

Worte, die Studierenden fehlen: eine Schlüsselwortanalyse der L1- und L2-Wissenschaftssprache

Tomáš MACH

Abstract

Words that students lack: a keyword analysis of L1 and L2 academic language

The article outlines the options for a pedagogically motivated keyword analysis in order to discover differences between the academic writing of Czech and Austrian students of German philology (or related fields). First, the paper presents a brief overview of the presently available metrics employed in keyword analyses, outlining some modifications that can be applied to make the results more relevant and easily interpretable. In the practical part, I explore which lexical items are frequently used by the L1 writers but at the same time are underused in the L2 subcorpus. The product of the study is a word list, which can be utilised by academic writing instructors or students themselves to provide them with an idea of what needs to be taught/learned.

Keywords: academic writing, DaF, corpus, quantitative analysis, keyword analysis

DOI: 10.15452/StudiaGermanistica.2022.31.0002

1. Einleitung

Die Fähigkeit, wissenschaftliche Texte zu verfassen, gehört ohne Zweifel zu einer der wichtigsten Kompetenzen, die innerhalb eines philologisch orientierten Studiums angestrebt werden. Ihr erfolgreicher Erwerb befähigt Studierende, im Rahmen der jeweiligen Disziplin diskursiv zu handeln und dadurch ihren Status von passiven Rezipienten zu aktiven Mitgestaltern zu ändern. Da allerdings nicht alle ihre Zukunft in der Wissenschaft und Forschung sehen, kommt die wissenschaftliche Schreibkompetenz hauptsächlich beim Verfassen von Abschlussarbeiten zum Tragen. Vor allem für fremdsprachliche Studierende kann sich der Schreibprozess als besonders anstrengend erweisen, denn er erfordert eine erfolgreiche Bewältigung der Wissenschaftssprache bzw. deren stilistischer Vielfalt. Um diesen Prozess möglichst viel zu erleichtern, ist es essenziell, die textsortenspezifische Gebrauchsweise umfassend zu charakterisieren und anschließend didaktische bzw. didaktisch relevante Materialien zu erstellen, die bei der Aneignung der wissenschaftssprachlichen Besonderheiten behilflich sein können. Dieses Ziel wird auch in diesem Beitrag verfolgt.

Zunächst bietet die vorliegende Studie einen allgemeinen Überblick über die Diskussion zur Rolle von Korpora und verschiedenen Metriken zur Bestimmung des didaktisch relevanten

Wortschatzes. Danach wird eine kontrastiv-korpusgesteuerte Untersuchung vorgenommen, die basierend auf Schlüsselwortanalyse lexikalische Lücken bei tschechischen Germanistikstudierenden zu enthüllen sucht. Daraus ergibt sich schließlich eine Wortliste, die sowohl von Studierenden selbst als auch von Lehrkräften/Kursleitenden in Anspruch genommen werden kann.

2. Wissenschaftssprache und ihre Entwicklung

Die Sprache der Wissenschaft umfasst die fachspezifische Terminologie auf der einen Seite und die textsorten- bzw. diskurspezifischen linguistischen Merkmale auf der anderen. Für Letzteres hat sich seit Jahrzehnten Ehlichs (1999) Begriff der „allgemeinen Wissenschaftssprache“ (AWS) durchgesetzt, die auch den Gegenstand dieser Untersuchung bildet. Was die AWS auszeichnet ist ihre Musterhaftigkeit, Abstraktion, Konventionsgebundenheit und Interdisziplinarität. Sie ist „Bestandteil, Resultat und zugleich Voraussetzung der Wissenschaftskommunikation, also unabdingbar für jeden, der sich am deutschen Wissenschaftsbetrieb beteiligen will“ (Ehlich/Graefen 2001:373, zit. n. Prutscher 2020:218). Sprachdidaktisch gesehen stellt sie jedoch kein isoliertes Phänomen dar, sondern ist mit der allgemeinen Sprachkompetenz eng verflochten, und hängt von deren Niveau ab. Unter diesem Gesichtspunkt ist es prinzipiell ausgeschlossen, dass Studierende in der Lage wären, die AWS kompetent zu benutzen, ohne gleichzeitig über ein höheres Sprachniveau zu verfügen. Dieses soll wiederum als eine Voraussetzung für die Aneignung der AWS verstanden werden, denn entscheidend dabei ist letztlich systematisches Üben, wodurch sich die einschlägige Schreiberfahrung erwerben lässt.

Um die einzelnen Stufen des Erwerbsprozesses beschreiben zu können, wird hier Steinhoffs (2007) Modell zur Entwicklung der wissenschaftlichen Textkompetenz herangezogen. Darin unterscheidet er zwischen vier Phasen: Transposition, Imitation, Transformation und kontextueller Passung. Transposition wird als eine Schreibstrategie aufgefasst, wobei Schreibende auf ihre Erfahrungen mit anderen, ihnen vertrauten Textarten zurückgreifen: „Ein möglicher Einflussbereich ist der schulische Aufsatzunterricht, ein anderer der Journalismus“ (Steinhoff 2007:140). Aus der Übernahme fremder Muster resultieren unzählige Konventionsbrüche, weswegen die auf diese Art produzierten Texte von Steinhoff als präkonventionell bezeichnet werden. Ähnliches gilt für Imitation, einen „Versuch zur Bildung (über)komplexer Ausdrücke, zur Überdehnung von Phrasen und zur übertriebenen Streckung des Satzgefüges durch eine Verschachtelung von Nebensätzen. Hinzu kommen einzelne dezidiert schriftsprachlich markierte Mittel, die offenbar fachliche Souveränität demonstrieren sollen“ (Steinhoff 2007:144). Charakteristisch für Imitation ist der Fokus auf Form – der Text soll von der Sprache her in erster Linie wissenschaftlich aussehen, seine kommunikative Funktion bzw. deren Erfüllung bleibt sekundär. Darin unterscheidet sie sich von Transformation, die einerseits einen quantitativen Aufbau des wissenschaftlichen Ausdrucksspektrums und ein erhöhtes Bewusstsein für dessen Funktionalität mit sich bringt, andererseits kann es nach Steinhoff zu folgendem Problem kommen:

„Lerner machen sich mit den neuen Mitteln vertraut und gebrauchen sie einige Zeit häufig, womöglich sogar zu häufig. Texte professioneller Wissenschaftler zeichnen sich dagegen vermutlich eher durch einen variablen, differenzierten Sprachgebrauch aus.“ (Steinhoff 2007:146)

Die letzte, den stilistischen Gepflogenheiten bereits entsprechende Phase ist die sog. kontextuelle Passung. Der Unterschied zu den vorherigen, präkonventionellen Sprachgebrauchsstufen besteht darin, dass „die Schreiber nunmehr grundsätzlich in der Lage sind, wissenschaftliche Schreibprobleme wissenschaftlich zu lösen“ (Steinhoff 2007:149).

