

Aufsätze

Digital und korpusbasiert: Wortschatzarbeit einmal anders

Der Spracherwerb moderner und historischer Sprachen¹ wird hinsichtlich seiner Zielsetzung und Methodik als different wahrgenommen. Während auf der einen Seite Interaktion und Immersion im Vordergrund stehen, zielt der Erwerb historischer Sprachen auf die Fähigkeit, aus Texten Informationen zu entnehmen und die Texte ggf. übersetzen zu können, wobei der Spracherwerb vor allem durch die (meta-)kognitive Auseinandersetzung mit der Sprache erfolgt. Doch ist diese Abgrenzung überhaupt mit dem Wissen über die Sprachverarbeitung im menschlichen Gehirn vereinbar? – Nein, denn für jede natürliche Sprache gibt es nur ein komplexes System der Sprachverarbeitung, das auf verschiedene, miteinander interagierende Areale des Gehirns aufgeteilt ist (Gage und Baars 2018, 192ff.). Also sollte die einer Sprache zugewiesene Funktion einen deutlich geringeren Einfluss auf die Methodenauswahl im Spracherwerb ausüben. Vielleicht ist es daher an der Zeit, den Unterricht für historische Sprachen stärker mit anderen Disziplinen zu vernetzen, insbesondere auf der Ebene der Methodik.

Auf der Basis dieser grundsätzlich offenen Einstellung hat das interdisziplinäre Forschungsprojekt CALLIDUS (Computer-Aided Language Learning: Lexikonerwerb im Lateinunterricht durch korpusgestützte Methoden)² an der Humboldt-Universität zu Berlin den Wortschatzerwerb im Lateinunterricht (LU) untersucht. Von zentralem Forschungsinteresse war hierbei die Frage, ob korpusbasierte Methoden des Spracherwerbs (Data-driven Learning,

DDL), die im modernen Fremdsprachenunterricht (computergestützt) Anwendung finden (Boulton 2017; Gilquin und Granger 2010), auch die Wortschatzarbeit des LU sinnvoll unterstützen können.

Ausgehend von der wissenschaftlichen Theorie zu Spracherwerb und Methodik des DDL wurde einerseits ein neues Modell zu einem weit gefassten Wortschatzbegriff erarbeitet, auf dessen Grundlage verschiedene Studien zur Wortschatzkompetenz der Lernenden entworfen, durchgeführt und ausgewertet wurden.³ Andererseits wurde eine Software (*Machina Callida*) entwickelt, die Methoden des Maschinellen Lernens (insbesondere Natural Language Processing, NLP) verwendet und es ermöglicht, direkt auf der Grundlage lateinischer Originalliteratur intralinguale Wortschatzübungen zu erstellen und online zu nutzen, als Datei (PDF- oder Word-Dokument) zu exportieren, in einer Datenbank für den späteren Aufruf zu speichern oder in die Lernplattform Moodle einzubinden.⁴

Ein weit gefasster Wortschatzbegriff

Im CALLIDUS-Projekt wird Wortschatz nicht nur als sprachtheoretisches Konstrukt (Lexikon), sondern als psychologische Gegebenheit verstanden, d. h. als mentales Lexikon, das als ein Zellverbund modelliert werden kann, in dem lexikalische Einheiten⁵ vielfach vernetzt gespeichert, verarbeitet und abgerufen werden. Der Zellverbund wiederum zerfällt in größere Teilnetze, z. B. Klang- oder Sachnetze (Neveling 2004, 42). Als kleinste Einheit gibt es im

mentalen Lexikon einzelne Zellnetze, die die verschiedenen Ebenen einer lexikalischen Einheit repräsentieren: Form, Funktion, Konzept (Höhle 2012, 56). Daraus ergibt sich, dass Wortschatz- und Grammatikerwerb im Sinne der Lexiko-Grammatik (vgl. *The lexical appro-*

ach von Lewis 2005) als eine Einheit betrachtet werden, weil unter einem kompetenten Umgang mit dem Wortschatz (Wortschatzkompetenz) der erfolgreiche Zugriff auf alle Repräsentationsebenen einer lexikalischen Einheit verstanden wird.

