

# **Sprachliche Konfigurationen im wissenschaftlichen Diskurs. Theoretische Prämissen und methodologische Herausforderungen der n-Gram-Analyse für die Erforschung der wissenschaftssprachlichen Typik**

## **Language Configurations in Scientific Discourse. Theoretical Premisses and Methodological Challenges of the N-gram Analysis for Research of Typicality in Scientific Language**

Markéta Hotařová, Iva Zündorf

### **Abstract**

In scientific text types the specific language configurations represent a register of language typicality which is regularly used in similar contexts of scientific discourse. It is obvious, that the appropriate knowledge about these structures is a condition for competent scientific outputs. The configured structures as specific patterns are manifested in different forms and functions both on the lexical and the textual level. The paper explains the theoretical premisses for researching these structures and discusses the methodological potential of the frequency-oriented n-gram analysis as a challenging base for the qualitative analysis of linguistic data.

### **Keywords**

n-gram-analysis; language typicality; lexical and textual; configurations in scientific discourse

## 1 Einleitung

Ein wissenschaftlicher Fortschritt verbirgt sich im Prinzip in den neuen Erkenntnissen und Ergebnissen, die von den Wissenschaftlern abschließend zu beschreiben und weiter zu diskutieren sind. Wie jedoch Weinrich (1994: 161) argumentiert: „Keine Wissenschaft kann heute bei Null anfangen“. Somit soll eine gewisse „Kontinuität“ der Wissenschaft gesichert werden (vgl. Weinrich 1994: 161).

Laut aktuellen Studien spielt sich ein ähnliches Szenario in der schriftlichen Fixierung des wissenschaftlichen Vorgehens ab, denn einzelsprachige wissenschaftliche Texte von verschiedenen Autorinnen und Autoren weisen häufig durchweg identische, mehr oder weniger feste Formeln auf. Ohne sie wäre beispielsweise ein wissenschaftlicher Artikel kein wissenschaftlicher Artikel und damit für Rezipienten nicht nachvollziehbar.

Im Folgenden wird die Aufmerksamkeit auf diese Formeln gelenkt, die hier begrifflich als sprachliche Konfigurationen gefasst werden. Neben der Erklärung ihrer Rolle und Bedeutung in wissenschaftlichen Texten wird gleichzeitig ein entsprechender Raum den methodisch-technischen Fragen der Datengewinnung gewidmet, die sich auf die aktuellen Erkenntnisse der korpuslinguistischen Forschung stützen und, im konzeptuellen Sinne, die Theorie der „Sprachgebrauchsmuster“ (Bubenhofer 2009) reflektieren.

## 2 Wissenschaftssprache und wissenschaftlicher Diskurs

Die sprachliche Vermittlung der theoretischen Behauptungen und Erkenntnisse erfolgt aus fachsprachlicher Sicht durch die Wissenschaftssprache, die neben der Sprache der Technik und der Sprache der Institutionen zu verorten ist (vgl. Roelcke 2010: 31). Diese Auffassung ist völlig im Einklang mit der Vorstellung der spezifischen Existenzweise der Wissenschaft (vgl. Wolf 2019: 28). Diese Existenzweise kann man sich im Wesentlichen als Gespräch, Austausch oder Diskussionen vorstellen, die die Wissenschaftler mündlich oder schriftlich führen, um die Wissenschaft voranzubringen (ibid.). In diesem Sinne können wir auch konzeptuell vom wissenschaftlichen Diskurs sprechen.

Der wissenschaftliche Artikel in Fachzeitschriften stellt eine Basisform dar, in der wissenschaftliche Erkenntnisse und neue Forschungsergebnisse kommuniziert werden. Somit ist der wissenschaftliche Artikel ein mehr als nur ein relevanter Bestandteil der wissenschaftlichen Diskurse. Die Erforschung der (einzel)sprachlichen Spezifika und der (einzel)sprachlichen Typik in diesem Sprachsegment ist demzufolge ein wichtiger Beitrag zur Reflexion der Wissenschaftssprache, zum Verständnis ihrer sprachsystemischen Spezifika, aber auch ihrer Kulturgebundenheit.

Die oben erwähnte Mehrzahl vom ‚wissenschaftlichen‘ Diskurs ist auf die Vielfalt in der Wissenschaft zurückzuführen, als ein konkretes Beispiel sind die geisteswissenschaftlichen gegenüber den naturwissenschaftlichen Texten zu nennen. Trotzdem sind unter den Spezifika gemeinsame Züge zu finden, die für den gemeinsamen Diskurs geltend gemacht werden. Zu erwähnen gehört der wissenschaftliche Stil, nach dem die Texte

verfasst werden (vgl. Dönninghaus 2005: 322). Ein gemeinsamer Nenner ist für sie „Exaktheit und Neutralität (Objektivität)“ (Auer/Bassler 2007: 12).

