

Daniela Prutscher

Universität Jena

Deutschland

<https://orcid.org/0000-0003-4693-5832>

Phraseme und Wortverbindungen in der deutschen Wissenschaftssprache¹

Abstract. Wissenschaftssprache zeichnet sich unter anderem durch die Verwendung von spezifischen Wortverbindungen und musterhaften Formulierungsweisen aus. Anhand eines Korpus, das aus 500 wissenschaftlichen Artikeln aus den Jahren 2010 bis 2017 besteht, werden aktuelle Kollokationen und Textroutinen für die Bereiche Germanistik, Glottodidaktik und Medizin ermittelt und nachgewiesen.

Schlüsselwörter: *Wissenschaftssprache, Wortverbindungen, Kollokationen, Phraseme, sprachliche Muster*

1. Einleitung

Das von der Deutsch-Polnischen Wissenschaftsstiftung geförderte und an der Universität Ermland-Masuren in Olsztyn und der Friedrich-Schiller-Universität in Jena angesiedelte Projekt "D/P-IPHRAS: Interphraseologie als Element der Wissenschaftssprache"² widmet sich dem Phänomen von (mehr oder weniger) festen Wortverbindungen in der deutschen und polnischen Wissenschaftssprache.

Für den Wissenschaftsdiskurs beider Sprachen relevante und typische Wendungen werden hier anhand aktueller Texte korpuslinguistisch ermit-

¹ Dieser Beitrag konnte durch die Unterstützung der Deutsch-Polnischen Wissenschaftsstiftung entstehen, die das Projekt *D/P-IPHRAS: Interphraseologie als Element der Wissenschaftssprache* (Projektnummer 2017-07) finanzierte.

² Leitung: Dr. Bettina Bock (Jena); Mitarbeiterinnen: Dr. Monika Czerepowicka, Dr. Magdalena Makowska und Dr. Joanna Targońska (Olsztyn) sowie Francis Gieseke-Golembowski und Daniela Prutscher (Jena).

telt und über die Datenbank des schon bestehenden Portals "IPHRAS – Interphraseologismen für Studien- und Berufsmobile" (www.iphras.eu) zugänglich gemacht. Dieses Portal stellt bereits für den Bereich "Studium und Beruf im Ausland" Phraseme zum Thema "Bewerbung" und allgemeine Routineformeln in den Sprachen Bulgarisch, Deutsch, Englisch, Griechisch, Romani, Rumänisch und Türkisch zum Üben und Lernen zur Verfügung. Das aktuelle Projekt ergänzt diese Datenbank um das Polnische als einer weiteren im Kontext der Mehrsprachigkeit wichtigen Sprache Europas.

Die für das D/P-IPHRAS-Projekt ermittelten Daten sollen mit einem praxisnahen Tool eine Hilfe für alle sein, die deutsche oder polnische wissenschaftliche Texte verfassen. Der Fokus liegt dabei sowohl auf (polnischen) Studierenden der Germanistik und (deutschen) Polnisch-als-Fremdsprache-Studierenden als auch den muttersprachlichen Studierenden dieser Fächer. Deshalb wurden für das Korpus, das aus jeweils 500 wissenschaftlichen Artikeln in deutscher und polnischer Sprache besteht, Texte aus den für die Studienrichtungen Germanistik und Polonistik besonders relevanten Bereichen Sprachwissenschaft, Literaturwissenschaft und Glottodidaktik³ gewählt. Als Kontrast zu den geisteswissenschaftlichen Beiträgen wurden für beide Sprachen Artikel aus dem Fachgebiet Medizin hinzugenommen.

2. Wissenschaftssprache

Die alltägliche oder allgemeine Wissenschaftssprache stellt über die Fachtermini der einzelnen Disziplinen hinausgehend den "Anteil der für wissenschaftliche Zwecke verwendeten Sprache, der in allen Fächern bekannt, verwendbar und – mehr oder weniger frequent – auch im Gebrauch ist" (Graefen 2001: 191) dar. Sie ist "Bestandteil, Resultat und zugleich Voraussetzung der Wissenschaftskommunikation, also unabdingbar für jeden, der sich am deutschen Wissenschaftsbetrieb beteiligen will" (Ehlich & Graefen 2001: 373). Zu ihr zählen "die fundamentalen sprachlichen Mittel [...], derer sich die meisten Wissenschaften gleich oder ähnlich bedienen" (Ehlich 1993: 33). Als relevante Merkmale der Wissenschaftskommunikation gelten beispielsweise die vier Gebote *Ökonomie*, *Präzision*, *Origo-Exklusivität* und *Dis-*

³ Der Begriff *Glottodidaktik* wird vor allem in Polen verwendet und dort einerseits synonym zu dem im deutschen Sprachraum verbreiteten Terminus Sprachlehr- und -lernforschung verwendet und bezeichnet andererseits auch den Prozess des Unterrichts einer Fremdsprache (vgl. Targońska 2014: Anmerkung 18).

kussion.⁴ Andere für die Wissenschaftssprache wichtige Handlungsmuster sind *Fragen, Begründen, Argumentieren* oder *Demonstrieren* (Ehlich 1993: 26 ff.).⁵

Fest steht auch, dass “[e]s [...] sich bei der allgemeinen Wissenschaftssprache um ein relativ festes Repertoire von Ausdrucksweisen [handelt], das zwecks Erhaltung seiner Funktionalität nicht beliebig spontan erweiterbar oder variierbar ist” (Fandrych & Graefen 2010: 512), denn “[h]och differenzierte kulturelle Ordnungen wie [...] die wissenschaftliche Domäne zeichnen sich durch den Gebrauch eines spezialisierten Ausdrucksspektrums aus. Die Verwendung dieser Ausdrücke ist für jeden, der an der Kommunikation in der Domäne partizipieren will, sozial bindend” (Steinhoff 2007: 100 f.)

Die Wissenschaftssprache zeichnet sich daneben durch bestimmte Formulierungsmuster aus. Diese Musterhaftigkeit der Ausdrucksweisen kann auch durch Kollokationen deutlich werden (vgl. Feilke 2012: 17).

