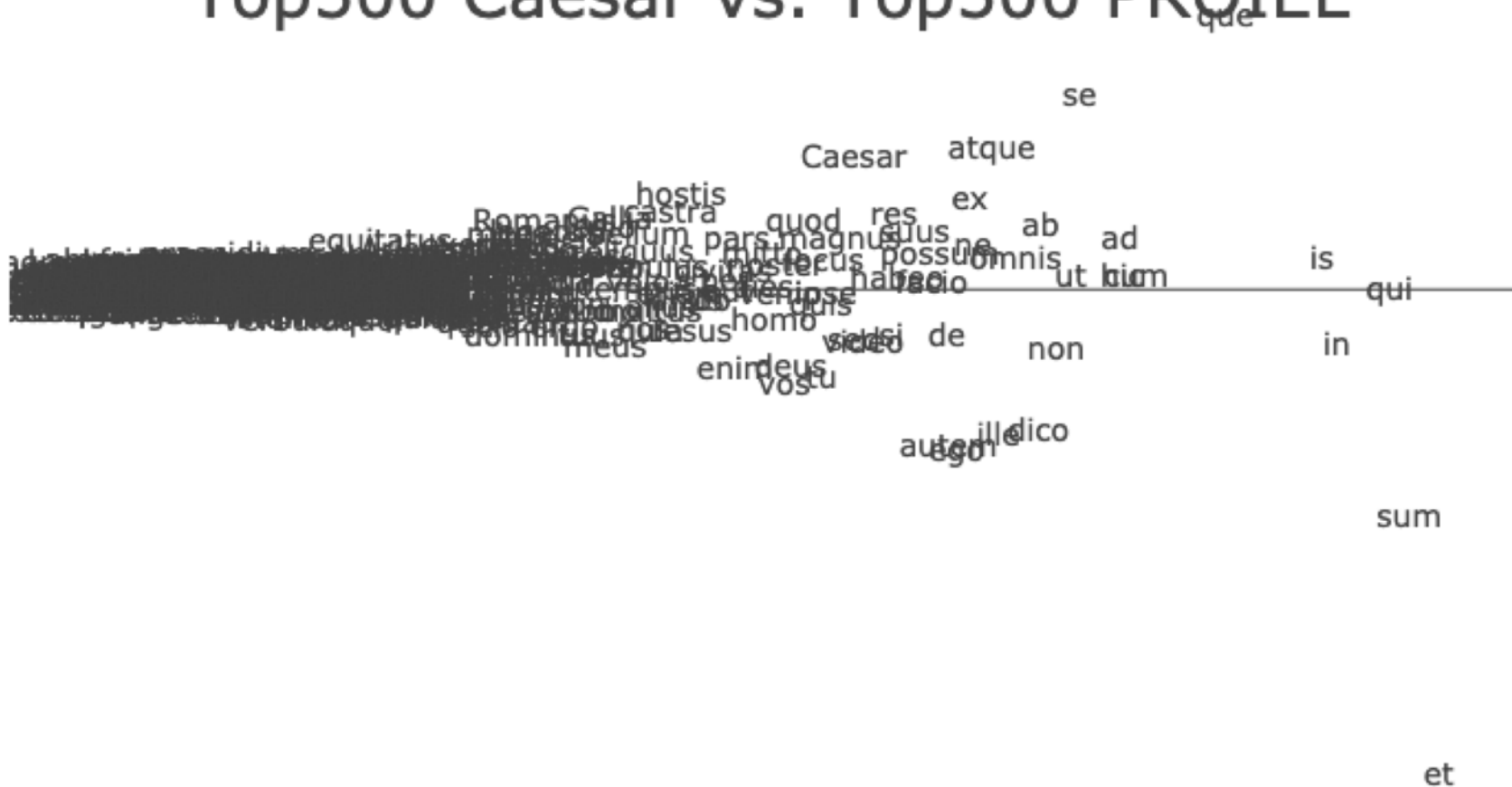
Konstruktionsorientierter Ansatz

Construction frequencies Ovid vs. AGLDT



Vokabular: Subkorpus vs. Referenzkorpus

Top500 Caesar vs. Top500 PROIEL



8. Übungsdesign

- Zuordnungsübungen: Potenzial zur Förderung der sprachübergreifenden lexikalischen Vernetzung
 - Ressourcen dazu noch nicht ausgeschöpft, z.B. Wikipedia / Wiktionary
- Variation der Interaktion: Lückentexte, Zuordnung, Keyword In Context, ... (demnächst: Erkennung)
- Input vs. Intake (vs. Retention)

Zusammenfassung

- ❤️ **Korpora** ❤️
- Lernt Vokabeln, aber bitte im **Kontext!**
- Übersetzung ist **nicht** Wortschatzerwerb.