Tab. 1 Die Repräsentationsebene eines Zellnetzes, übertragen auf die Wissensbereiche des lat. Spracherwerbs

Repräsentationsebene	Wissensbereiche
Konzept (Bedeutung)	<ul style="list-style-type: none"> • (allgemeines) Welt-/Sachwissen • fachspezifisches Kontextwissen (z.B. Textsortenwissen, Alltagskonzepte)
Funktion (lexikalisch-syntaktische Eigenschaften)	<ul style="list-style-type: none"> • syntaktisches Wissen: Syntagmen (z. B. präpositionale Wendungen, satzwertige Konstruktionen, Funktionsverbgefüge) • semantisches Wissen: Kollokationen, Phrasen, Idiome • pragmatisches Wissen: kontextabhängiges sprachliches Wissen (z.B. Ironie, Andeutungen, Metaphern)
Form (lautliche Struktur und Schreibweise)	<ul style="list-style-type: none"> • phonetisch-phonologisches Wissen: Betonungsregeln, Verkürzungen, Verschleifungen • graphematisch-orthographisches Wissen: k/c, i/j, Homographie (z.B. <i>eo</i>) • morphologisches Wissen: Wortbildung, Bedeutung von Morphemen, Assimilation (z.B. <i>ad-/ac-cedere</i>)

DDL: Wortschatzarbeit mit Kontexten

Aus dem weit gefassten Wortschatzbegriff folgt, dass Wortschatzarbeit viel mehr als bisher auf die verschiedenen Repräsentationsebenen einer lexikalischen Einheit eingehen muss. Dazu ist es essentiell, den Kontext der lexikalischen Einheiten bei der Einführung, Wiederholung und Vertiefung von (unbekannten) Wörtern einzu beziehen, damit der Wortschatzerwerb, d. h. das Speichern im und das (fehlerfreie und schnelle) Abrufen von Informationen aus dem mentalen Lexikon, erfolgreich und nachhaltig ist.⁶ Hierbei wird vereinfacht alles als Kontext angesehen, das zusätzliche Informationen zu einer lexikalischen Einheit liefert. Im CALLIDUS-Projekt wurde dieser Kontext analog durch spezifische Ergänzungsmaterialien, z. B. eine kontextbasierte

Einführung neuer Vokabeln und entsprechende Testformate (Lehrbuch VIVA) oder die Integration von vielfältigen Wortschatzübungen in eine Textausgabe (Ovid, *Pyramus und Thisbe*), und digital durch das Textkorpus der Software und entsprechende intralinguale Übungsformate (*mark words, cloze, matching, keyword in context*) bereitgestellt. Die bisherigen Studienergebnisse legen zum einen nahe, dass die Lernenden Zeit benötigen, um sich auf DDL einzustellen, dass dies aber auch schon von sehr jungen Lernenden geleistet werden kann. Zum anderen zeichnet es sich ab, dass eine am Kontext ausgerichtete Wortschatzarbeit erfolgreicher ist als das immer noch vorherrschende Listenlernen (Beyer und Schulz 2020b).

Einsatzmöglichkeiten der Software *Machina Callida*

Die Software bietet einen Übungsgenerator, eine Übungsdatenbank und eine beispielhafte Wortschatzeinheit mit Testfunktion. Da sie dem DDL-Ansatz folgt, nutzt sie authentische Texte der lateinischen Literatur, die frei zugänglich sind und ohne Rechtsverletzung in einem Korpus zusammengefasst werden konnten. Das so entstandene Korpus beinhaltet Texte von über 100 Autoren (z. B. Cicero, Caesar, Ovid), deren Qualität jedoch im Einzelnen sehr unterschiedlich ausfällt, da es sich nur zu einem geringen Anteil um manuell kuratierte Textfassungen handelt. Dies wie auch die im Rahmen der Software angewandten Methoden der automatischen maschinellen Sprachverarbeitung (NLP) erklären gelegentliche systematische Fehler, die beim Erstellen der Übungen auftreten können.