Kollokationen oder feste Wortverbindungen⁶ gehören auch in den Bereich der Phraseologie, in der es jedoch eine Vielzahl von Termini gibt, die – abhängig vom jeweiligen Forschungsschwerpunkt – unterschiedliche Phänomene bezeichnen können. Bereits 1981 hat Klaus-Dieter Pilz konstatiert, dass

[b]isher [...] über die Terminologie der Phraseologie sehr wenig reflektiert worden [ist]. Deshalb konnte sie (vor allem die deutschsprachige) so katastrophal ausufern. Überspitzt gesagt: Jeder, der sich mit phras.[eologischen] Phänomenen befasste, schuf sich eine eigene Terminologie. (Pilz 1981: 27)

Auch 20 Jahre später hat sich daran nicht viel geändert:

Wortübergreifende Phänomene werden als Phraseologismen **oder** als Kollokationen **oder** als Valenz **oder** als syntaktische Konkurrenz **oder** als Funktionsverbgefüge **oder** als Regelsystem semantischer Vereinbarkeiten **oder** als kommunikative Formeln betrachtet. (Steyer 2000: 102; Hervorhebungen im Original)

⁴ Die Umsetzung dieser Gebote kann durch Attribuierungen, nominale und pronominale Wiederaufnahmen, Passivkonstruktionen zur Deagentivierung und semantische Relationen mit für die Wissenschaftskommunikation typischen syntaktischen Junktionstechniken erfolgen (vgl. Czicza et al. 2012: 5–6).

⁵ Ein allgemeines *Ich*-“Verbot” (neben einem *Erzähl*- und einem *Metaphern*-Verbot), wie es Harald Weinrich (1989: 132 ff.) postulierte und Kretzenbacher (1994: 26) spezifizierte, kann im vorliegenden Korpus nicht bestätigt werden: Dieses Personalpronomen wird in 307 der 500 Aufsätze 2.403 mal verwendet, wobei die Treffer auf jeweils 99 Dateien der Sprach- und Literaturwissenschaft, auf 82 der Glottodidaktik und auf lediglich 27 der Medizin verteilt sind (die geringe Frequenz in den medizinischen Artikeln lässt sich auch darauf zurückführen, dass in diesem Fachgebiet hauptsächlich AutorInnenkollektive publizieren). Bei der Betrachtung der *ich*-Verwendung sollte allerdings immer zwischen Verfasser-*ich*, Forscher-*ich* und Erzähler-*ich* unterschieden werden (vgl. z. B. Deml 2015: 56–61).

⁶ Das von Kathrin Steyer herausgegebene *Jahrbuch des Instituts für deutsche Sprache* 2003 trägt den Titel “Wortverbindungen – mehr oder weniger fest”.

Als ein Oberbegriff bietet sich “Phrasem” an. In der weiten Auffassung Harald Burgers aus dem Jahr 2015 bestehen Phraseme

[erstens] aus mehr als einem Wort, zweitens sind die Wörter nicht für dieses eine Mal zusammengestellt, sondern es handelt sich um Kombinationen von Wörtern, die uns als Deutschsprechenden genau in dieser Kombination (eventuell mit Varianten) bekannt sind, ähnlich wie wir die deutschen Wörter (als einzelne) kennen. (Burger 2015: 11)⁷

Während lange Zeit im Allgemeinen Einigkeit darüber bestand, dass sich der Untersuchungsgegenstand der Phraseologie durch *Polylexikalität*, *Einheitsstatus*⁸, *Stabilität*, und *Idiomatizität* auszeichne, erfolgte (auch unter dem Einfluss der Korpuslinguistik) in den letzten Jahren ein Perspektivwechsel “weg von den auffälligen Idiomen hin zu den unauffälligen Kollokationen und musterhaften Konstruktionen” (Handwerker 2010: 249). Sowohl die Konstruktionsgrammatik als auch Erfahrungen aus der Fremdsprachendidaktik haben zusammen mit empirischen Ergebnissen zu der Erkenntnis geführt, dass “[m]ehr oder weniger feste Wortverbindungen [...] keine Sonder-, sondern Normalfälle sprachlicher Zeichenbildung dar[stellen]” (Ágel 2004: 65).

Vor der Arbeit mit großen Korpora wurde Wortverbindungen die Stabilität häufig intuitiv zugeschrieben, nun können tatsächliche Frequenzen ermittelt werden (die unter Umständen dem eigenen Sprachgefühl widersprechen).

In den letzten Jahren wurden auch Forschungsfragen aus dem Bereich der Wissenschaftskommunikation vermehrt korpuslinguistisch bearbeitet und somit auf eine breitere empirische Basis gestellt.⁹ Mit solchen

⁷ Kürzer formuliert es Ken Farø: “Phraseme sind lexikalisierte (spatial) polylexikalische Sprachzeichen” (Farø 2006: 53).

⁸ Phraseme werden mental als Einheiten gespeichert.

⁹ Es muss jedoch angemerkt werden, dass bei bisherigen Korpusuntersuchungen fast ausschließlich Texte aus den Geisteswissenschaften und sogar oft nur aus der germanistischen Sprachwissenschaft bzw. Deutsch als Fremdsprache untersucht wurden. Eine Ausnahme ist jedoch beispielsweise die Untersuchung von Mirjam Weder, deren Korpus aus 100 Wissenschaftstexten die “Disziplinen Geschichte, Philosophie, germanistische Linguistik und Literatur sowie Medien-/Kulturwissenschaften” (Weder 2015: 206) umfasst. In Sarah Brommers Arbeit zu sprachlichen Mustern (2018) werden über 2.000 Texte aus der germanistischen Sprachwissenschaft und der Medizin untersucht. Beispiele für weitere aktuelle korpuslinguistische Untersuchungen sind etwa Franziska Wallners “Kollokationen in Wissenschaftssprachen” (aus dem Jahr 2014) oder die Arbeit Cordula Meißners zu “Figurative[n] Verben in der allgemeinen Wissenschaftssprache des Deutschen” aus dem gleichen Jahr. Aktuelle Projekte wie *GeWiss* (hier handelt es sich um ein mehrsprachiges Korpus zu den Bereichen wissenschaftliche und studentische Vorträge sowie Prüfungsgespräche; siehe <https://gewiss.uni-leipzig.de>; vgl. z.B. Fandrych 2017: 13–32 und Fandrych et al. 2012: 319–337) und *eurowiss* (Untersuchung deutscher

Arbeiten wird dem noch 2010 von Christian Fandrych und Gabriele Graefen bemängelten "Problem einer unzureichenden Datenlage"¹⁰ bei der Untersuchung von Wissenschaftssprache entgegengearbeitet.