→ Wir brauchen bessere **Evaluationen** und **Übungen** für lateinischen Wortschatz.

Bibliografie (I)

- Aijmer, K. (2009). *Corpora and language teaching* (Vol. 33). Retrieved from <https://books.google.de/books?hl=de&lr=&id=Il-hYaUJLuOC>.
- Almas, B. (2017). Perseids: Experimenting with Infrastructure for Creating and Sharing Research Data in the Digital Humanities. *Data Science Journal*, 16(19), 1–17. Retrieved from <https://datascience.codata.org/articles/10.5334/dsj-2017-019/galley/663/download/>.
- Bamman, D., & Crane, G. (2011). The Ancient Greek and Latin Dependency Treebanks [AGLDT]. In *Language Technology for Cultural Heritage* (pp. 79–98). Retrieved from https://perseusdl.github.io/treebank_data/.
- Bartoszek, V., Datené, V., Lösch, S., Mosebach-Kaufmann, I., Nagengast, G., Schöffel, C., ... Schröttel, W. (2013). *VIVA 1 Lehrerband* (Vol. 1). Retrieved from https://books.google.de/books?hl=de&lr=&id=b_M4oV8VINgC.
- Blackwell, C., & Smith, N. (Eds.). (2015). *The Canonical Text Services URN Specification, Version 2.0.rc.1 [CITE / URN]*. Retrieved from http://cite-architecture.github.io/cturn_spec/.
- Bruza, P., Kitto, K., Nelson, D., & McEvoy, C. (2009). Is there something quantum-like about the human mental lexicon? *Journal of Mathematical Psychology*, 53(5), 362–377. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2834425/pdf/nihms160244.pdf>.
- Brysaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. Retrieved from <https://link.springer.com/content/pdf/10.3758/BRM.41.4.977.pdf>.
- Campoy, M. C., Cubillo, M. C. C., Belles-Fortuno, B., & Gea-Valor, M. L. (2010). *Corpus-Based Approaches to English Language Teaching*. Retrieved from <https://books.google.de/books?hl=en&lr=&id=L6zzqz8JPUgC>.
- Cecchini, F. M., Passarotti, M., Marongiu, P., & Zeman, D. (2018). Challenges in Converting the Index Thomisticus Treebank into Universal Dependencies [ITTB]. *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, 27–36. Retrieved from <https://www.aclweb.org/anthology/W18-6004>.
- Daum, M. (2016). *Wortschatz und Lehrbuch: Ein Kriterienkatalog für die Wortschatzkonzeption in Lateinlehrwerken*. Retrieved from <https://books.google.de/books?id=CSv8CwAAQBAJ>.
- East, M. (2019). Sustaining innovation in school modern foreign language programmes: teachers' reflections on task-based language teaching three years after initial teacher education. *The Language Learning Journal*, 47(1), 105–115. <https://doi.org/10.1080/09571736.2016.1221440>.
- Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122–139). Retrieved from <https://pdfs.semanticscholar.org/38b5/7a07241bbb22d1d0bf404f7d5e346f41e078.pdf>.
- Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19(3), 221–246. Retrieved from <https://pdfs.semanticscholar.org/2866/fad24dcd59848acb08a5468bacbecc02bf8a.pdf>.
- Evert, S. (2005). *The statistics of word cooccurrences: word pairs and collocations* (Dissertation, Universität Stuttgart). Retrieved from <https://elib.uni-stuttgart.de/bitstream/11682/2573/1/Evert2005phd.pdf>.
- Foley, J., & Thompson, L. (2017). *Language learning: A lifelong process*. Retrieved from <https://content.taylorfrancis.com/books/download?dac=C2012-0-11240-8&isbn=9781317835912&format=googlePreviewPdf>.