Zum wissenschaftlichen Stil auf der Textebene gehören bestimmte Routinen:

„Um sicheres und neues Wissen darzustellen, greifen Forscher auf ein spezifisches Repertoire von domänentypischen Mustern und Routinen zurück“ (Steinhoff 2009: 101).

An dieser Stelle muss die Frage gestellt werden, was alles zu diesem wissenschaftssprachlichen „Repertoire“ gehört. Um diese Frage beantworten zu können, sind mehrere Studien nötig, deren Ziel und Untersuchungsgegenstand festgelegt werden müssen. Bereits mehrere Autoren haben auf diese Besonderheit aufmerksam gemacht, indem sie das wissenschaftliche „Repertoire“ zu beschreiben versuchten. Auer/Bassler (vgl. 2007: 16; 20) weisen auf die häufige Nominalisierung, Modalverben und Infinitivkonstruktionen hin oder Heinemann (vgl. 2000: 705) konzentriert sich in diesem Zusammenhang auf die Strukturierung der Texte, Fachterminologie, häufige Zitationen und spricht auch von dominantem Nominalstil und Passiv- bzw. Infinitivkonstruktionen. Im Folgenden wird aber versucht, auf ein anderes Phänomen im Zusammenhang mit der Wissenschaftssprache aufmerksam zu machen, das gerade in der Wissenschaftssprache ein Forschungspotenzial darstellt, nämlich rekurrente Mehrwortverbindungen, mehr oder weniger feste Formeln oder ganze textrelevante Strukturen, durch die der jeweilige wissenschaftliche Diskurs an sich verbalisiert und kommuniziert wird.

### 3 Die sprachliche Konfiguriertheit als Untersuchungsgegenstand

Der Einsatz der korpuslinguistischen Methoden reflektiert nicht nur diese Tendenz, sondern legt damit den Fokus auf sprachliche Äußerungen und Sprechhandlungen (vgl. Lemnitzer/Zinsmeister 2010: 28). Der ‚Kontext‘ manifestiert sich in diesen Äußerungen auf der Textoberfläche, die empirisch durch korpuslinguistische Instrumente erkennbar gemacht werden können (vgl. Bubenhofer/Scharloth 2010: 85).

Bubenhofer/Scharloth (2010: 86) halten den Kontext für eine „Folge von Text“ und betrachten ihn nicht „als Grundlage für das Verstehen“. Die empirischen Forschungsergebnisse aus der sprachlichen Oberfläche in Form von rekurrenten sprachlichen Einheiten sind nicht als einziger Aspekt für das eigentliche Verstehen der Autorenintension ausschlaggebend, sie konstituieren sich aber auch im Kontext und die erfassten Frequenzen können im Verstehensprozess einen gewissen Beitrag leisten – und zwar auch in kontrastiven Untersuchungen, in denen übergreifende, gemeinsame oder hingegen verschiedene Tendenzen beobachtet werden können.

Eine zentrale Frage an dieser Stelle ist jedoch die nach der Signifikanz der Ergebnisse, die aus statistischer Perspektive möglichst in Verbindung mit der Korpusgröße, der untersuchten Fragestellung und den technischen Möglichkeiten zu beantworten ist. Bubenhofer/Scharloth (2010: 86) stellen klar, dass der in den letzten Jahrzehnten untersuchte Kontext mit dem „pragmatische[n] Wert eines Textes“ zusammenhängt.

Perkuhn/Keibel/Kupietz (2012: 15) sprechen in diesem Zusammenhang über einen „echte[n] Sprachgebrauch“, der als ein Prozess „innerhalb einer Sprachgemeinschaft“ zu verstehen ist. Auch Feilke (vgl. 1993: 12) sieht in den auf sprachlicher Oberfläche „regulierten Ausdrücken“ ihr Potenzial für den Verstehensprozess; es steckt nämlich „die konventionelle Strukturierung des Verstehens“ dahinter, nach welcher wir den Text rezipieren und ihn nicht nur anhand unseres Weltwissens interpretieren (Feilke 1993: 12). Diese sprachlichen Einheiten, die in einem Text rekurrent auftreten, werden vom Rezipienten erwartet bzw. vorausgesetzt. Feilke (vgl. 1993: 11–12) bezeichnet dieses Phänomen des Sprachgebrauchs zutreffend als ‚sprachliche Common-sense-Kompetenz‘.