Angemerkt sei, dass Steinhoffs Entwicklungsmodell und somit auch die Einstufung im Hinblick auf muttersprachliche Verfasser formuliert wurden. Es fragt sich also, inwiefern sie auf L2-Lernende¹ übertragen werden können und wie sich tschechische Nichtmuttersprachler innerhalb des

¹ Als L2 werden in diesem Beitrag Lernende bezeichnet, die eine andere Muttersprache als Deutsch haben.

Modells verorten lassen. Dafür wird ein kurzer Einblick in den tschechischen Bildungskontext benötigt. Meinen Ausgangspunkt hier bildet die Annahme, dass die voruniversitäre Bildung in ihrer aktuellen Form außerstande ist, die Lernenden mit genügenden Schreibkompetenzen auszustatten, die sich zur Lösung wissenschaftlicher Aufgaben verwerten ließen. Konkret wird damit gemeint, dass im muttersprachlichen Unterricht das Spektrum möglicher Textsorten dermaßen breit ist, dass denen, die sich mit wissenschaftlichem Stil überlappen (z. B. ein Essay oder Referat), nicht genug Aufmerksamkeit gewidmet werden kann. Dasselbe ist auch im Fremdsprachenunterricht der Fall; darüber hinaus werden aufgrund der Zielniveaus (B2 für Englisch bzw. B1 für die zweite Fremdsprache) erwartungsgemäß kaum längere und komplexe Aufsätze geübt.

Während ihres Bachelor-Studiums werden folglich L2-Lernende vor die Aufgabe gestellt, sich die alltägliche Wissenschaftssprache zu eigen zu machen, ohne genau zu wissen, was die Alltagssprache überhaupt ausmacht. Das Erlernen der AWS erfolgt in vielen Fällen vermutlich ohne jeglichen Bezug auf das bei Muttersprachlern vorausgesetzte Vorwissen. Während bei L1-Schreibern nur von mangelnder Schreiberfahrung im Bereich Wissenschaftssprache die Rede sein kann, haben tschechische L2-Verfasser hingegen mit

- mangelnder Schreiberfahrung im allgemeinen Sinne,
- mangelnder Schreiberfahrung in einer Fremdsprache und
- mangelnder Erfahrung mit der jeweiligen Sprache selbst zu kämpfen.

So betrachtet steht fest, eine ganzheitliche Entwicklung der Schreibfähigkeit wird über alle vier Stufen hinweg in drei (Bachelor) oder fünf (Bachelor + Master) Jahren vielleicht nur einigen, jedenfalls nicht allen gelingen. Den Phasen nach kann sich die Entwicklung für Nichtmuttersprachler folgendermaßen gestalten: Transponiert wird nicht besonders viel, denn zu Beginn des Studiums erweist sich die Vertrautheit mit dem journalistischen Stil sowie der Aufsatzproduktion wohl ebenfalls als verschwindend gering. Es wird jedoch Gebrauch von vertrauten, alltäglichen Sprachmustern, Wortschatz und Grammatik gemacht. Eine große Rolle spielt m. E. die Imitation; imitative Strategien zielen auf die Aneignung der Basislexik und der usuellen Konstruktionen des wissenschaftlichen Schreibstils und ermöglichen eine Partizipation am Diskurs. Zum Auftreten von überkomplexen Strukturen (wie bei L1-Verfassern) kommt es meiner Erfahrung nach relativ selten. Umso mehr häufen sich immerhin Probleme stilistischer Art an (vgl. beispielsweise Graefen 2002:10–16). Die Transformationsphase kennzeichnet sich besonders bei L2-Lernenden durch die Neigung zur Wiederholung. Hasselgren (1994) spricht von den sog. ‚lexical teddy bears‘, d. h. den auf der Ebene des Idiolekts präferierten und vergleichsweise zu häufig verwendeten Ausdrücken. Angesichts ihres Umfangs eignen sich daher Abschlussarbeiten zu Analysen der sprachlichen Variation, bzw. deren Absenz besonders gut.

Dass die L2-Schreibproduktion in vielerlei Hinsicht Lücken aufweist, wird sowohl in den Lehrmaterialien als auch in der Forschung reflektiert. Dabei werden im Wesentlichen der Wortschatz und Funktionen fokussiert, sei es in Form einzelner Lexeme, Kollokationen oder größerer Einheiten (vgl. Fügler/Richter 2017; Graefen 2009; Graefen/Moll 2009; Prutscher 2020). Tendenzen zur Erforschung wissenschaftssprachlicher Lexik und Pragmatik spiegeln sich auch in den kontrastiven Studien in Tschechien wider (siehe z. B. Goldhahn 2016; Hotařová 2019; Mach 2020a, 2020b; Polášková 2016). Gemeinsam diesen Studien ist ihr korpusbasiertes Design, d. h. die eventuellen Unterschiede erschließen sich als sekundäres Ergebnis einer konkreten, vorgegebenen Fragestellung. Da diese den Forschungsgegenstand darstellt und nicht die Unterschiede selbst, kann es mitunter schwerfallen, die Bedeutung der festgestellten Differenzen im Rahmen der gesamten Textmenge einzuschätzen. Dem lässt sich zum Teil entgegenreten, indem kein eng definiertes Ziel angekündigt und verfolgt wird. Da bisher keine explorative, korpusgeleitete Untersuchung zu L1-L2-Kontrasten in der Wissenschaftssprache vorliegt, wird hier ein solcher Versuch unternommen. Im nächsten Abschnitt beleuchte ich hierzu die fundamentalen Prinzipien der Schlüsselwortanalyse, stelle verschiedene Metriken zu ihrer Auswertung vor und reflektiere kurz über ihr Potenzial für die Didaktik der Wissenschaftssprache, wobei ich die konkrete Herangehensweise in dieser Untersuchung darlege.