Am richtigen Gebrauch von Kollokationen kann nicht nur der Grad der Beherrschung einer Fremdsprache festgemacht werden, auch im Bereich der Wissenschaftssprache geben feste Ausdrücke "wichtige Hinweise auf die wissenschaftssprachliche Kompetenz des Schreibers, indem sie die fachspezifischen Themen und Begriffe indizieren und den fachübergreifenden textlichen, sprachlichen und kognitiven Prozeduren Gestalt geben" (Feilke & Steinhoff 2003: 115). Diese Kompetenz muss in besonderem Maße von L2-Sprecherinnen und -sprechern erworben werden, aber auch MuttersprachlerInnen beherrschen diese Verbindungen nicht automatisch, da sie nicht immer den alltagssprachlichen entsprechen.

3. Das D/P-IPHRAS-Korpus

Für das D/P-IPHRAS-Projekt wurde ein Korpus erstellt, das aus insgesamt 1.000 wissenschaftlichen Artikeln in polnischer und deutscher Sprache aus den Jahren 2010 bis 2017 besteht. Für jede Sprache wurden je 125 Beiträge aus den Bereichen germanistische Sprachwissenschaft, germanistische Literaturwissenschaft, Deutsch als Fremdsprache bzw. Glottodidaktik und Medizin zusammengetragen, die aus jeweils 6 online zugänglichen Zeitschriften stammen – nur die medizinischen Artikel wurden aus 5 Fachmagazinen zusammengestellt. Die Texte wurden in der Korpusvorverarbeitung bereinigt, d.h. Abstracts, Fußnoten und Literaturverzeichnisse wurden ebenso wie längere Zitate und Literaturverweise im Fließtext entfernt. Die Metainformationen zu jedem Artikel beinhalten Zeitschrift, Jahr, Aufsatztitel, AutorInnenamen und – soweit ermittelbar – auch das Geschlecht¹¹. Das bereinigte deutsche Korpus hat einen Umfang von über 2 Millionen (genau 2.281.135) Token.

und italienischer Lehrveranstaltungen aus verschiedenen Wissenschaftsdisziplinen; vgl. Thielmann et al. 2014: 7–17) beschäftigen sich mit der gesprochenen, *GeSig* mit der geschriebenen Wissenschaftssprache – für alle drei wurden umfangreiche Korpora erhoben. *GeSig* verfolgt dabei z.B. das Ziel, "[d]as gemeinsame sprachliche Inventar der Geisteswissenschaften" (siehe <http://research.uni-leipzig.de/gesig/>) korpusmethodisch zu beschreiben.

¹⁰ "Zwar werden Korpora verwendet, diese stehen aber oft nicht für andere Analysen zur Verfügung, sind meist klein und insgesamt heterogen, also untereinander kaum vergleichbar" (Fandrych & Graefen 2010: 510).

¹¹ Vor allem bei den medizinischen Artikeln sind die Vornamen der AutorInnen häufig abgekürzt und das Geschlecht nicht immer ermittelbar.

Tabelle 1: Korpusaufbau

Bereich	Zeitschrift	Beiträge
Glottodidaktik	Acta Philologica	5
	Fadaf – Materialien Deutsch als Fremdsprache	5
	GFL – German as a foreign language	35
	InfoDaF – Informationen Deutsch als Fremdsprache	20
	ZiF – Zeitschrift für interkulturellen Fremdsprachenunterricht	40
	ZVPG – Zeitschrift des Verbandes Polnischer Germanisten	20
Literaturwissenschaft	Acta Philologica	10
	GJB – Goethe-Jahrbuch	20
	LiTheS – Zeitschrift für Literatur- und Theatersoziologie	25
	PBB – Beiträge zur Geschichte der deutschen Sprache und Literatur	25
	TP – textpraxis. Digitales Journal für Philologie	20
	ZVPG – Zeitschrift des Verbandes Polnischer Germanisten	25
Sprachwissenschaft	FRAGL – Freiburger Arbeitspapiere zur Germanistischen Linguistik	10
	PBB – Beiträge zur Geschichte der deutschen Sprache und Literatur	10
	Gesuss – Sprache & Sprachen	15
	tekst i dyskurs – text und diskurs	30
	ZGL – Zeitschrift für germanistische Linguistik	35
	ZVPG – Zeitschrift des Verbandes Polnischer Germanisten	25
Medizin	Der Ernährungsmediziner	10
	Deutsche Zeitschrift für Sportmedizin	30
	GMS Journal for Medical Education	30
	Schmerzmedizin. Angewandte Schmerztherapie und Palliativmedizin	25
	ZfMER – Zeitschrift für Medizin-Ethik-Recht	30

Quelle: eigene Forschung.

Das Korpus ermöglicht es einerseits, *korpusbasiert (corpus-based)* zu arbeiten, d. h. für die Wissenschaftssprache als typisch angesehene Formeln auf ihr Vorhandensein hin zu überprüfen. Andererseits können damit auch Untersuchungen angestellt werden, die *korpusgesteuert* oder *datengeleitet (corpus driven)*¹² sind, denn so können beispielsweise bisher auch unbekannte Wortver-

¹² "Korpusgesteuertes und korpusbasiertes Vorgehen stellen [...] zwei methodische Zugänge dar, die sich ergänzen: Aus datengeleitet ermittelten Beschreibungskategorien lassen sich gezielt

bindungen ermittelt werden. Von Interesse sind neben hochfrequenten Kollokationen auch bisher unbekannte formelhafte Sequenzen oder Paarformeln.

4. Datenanalyse

Die nachfolgenden Daten wurden zum einen mithilfe des Korpus-Analyse-Tools **LancsBox** (Lancaster University corpus toolbox)¹³ erhoben, das 2015 an der Universität Lancaster entwickelt wurde. Dieses Programm kann beispielsweise auch bei deutschen Texten ein Part-of-Speech-Tagging mit dem TreeTagger¹⁴ vornehmen, Sprachdaten visualisieren und Kookkurrenzen mit verschiedenen statistischen Parametern errechnen.