Die *Machina Callida* richtet sich aufgrund ihres authentischen Kontextes und der intralingualen Übungen an fortgeschrittene Lernende in Schule und Universität. Die Einführung in die Softwarenutzung sowie die Erstellung der Übungen wird im schulischen Kontext vermutlich zunächst in der Hand der Lehrkraft liegen, da die Lernenden u. U. an der Auswahl des gesuchten Textabschnittes oder der (linguistischen) Übungsparameter scheitern könnten. Doch nach einer expliziten Erklärung sollte die eigenständige Erprobung und sinnvolle Nutzung der Software durch Lernende kein Problem darstellen. Im Gegenteil, es darf ggf. erwartet werden, dass sie den Lehrenden ihrerseits Anwendungsoptionen aufzeigen können.

1. Textauswahl

Die Textauswahl erfolgt üblicherweise mittels einer Liste von Autorinnen und Autoren, denen

wiederum die zugehörigen Werke untergeordnet sind. Einzelne Schriften folgen sodann der üblichen strukturellen Einteilung, z. B. in Buch, Kapitel und Abschnitt (Blackwell und Smith 2015). Größere Textmengen können stärker eingegrenzt werden, indem ihre Textkomplexität (Dascalu et al. 2017) und das enthaltene Vokabular automatisch geprüft werden. Daraus ergibt sich die Möglichkeit, nur diejenigen Texte zu verwenden, deren syntaktische Gestaltung gut zugänglich ist und deren Inhalt sich möglichst flächendeckend durch den Bamberger Wortschatz (Utz 2000) erschließen lässt. Der Abgleich eines Textes mit einem bestimmten Vokabular wird bei Bedarf visualisiert, sodass etwaige unbekannte Wörter optisch leicht identifiziert werden können. Um sie anschließend mit Übersetzungshilfen zu versehen, brauchen Lehrkräfte lediglich den Text mit seinen Markierungen als Word-Dokument herunterzuladen. Dort können sie dann alle weiteren Verarbeitungen vornehmen, die sie in ihrem Alltag sonst auch anwenden würden. Darüber hinaus sind nach demselben Verfahren auch Fehler in fertigen Übungen korrigierbar, z. B. wenn ein Adjektiv von der Maschine fälschlicherweise als Substantiv identifiziert wurde.

2. Übungsgenerator

Bei der Gestaltung der Übungen kommt es jedoch nicht nur auf die korrekte linguistische Analyse, sondern auch auf eine geeignete kognitive Progression (Crossley et al. 2010) an. Dementsprechend können mehrere Übungen als feste Sequenz geplant werden, in der Lernende zuerst basale, eher einfache Interaktionen zeigen müssen, um dann schrittweise anspruchsvollere Aufgaben zu lösen. Zu den Grundlagen gehört hierbei das Identifizieren von Wortformen nach vorgegebenen Kriterien, z. B. durch das

Anklicken aller Formen von *esse* in einem Text. Fortgeschrittene Lernende ordnen dann in einem memory-artigen Verfahren beispielsweise attributive Adjektive den morphologisch und semantisch passenden Substantiven zu.

Noch komplizierter wird es, wenn Lückentexte mit beinahe identischen Konjunktionen gefüllt werden sollen, etwa *et / atque* oder *vel / aut*. Hier reicht es nicht mehr aus, nur auf die äußere Gestalt oder die grobe Bedeutung der Wörter zu achten, sondern der gesamte Kontext und auch etwaiges Hintergrundwissen müssen herangezogen werden, um eine Lücke korrekt zu bestücken. Eine etwas anders gelagerte, aber strategisch ähnliche Herausforderung bietet die Keyword-in-Context-Ansicht (Helm 2009), mit der die Verwendung bestimmter Wörter über mehrere beispielhafte Textpassagen hinweg gezielt untersucht werden kann. So wird es möglich, etwa die Unterschiede zwischen *cum* als Präposition und als Konjunktion oder sogar die noch feineren Unterschiede zwischen *cum*-Sätzen mit Konjunktiv oder Indikativ zu erforschen.