3. Schlüsselwortanalyse, ihre Potenziale und Auswertung

Die Schlüsselwortanalyse (*keyword analysis*) setzt sich zum Ziel, für ein Korpus typische Einheiten (Wortformen, Lemmata, N-Gramme u. Ä.) ausfindig zu machen, und zwar mithilfe eines Referenzkorpus, gegenüber dem die Frequenzen verglichen werden können. Anders ausgedrückt: „[w]enn sich die Frequenz eines Wortes in einem Korpus A im Vergleich zu einem oder mehreren anderen Korpora (B, C, ...) erheblich voneinander unterscheidet, ist dieses Wort für Korpus A ein Keyword“ (Bubenhöfer/Scharloth 2015:7). Der relative Ausdruck *erheblich unterscheiden* deutet auf die Notwendigkeit einer Quantifizierung des Verhältnisses zwischen den verglichenen Einheiten hin. Zu diesem Zweck stehen uns verschiedene, auf Effektgröße basierende Keyness-Maße zur Verfügung, darunter z. B. *Ratio*, *Odds-Ratio*, *Log-Ratio*, *%DIFF* oder *Difference Coefficient* (für eine genaue Beschreibung siehe Gabrielatos 2018). Das für diese Analyse eingesetzte korpusanalytische Instrument Sketch Engine (Kilgarriff et al. 2014) bedient sich eines weiteren Keyness-Maßes: *simple maths* (Kilgarriff 2009). Dabei lässt sich der Häufigkeitsbereich der gesuchten Einheiten spezifizieren, indem zu ihrer relativen Frequenz immer 10^x ($x \in \{1,2,\dots,6\}$) addiert wird (siehe Kilgarriff 2009 für Beispiele und Begründung).

Was *simple maths* jedoch nicht berücksichtigt, ist die Streuung (Dispersion). Über den zunehmend hohen Stellenwert dieses Parameters scheint in der Korpuslinguistik ein allgemeiner Konsens zu herrschen (Brommer 2018; Egbert/Biber 2019; Gries 2008). Die Tatsache, dass Streuungsmaße zwar unentbehrlich sind, dennoch deutlich weniger zum Einsatz kommen als andere quantitative Eigenschaften, ist aus dem folgenden Zitat ersichtlich:

„One underlying – but usually unrecognised – assumption of this approach is that the corpus is homogeneous, and thus words are evenly distributed across the corpus. In actual fact, though, corpus frequency keywords can be, and often are, frequent in a corpus, but are not widely dispersed across the texts of that corpus. As a result, such words are not truly typical of the discourse domain represented by the corpus.“ (Egbert/Biber 2019:78)

Ähnlich wie Keyness lässt sich Dispersion auf verschiedene Art und Weise operationalisieren, z. B. durch Maße wie *range*, Standardabweichung, Juilland's *D*, *max-min difference*, um nur einige zu nennen. Der Artikel von Gries (2008) bietet einen exhaustiven Überblick von möglichen Dispersionsmaßen mit ihren Vor- und Nachteilen sowie einen Entwurf eines neuen (*DP*). In der Schlüsselwortanalyse in Sketch Engine werden folgende Maße berechnet: *range* (DOCF genannt), ARF und ALDF.² DOCF (*document frequency*) steht für Dokumenthäufigkeit und wird sowohl in absoluten als auch relativen Zahlen angegeben. Es handelt sich um die Anzahl bzw. den prozentualen Anteil von Dokumenten im Korpus, wo die Einheit mindestens *n*-mal zu finden ist. ARF (*Average Reduced Frequency*) (Savický/Hlaváčová 2002) entstand als Korrektur von RF (Hlaváčová/Rychlý 1999). Bei der Berechnung von ARF wird das Korpus zunächst in *f* gleiche Segmente aufgeteilt (*f* bezeichnet die absolute Frequenz der gesuchten Einheit). RF ist dann die Anzahl von Segmenten, wo die untersuchte Einheit vorhanden ist. Da sich ein Korpus allerdings vielfach segmentieren lässt – je nachdem, an welcher Position man beginnt, wird RF für all diese Segmentierungsmöglichkeiten ermittelt und daraus ein arithmetisches Mittel berechnet. Diese Zahl entspricht der ARF. ALDF (*Average Logarithmic Distance Frequency*) verhält sich in der Praxis ähnlich wie ARF; die Werte der analysierten Einheit im Fokus- und Referenzkorpus stehen in vergleichbarem Verhältnis zueinander (für Details der Berechnung und einige geringfügige Unterschiede zu ARF sei auf Savický/Hlaváčová 2002 verwiesen).

Durch Korpora gewonnene Erkenntnisse finden immer häufiger Anwendung in der Erstellung von DaF-Materialien (vgl. Perkuhn 2021). Die Fragen nach dem Unterrichtsfokus (d. h. *Was sind die „key“ Wortschatzeinheiten?*) und dessen allgemeinen Relevanz (*Sind die Einheiten generell genug/nicht allzu spezifisch?*) oder Reproduzierbarkeit (Targońska 2019) bilden keine Ausnahme.

² Mehr Informationen zu den angegebenen Metriken bietet die Webseite von Sketch Engine (URL 1: <https://www.sketch-engine.eu/documentation/statistics-used-in-sketch-engine/>). Eine gute Einführung in die Problematik der Häufigkeitsmaßen ist in Perkuhn (2021) zu finden.

Wie nützlich und informativ die genannten korpuslinguistischen Begrifflichkeiten auch sein mögen, sollen bei ihrer Anwendung auf Lösung sprachdidaktischer Fragen praktische Bedürfnisse nicht außer Acht gelassen werden. Was genau sich hinter dieser Forderung verbirgt, soll die Analyse und die Erklärung der einzelnen Schritte davor im folgenden Abschnitt anschaulich machen.

4. Die Analyse

Die Ergebnisse einer didaktisch motivierten Untersuchung sollen in erster Linie verständlich und leicht interpretierbar sein (möglicherweise auch zulasten der Genauigkeit), sodass Lehrkräfte und Studierende sie ohne umfangreiches korpuslinguistisches Wissen problemlos in die Unterrichtspraxis umsetzen können. Des Weiteren soll Relevantes, d. h. regelmäßig Verstreutes, Häufiges und Allgemeines in den Vordergrund gerückt werden. Im Mittelpunkt sollen Unterschiede stehen, denn das Ziel besteht hier darin, Wissenslücken (nicht Lernerfolge) aufzudecken.