Daneben wurde auch das vom Institut für Deutsche Sprache (IDS) für das Projekt "Usuelle Wortverbindungen" entwickelte Analysewerkzeug zur Untersuchung syntagmatischer Strukturen auf der Basis von Korpusdaten **lexpan** (Lexical Pattern Analyzer)¹⁵ verwendet. Über KWIC-Listen können mit dem Programm explorative, korpusgesteuerte Untersuchungen von Mehrwortstrukturen und sprachlichen Mustern vorgenommen werden.

4.1. Fallanalyse 1: Das hochfrequente Lemma *Frage*

Im Korpus kann nun nach für die Wissenschaftssprache als typisch angesehenen Wortverbindungen gesucht werden. Als ein Beispiel sollen hier verbale Kollokationen zu der Basis *Frage* vorgestellt werden, die es ermöglichen, "Textäußerungen wissenschaftssprachlich angemessen zu artikulieren" (Feilke & Steinhoff 2003: 116). Es gelten jedoch nicht alle denkbaren verbalen Partner als "domänentypische Möglichkeiten" (ebd.):

Es ist nicht verständlich zu machen, warum "sich" in wissenschaftlichen Texten "Fragen stellen", sie "sich" aber nicht, was ja ebenfalls möglich wäre, "ergeben", warum man dort "Fragen nachgeht", ihnen aber nicht "folgt", oder auch, warum "sich" dort "Fragen aufdrängen", aber nicht "aufzwingen". (Steinhoff 2007: 88)

Hypothesen formulieren und korpusbasiert untersuchen. [...] Am Beginn der Untersuchung steht dabei die Beobachtung der Sprachdaten, wobei zunächst alle Evidenzen zu akzeptieren sind. Auf der Basis dieser Evidenzen werden dann Hypothesen gebildet und überprüft. Hierbei stehen induktives und deduktives, also korpusgesteuertes und korpusbasiertes Vorgehen im Wechselspiel" (Wallner 2014: 91).

¹³ <http://corpora.lancs.ac.uk/lancsbox/> (Brezina et al. 2015: 139–173).

¹⁴ 1995 von Helmut Schmid am Institut für maschinelle Sprachverarbeitung der Universität Stuttgart entwickelt: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> (vgl. Schmid 1995).

¹⁵ <http://uwv.ids-mannheim.de/lexpan/>.

“Verständlich” werden solche Präferenzen durch ihre Verwendung, denn “gebrauchte Formen werden zu Formen des Gebrauchs” (Feilke 2003: 2016). Oder um es anders zu formulieren: “Kollokationen sind Mehrworteinheiten, die sich durch usuellen Gebrauch konventionell verfestigt haben” (Belica & Perkuhn 2015: 213).

Was zeigt das Korpus? In 409 der 500 Artikel kommt das Lemma *Frage* 2.321-mal vor, wobei die meisten Treffer im sprachwissenschaftlichen Korpus (941-mal in 109 Texten) zu finden sind, gefolgt von der Glottodidaktik (655-mal in 114 Texten) und der Literaturwissenschaft (424-mal in 104 Texten). Am seltensten taucht das Lemma in medizinischen Artikeln auf (301-mal in 82 Aufsätzen). Die Kookkurrenz *Frage stellen* ist mit 289 Treffern (das entspricht 12,5% aller Verbindungen mit *Frage*) die frequenteste. Mit Abstand am häufigsten wird diese Verbindung in sprachwissenschaftlichen Texten verwendet, nämlich 119-mal in 60 Artikeln. In Beiträgen der Glottodidaktik kommt diese Kombination noch 69-mal (in 49 Artikeln) vor, in denen der Literaturwissenschaft 64-mal (41 Artikel). Nur 26 der 125 medizinischen Aufsätze bedienen sich dieser Formulierungssequenz 37-mal.

Einer *Frage nachgegangen* wird 57-mal; das sind nur noch 2,5%, dennoch spricht die Häufigkeit für eine Kollokation. Auch hier versammeln sich die meisten Treffer in Artikeln der Sprachwissenschaft und der Glottodidaktik (nämlich jeweils 26). Lediglich vier Treffer gibt es in der Literaturwissenschaft und nur einen in der Medizin.

In nur 8 Fällen *drängt sich eine Frage auf* (0,3%). Wie von Steinhoff postuliert, wird den Fragen nicht *gefolgt* und sie *zwingen sich* auch nicht *auf*. 5-mal *ergeben sich* jedoch Fragen, das sind zwar lediglich 0,2% der Treffer; nichtsdestotrotz dürfte eine wissenschaftssprachliche Kollokation vorliegen. Diese Annahme wird auch dadurch gestützt, dass in den Wissenschaftstexten des DWDS¹⁶-Kernkorpus aus den Jahren 1900 bis 1999 die Verbindung immerhin 70-mal¹⁷ vorkommt.

Wird nach den statistisch signifikanten verbalen Kookkurrenz-Partnern¹⁸ für das Lemma *Frage* gesucht, steht auf Platz 1 ebenfalls *stellen*, gefolgt von *beantworten* – dieser Kollokator erzielt 97 Treffer im Korpus (4,2%).

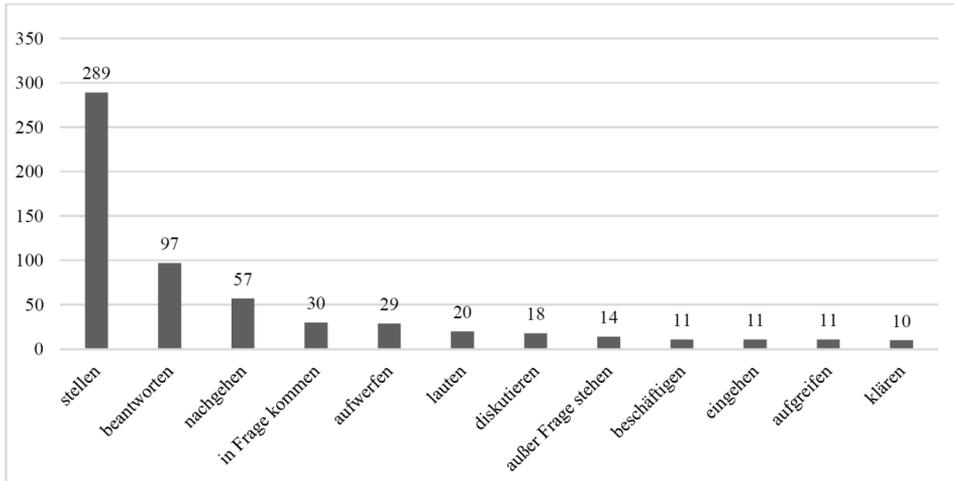
Einen Überblick über verbale Kollokationen, die mindestens 10-mal im Korpus auftreten¹⁹, gibt Abbildung 1:

¹⁶ Digitales Wörterbuch der deutschen Sprache (www.dwds.de/r).