3. Wiederverwendung der einmal erstellten Übungen

Neben den erwähnten Exportformaten (PDF- und Word-Dokument) stehen alle Übungen auch im XML-Format oder als komprimierte Archive zur Verfügung, so dass sie maschinell verarbeitbar sind. Diese Verarbeitung kann dann zu längerfristigen didaktischen Analysen beitragen, wie es in den Untersuchungen des CALLIDUS-Projekts der Fall war. Sie dient aber gleichzeitig auch zur Nachnutzung der Übungen in fremder Software, z. B. in der Lernplattform Moodle (Moodle Pty Ltd. 2020). So können Übungen dort problemlos hochgeladen, modifiziert und ausgewertet werden.

Wer hingegen eine eigene Website hat, aber das Übungsmaterial nicht auf einem eigenen Server bereitstellen möchte, kann stattdessen implizit auf die öffentliche Instanz der *Machina Callida* an der Humboldt-Universität zu Berlin verweisen. Diese Art der Einbettung hat den Vorteil, dass die Übungen mit ihrer Interaktivität auf jeder beliebigen Website einsehbar und bearbeitbar werden. Dabei benötigen sie jedoch keinen zusätzlichen Speicherplatz mehr, weil alle Daten weiterhin zentral von der Universität bereitgestellt werden. Der simpelste Weg für Lehrkräfte ist allerdings das Verschicken von Links, die direkt auf bestimmte Übungen aus der öffentlichen Übungsdatenbank verweisen. Dadurch profitieren die Lernenden auch vom einheitlichen Layout und können selbständig weitere Übungen erstellen, abrufen oder lösen.

Die Wortschatzarbeit der Zukunft

Wie seit Langem immer wieder erörtert und auch in den ersten beiden CALLIDUS-Studien belegt, gelingt der Wortschatzerwerb im LU nicht in dem Maße, wie es sich alle Beteiligten wünschen. Die mannigfaltigen Vorschläge haben die Unterrichtspraxis über die Jahrzehnte hinweg mehr oder weniger bereichert, doch letztlich wurden weder das Listenlernen bzw. das Abfragen von Wortgleichungen abgeschafft noch hat sich in den Lehrbüchern der Fokus vom Grammatik- zum Wortschatzerwerb verschoben. Auch die zunehmende Anzahl an Vokabellernmaterialien (Vokabelhefte, Vokabeltrainer, Apps) oder Wortschatzübungen in den Lehrbüchern scheint durch spracherwerbstheoretische Kenntnisse (z. B. mentales Lexikon, Modelle zum fremdsprachlichen Lexikonerwerb) und Methoden des Unterrichts in modernen Fremdsprachen (z. B. DDL, Lernstrategien) kaum beeinflusst worden zu sein. Durch die