Für die Analyse wurde das DESTWIKO (v0.2) Korpus verwendet, das im Rahmen von dem Förderprogramm ELEXIS-MU in Sketch Engine frei zugänglich ist. Die Datenbank umfasst ca. 5,6 Mio. Wortformen und setzt sich aus 280 germanistischen Masterarbeiten zusammen. Davon sind 100 an der Universität Wien und 180 an der Masaryk Universität (Institut für Germanistik, Nordistik, und Niederlandistik) eingereicht worden (für weitere Informationen zum Korpus und seinem Aufbau siehe Mach, 2020b). Die Auswahl des Korpus liegt in den Ähnlichkeiten zwischen dem muttersprachlichen und nichtmuttersprachlichen Teil begründet, denn wesentliche Unterschiede zwischen den verglichenen Korpora sind bei Schlüsselwortanalysen unerwünscht, so z. B. die Warnung Kilgarriffs (2009): „[t]he list may well be dominated by other differences, which we are not at all interested in. Keyword lists tend to work best where the corpora are very well matched in all regards except the one in question“ (Kilgarriff 2009:1). Es ist deswegen vorteilhaft, dass

- die Texte in DESWIKO ähnliche Themen behandeln,
- die Verfassernden gleichen Alters sind und praktisch das Gleiche studieren und
- literarische und sprachwissenschaftliche Arbeiten ausgewogen repräsentiert sind.

Trotz der zuweilen unterschiedlichen Länge der einzelnen Dokumente über die zwei Korpusteile hinweg, sind beide Subkorpora gut vergleichbar, solange dies bei der Entscheidung über die zu verwendenden Metriken und Ergebnisinterpretation berücksichtigt wird.

Nun kurz zu dem Verfahren und den Metriken. Die Funktion *Keywords* in Sketch Engine ermöglicht nicht nur Vergleiche zwischen Korpora, sondern auch einzelner Subkorpora, wobei das Erste als Fokus- und das Zweite als Referenzkorpus fungiert. Für diese Studie wurde also der muttersprachliche Teil zum Fokuskorpus während der Nichtmuttersprachliche als Referenz konzipiert war. Wie bereits oben erwähnt, ist für die Berechnung von Kilgarriffs (2009) *simple score* eine Entscheidung hinsichtlich des Frequenzbereichs erforderlich. Die Konstanten 10 und 100 waren im Hinblick auf die relativen Häufigkeiten der gewünschten Ergebnisse (AWS) m. E. am anschaulichsten. Auf der einen Seite produzierte die Konstante 100 eine relevantere Liste, indem nur wenige niederfrequente Lemmata hohe Scores hatten. Auf der anderen Seite waren sich die *simple score* Werte sehr ähnlich und fielen somit nicht besonders eindeutig aus. Daher wurden die unten angezeigten Werte mit 10 ausgerechnet; auch in Anbetracht dessen, dass neben *simple score* andere Metriken miteinbezogen werden, wodurch sich die zu spezifischen Termini ausfiltern lassen. Zu diesem Zweck kamen folgende Angaben in Frage: ARF, DOCF und relative DOCF (siehe oben). Aus der didaktischen Sicht war ARF an sich nicht besonders informativ – als ein Frequenz und Häufigkeit kombinierendes Maß kann sie jedoch beim Vergleich und bei der Prioritätensetzung in der endgültigen Wortliste helfen. Da sich die Anzahl von Dokumenten in den Subkorpora unterscheidet, wurde mit dem relativen Wert von DOCF operiert. Wie bereits erwähnt, drückt dieses Dispersionsmaß den prozentualen Anteil der Dokumente aus, in denen das Wort mindestens n -mal zu finden ist. Um einmalige Vorkommen einer Wortform auszuschließen (beispielsweise in einem übernommenen Zitat), wird der Wert von n in DOCF nicht wie gewöhnlich auf 1, sondern auf 2 gesetzt.

DOCF scheint sich vortrefflich zur Klärung der Frage zu eignen, welche Wörter den L2-Verfassenden am meisten fehlen. Diese Frage muss immerhin operationalisiert werden. Im Grunde genommen geht es darum, bei welchen Lemmata die größten Unterschiede zwischen den DOCF-Werten festzustellen sind. Der Übersichtlichkeit halber habe ich beide Werte voneinander subtrahiert, woraus sich diffDOCF ergibt. Dazu ein illustrierendes Beispiel: falls DiffDOCF bei dem Wort *Methode* 30 beträgt, bedeutet dies, dass der Ausdruck in 30 % mehr L1-Arbeiten gebraucht wird. Läge der Wert auf 0, würde der Ausdruck im gleichen Anteil von Arbeiten gebraucht. Auf diese Weise gelangen wir zwar zu Informationen über ungleichmäßig verteilten Gebrauch, der allgemeinen Natur von AWS wird aber keine Rechnung getragen, denn zu diffDOCF = 30 führt 100–70 genauso wie 40–10. Erstes verteilt sich über alle Arbeiten im L1-Subkorpus und kann zurecht als allgemeinsprachlich bezeichnet werden, wohingegen bei Letzterem mit 40 % wir auf solche Beurteilung entweder vollkommen verzichten würden oder sie deutlich vorsichtiger formulieren müssten. Damit also AWS im Mittelpunkt bleibt, legte ich drei Stufen von „Allgemeinheit“ mit Schwellenwerten 90, 80 und 70 fest. In diesem Zusammenhang macht Baker (2004 zit. n. Egbert/Biber 2019:82) auf Folgendes aufmerksam: „Among studies that use range to account for keyword dispersion there is no consensus on the ideal text frequency or percentage to use as a threshold.“ Daraus folgt, dass die Werte arbiträr und intuitiv gewählt werden mussten. Um ihre Bedeutung nochmals zu illustrieren, entspricht 90 dem Anteil (90 %) des L1-Subkorpus, in dem wenigstens zwei Wortformen des jeweiligen Lemmas anzutreffen sind. Werden nur Lemmata aus den drei Stufen in die Analyse eingeschlossen, wird die didaktische Relevanz der so aufbereiteten Daten sichergestellt und wir können uns zugleich nur auf ein Maß (diffDOCF) orientieren (im Falle niedrigerer Werte lohnt es sich immerhin, auch ARF und *simple score* in Betracht zu ziehen, weshalb beide auch angezeigt werden). Darüber hinaus lässt sich dadurch größtenteils das Problem der unterschiedlichen Textlänge umgehen. Falls mehr als 90/80/70 % aller Texte in einem Subkorpus zwei konkrete Spracherscheinungen (oder sogar noch mehrere Instanzen davon) beinhalten, erscheint sinnvoll anzunehmen, jene Erscheinung sei in gewissem Maße typisch für die Gesamtheit dieser Texte.