¹⁷ Mit einem rechten und linken Suchabstand zwischen den beiden Lemmata von 15 Token.

¹⁸ Mit dem statistischen Kollokationsmaß der *Mutual Information* und einer Kollokationsweite von jeweils 5 Token auf der rechten und der linken Seite sowie einem statistischen Wert von 5.

¹⁹ Die in einem Korpus mit wissenschaftlichen Texten in hoher Frequenz zu erwartende

Abbildung 1. Verbale Kollokationen zum Lemma *Frage*

Quelle: eigene Forschung.

Als hochfrequentes Lemma erscheint “Frage” oft in ähnlichen oder gleichen Wendungen, die über die reine verbale Kollokation hinausgehen. Beispiele für solche Formulierungsmuster aus dem Korpus, die auch explizit auf das jeweilige Forschungsinteresse des Beitrags Bezug nehmen (vgl. Weder 2015: 208), sind:

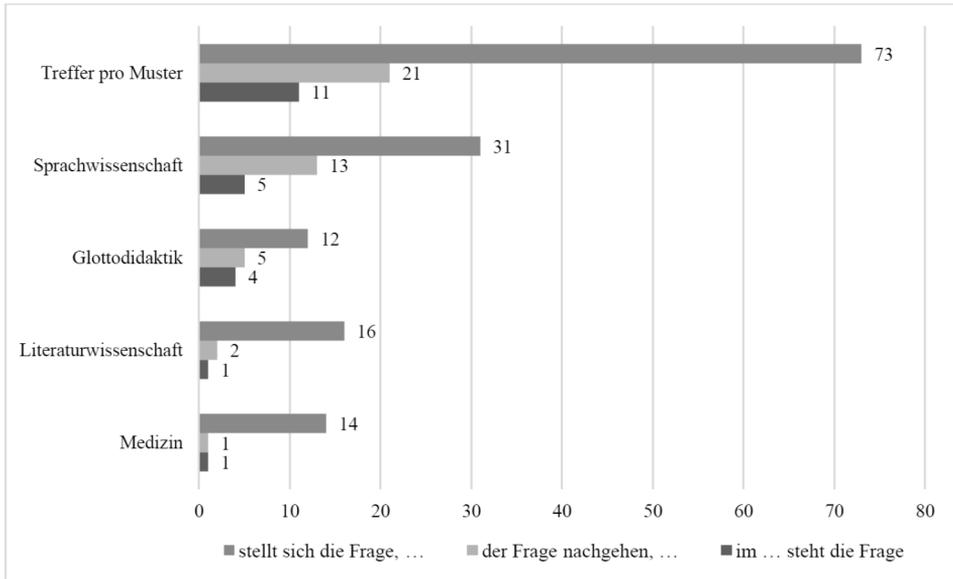
- [es] **stellt sich die Frage**, ob/welche/wie ...
- im Fokus/Mittelpunkt/Zentrum **steht die Frage** ...
- [es] **wird der Frage nachgegangen**, inwieweit/ob/was/warum/welche/wie ...

Hier zeigt sich, dass feste Formulierungssequenzen an den rechten und linken Rändern wiederum formelhaft aufgefüllt werden können.

Knapp die Hälfte dieser musterhaften Verwendungen kommt im Unterkorpus mit den sprachwissenschaftlichen Artikeln vor (49 von 105). Abbildung 2 zeigt die Verteilung.

Bei der Suche nach diesen Mustern in den Textsorten Belletristik, Zeitung und Gebrauchsliteratur im DWDS-Kernkorpus des 20. Jahrhunderts (das ca. 92.196.000 Token umfasst, also über 40-mal größer als das vorliegende Wissenschaftssprachkorpus ist) ergab die Sequenz “es wird der Frage nachgegangen” überhaupt keine Übereinstimmung, die Formulierung “im Fokus/Mittelpunkt/Zentrum steht die Frage” wurde nur 3-mal gefun-

Kollokation *eine Frage behandeln* hat lediglich 5 Treffer (davon 3 in literaturwissenschaftlichen Texten und jeweils 1 aus der Sprachwissenschaft und aus der Glottodidaktik).

Abbildung 2. Musterhafte Verwendungen mit *Frage*

Quelle: eigene Forschung.

den und “[es] stellt sich die Frage” mit den oben erwähnten nachfolgenden Lexemen 28-mal. In den Wissenschaftstexten des DWDS-Kernkorpus (ca. 27.780.000 Token) wird 1-mal “der Frage nachgegangen” und 3-mal “steht die Frage im Fokus/Mittelpunkt/Zentrum”. 30-mal “stellt sich die Frage”. Dieser Befund unterstützt die Annahme, dass diese Muster tatsächlich wissenschaftssprachlich sind und in anderen Textsorten selten oder gar nicht auftreten.