Vorstellung des Forschungsprojektes sowie der darin entstandenen Materialien und Software soll gezeigt werden, dass eine Veränderung der täglichen Wortschatzarbeit im LU nicht nur notwendig ist, sondern durch die Inspiration aus anderen wissenschaftlichen Disziplinen nicht so schwierig ist, wie es erscheinen mag: Warum sollte der LU nicht vom Wortschatz her, d. h. von den kontextsensitiven Bedeutungen der lexikalischen Einheiten gedacht werden? ‚Wörter‘ eröffnen nicht nur den Zugang zu der unbekanntem Welt des LU (Waiblinger 2002), sondern bieten auch wichtige Anknüpfungspunkte an individuelles Vorwissen, das für den Lexikonerwerb von großer Bedeutung ist (Schneider und Lindenberger 2018, 438). Hinzu kommt, dass die lexikalischen Einheiten je nach Kontext zusätzliche wissenswerte und lernförderliche Details bieten, die dazu beitragen können, dass die Wortschatzarbeit im LU abwechslungsreich und spannend wird. Wenn außerdem der neue Wortschatz durch zusätzliche Komponenten wie (gestützte) Sprachproduktion, digitale intralinguale Übungsformate und metakognitive Strategien ‚erlebt‘ wird, verlore die (gern in die Hausaufgabe verlagerte) Wortschatzarbeit vielleicht etwas von ihrem negativen Beigeschmack eines mühevollen und wenig freudvollen Vorganges. Dass durch alle diese Veränderungen auch das nachhaltige Lernen der Vokabeln positiv beeinflusst wird, darf dann als angenehmer Nebeneffekt bezeichnet werden.

Literaturverzeichnis

- Beyer, A. (2020). *New Insights and Methods of Vocabulary Acquisition in Latin Classes (Preprint)*, Forma y Función. <https://zenodo.org/record/3944100>.
- Beyer, A. & Schulz, K. (2020a). CALLIDUS – Korpusbasierte, digitale Wortschatzarbeit im Lateinunterricht. In S. Chronopoulos, F. K. Maier & A. Novokhatko (Hrsg.), *Digitale Alter-*

tumswissenschaften. Thesen und Debatten zu Methoden und Anwendungen (Digital Classics Books, Bd. 4, 1. Aufl., S. 149-167). Heidelberg: Propylaeum.

- Beyer, A. & Schulz, K. (2020b). Using NLP to Create Corpus-based Vocabulary Exercises in Latin Classes. In L. Gómez Chova, A. López Martínez & I. Candel Torres (Hrsg.), *INTED2020 Proceedings* (INTED Proceedings, S. 1750-1757). IATED.
- Blackwell, C. & Smith, N. (2015). *The Canonical Text Services URN Specification, Version 2.0.Rc.1*.
- Boulton, A. (2017). Data-Driven Learning and Language Pedagogy. In S. L. Thorne & S. May (Hrsg.), *Language, Education and Technology* (Encyclopedia of Language and Education, 3rd, S. 181-192). Cham: Springer International Publishing.
- Crossley, S. A., Salsbury, T. & McNamara, D. S. (2010). The development of semantic relations in second language speakers: A case for latent semantic analysis. *Vigo International Journal of Applied Linguistics* 7, 55-74. <https://pdfs.semanticscholar.org/5f05/7cfe49a7496b-453489f03029440e44d804ef.pdf>.
- Dascalu, M. A., Gutu, G. S., Ruseti, S. S., Cristian Paraschiv, I. S., Dessus, P., Mcnamara, D. A., Crossley, S. A. & Trausan-Matu, S. A. (2017). ReaderBench: A Multi-Lingual Framework for Analyzing Text Complexity. In É. Lavoué, H. Drachler, K. Verbert, J. Broisin & M. Pérez-Sanagustín (Hrsg.), *Data Driven Approaches in Digital Education, Proc 12th European Conference on Technology Enhanced Learning, EC-TEL 2017* (Data Driven Approaches in Digital Education 12th European Conference on Technology Enhanced Learning, EC-TEL 2017, Tallinn, Estonia, September 12-15, 2017, Proceedings, S. 606-609). Tallinn, Estonia: Springer.
- Gage, N. M. & Baars, B. J. (2018). Language and Thought. In N. M. Gage & B. J. Baars (Hrsg.), *Fundamentals of Cognitive Neuroscience. A Beginner's Guide* (2. Aufl., S. 185-214). Academic Press.
- Gilquin, G. & Granger, S. (2010). How Can Data-driven Learning Be Used in Language Teaching. In A. O'Keeffe & M. McCarthy (Hrsg.), *The Routledge handbook of corpus linguistics* (1st, Bd. 359370, S. 359-370). New York: Routledge.
- Harm, V. (2015). *Einführung in die Lexikologie*. Darmstadt: WBG.