Darüber hinaus setzt er diese Kompetenz mit der ‚Kontextualisierungskompetenz‘ gleich, weil sie „sich in der Verfügbarkeit strukturell komplexer Ausdruckseinheiten auf den verschiedenen Ebenen sprachlicher Struktur (phonotaktische Cluster, usuelle Komposita, geprägte Syntagmen, syntaktische Pattern, Routineformeln, Text-Pattern) manifestiert“ (Feilke 1993: 19). Das Kontextualisierungspotenzial einer in anderen Kontexten auch freien Kombination bedeutet, dass derartige komplexe Ausdrücke, z. B. im sprachwissenschaftlichen Diskurs, ein konkretes, wiedererkennbares Schema (Hintergrund) indizieren. Dadurch bekommen sie einen speziellen Signalwert, da sie gerade in ähnlichen Kontexten oder in der Interaktion als Ganzes abgerufen werden (vgl. Feilke 2004: 52f.).<sup>1</sup>

Bisher wurde die Rolle der rekurrenten sprachlichen Einheiten für den Verstehensprozess erklärt. Des Weiteren wird in diesem Artikel zum Ziel gesetzt, die empirischen Möglichkeiten der aktuellen korpuslinguistischen Forschung anzudeuten, die diese Einheiten zu ihrem Untersuchungsgegenstand macht. Obwohl die Eigenschaften des Untersuchungsgegenstandes bereits mehr oder weniger klar sind, darunter wird vor allem der rekurrente Auftritt verstanden, besteht in diesem Forschungsbereich definitorische Uneinheitlichkeit, wie es aber auch sonst in sprachwissenschaftlichen Arbeiten häufig der Fall ist.

Somit kann festgestellt werden, dass sich die oben dargestellten Ausführungen mit dem Begriff ‚Muster‘ (Bubenhofer/Scharloth 2010: 90) identifizieren lassen. Unter dem Sprachgebrauch in diesem Sinne wird „rekurrentes Auftreten von textuellen Einheiten in bestimmten Sprachausschnitten“ verstanden (Bubenhofer/Scharloth 2010: 90). Die rekurrenten sprachlichen Einheiten, die auch als Mehrwortverbindungen bezeichnet werden können, treten in unterschiedlichen Formen auf. Im Folgenden werden einige Bezeichnungen angeführt, natürlich ohne Anspruch auf ihre vollständige Erfassung und Erklärung, mit denen die jeweiligen MWV-Forschungsansätze arbeiten:

- a) rekurrente Formulierungsmuster (vgl. Steinhoff 2009: 102),
- b) Sprachgebrauchsmuster‘ (vgl. Bubenhofer 2009),
- c) Handlungs- und Textmuster (vgl. Heinemann 2000: 704),
- d) usualisierte Mehrworteinheiten (vgl. Steyer 2011: 13–14),
- e) Kookkurrenz (vgl. Bubenhofer 2009: 114),
- f) Kollokation (vgl. Firth 1957 et. al; Evert 2009: 1213)
- g) sprachliche Konfigurationen (vgl. Kratochvílová-Zündorf 2011: 250)

<sup>1</sup> Mehr dazu siehe bei Kratochvílová 2011: 94.

Es steht außer Frage, dass wir hier mit Übergangseinheiten rechnen müssen, die uns vor eine Reihe offener Fragen stellen. Zu solchen Fragen gehört zum einen das Problem der systemischen Abgrenzung von Formulierungen, die in wissenschaftlichen Texten als erwartbare, mehr oder weniger feste Strukturen auftreten. Das Spezifikum solcher sprachlichen Mittel, seien es fachtextbezogene Formulierungsroutinen und Konfigurationen oder sogar ganze vorgeformte Textstrukturen, besteht darin, dass sie in der geschriebenen wissenschaftlichen Kommunikation unter ähnlichen kommunikativen Bedingungen repetitiv eingesetzt werden (vgl. Zündorf 2016: 65).

Zum anderen, eng mit dem ersten komplexen Thema verbunden, stellt sich hier die Frage der methodologischen Vorgehensweise bei der Erforschung solcher Strukturen. Es fällt auf, dass die einzelsprachlich spezifischen Sprachgebrauchsmuster und Argumentationsketten häufig über die Wortgrenze hinaus bis auf die Textebene gehen. Außer lexikalischen Kriterien sollten demzufolge auch textlinguistische Aspekte herangezogen werden, da diese komplexen Einheiten als charakteristische Wortkombinationen, im engen Zusammenhang mit der pragmatischen Funktion des Textes, seine textinterne Konfiguration reflektieren. Als Argument kommt die Tatsache, dass bestimmte Konfigurationen in entsprechenden Texten auch eine textorientierende Funktion erfüllen; sie können für bestimmte Textsorten jeweils sogar signifikant sein. Durch den wiederholten Rückgriff auf solche Strukturen in konkreten wissenschaftssprachlichen Textsorten wird die Fachtextualität und somit auch die Fachtextrezeption wesentlich unterstützt. Hier können sich natürlich einzelsprachliche Konventionen unterschiedlich auswirken. Das textlinguistische Wissen, d.h. das Wissen über die jeweilige Textsorte und ihre Textaufbauprinzipien, vor allem aber das Wissen über Konventionen in der Gedankenführung (vgl. Blumenthal 1997:113), die sich gerade in solchen einzelsprachlich typischen Strukturen zeigen, sollen daher zu den wichtigen Grundlagen des textsortenbezogenen Wissens gehören.