Bibliografie (II)

- Gardner, D. (2007). Validating the Construct of Word in Applied Corpus-based Vocabulary Research: A Critical Survey. *Applied Linguistics*, 28(2), 241–265. <https://doi.org/10.1093/applin/amm010>.
- Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., ... Weintraub, S. (2013). NIH TOOLBOX COGNITION BATTERY (CB): MEASURING LANGUAGE (VOCABULARY COMPREHENSION AND READING DECODING). *Monographs of the Society for Research in Child Development*, 78(4), 49–69. Retrieved from <https://www.istor.org/stable/43772790>.
- Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching. In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (Vol. 359370, pp. 359–370). Retrieved from https://sites.uclouvain.be/cecl/archives/Gilquin_Granger_2010_How_can_DDL_be_used_in_language_teaching.pdf.
- González-Fernández, B., & Schmitt, N. (2019). Word Knowledge: Exploring the Relationships and Order of Acquisition of Vocabulary Knowledge Components. *Applied Linguistics*. <https://doi.org/10.1093/applin/amy057>.
- Gries, S. Th. (2013). 50-something years of work on collocations. *International Journal of Corpus Linguistics*, 18(1), 137–166. Retrieved from <https://pdfs.semanticscholar.org/b0a1/c74fb94d8da6c2f0f0a219755665936fdb4.pdf>.
- Gries, S. T., & Durrant, P. (2019). *Analyzing co-occurrence data*. Retrieved from http://stgries.faculty.linguistics.ucsb.edu/research/ToApp_STG_PD_CooccData_PHCL.pdf.
- Große, M. (2015). Pons Latinus: Latein als reflexionsbasierte Brückensprache im Rahmen eines sprachsensiblen Lateinunterrichts. In E. M. F. Ammann, A. Kropp, & J. Müller-Lancé (Eds.), *Herkunftsbedingte Mehrsprachigkeit im Unterricht der romanischen Sprachen* (Vol. 17, pp. 185–206). Retrieved from <https://books.google.de/books?hl=de&lr=&id=0EHWBgAAQBAJ>.
- Haug, D. T., & Jøhndal, M. (2008). Creating a parallel treebank of the old Indo-European Bible translations [PROIEL]. *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, 27–34. Retrieved from <http://www.hf.uio.no/ifikk/english/research/projects/proiel/Activities/proiel/publications/marrakech.pdf>.
- Hosseini, S. M. H. (2019). Cognitive Socio-political Language Learning Theory, Multiple Input-output Hypothesis and Competitive Team-based Learning. *Theory and Practice in Language Studies*, 9(4), 411–423. Retrieved from <https://www.academypublication.com/ojs/index.php/tpls/article/download/tpls0904411423/1875>.
- Hundt, M., & Leech, G. (2012). “Small is beautiful”: On the value of standard reference corpora for observing recent grammatical change. In T. Nevalainen & E. C. Traugott (Eds.), *The Oxford handbook of the history of English* (pp. 175–188). Retrieved from <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199922765.001.0001/oxfordhb-9780199922765-e-17>.
- Kersten, S. (2010). *The mental lexicon and vocabulary learning: Implications for the foreign language classroom*. Retrieved from <https://books.google.de/books?hl=de&lr=&id=NEOx13zHKVoC>.
- Kipf, S. (2015). ... und wo bleibt die Literatur? Gedanken zum Kompetenzerwerb im altsprachlichen Unterricht. *Forum Classicum. Zeitschrift Für Die Fächer Latein Und Griechisch an Schulen Und Universitäten*, (2), 70–82. Retrieved from <https://www.altphilologenverband.de/forumclassicum/pdf/FC2015-2.pdf>.
- Krause, T. (2019). ANNIS: A graph-based query system for deeply annotated text corpora. <https://doi.org/10.5281/zenodo.2598164>.
- Lijffijt, J., Papapetrou, P., Puolamäki, K., & Mannila, H. (2011). Analyzing word frequencies in large text corpora using inter-arrival times and bootstrapping. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 341–357. Retrieved from <https://link.springer.com/content/pdf/10.1007%2F978-3-642-23783-6.pdf>.
- Lüdeling, A., & Kytö, M. (2008). *Corpus linguistics: An international handbook*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.455.9749&rep=rep1&type=pdf>.
- McGillivray, B. (2013). *Methods in Latin computational linguistics*. Retrieved from <https://books.google.de/books?hl=de&lr=&id=33dfAgAAQBAJ>.
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Retrieved from <https://books.google.de/books?id=KjYuDwAAQBAJ>.
- Newman, G. (2014). Citizen CyberScience—New directions and opportunities for human computation. *Human Computation*, 1(2), 103–109. Retrieved from https://www.researchgate.net/profile/Newman_Greg/publication/287704099_Citizen_CyberScience_New_Directions_and_Opportunities_for_Human_Computation/links/57c74ebf08aec24de042aa9c.pdf.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Retrieved from <https://books.google.de/books?hl=en&lr=&id=O3k0tWLP5LYC>.