Aufbauend auf den methodischen Vorüberlegungen wurde in Sketch Engine eine Liste von 10,000 Lemmata generiert. Diese wurde anschließend dem oben skizzierten (Ausschluss)Verfahren unterzogen, woraus die folgenden drei Tabellen entstanden sind. Nur Lemmata mit diffDOCF > 20 sind in den Tabellen aufgeführt.

100–90 %				
Lemma	diffDOCF _{rel}	simple maths	ARF(Fok)	ARF(Ref)
wiederum	60,89	4,29	393,95	73,48
durchaus	59,44	4,88	315,09	52,65
hingegen	57,67	4,57	412,65	67,51
jeweils	53,22	4,46	299,68	56,51
eingehen	51,22	3,45	336,50	80,15
somit	47,67	5,89	963,60	125,54
zumindest	47,22	4,41	366,96	76,16
dennoch	46,56	4,70	535,16	89,74
vielmehr	46,56	2,76	255,56	88,88
begründen	42,78	2,57	211,25	81,76
annehmen	41,89	2,24	318,83	128,77
vorhanden	41,11	2,16	281,02	108,85

jeweilig	41,00	2,70	385,88	126,28
eben	40,89	3,11	386,51	118,74
tatsächlich	40,56	3,76	445,45	117,20
anhand	38,89	2,36	425,75	147,36
ebenfalls	38,11	2,99	668,21	175,45
indem	37,78	2,32	566,83	181,68
daher	36,00	3,21	703,48	182,11
zusätzlich	35,44	2,77	334,40	110,34
jene	35,11	6,59	1089,04	163,74
hinaus	34,89	1,80	271,17	136,91
gesamt	34,44	2,24	327,93	134,91
Beginn	33,44	3,72	430,80	118,91
zentral	32,78	2,68	315,41	123,92
Punkt	31,78	2,26	362,08	155,32
weder	30,11	1,70	265,46	145,30
andererseits	29,67	1,78	421,35	202,44
Bezug	28,33	2,64	695,85	243,06
rein	28,33	1,63	351,78	190,39
allerdings	27,22	4,53	1150,48	227,12
wohl	27,22	2,41	553,55	200,32
ebenso	26,44	2,91	658,44	216,84
nun	26,11	3,33	1074,88	273,82
notwendig	25,89	1,64	242,93	147,35
aufgrund	25,11	2,46	835,64	310,17
untersuchen	24,78	1,63	322,52	201,60
Rahmen	24,44	1,28	378,56	280,98
eindeutig	24,22	2,02	381,22	200,62
gering	23,89	1,98	264,40	144,15
innerhalb	22,89	3,30	754,15	235,90
Verwendung	22,78	1,33	408,71	314,19
schließlich	21,89	2,31	586,89	257,57
Kontext	21,78	1,43	394,38	262,01
kaum	21,67	1,73	386,61	216,56
schließen	21,56	1,66	331,22	198,14
einerseits	21,44	1,85	442,00	223,10
laut	21,33	1,63	519,06	264,41
trotz	21,33	1,39	336,04	229,39
teilweise	20,78	1,33	264,93	206,61

daran	20,67	1,22	267,58	213,65
wenige	20,56	1,50	285,02	192,68
Gegensatz	20,33	1,56	506,78	288,23
scheinen	20,11	2,54	942,25	376,17

Tab. 1: Einheiten, die in 100–90 % der L1-Texte vertreten sind (nach diffDOCF_{rel} geordnet)

89–80 %				
Lemma	diffDOCF _{rel}	simple maths	ARF(Fok)	ARF(Ref)
ausmachen	64,44	3,90	145,64	23,17
insofern	61,67	4,57	185,80	30,97
demnach	61,44	7,60	279,74	28,15
aufzeigen	61,33	3,96	171,41	30,32
Ansatz	57,11	5,03	261,83	50,85
Annahme	56,67	5,98	230,42	35,65
zuvor	55,56	4,38	237,45	47,13
festhalten	55,22	3,89	211,72	44,08
verdeutlichen	54,89	3,69	192,38	44,28
explizit	50,44	1,60	206,52	82,90
einnehmen	50,33	2,59	193,55	63,92
Betrachtung	48,00	4,07	229,35	54,12
relevant	47,22	3,25	212,64	57,56
anwenden	44,33	1,92	151,55	69,82
Nahe	43,22	3,06	197,24	71,07
beispielsweise	42,11	3,79	518,36	97,71
Zeitpunkt	42,00	2,98	212,17	75,00
all	41,56	2,27	261,85	94,48
formulieren	38,67	2,37	178,89	84,00
erläutern	37,78	2,40	202,93	83,85
stattfinden	36,67	1,89	154,94	78,17
beschränken	36,44	1,49	158,88	103,21
verweisen	36,11	3,06	347,13	104,15
erfolgen	35,67	2,32	306,21	122,45
unbedingt	35,56	1,47	117,79	75,36
Verlauf	34,89	2,04	180,86	91,38
Phänomen	34,56	2,27	280,16	132,86
speziell	34,56	1,49	208,33	119,27
Erkenntnis	34,56	1,35	196,98	125,32
grundlegend	33,22	1,65	151,60	95,02

ausgehen	32,56	2,26	256,29	113,20
Theorie	32,22	1,53	223,76	142,06
vollständig	32,00	1,29	143,04	98,83
Kategorie	31,89	2,63	459,08	190,21
sogenannt	31,78	1,71	187,28	104,79
voneinander	31,56	1,73	139,43	86,84
einsetzen	31,00	1,92	222,39	124,83
Methode	30,56	1,41	169,45	121,69
bevor	30,22	1,39	189,14	123,36
ausschließlich	29,78	2,11	264,31	121,22
zunächst	29,56	3,58	495,15	127,86
Verständnis	28,11	1,50	169,13	109,60
unabhängig	26,67	1,64	149,19	101,17
schlagen	26,67	1,18	178,22	137,10
zuordnen	26,33	1,33	209,41	146,54
kritisch	25,44	1,21	159,22	130,40
Untersuchung	25,33	2,52	516,52	215,89
nachdem	25,22	1,33	237,42	163,04
verfolgen	24,89	1,28	154,65	122,52
aufweisen	24,22	2,24	315,02	149,38
Definition	23,89	1,96	296,47	164,52
Eindruck	23,89	1,36	183,80	136,82
heraus	23,22	1,56	216,48	132,40
Bedingung	23,11	1,37	167,56	142,50
relativ	22,67	1,74	260,73	158,81
eins	22,56	1,29	131,92	114,64
Zug	21,56	1,84	340,27	185,17
komplex	21,44	1,97	223,19	133,02
daraus	21,22	1,68	298,19	174,70
entscheidend	21,11	1,31	152,14	123,86
vorliegend	20,33	2,08	222,40	154,15