Alle genannten Argumente bekommen jedoch in empirischen Arbeiten unter Einbezug von verschiedenen technischen Hilfsmitteln eine neue Dimension (vgl. Evert 2009: 1213). Diese scheinbare Diskrepanz ist lediglich ein Beweis dafür, dass kein Zugang, auch nicht der frequenzorientierte, Anspruch auf eine vollständige und eindeutige Interpretation der Ergebnisse geltend machen kann. Hier ist und bleibt die sprachreflektive Sicht der Sprachwissenschaftlerin bzw. des Sprachwissenschaftlers weiterhin wichtig.

Im Folgenden werden einige Vor- und Nachteile frequenzorientierten Ansatzes aus dem methodischen Repertoire der Korpuslinguistik am Beispiel der wissenschaftssprachlichen Einheiten dargestellt, der all die früher genannten Ideen und Notwendigkeiten in Betracht ziehen will.

#### **4 Methodologisch technische Fragen: Die n-Gramm-Analyse als eine effiziente Ausgangsbasis**

Aus den bisherigen Ausführungen ist ersichtlich, dass die unterschiedlichen Betrachtungsweisen und die damit zusammenhängenden Implikationen bestimmte textlinguistische und stilistische Anforderungen mit sich bringen. Im Folgenden soll vor allem

auf die n-Gramme aufmerksam gemacht werden, da sie eine statistische Auffassung ermöglichen. In diesem Sinne stellen sie eine gute Basis für weitere Untersuchungen zu Mehrwortverbindungen und Mehrwortstrukturen im Allgemeinen dar.

Ein n-Gramm ermöglicht aus korpuslinguistischer Sicht eine gewisse Freiheit in den Einstellungsmöglichkeiten im Vergleich zu anderen Arten der Mehrwortverbindungen, wie z. B. bei Kollokationen. Diese Einstellungen stellen Anzahl und Art der Kombinationsmöglichkeiten dar und es können dabei Wort- und Grundformen in einer selbst gewählten Spannweite unterschieden werden (vgl. Bubenhofer 2015: 496). Kollokationen sind semantisch, lexikalisch und nicht zuletzt auch grammatisch geprägt, bei n-Grammen handelt es sich um eine Segmentierung des Textes (vgl. Pořízka 2014: 37). Ausschlaggebend für n-Gramme ist ihre Häufigkeit. Nach Banerjee/Pedersen (vgl. 2003: 370) sind solche n-Gramme interessant, wenn sie häufiger vorkommen, als erwartbar wäre. Wie Bubenhofer/Scharloth (vgl. 2010: 90) bemerken, bedeutet „typisch“ aber nicht dasselbe wie „häufig“. Die Typizität wird dann vor allem auf die wissenschaftssprachlichen Mehrwortverbindungen bezogen. Im Folgenden wird aber die Aufmerksamkeit auf die aktuellen technischen Möglichkeiten der Korpuslinguistik gerichtet.

Die selbst erstellten Korpora oder die bereits verfügbaren Korpora können mit einer Reihe von korpuslinguistischen Softwaretools untersucht werden. Im Folgenden werden drei vorgestellt, denen die grundlegenden Konkordanzfunktionen und die Funktion der Extraktion von n-Grammen gemeinsam sind.

#### a) N-gram Statistics Package (NSP)

Dieses Softwaretool eignet sich gut auch für größere Korpora in Größenordnung von Millionen Tokens, die entweder selbst erstellt werden oder die online zur Verfügung stehen (vgl. Banerjee/Pedersen 2003: 370). NSP ist ein kostenfreies Tool, das nach der Berechnung der Assoziationsmaße Mehrwortverbindungen extrahiert (vgl. Glynn 2014: 315–316). Nach Bubenhofer (vgl. 2009: 182) ist es vor allem bei korpusgeleiteten Analysen anwendbar. Seine „vielseitige“ Anwendbarkeit ist nur von gewissen IT-Kenntnissen bedingt.