4.2. Fallanalyse 2: Paar- und Mehrlingsformeln

Auch Paar- oder Mehrlingsformeln sind typische Phrasemtypen. Sie sind schon lange Forschungsgegenstand der Phraseologie; durch Erkenntnisse aus der Konstruktionsgrammatik rückt die Musterhaftigkeit jedoch besonders in den Blick: Sie werden immer nach dem Muster gebildet, dass “[z]wei Wörter der gleichen Wortart oder auch zweimal dasselbe Wort [...] mit *und*, einer anderen Konjunktion oder einer Präposition zu einer paarigen Form verbunden” (Burger 2015: 55) werden. Die Formelhaftigkeit zeigt sich hier nicht nur in der festen oder einer bevorzugten Reihenfolge (z.B. *klipp und klar* vs. **klar und klipp*), sondern auch in der lexikalischen Festigkeit (z.B. *ganz und gar* vs. **völlig und gar*; vgl. Gawel 2017: 25–43). Ob solche in der Umgangssprache

häufig verwendeten sprachlichen Muster auch in der Wissenschaftssprache Verwendung finden, kann ebenfalls korpuslinguistisch untersucht werden. Wird nicht nach den verschiedenen Arten von Paarformeln unterschieden, sondern lediglich eine Frequenz von mindestens 10 Treffern vorausgesetzt, ergibt sich für die Paarformeln des Korpus das in Tabelle 2 dargestellte Ergebnis (wobei die jeweils meisten Treffer grau hinterlegt sind):

Tabelle 2: Frequente Paarformeln

Paarformel	Gesamt	Glottodidaktik	Literaturwissenschaft	Sprachwissenschaft	Medizin
<i>mehr oder weniger</i>	182	61	40	78	3
<i>nach wie vor</i>	82	28	16	27	11
<i>nach und nach</i>	43	11	19	9	4
<i>mehr oder minder</i>	31	4	16	11	0
<i>mit und ohne</i>	24	7	1	8	8
<i>mehr und mehr</i>	19	6	9	3	1
<i>[im] Großen und Ganzen</i>	17	8	3	5	1
<i>Schritt für Schritt</i>	14	9	4	0	1
<i>hin und her</i>	14	3	10	1	0
<i>Hand in Hand</i>	11	4	5	2	0
<i>ganz und gar</i>	10	1	5	4	0
<i>Theorie und Praxis</i>	10	3	3	2	2
gesamt	457	145	131	150	31

Quelle: eigene Forschung.

Die mit Abstand häufigste Verwendung findet die mit dem disjunktiven Konnektor *oder* verbundene Formel *mehr oder weniger*, die allein schon 40% aller Treffer ausmacht. Insgesamt finden sich vergleichbar viele Ergebnisse in den Artikeln der Sprachwissenschaft, der Glottodidaktik und der Literaturwissenschaft (33, 32 und 29% der Treffer) – in den medizinischen Beiträgen werden die Paarformeln fast gar nicht verwendet (nur 7%).

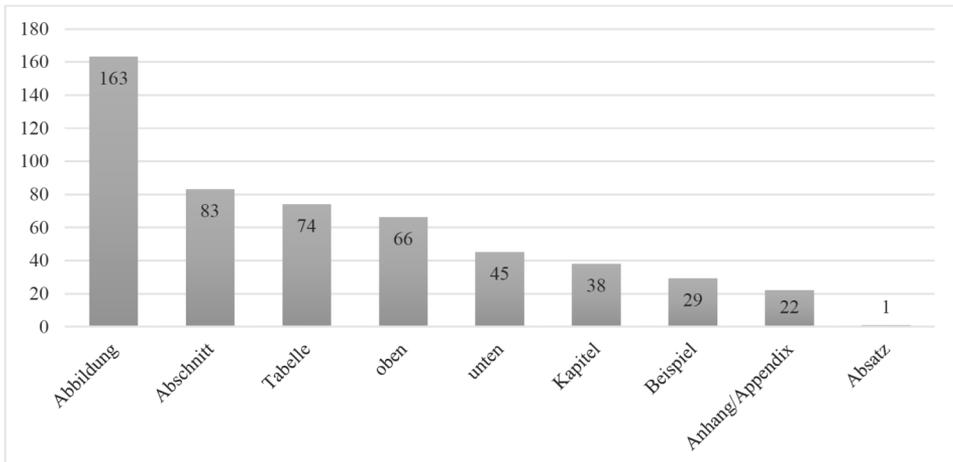
4.3. Fallanalyse 3: Metadiskursive Textroutinen

Als typisch für wissenschaftliche Beiträge gelten auch metadiskursive Textroutinen oder Metakomentierungen. Diese sprachlichen Routinen kommentieren die Struktur und Ziele des Textes und sind häufig musterhaft aufgebaut. Sie sind somit wichtige "Bausteine" der Wissenschaftssprache, die zum einen der Verständnissicherung beim Rezipieren dienen, zum anderen

aber auch den AutorInnen durch ihre textkompositorischen Funktionen bei der Textproduktion helfen (vgl. Olszewska 2013: 80).

Metadiskursive Verfahren wie das intra- und intertextuelle Verweisen treten in der Wissenschaftssprache häufiger auf als im nichtwissenschaftlichen Sprachgebrauch (vgl. Hyland 2005). Die einfachste Art eines Metakommentars ist die Verbindung mit *siehe* oder *vergleiche*. Diese Verben können zwar auch auf Sachverhalte außerhalb des Textes verweisen (z.B. auf Literatur), sie dienen aber ebenso der internen Textorganisation, indem auf oben oder unten Erwähntes oder auf zurückliegende und folgende Kapitel, Abschnitte, Seiten, Tabellen oder Abbildungen verwiesen wird. Im Korpus kommen 707 solcher Verweise vor. 45% davon werden in den sprachwissenschaftlichen Aufsätzen verwendet, 25 bzw. 24% in der Medizin und der Glottodidaktik und lediglich 6% in literaturwissenschaftlichen Beiträgen. Am häufigsten werden vorhergehende oder nachfolgende Abbildungen genannt, gefolgt von Abschnitten und Tabellen, die allgemeineren Verortungen *oben* und *unten* finden sich auf den Plätzen 4 und 5 (siehe Abbildung 3).

Abbildung 3. Intratextuelle Verweise mit *siehe* und *vergleiche*



Quelle: eigene Forschung.

Komplexere Verweise können metakommunikative/textkommentierende Äußerungen wie *“in diesem Abschnitt soll gezeigt werden ...”*, *“das folgende Kapitel geht auf X ein”*, *“im Folgenden geht es um ...”* und andere (mehr oder weniger) feste Wortverbindungen sein.

Textorganisierende Ausdrücke (vgl. Völz 2016: 14) dienen bei wissenschaftlichen Beiträgen oft dem *Initiieren*, *Sequenzieren* und *Schließen* (vgl. Olszewska 2013: 91) des gesamten Textes oder eines Textabschnitts.