Tab. 2: Einheiten, die in 89–80 % der L1-Texte vertreten sind (nach diffDOCFrel geordnet)

79–70 %				
Lemma	diffDOCF _{rel}	simple maths	ARF(Fok)	ARF(Ref)
heranziehen	61,89	5,61	174,02	22,32
zurückführen	60,44	2,64	123,39	26,64
Fokus	58,67	4,82	161,47	20,07

hierbei	57,33	6,80	306,73	34,55
Ausführung	56,00	4,93	163,09	27,88
hinsichtlich	52,11	6,05	240,61	35,23
vornehmen	50,56	3,11	129,50	35,08
ausgehend	50,56	2,89	104,39	27,39
bestehend	49,67	2,40	126,94	41,74
abschließend	49,44	2,98	128,43	31,69
weshalb	48,33	5,14	216,80	33,99
klären	47,44	2,57	104,74	36,22
bedingt	46,67	2,29	125,28	45,73
Unterscheidung	45,22	3,06	165,00	54,81
anschließend	45,22	2,22	137,43	54,48
bezüglich	44,89	3,91	202,83	45,81
Abgrenzung	44,44	1,93	116,24	45,56
beschrieben	44,11	2,68	173,30	59,94
ersichtlich	43,33	3,02	143,32	41,77
lediglich	43,00	4,40	342,22	63,90
zugrunde	42,78	2,41	99,92	36,64
generell	42,56	2,75	165,21	50,87
fungieren	41,33	2,47	122,57	46,94
vollziehen	39,67	2,31	122,82	48,65
Überlegung	39,67	2,10	132,03	61,21
eigenständig	39,44	2,40	108,56	43,21
thematisieren	39,22	2,61	186,71	64,14
wodurch	38,67	3,38	250,29	70,10
bedienen	38,33	1,68	124,94	57,44
Konsequenz	38,33	1,61	94,62	53,28
Hinblick	37,78	2,95	166,45	55,37
insbesondere	37,44	3,51	237,33	70,45
zunehmend	36,89	2,64	157,98	57,92
Konzept	35,22	3,19	230,57	77,75
stehend	35,00	1,51	94,91	54,81
ausschließen	34,33	1,91	112,73	55,67
liefern	33,67	2,12	169,17	77,97
stets	32,67	2,98	231,87	85,82
Diskussion	32,67	1,38	146,62	97,65
rücken	32,56	1,89	144,64	70,09
erkennbar	32,44	1,71	124,86	71,66
Anspruch	32,11	1,94	142,76	79,75

greifen	31,89	1,29	141,08	96,67
grundsätzlich	31,67	2,21	168,42	79,41
Modell	31,56	4,42	203,67	74,77
gestalten	31,44	1,38	139,02	94,58
primär	31,33	1,68	159,64	92,56
gelangen	31,11	1,56	108,81	63,77
erweisen	30,56	2,63	206,36	87,45
außerhalb	30,11	1,65	124,14	71,50
erheben	30,00	1,45	102,72	68,49
individuell	29,56	1,76	164,39	100,06
beitragen	29,44	1,27	129,02	91,49
ansprechen	28,78	1,51	148,06	95,97
herstellen	28,67	1,45	114,97	74,16
Natur	28,44	1,21	216,82	160,48
Auseinandersetzung	28,33	2,26	228,77	113,39
Forschung	28,22	1,53	148,01	103,00
darunter	27,33	1,44	115,55	85,47
interpretieren	26,89	1,76	164,74	98,30
Fähigkeit	26,56	1,44	156,91	114,42
Faktor	25,78	1,76	173,91	101,78
Voraussetzung	25,00	1,26	112,09	89,88
übrig	24,67	1,57	138,66	92,68
erlauben	24,56	1,28	115,30	90,18
bisher	24,56	1,21	143,21	121,31
Perspektive	24,00	1,66	211,85	126,52
beantworten	23,33	1,24	120,14	98,32
berücksichtigen	23,11	1,60	158,45	101,81
Hinweis	22,78	1,47	144,24	106,70
verbunden	22,56	1,41	126,02	91,39
deuten	22,56	1,38	163,27	120,40
Existenz	22,44	1,35	173,44	115,97
inhaltlich	22,11	1,42	177,41	134,34
Zentrum	21,78	2,38	157,16	90,56
zusammenfassen	21,67	1,25	130,58	107,20
zeitlich	21,67	1,25	140,46	112,31
absolut	21,11	1,85	134,88	90,94
insgesamt	20,89	2,19	240,19	125,95
werfen	20,89	1,30	145,96	108,76

Tab. 3: Einheiten, die in 79–70 % der L1-Texte vertreten sind (nach diffDOCFrel geordnet)

Da Interpretation und Kontextualisierung aller Lemmata weit über den Rahmen dieses Beitrags hinausgehen würden, beschränke ich mich nur auf Tendenzen und Besonderheiten. Wie der folgenden Übersichtstabelle zu entnehmen ist, verteilen sich die Lemmata gleichmäßig unter allen offenen Wortarten.