#### b) Sketch Engine

Dieses Softwaretool ist über eine Webseite zugänglich und wird ständig von seiner Entwicklungsgruppe verwaltet. Die Nutzer haben somit die Möglichkeit, sich jederzeit an ihren Web-Service zu wenden. Dieses ursprünglich lexikographische Tool nimmt innerhalb der korpuslinguistischen Softwares eine zentrale Stellung ein (vgl. Kilgarriff et al. 2014: 7-8). Da Sketch Engine sowohl Semantik als auch Syntax in Verbindung setzt (vgl. Bubenhofer/Ptashnyk 2010: 12; vgl. Cvrček 2013: 26), muss diese Tatsache in der abschließenden Interpretation betrachtet werden. Auch dieses Tool eignet sich für die Berechnung von größeren Datenmengen. Weitere Vorteile betreffen die Vielfalt an Dateiformaten, die akzeptiert werden, als auch an verschiedenen Sprachen, die dort unterstützt werden. Bei n-Grammen müssen die Leerstellen in den Buchstabensequenzen beachtet werden, dies ermöglicht eine benutzerdefinierte Einstellung. Als offene Frage bleibt jedoch der weitere in erster Reihe akademische Zugang, der zurzeit nur bis März 2022 aufgrund der Projektdauer von ELEXIS garantiert ist (vgl. Kilgarriff et al. 2004).

**c) AntConc**

AntConc reiht sich auch in die kostenfreien Softwaretools, die sich für korpuslinguistische Textanalysen eignen, ein (vgl. Anthony 2019). Auch hier kann ein eigenes (jedoch nicht-annotiertes) Korpus erstellt werden. Obwohl dieses Softwaretool ein fertiges Produkt ohne weitere Verwaltung darstellt, kann es zuverlässige, signifikante Ergebnisse liefern, jedoch wie Pořízka (vgl. 2014: 165) bemerkt, eher aus kleinen Datenmengen.

Wenn aber der Hauptfokus auf die übergreifenden Erkenntnisse aus den Untersuchungen mithilfe von n-Grammen gelegt wird, lässt sich Folgendes zusammenfassen: N-Gramme stellen aufgrund ihrer Kombinationsmöglichkeiten eine gute Untersuchungsbasis für weitere Interpretationen. Dabei müssen aber folgende Punkte beachtet werden:

- a) Repräsentativität des Korpus bzw. der ausgewählten Stichprobe;
- b) relative und absolute Häufigkeit (vgl. Perkuhn/Keibel/Kupietz 2012: 71–72; 78–82);
- c) statistische Assoziationsmaße und Signifikanztests;
- d) verfügbare korpuslinguistische Softwaretools, ihre spezifischen Funktionen und ihre technischen Beschränkungen bzw. Kompatibilitätsansprüche.

Aus den beschriebenen Beobachtungen resultiert die Frage nach der Vergleichbarkeit der so erworbenen Ergebnisse und Interpretationen. Unterschiedliche Studien, die zwar gleichermaßen korpuslinguistisch angelegt sind und die aufgrund der schnellen Entwicklung der Computertechnologien und des unterschiedlichen Einsatzes seitens der Sprachwissenschaftlerinnen und Sprachwissenschaftler zweifellos signifikante Erkenntnisse liefern, stellen aber gleichzeitig einen klaren Beweis dafür dar, dass es nicht nur ein ‚Kochrezept‘ mit klar definierten ‚Kochwerkzeugen‘ gibt und dass hier ein weiterer Entwicklungsbedarf hinsichtlich der methodisch-technischen Anforderungen besteht.

## 5 Statt einer Zusammenfassung weitere Forschungsperspektiven

Das Interesse an rekurrenten Mehrwortverbindungen im Zusammenhang mit modernen korpuslinguistischen Werkzeugen wurde in der Forschung vor allem durch die korpuslinguistischen Arbeiten von Bubenhofer (2009) geweckt. Sie gelten als „Vorreiter“ für die „Wiederentdeckung“ der induktiven korpusgeleiteten Methode, die in letzten Jahren durch die deduktive korpusbasierte Methode etwas verdrängt wurde. Eine gewisse Ausweitung bzw. ein Perspektivenwechsel findet im Bereich der Kollokationsforschung statt. Zu der lexikographischen, phraseologischen oder pragmatischen Sicht (Hausmann 1976; Burger 2003; Ďurčo 2014 und 2016; Kratochvílová-Zündorf 2006 und 2011) reiht sich zunehmend auch die korpuslinguistische Perspektive (vgl. Scharloth/Eugster/Bubenhofer 2013: 348), die aus den methodischen Beobachtungen stammt (vgl. Evert 2009: 1214; vgl. Bubenhofer 2009: 112). Neue sprachwissenschaftliche Tendenzen zeigen, dass aufgrund der neuen Computertechnologien diese Forschungsperspektive einen Schritt weiter geht, indem der Blick auf die Ausgangsbasis von Daten auf n-Gramme ausgeweitet wird.