Als Beispiel sollen hier sprachliche Muster für das Sequenzieren vorgestellt werden. Danuta Olszewska nennt hierfür die drei relationalen Positionen *zunächst*, *dann* und *schließlich*.²⁰ Als Beispiele für die erste Position in einer Sequenz sind etwa folgende Muster im Korpus zu finden:

- **zunächst** soll[en] [X] beschrieben/betrachtet/diskutiert/überprüft ... werden
- [es] ist **zunächst** zu bemerken/klären/konstatieren/zeigen, dass/was/wie ...
- **zu Beginn** des Abschnitts/Artikels/Aufsatzes/Beitrags/Kapitels ... wird/soll ...
- **am Anfang** der Analyse/des Beitrags/Kapitels/Textes ... steht

Beispielhafte Metatexteme²¹ für die nächste Position der Sequenz sind

- der **nächste** Abschnitt/Punkt/Schritt widmet sich ...
- das **nächste** Beispiel/Kapitel zeigt ...
- wir wollen/ich möchte **nun** auf [X] eingehen
- im **folgenden** Abschnitt/Beispiel/Beleg wird argumentiert/dargelegt/hergeleitet ...

Geschlossen wird eine Sequenz durch Muster wie

- **zusammenfassend** kann bestätigt/festgehalten/gesagt werden, dass ...
- **zusammenfassend** lässt sich festhalten/feststellen/sagen, dass ...
- **abschließend** ist zu betonen/festzustellen/zu konstatieren ...

Auch hier zeigt sich, dass es formelhafte Verwendungen gibt, die entweder identisch oder mit lexikalischen Varianten gebraucht werden. Solche Textschablonen können auch unterschiedlich komplex aufgefüllt werden, sie ändern ihre Struktur dabei aber nicht.

5. Zusammenfassung

Die Beispiele haben gezeigt, dass mit korpuslinguistischer Analyse und den verwendeten Programmen sowohl wissenschaftssprachliche Kollokationen als auch musterhafte Verwendungen im gesamten Korpus der Wissenschaftssprache nachgewiesen und entdeckt werden können.

²⁰ "Während des gesamten Textherstellungsprozesses hat der Autor zu entscheiden, in welcher Reihenfolge er das zu vermittelnde Wissen präsentieren soll. Diese Entscheidungen manifestieren sich häufig an der Textoberfläche in Form von Metatextemen mit zahlreichen ablaufkonstituierenden Ausdrücken, die die Position der einzelnen Wissenskomponenten in der linearen Abfolge bestimmen und ihre Relationierung verdeutlichen" (Olszewska 2013: 93).

²¹ "In der deutschen linguistischen Literatur werden dafür verschiedene Bezeichnungen gebraucht, z.B. *metakommunikative Mittel* (Göpferich 1995), *Textkommentare* (Graefen 1997), *metadiskursive Signale* (Mautner 2011). Aus ökonomischen Gründen verwende ich den Einzelwortterminus *Metatexteme*, der eine Übersetzung der polnischen Bezeichnung *metatekstemy* ist, die Gajewska (2004) in die polnische Linguistik eingeführt hat" (Olszewska 2013: 80).

Die Festigkeit der Wortverbindungen besteht zum einen in der genauen Wiederholung lexikalischer Abfolgen, aber auch in der Struktur von Mustern, die dann mit lexikalischen Varianten gefüllt werden (die wiederum aus einem relativ festen Repertoire bestehen). Auch wenn alle Unterkorpora Treffer aufweisen, sind besonders die sprachwissenschaftlichen Texte von Musterhaftigkeit und formelhaften Verwendungen geprägt.

Literatur

- Ágel, Vilmos. 2004. "Phraseologismus als (valenz)syntaktischer Normalfall". In: Steyer, Kathrin (Hg.). *Wortverbindungen – mehr oder weniger fest*. Berlin / New York: de Gruyter, 65–86.
- Belica, Cyril & Perkuhn, Rainer. 2015. "Feste Wortgruppen / Phraseologie I: Kollokationen und syntagmatische Muster". In: Haß, Ulrike & Storjohann, Petra (Hgg.): *Handbuch Wort und Wortschatz*. Berlin / New York: de Gruyter, 201–225.
- Brezina, Vaclav / McEnery, Tom & Wattam, Stephen 2015. "Collocations in context: A new perspective on collocation networks." In: *International Journal of Corpus Linguistics* 20 (2), S. 139–173.
- Brommer, Sarah. 2018. *Sprachliche Muster. Eine induktive korpuslinguistische Analyse wissenschaftlicher Texte*. Berlin / Boston: de Gruyter.
- Burger, Harald. 2015. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 5., neu bearbeitete Auflage. Berlin: Erich Schmidt.
- Burger, Harald / Dobrovol'skij / Kühn, Peter & Norrick, Neal R. 2007. "Phraseologie: Objektbereich, Terminologie und Forschungsschwerpunkte". In: Burger, Harald / Dobrovol'skij, Dmitrij / Kühn, Peter & Norrick, R. Neal (Hgg.): *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung* (Handbücher zur Sprach- und Kommunikationswissenschaft 35.1). Berlin / New York: de Gruyter, 1–10.
- Czicza, Dániel / Hennig, Mathilde / Emmerich, Volker & Niemann, Robert. 2012. "Zur Verortung von Texten zwischen den Polen maximaler und minimaler Wissenschaftlichkeit. Ein Operationalisierungsvorschlag". In: *Fachsprache* 1–2, 2–44.
- Deml, Isabell. 2015. *Gebrauchsnormen der Wissenschaftssprache und ihre Entwicklung vom 18. bis zum 21. Jahrhundert*. Dissertation. Universität Regensburg.
- Ehlich, Konrad. 1993. "Deutsch als fremde Wissenschaftssprache". In: *Deutsch als Fremdsprache* 19, 13–42.
- Ehlich, Konrad & Graefen, Gabriele. 2001. "Sprachliches Handeln als Medium diskursiven Denkens. Überlegungen zur sukzursiven Einübung in die deutsche Wissenschaftskommunikation". In: *Jahrbuch Deutsch als Fremdsprache 2001* (27). München: iudicium, 351–378.
- Fandrych, Christian. 2017. "Gesprochene-Sprache-Forschung und Korpuserschließung am Beispiel von GeWiss digital". In: Fandrych, Christian / Meißner, Cordula & Wallner, Franziska (Hgg.): *Gesprochene Wissenschaftssprache – digital. Ver-*

- fahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg, S. 13–32.
- Fandrych, Christian & Graefen, Gabriele. 2010. "Wissenschafts- und Studiensprache Deutsch". In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta & Riemer, Claudia (Hgg.): *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch* (Handbücher zur Sprach- und Kommunikationswissenschaft 35.1). Berlin / New York: de Gruyter, 509–517.
- Fandrych, Christian / Meißner, Cordula & Slavcheva, Adriana. 2012. "The GeWiss corpus: Comparing spoken academic German, English and Polish". In: Schmidt, Thomas & Wörner, Kai (Hgg.): *Multilingual corpora und multilingual corpus analysis*. Amsterdam: Benjamins, 319–337.
- Farø, Ken. 2015. "Feste Wortgruppen / Phraseologie II: Phraseme". In: Haß, Ulrike & Storjohann, Petra (Hgg.): *Handbuch Wort und Wortschatz*. Berlin / New York: de Gruyter, 226–247.
- Farø, Ken. 2006. *Idiomatizität, Ikonizität, Arbitrarität – Beitrag zu einer funktionalistischen Theorie der Idiomäquivalenz*. Kopenhagen: Museum Tusulanum.
- Feilke, Helmuth. 2012. "Was sind Textroutinen? – Zur Theorie und Methodik des Forschungsfeldes". In: Feilke, Helmuth & Lehnen, Kathrin (Hgg.): *Schreib- und Text-routinen. Theorie, Erwerb und didaktisch-mediale Modellierung*. Frankfurt/Main u.a.: Peter Lang, 1–31.
- Feilke, Helmuth & Steinhoff, Torsten. 2003. "Zur Modellierung der Entwicklung wissenschaftlicher Schreibfähigkeiten". In: Ehlich, Konrad & Steets, Angelika (Hgg.): *Wissenschaftlich schreiben – lehren und lernen*. Berlin / New York: de Gruyter, 112–128.
- Gaweł, Agnieszka. 2017. "Zur Ikonizität deutscher Zwillingformeln". In: *Linguistik online* 81 (2), S. 25–43.
- Graefen, Gabriele. 2001. "Einführung in den Gebrauch der Wissenschaftssprache". In: *Materialien Deutsch als Fremdsprache* 58 ("Wie schwer ist die deutsche Sprache wirklich?"), 191–210.
- Handwerker, Britta. 2010. "Phraseologismen und Kollokationen". In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta & Riemer, Claudia (Hgg.): *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch* (Handbücher zur Sprach- und Kommunikationswissenschaft 35.1). Berlin/New York: de Gruyter, 246–255.
- Hyland, Ken. 2005. *Metadiscourse. Exploring Interacting in Writing*. London/New York: Bloomsbury Publishing.
- Kretzenbacher, Heinz. 1994. "Wie durchsichtig ist die Sprache der Wissenschaften?". In: Kretzenbacher, Heinz & Weinrich, Harald (Hgg.): *Linguistik der Wissenschaftssprache* (= Forschungsbericht der Akademie der Wissenschaften zu Berlin Nr. 10). Berlin / New York: Walter de Gruyter.
- Olszewska, Danuta. 2013. "Auf der Suche nach einem *tertium comparationis*: Wissenschaftliche Texte im deutsch-polnischen Vergleich". In: *Studia Germanica Gedan-sia* 29, 79–99.

- Pilz, Hans-Dieter. 1981. *Phraseologie, Redensartenforschung*. Stuttgart: Metzler.
- von Polenz, Peter. 1981. "Über die Jargonisierung von Wissenschaftssprache und wider die Deagentivierung". In: Bungarten, Theo (Hg.): *Wissenschaftssprache. Beiträge zur Methodologie, theoretischen Fundierung und Deskription*. München: Fink, 85–110.
- Schmid, Helmut. 1995. Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.
- Steinhoff, Torsten. 2007. *Wissenschaftliche Textkompetenz: Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten*. Tübingen: Niemeyer.
- Steyer, Kathrin (Hg.). 2004. *Wortverbindungen – mehr oder weniger fest* (Jahrbuch des Instituts für Deutsche Sprache 2003). Berlin et al.: de Gruyter.
- Steyer, Kathrin. 2000. "Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikographische Möglichkeiten". In: *Deutsche Sprache* 28, 101–125.
- Targońska, Joanna. 2014. "Kollokationen – ein vernachlässigtes Gebiet der DaF-Didaktik?". In: *Linguistik online* 68 (6), 127–149.
- Thielmann, Winfried / Redder, Angelika & Heller, Dorothea. 2014. "Akademische Wissensvermittlung im Vergleich". In: Redder, Angelika / Heller, Dorothea & Thielmann, Winfried (Hgg.): *Eristische Strukturen in Vorlesungen und Seminaren deutscher und italienischer Universitäten*. Heidelberg: Synchron, S. 7–17.
- Völz, Irina. 2016. *Lexikalische Textgliederung beim wissenschaftlichen Schreiben in der Fremdsprache Deutsch. Eine empirische Untersuchung zum Erwerb und Gebrauch textorganisierender Ausdrücke durch internationale DaF-Studierende*. Kassel: university press.
- Wallner, Franziska. 2014. *Figurative Verben in der allgemeinen Wissenschaftssprache des Deutschen. Eine Korpusstudie*. Tübingen: Stauffenburg.
- Weder, Mirjam. 2015. "Fragen, zeigen, argumentieren? Metadiskursive Textroutinen zur Etablierung von Forschungspraktiken in wissenschaftlichen Einleitungen". In: Schmidlin, Regula / Behrens, Heike & Bickel, Hans (Hgg.): *Sprachgebrauch und Sprachbewusstsein. Implikationen für die Sprachtheorie*. Berlin: de Gruyter, 199–220.
- Weinrich, Harald. 1989. "Formen der Wissenschaftssprache". In: *Jahrbuch 1988 der Akademie der Wissenschaften zu Berlin*. Berlin et al.: de Gruyter, 119–158.

Phrasemes and Word Combinations in the German Scientific Language

Summary

Scientific language is characterized, among other things, by the use of specific phrases and exemplary formulations. Current collocations and text routines can be determined and proven using a corpus consisting of 500 scientific articles from the years 2010 to 2017 for the areas of German Studies, Glottodidactics and Medicine.