	100–90	89–80	79–70	gesamt	Anteil
Substantiv	7	18	21	46	23%
Verb	6	18	25	49	25%
Adjektiv	11	12	18	41	21%
Adverb	20	10	13	43	22%
Konjunktion	3	0	1	4	2%
Pronomen	1	2	1	4	2%
Präposition	6	1	2	9	5%
gesamt	54	61	81	196	100%

Tab. 4: Verteilung der generierten Einheiten nach Wortart

Semantisch und funktional betrachtet handelt es sich in vielen Fällen um textorganisierende Kohäsionsmittel (*insofern, demnach, beispielsweise, dennoch, vielmehr, daher, bezüglich, insbesondere, wodurch, zunächst, weshalb* usw.). An der Textorganisation sind auch abstrakte Substantive wesentlich beteiligt und ähnlich wie in den Arbeiten von Mach (2020a; 2020b) traten L1-L2-Unterschiede in diesem Bereich auch hier zutage. Allerdings zeigt sich, dass Machs Auseinandersetzung mit satzbezogenen Verweisen (hier *Annahme, Erkenntnis, Definition, Bedingung, Konsequenz, Fähigkeit*) zwar informativ war, die Gebrauchsunterschiede sich jedoch nicht lediglich auf diese konkrete Kategorie limitieren lassen. Angesichts dessen ist von Vorteil, dass in dieser Studie korpusgeleitet vorgegangen wurde und die Aufmerksamkeit dementsprechend auf weitere Substantive ausgedehnt werden konnte.

Einige, vor allem die auf der ersten Stufe angesiedelten Substantive treten fast ausschließlich als Teil von festen Wendungen auf, wie z. B. *[zu] Beginn, Verwendung [finden], [im] Gegensatz [(da) zu], [im] Rahmen [von|GEN], [im] Hinblick [auf], [zu diesem] Zeitpunkt* usw. Diese wurden ausführlich in Hotářovás (2019) N-Gramm-Analyse dokumentiert. Andere wie *Ansatz, Theorie, Konzept, Anspruch, Auseinandersetzung, Forschung, Hinweis, Faktor, Ausführung, Unterscheidung, Abgrenzung, Überlegung* u. Ä. variieren mehr in ihrem Gebrauch und benennen grundlegende Konzepte/Verfahren/Prozesse, die in der wissenschaftlichen Tätigkeit gang und gäbe sind. In Anlehnung an das berühmte Zitat von Wittgenstein, dass „die Grenzen meiner Sprache die Grenzen meiner Welt bedeuten“, drängt sich folglich die Frage auf, wie Schreibende im wissenschaftlichen Diskurs ohne diese Ausdrücke umhinkommen können, bzw. wodurch sie diese ersetzen.

Mit Blick auf die Verben zeichnet sich ein ähnliches Bild ab, denn *eingehen, begründen, annehmen, untersuchen, scheinen, ausmachen, aufzeigen, festhalten, verdeutlichen, anwenden, formulieren, erläutern, verweisen, erfolgen, ausgehen, einsetzen, verfolgen, aufweisen* gehören ohne Zweifel zum zentralen Wortschatz der AWS. Ihre geringere Frequenz deutet erstens wohl auf die variable Breite des alltäglichen Wortschatzes von nichtmuttersprachlichen Schreibenden, zweitens gleichzeitig auf die mangelnde Schreib- bzw. Leseerfahrung hin. Mithilfe von Verben wie beispielsweise *eingehen, festhalten, verweisen, ausgehen* werden nämlich wenig variierende und mithin konventionelle Routinen (Spezifikation/Vorstellung des Themas, Zusammenfassen, Quellenangabe usw.) realisiert (vgl. Brommer 2018).

Im Fall von isolierten Adjektiven kann bis auf einige Besonderheiten (*vorhanden, ausgehend, unabhängig, ersichtlich*) hingegen keine Rede von Mustern sein. Auffällig ist allerdings die Tatsache, dass viele der Adjektive (*zentral, notwendig, eindeutig, gering, wenig, explizit, relevant,*

speziell, grundlegend, sogenannt, kritisch, relativ, komplex, entscheidend, erkennbar) in gewisser Hinsicht subjektive Positionen der Verfassenden darlegen. Durch diese Optik lässt sich die These wagen, dass die (tschechischen) Nichtmuttersprachler etwas zurückhaltender sind, ihrer Autorenstimme Ausdruck zu verleihen. Dies steht zwar im Einklang mit Machs (2020a:31) Beobachtung, in der er sich über zu viele nicht kommentierte Beispiele in tschechischen Abschlussarbeiten beklagt, die Unterschiede in der auktorialen Präsenz würden aber einer separaten Untersuchung bedürfen.

„Obwohl [] [der] Gebrauch [von Beispielen] an sich gewiss nicht zu kritisieren ist, wird es dennoch problematisch, wenn man auf sie als primäre Textgestaltungsstrategie zurückgreift, denn eine Abschlussarbeit bedarf komplexer Ideen, die sich in der Regel nicht mithilfe bloßer Auflistungen, sondern eher mit Fließtext effektiv übermitteln lassen. In vielen Fällen mangelt es zudem an Kommentaren, die die angeführten Beispiele bzw. Kategorisierungen in die Textmenge integrieren würden.“
(Mach 2020a:31)

5. Fazit

In dem Beitrag wurde versucht, die Schlüsselwortanalyse zur Bestimmung des didaktisch relevanten Wortschatzes anzuwenden. Die hier vorgestellte Kombination von Frequenz- und Dispersionsmaßen hatte zur Folge, dass in den Wortschatzlisten mehrheitlich Beispiele aus der AWS und keine Fachtermini enthalten waren. Die Ergebnisse der vergleichenden Analyse deuten darauf hin, dass die Unterschiede (Lücken) in der Produktion von tschechischen Germanistikstudierenden nicht als isolierte Lapsus zu betrachten sind. Vielmehr liegt die Vermutung nahe, dass sie mit dem Niveau der Alltagssprache einhergehen, von der mangelnden Schreiberfahrung in der Mutter- sowie Fremdsprache herrühren und in diesem Licht als systematisch angesehen werden sollten. Dies soll wiederum nicht heißen, dass jede(r) Schreibende immer alle oben aufgeführten Ausdrücke zu verwenden hat oder dass diese in allen L1-Arbeiten zu dokumentieren sind. Auf der Makroebene sind die Diskrepanzen zwischen Mutter- und Fremdsprache jedoch so eklatant, dass sie nicht ignoriert werden können. Zusammenfassend sei festgehalten, dass die Schlüsselwortanalyse kombiniert mit sorgfältig ausgewählten Metriken ein wertvolles Instrument für die Erforschung von Lernerkorpora darstellt. Ich möchte hoffen, dass die daraus entstandene Wortliste den Lernenden ihren Einstieg in die Wissenschaftssprache erleichtert und den Lehrkräften eine weitere Idee gibt, worauf sie sich bei der Vermittlung akademischer Schreibfertigkeiten konzentrieren könnten.