Als Prämisse für weitere Forschungsaufgaben gilt, dass die konfigurierten sprachlichen Mittel in der geschriebenen wissenschaftlichen Kommunikation zur Texttypik dieses Sprachsegments gehören. Die Erforschung der Wissenschaftssprache ist hier mit festen polylexikalen Strukturen konfrontiert, auf die bei der Textproduktion in unterschiedlichem Maße zurückgegriffen wird und die in vielen Fällen als ein lexikalisches oder sogar textgestaltendes Ganzes abgerufen werden. „Sprachliche Konfigurationen in Texten evozieren Verwendungskontexte mit besonderen Anforderungen an die Textortgestaltung sowie an entsprechende sprachliche Realisierung“ (Zündorf 2018: 60). Im Hinblick auf einzelne Arbeiten zu Nationalsprachen der Wissenschaft repräsentiert diese kontrastiv angelegte Forschungsaufgabe eine Fortsetzung und weitere Entwicklung des Forschungsbereiches.

Mit anderen Worten – die einzelsprachlichen usuellen Wortverbindungen, Mehrwortstrukturen im ganzen Spektrum der fachsprachlichen Phraseologie, Kollokationen, Formulierungsroutinen, Argumentationsverfahren, Konnektoren, ganze Texte. Diese Art der Konfiguriertheit im wissenschaftlichen Diskurs verbunden mit dem Potenzial der korpuslinguistischen Methoden und Instrumente stellen die Forschung, in unserem Falle den Sprachvergleich Deutsch -Tschechisch, vor eine größere Herausforderung denn je.

## Wissenschaftliche Quellen und Literatur

- Auer, Peter – Bassler, Harald (2007): Der Stil der Wissenschaften. In: Auer, Peter – Bassler, Harald (Hg.): Reden und Schreiben in der Wissenschaft. Frankfurt am Main: Campus Verlag, S. 9–30.
- Banerjee, Satyanjee – Pedersen, Ted (2003): The Design, Implementation, and Use of the Ngram Statistics Package. In: Gelbukh, Alexander (Hg.): Computational Linguistics and Intelligent Text Processing. CICLing 2003. Lecture Notes in Computer Science, vol 2588, Berlin/Heidelberg: Springer, S. 370–381.
- Blumenthal, Peter (1997): Sprachvergleich Deutsch-Französisch. 2. neu erarbeitete und ergänzte Auflage. Tübingen: Niemeyer.
- Bubenhofer, Noah (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin/New York: de Gruyter.
- Bubenhofer, Noah (2015): 21. Muster aus korpuslinguistischer Sicht. In: Dürscheid, Christa – Schneider, Jan Georg. (Hg.): Handbuch Satz, Äußerung, Schema. Berlin/New York: De Gruyter, S. 485–502.
- Bubenhofer, Noah – Ptashnyk, Stefaniya (2010): Korpora, Datenbanken und das Web: State of the Art computergestützter Forschung in der Phraseologie und Lexikographie. In: Ptashnyk, Stefaniya – Hallsteinsdóttir, Erla – Bubenhofer, Noah (Hg.): Korpora, Web und Datenbanken. Computergestützte Methoden in der modernen Phraseologie und Lexikographie/Corpora, Web and Databases. Computer-Based Methods in Modern Phraseology and Lexicography. Phraseologie & Parömiologie. Hohengehren: Schneider Verlag, S. 7–20.
- Bubenhofer, Noah – Scharloth, Joachim (2010): Kontext korpuslinguistisch: Die induktive Berechnung von Sprachgebrauchsmustern in großen Textkorpora. In: Klotz, Peter– Portmann-Tselikas, Paul–Weidacher, Georg (Hg.): Text-Zeichen und Kon-Texte. Studien zu soziokulturellen Konstellationen literalen Handelns. Tübingen: Narr, S. 85–108.