- Helm, F. (2009). Language and Culture in an Online Context: What Can Learner Diaries Tell Us about Intercultural Competence? *Language and Intercultural Communication* 9 (2), 91-104. doi:10.1080/14708470802140260
- Höhle, B. (Hrsg.). (2012). *Psycholinguistik*. Berlin: Akademie Verlag.
- Lewis, M. (2005). *The Lexical Approach. The State of ELT and a Way forward*. Ismaning: Hueber.
- Moodle Pty Ltd. (2020). Moodle Documentation. https://docs.moodle.org/39/en/About_Moodle.
- Neveling, C. (2004). *Wörterlernen mit Wörternetzen. Eine Untersuchung zu Wörternetzen als Lernstrategie und Forschungsverfahren* (Gießener Beiträge zur Fremdsprachendidaktik). Tübingen: Narr (Zugl.: Berlin, Freie Univ., Diss, 2004).
- Schneider, W. & Lindenberger, U. (2018). Gedächtnis. In W. Schneider & U. Lindenberger (Hrsg.), *Entwicklungspsychologie* (8., überarbeitete Auflage, S. 423-444). Weinheim: Beltz.
- Schulz, K. (2020). *Building a Database of Exercises for Learning Latin (Preprint)*, New Classicists. <https://zenodo.org/record/3735032>.
- Schulz, K. & Beyer, A. *A data-driven platform for creating educational content in language learning*.
- Utz, C. (2000). Mutter Latein und unsere Schüler – Überlegungen zu Umfang und Aufbau des Wortschatzes. In P. Neukam (Hrsg.), *Antike Literatur – Mensch, Sprache, Welt* (Dialog – Klassische Sprachen und Literaturen, Bd. 34, S. 146-172). München.
- Waiblinger, F. P. (2002). Wortschatzerwerb im Lateinunterricht. https://www.fachdidaktik.klassphil.uni-muenchen.de/forschung/didaktik_waiblinger/wortschatzerwerb.pdf.

Anmerkungen:

- 1) Unter einer ‚historischen Sprache‘ werden in diesem Beitrag nur die Sprachformen verstanden, die im Alltag nicht mehr gesprochen werden und somit nur noch in schriftlicher Form erhalten sind, neben Latein z. B. das Mittelhochdeutsche.
- 2) An dem Projekt sind die Fachdidaktik Latein, die Korpuslinguistik und der Computer- und Medienservice der HU Berlin beteiligt. Es wird durch die Deutsche Forschungsgemeinschaft (DFG) unter der Projektnummer 316618374 gefördert.
- 3) Die Ergebnisse werden zusammengefasst in Beyer (2020).
- 4) Über die Software informieren folgende Beiträge: Beyer und Schulz (2020a), (2020b); Schulz (2020); Schulz und Beyer (2020).
- 5) Aus linguistischer Sicht ist laut Harm (2015, S.16) „Wort“ (~ Lexikoneintrag) als ein Konzept mit unscharfen Rändern zu verstehen, das eine lexikalische Einheit umfasst, deren Größe durch syntaktische Regeln oder gedankliche Konzepte bestimmt wird. Eine lexikalische Einheit kann daher aus einem Wort oder mehreren Wörtern bestehen, z. B. *res publica*, das als Einheit einen anderen Sinn bietet als die Summe der einzelnen Wortbedeutungen.
- 6) Wichtig für das erfolgreiche Lernen einer lexikalischen Einheit, d. h. für das Abspeichern im Langzeitgedächtnis, sind neben der kontextuellen Einbettung die angewandten (Lern-)Strategien und das Vorwissen. Auf beide Aspekte wird in diesem Beitrag nicht weiter eingegangen, sondern auf die Literatur verwiesen, z. B. Neveling (2004).

ANDREA BEYER UND KONSTANTIN SCHULZ