Literaturverzeichnis

- BROMMER, Sarah (2018): Korpuslinguistische Methodologie und Mustererkennung aus statistischer Sicht. In: BROMMER, Sarah: *Sprachliche Muster: Eine induktive korpuslinguistische Analyse wissenschaftlicher Texte*. Berlin, S. 91–114.
- BUBENHOFER, Noah / SCHARLOTH, Joachim (2015): Maschinelle Textanalyse im Zeichen von Big Data und Data-driven Turn – Überblick und Desiderate. In: *Zeitschrift für germanistische Linguistik*, Nr. 43, S. 1–26.
- EGBERT, Jesse / BIBER, Douglas (2019): Incorporating text dispersion into keyword analyses. In: *Corpora*, Nr. 14, S. 77–104.
- EHLICH, Konrad (1999): Alltägliche Wissenschaftssprache. In: *Info DaF*, Nr. 26, S. 3–24.
- FÜGERT, Nadja / RICHTER, Ulrike (2017): *Wissenschaftssprache verstehen: Wortschatz – Grammatik – Stil – Lesestrategien*. Stuttgart.
- GABRIELATOS, Costas (2018): Keyness analysis: Nature, metrics and techniques. In: TAYLOR, Charlotte / MARCHI, Anna (Hrsg.): *Corpus Approaches To Discourse: A Critical Review*. New York. S. 225–258.
- GOLDHAHN, Agnes (2016): *Tschechische und deutsche Wissenschaftssprache im Vergleich. Eine kontrastive Analyse tschechischer und deutscher Wissenschaftlicher Artikel der Linguistik*.

- Brno: Masaryk Universität, (Dissertation). Zugänglich unter: <https://is.muni.cz/th/ukgos> [13.01.2022].
- GRAEFEN, Gabriele (2002): Probleme mit der alltäglichen Wissenschaftssprache in Hausarbeiten ausländischer StudentInnen. In: REHBEIN, Jochen (Hrsg.): *Lernen in der zweiten Sprache*. Münster, S. 1–20. Zugänglich unter: <https://epub.ub.uni-muenchen.de/13669/> [24.03.2022].
- GRAEFEN, Gabriele (2009): Die Didaktik des wissenschaftlichen Schreibens: Möglichkeiten der Umsetzung. In: *German as a foreign language (GFL)*, Nr. 10, S. 106–127.
- GRAEFEN, Gabriele / MOLL, Melanie (2011): *Wissenschaftssprache Deutsch: lesen – verstehen – schreiben: ein Lehr- und Arbeitsbuch*. Frankfurt am Main.
- GRIES, Stefan T. (2008): Dispersions and adjusted frequencies in corpora. In: *International Journal of Corpus Linguistics*, Nr. 13, S. 403–437.
- HASSELGREN, Angela (1994): Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. In: *International Journal of Corpus Linguistics*, Nr. 4, S. 237–258.
- HLAVÁČOVÁ, Jana / RYCHLÝ, Pavel (1999): Dispersion of Words in a Language Corpus. In: MATOUŠEK, Václav / MAUTNER, Pavel / OCELÍKOVÁ, Jana / SOJKA, Petr (Hrsg.): *Text, Speech and Dialogue. TSD 1999. Lecture Notes in Computer Science*. Berlin, S. 321–324.
- HOTAŘOVÁ, Markéta (2019): *Typische Wortverbindungen in der Wissenschaftssprache. Eine korpuslinguistische kontrastive n-Gramm-Analyse*, Dissertation. Zugänglich unter: <https://is.muni.cz/auth/th/v2rwb> [12.08.2022].
- KILGARRIFF, Adam (2009): Simple maths for keywords. In: MAHLBERG, Michaela / GONZÁLEZ-DÍAZ, Victorina / SMITH, Catherine (Hrsg.): *Proceedings of Corpus Linguistics Conference CL2009*. Liverpool, S. 1–6.
- KILGARRIFF, Adam / BAISA, Vít / BUŠTA, Jan / JAKUBÍČEK, Miloš / KOVÁŘ, Vojtěch / MICHELFEIT, Jan / RYCHLÝ, Pavel / SUCHOMEL, Vít (2014): The Sketch Engine: ten years on. In: *Lexicography*, Nr. 1, S. 7–36.
- MACH, Tomáš (2020a): *Analyse von Kollokationen ausgewählter Abstrakta in der studentischen Wissenschaftssprache*. Masterarbeit. Brno. Zugänglich unter: <https://is.muni.cz/auth/th/rpnfc> [24.09.2022].
- MACH, Tomáš (2020b): Zur Leistung der Abstrakta in der studentischen Wissenschaftssprache: ein quantitativer L1-L2-Vergleich. In: *Linguistik Online*, Nr. 102, S. 29–57. Zugänglich unter: <https://bop.unibe.ch/linguistik-online/article/view/6826> [07.03.2022].
- PERKUHN, Rainer (2021): Korpusfrequenzen und andere Metriken zur Strukturierung von DaF-Lehrmaterial. In: *Korpora Deutsch als Fremdsprache*, Nr. 1, S. 116–136.
- POLÁŠKOVÁ, Eva (2016): Zur Entwicklung der Paraphrasierungs- und Zusammenfassungskompetenz bei Studierenden. Ausgewählte Aspekte der Arbeit mit Fachtexten. In: *Acta Facultatis Philosophicae Universitatis Ostraviensis. Studia Germanistica*, Nr. 19, S. 115–138.
- PRUTSCHER, Daniela (2020): Phraseme und Wortverbindungen in der deutschen Wissenschaftssprache. In: SZERSZUNOWICZ, Joanna / AWIER, Martyna (Hrsg.): *Reproducible Multiword Expressions from a Theoretical and Empirical Perspective*. Białystok. S. 217–232.
- SAVICKÝ, Petr / HLAVÁČOVÁ, Jana (2002): Measures of Word Commonness. In: *Journal of Quantitative Linguistics*, Nr. 9, S. 215–231.
- STEINHOFF, Torsten (2007): *Wissenschaftliche Textkompetenz: Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten*. Berlin.
- TARGOŃSKA, Joanna (2019): Reproduzierbare Elemente in der alltäglichen Wissenschaftssprache Deutsch. In: *Germanica Wratislaviensia*, Nr. 144, S. 261–279.

Internetquellen

- URL 1: Statistics used in Sketch Engine. *Sketch Engine*. <https://www.sketchengine.eu/documenta-tion/statistics-used-in-sketch-engine> [07.12.2022]