- Burger, Harald (2003): *Phraseologie: eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmidt Verlag.
- Cvrček, Václav (2013): *Kvantitativní analýza kontextu*. Praha: NLN Nakladatelství lidové noviny.
- Dönninghaus, Sabine (2005): *Die Vagheit der Sprache. Begriffsgeschichte und Funktionsbeschreibung anhand der tschechischen Wissenschaftssprache*. Wiesbaden: Harrassowitz Verlag.
- Đurčo, Peter (2014): *Feste Wortverbindungen mit Adjektiven: Korpuslinguistischer Ansatz als Grundlage für bilinguale Lexikographie*. In: Đurčo, Peter et al. (Hg.): *Valenz und Kookkurrenz. Grammaticale und lexikologische Ansätze*. Wien: Lit Verlag, 147–161.
- Đurčo, Peter (2016): *Zum Konzept der Kollokationsdidaktik und des Kollokationslernens bei Germanistikstudenten*. In: Đurčo, Peter (Hg.): *Kollokationsforschung und Kollokationsdidaktik*. Wien: Lit Verlag, 147–172.
- Evert, Stefan (2009). *Corpora and Collocations*. In: Lüdeling, Anke; Kytö, Merja (Hg.): *Corpus Linguistics. An International Handbook*. Berlin/New York: Mouton de Gruyter, 1212–1248.
- Feilke, Helmut (1993): *Common-sense-Kompetenz. Überlegungen zu einer Theorie des „sympathischen“ und „natürlichen“ Meinens und Verstehens*. Frankfurt am Main: Suhrkamp.
- Feilke, Helmut (1996): *Sprache als soziale Gestalt: Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt am Main: Fischer.
- Feilke, Helmut (2004): *Kontext-Zeichen-Kompetenz. Wortverbindungen unter sprachtheoretischem Aspekt*. In: Steyer, Kathrin (Hrsg.): *Wortverbindungen mehr oder weniger fest. Jahrbuch des IDS 2003*. Berlin: Walter de Gruyter, 41–64.
- Feilke, Helmut (1993): *Sprachlicher Common sense und Kommunikation. Über den „gesunden Menschenverstand“, die Prägung der Kompetenz und die idiomatische Ordnung des Verstehens*. In: *Der Deutschunterricht* 45/6, 6–21.
- Firth, John Rupert (1957): *Modes of Meaning*. In: *Papers in Linguistics 1934–1951*. London: Oxford University Press, 190–215.
- Glynn, Dylan (2014): *Techniques and tools. Corpus methods and statistics for semantics*. In: Glynn, Dylan – Robinson, Justyna: *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins, 307–342.
- Hausmann, Franz Josef (1976): *Strukturelle Wortschatzbetrachtung von Saussure*. In: *Romanische Forschungen* 88, 331–354.
- Heinemann, Margot (2000): *Textsorten des Bereichs Hochschule und Wissenschaft*. In: Brinker, Klaus – Antos, Gerd; Heinemann, Wolfgang – Sager, Sven F. (Hg.): *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. Handbücher zur Sprach- und Kommunikationswissenschaft, Bd. 16.1*. Berlin/New York: de Gruyter, 702–709.
- Kilgariff, Adam – Rychlý, Pavel – Smrž, Pavel – Tugwell, David (2004): *The Sketch Engine. Information Technology*. In: *Proceedings of EURALEX 2004*. Lorient. France. 42.
- Kratochvílová (Zündorf), Iva (2006): *Zum aktuellen Thema: Die Korpuslinguistik im deutsch-tschechischen Sprachvergleich*. In: *Acta Facultatis Philosophicae Universitatis Ostraviensis*. Ostrava: Ostravská univerzita v Ostravě. Filozofická fakulta, 35–38.
- Kratochvílová (Zündorf), Iva (2011): *Kollokationen im Lexikon und im Text. Mehrwortverbindungen im Deutschen und Tschechischen*. Berlin: LIT Verlag. *Studien und Quellen zur Sprachwissenschaft*.
- Lemnitzer, Lothar – Zinsmeister, Heike (2010): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr Francke Attempto Verlag.
- Perkuhn, Rainer – Keibel, Holger – Kupietz Marc (2012): *Korpuslinguistik*. Paderborn: Wilhelm Fink.