Bibliografie (III)

- [RLP 2006] Senatsverwaltung für Bildung, Jugend und Sport Berlin. (2006). *Rahmenlehrplan für die gymnasiale Oberstufe. Gymnasien, Gesamtschulen mit gymnasialer Oberstufe, Berufliche Gymnasien, Kollegs, Abendgymnasien. Latein*(RLP 2006). Retrieved from https://www.berlin.de/sen/bildung/unterricht/faecher-rahmenlehrplaene/rahmenlehrplaene/mdb-sen-bildung-unterricht-lehrplaene-sek2_latein.pdf.
- Römer, U. (2009). Corpus research and practice. *Corpora and Language Teaching*, Amsterdam: John Benjamins, 83–98. Retrieved from <https://books.google.de/books?hl=en&lr=&id=3yfOJSMTRhoC>.
- Sass, A. (2015). *Sprachenübergreifendes Vokabellernen: eine qualitativ-interpretative Studie zur Vernetzung der Fächer Englisch und Latein* (Dissertation, Staats- und Universitätsbibliothek Bremen). Retrieved from <https://elib.suub.uni-bremen.de/edocs/00105616-1.pdf>.
- Schirok, E. (2010). „Wortschatzarbeit“. In T. Doepner & M. Keip (Eds.), *Interaktive Fachdidaktik Latein* (pp. 13–34). Retrieved from <https://static.onleihe.de/content/vandenhoeck/20141125/978-3-647-26411-0/v978-3-647-26411-0.pdf>.
- Schmid, H.-J., & Küchenhoff, H. (2013). Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics*, 24(3), 531–577. Retrieved from http://www.anglistik.uni-muenchen.de/personen/professoren/schmid/schmid_publ/collostructional-analysis.pdf.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951. Retrieved from http://eprints.nottingham.ac.uk/32299/1/Schmitt_final_version_LL.pdf.
- Siebel, K. (2011). Lateinischer Wortschatz als Brücke zur Mehrsprachigkeit? Eine Durchsicht des Aufgabenspektrums aktueller Lateinlehrwerke. *Pegasus-Onlinezeitschrift*, XI(1), 102–132. Retrieved from www.pegasus-onlinezeitschrift.de/2011_1/pegasus_2011-1_siebel.pdf.
- Siebel, K. (2017). *Mehrsprachigkeit und Lateinunterricht: Überlegungen zum lateinischen Lernwortschatz* (Vol. 4). Retrieved from <https://books.google.de/books?hl=de&lr=&id=YOMsDwAAQBAJ>.
- Straková, J., Straka, M., & Hajič, J. (2014). Open-source tools for morphology, lemmatization, POS tagging and named entity recognition. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 13–18. Retrieved from <http://www.aclweb.org/anthology/P14-5003>.
- Tiepmar, J., Teichmann, C., Heyer, G., Berti, M., & Crane, G. (2014). A new implementation for canonical text services [CTS]. *Proceedings of the 8th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, 1–8. Retrieved from <https://www.aclweb.org/anthology/W14-0601>.
- Utz, C. (2000). Mutter Latein und unsere Schüler — Überlegungen zu Umfang und Aufbau des Wortschatzes [BWS]. *Antike Literatur–Mensch, Sprache, Welt. Dialog Schule Und Wissenschaft*, 34, 146–172. Retrieved from https://www.ccbuchner.de/files_media/mediathek/downloads/267.pdf.
- Van de Loo, T. (2016). Wortschatzarbeit – neuere Perspektiven und schulische Praxis. *Pegasus-Onlinezeitschrift*, 16, 131–151. Retrieved from <https://journals.ub.uni-heidelberg.de/index.php/pegasus/article/viewFile/35254/28912>.
- Vincze, V., Szauter, D., Almási, A., Móra, G., Alexin, Z., & Csirik, J. (2010). *Hungarian dependency treebank*. 1855–1862. Retrieved from http://publicatio.bibl.u-szeged.hu/8749/1/465_Paper.pdf.
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20(2), 232–245. Retrieved from <https://files.eric.ed.gov/fulltext/EJ815123.pdf>.
- Xie, Z. (2017). *Neural Text Generation: A Practical Guide*. 1–21. Retrieved from <https://arxiv.org/abs/1711.09534v1>.