- Pořízka, Petr (2014): *Tvorba korpusů a vytěžování jazykových dat: metody, modely, nástroje*. Olomouc: Vydavatelství Filozofické fakulty Univerzity Palackého v Olomouci.
- Roelcke, Thorsten (2015): *Besondere Wörter II: Fachwörter, Termini*. In: Haß, Ulrike – Storjohann, Petra (Hg.): *Handbuch Wort und Wortschatz. Handbücher Sprachwissen 3*. Berlin/Boston: de Gruyter, 371–393.
- Scharloth, Joachim – Eugster, David; Bubenhofer, Noah (2013): *Das Wuchern der Rhizome. Linguistische Diskursanalyse und Data-driven Turn*. In: Busse, Dietrich – Teubert, Wolfgang (Hg.): *Linguistische Diskursanalyse. Neue Perspektiven*. Wiesbaden: VS Verlag, 345–380.
- Steinhoff, Torsten (2009): *Alltägliche Wissenschaftssprache und wissenschaftliche Textprozeduren. Ein Vorschlag zur kulturvergleichenen Untersuchung wissenschaftlicher Texte*. In: Dalmas, Martine – Foschi Albert, Marina – Neuland, Eva (Hg.): *Wissenschaftliche Textsorten im Germanistikstudium deutsch-italienisch-französisch kontrastiv. Trilaterales Forschungsprojekt in der Villa Vigoni (2008–2008)*. Villa Vigoni.
- Steyer, Kathrin (2011): *Von der sprachlichen Oberfläche zum Muster. Zur qualitativen Interpretation syntagmatischer Profile*. In: Elminger, Daniel – Kamber, Alain (Hg.): *La linguistique de corpus – de l'analyse quantitative à l'interprétation qualitative. Korpuslinguistik – von der quantitativen Analyse zur qualitativen Interpretation. TRANEL Travaux neuchâtelois de linguistique 55*. Neuchâtel: Université de Neuchâtel, 219–239.
- Weinrich, Harald (1994): *Sprache und Wissenschaft*. In: Kretzenbacher, Heinz Leonard – Weinrich, Harald (Hg.): *Linguistik der Wissenschaftssprache*. Berlin: de Gruyter, 3–14.
- Wolf, Norbert Richard (2019): *„Endlich die rechten Anstalten sind billig zu knüpfen. Zusammensetzung vortrefflicher Leute auszusetzen“*. Sprachkultur, Sprachakademien und Fachterminologie. In: Vaňková, Lenka (Hg.): *Fach(sprach)lichkeit in fachlichem und nichtfachlichem Kontext*. Berlin: Frank & Timme.
- Zündorf (Kratochvílová), Iva (2015): *Kollokationen aus deutsch-tschechischer Perspektive: Kriterien und Optionen für die Erfassung der konfigurierten Mehrwortlexik*. In: Ďurčo, Peter: *Kollokationsforschung und Kollokationsdidaktik*. Wien: Lit Verlag, 2016, S. 65–80.
- Zündorf (Kratochvílová), Iva (2016): *Formulierungsroutinen und Konfigurationen der fachinternen Wirtschaftskommunikation als Spezialgebiet der fachsprachlichen Textlinguistik und Phraseologie*. In: Satzger, Alex – Vaňková, Lenka – Wolf, Norbert Richard (Hg.): *Fachkommunikation im Wandel. The changing landscape of professional discourse*. Ostrava: Ostravská univerzita v Ostravě, 2015, 65–78.
- Zündorf (Kratochvílová), Iva (2018): *Korpuslinguistik und Übersetzung von Konnektoren: Zur Bedeutung von repräsentativen Datensammlungen und korpusbasierten sprachwissenschaftlichen Analysen für die Übersetzungspraxis*. *Brünner Beiträge zur Germanistik und Nordistik* 32, 1, 61–73.

## **Elektronische Quellen**

- Anthony, Laurence (2018): *AntConc (Version 3.5.7) [Computer Software]*. Japan: Waseda University. (online), abgerufen am 30.4.2019 <https://www.laurenceanthony.net/software/antconc/>.
- Banerjee, Satanjeev – Pedersen, Ted (2003): *The Design, Implementation, and Use of the Ngram Statistic Package*. In: *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City. (online), abgerufen am 30.4.2019 <http://ngram.sourceforge.net/>.

Kilgarriff, Adam – Baisa, Vít – Bušta, Jan – Jakubíček, Miloš – Kovář, Vojtěch – Michelfeit, Jan – Rychlý, Pavel – Suchomel, Vít (2014): The Sketch Engine: ten years on. *Lexicography ASIALEX* (2014) 1. S. 7–36. Berlin/Heidelberg: Springer. (online), abgerufen am 30.4.2019, <https://www.sketchengine.eu/>.

Kilgarriff, Adam – Rychlý, Pavel – Smrž, Pavel – Tugwell, David (2004): the sketch engine. *Information Technology, Itri-04-08*. (online), abgerufen am 30.4.2019, <https://www.sketchengine.eu/>.

---

**Mgr. Markéta Hotařová** / [marketa.hotarova@seznam.cz](mailto:marketa.hotarova@seznam.cz)

Masarykova univerzita, Filozofická fakulta, Ústav germanistiky, nordistiky a nederlandistiky  
Arna Nováka 1, 602 00 Brno, CZ

---

**prof. PhDr. Iva Zündorf, Ph.D.** / [izundorf@phil.muni.cz](mailto:izundorf@phil.muni.cz)

Masarykova univerzita, Filozofická fakulta, Ústav germanistiky, nordistiky a nederlandistiky  
Arna Nováka 1, 602 00 Brno, CZ



This work can be used in accordance with the Creative Commons BY-SA 4.0 International license terms and conditions (<https://creativecommons.org/licenses/by-sa/4.0/legalcode>). This does not apply to works or elements (such as image or photographs) that are used in the work under a contractual license or exception or limitation to relevant rights

---

