

# Überprüfung eines phraseologischen Grundwortschatzes – eine explorative Datenanalyse

**Lea Hoffmann**

Institute for German Studies  
University of Duisburg-Essen  
Universitätsstr. 12  
D-45141 Essen, Germany  
E-Mail: [lea.hoffmann@uni-due.de](mailto:lea.hoffmann@uni-due.de)

---

## Review of a Basic Phraseological Vocabulary – an Explorative Data Analysis

**Abstract.** This article addresses the question of which possibilities and limitations of frequency-based studies on the relevance of multi-word expressions open up for applied purposes. For this purpose, the corpus *Ref10* of the project *Wortschatzwissen.de* was exploratively examined. After the development of a category system for multi-word expressions, a sample of the corpus was examined and assigned to the different categories. Subsequently, the identified multi-word expressions were compared with a phrase list of Hallsteinsdóttir, Šajánková & Quasthoff (2006). Findings suggest that the proportion of collocations is particularly high in all subcorpora and that, in addition, idioms and light verb constructions are predominant. Moreover, a large proportion of the idioms identified in the *Ref10* corpus sample does not occur at all or occurs only partially, i.e. in an unlisted variant, in the phraseological optimum of Hallsteinsdóttir, Šajánková & Quasthoff (2006). This raises above all the question of how phrase variance is to be evaluated in corpus analyses and to what extent corpus linguists should rely only on basic vocabulary from the perspective of Applied linguistics.

**Keywords:** phraseology, exploratory analysis, basic vocabulary, fixed words, idioms, multiword expressions, Applied Linguistics

---

Submitted: 05/06/2020. Accepted: 11/09/2020

Copyright © 2020 Lea Hoffmann. Published by Vilnius University Press

This is an Open Access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 1 Einleitung

Wie üblich muss in einer phraseologischen Untersuchung die Terminologie geklärt werden. In diesem Beitrag wird Phrasem als jegliche feste Wortverbindung verstanden (zum Problem der Terminologie vgl. z. B. Burger, Dobrovol'skij, Kühn & Norrick 2007, zur genaueren Ausführung der terminologischen Festlegung in diesem Beitrag siehe Kap. 3.3.1)

Phraseologische Grundwortschätze können im Bereich der angewandten Linguistik und in verschiedenen Anwendungsgebieten unterschiedliche Anknüpfungspunkte bieten. Einen solchen Fall bildet die schulische Praxis, speziell im deutschen Schulsystem. Inwieweit wird die Bedeutung von Phrasemen für das Sprachvermögen von Schülerinnen und Schülern grundsätzlich im Unterricht und in der Forschung beachtet? Wenke Mückel (2014) beschreibt eine Entwicklung in der Phraseologie, bei der die generelle Bedeutsamkeit von Phrasemen im Sprachgebrauch in den Fokus rücke, wodurch das Interesse an einer Didaktisierung der Phraseologie wachse. Daraus entstehe eine eigenständige primärsprachliche Didaktik, die laut Mückel zum Ziel haben solle, die Schnittstellen mit der Fremdsprachendidaktik zu nutzen und gleichzeitig die Besonderheiten für den primärsprachlichen Unterricht herauszuarbeiten (vgl. Mückel 2014, 395). Einen Ansatz hierfür stellt die Entwicklung eines Phraseminventars dar, wodurch der Unterrichtsgegenstand, genauer die Phraseme selbst, näher für den Einsatz im Unterricht ausdifferenziert werden können.

Dieser Beitrag wird von der Frage geleitet, welche Möglichkeiten und Beschränkungen von frequenzbasierten Untersuchungen zur Relevanz von Phrasemen sich für den Einsatz im Regelunterricht auf tun. Zu diesem Zwecke wurde das Korpus *Ref10* aus dem Projekt *Wortschatzwissen* (siehe Kapitel 3.1.1), dessen Textauswahl Jugendliche im Alter von 11–16 als mögliche Rezipienten im Blick hat, explorativ untersucht, indem ein Kategoriensystem für Phraseme deduktiv abgeleitet und die Phraseme einer Stichprobe des Korpus diesen Kategorien zugeordnet wurden. In einem zweiten Schritt wurden die identifizierten Phraseme mit dem phraseologischen Optimum von Hallsteinsdóttir, Šajánková & Quasthoff (2006) verglichen, das in dieser Analyse als Referenzliste herangezogen wird. Auf diese Weise konnte zum einen ein Überblick über die Phrasem ausgewogenheit in der Stichprobe des Korpus gewonnen werden und außerdem überprüft werden, inwieweit ein auf das gesamte Korpus ausgeweiteter korpuslinguistischer Abgleich des Korpus *Ref10* mit einer Phrasemliste aus der Fremdsprachendidaktik als einem phraseodidaktischen Ansatz im deutschen Bildungssystem<sup>1</sup> fruchtbar wäre.

<sup>1</sup> Der Fokus des Korpus *Ref10* und damit auch dieser Untersuchung liegt auf der nationalen Varietät des Bundesdeutschen, da das deutsche Bildungssystem die Grundlage für die didaktischen Überlegungen bildet. Andere nationale Varietäten des Deutschen werden somit nicht berücksichtigt. Ob die Ergebnisse übertragen werden könnten, müsste überprüft werden.

## 2 Phraseodidaktik und Grundwortschätze

Das folgende Kapitel reißt zunächst die Wortschatz- und Phraseodidaktik des Deutschunterrichts an und geht anschließend auf Grundwortschätze und Phraseminventare in didaktischen Zusammenhängen ein.

### 2.1 Wortschatz und Phraseologie in der Sekundarstufe I

Die Phraseodidaktik als eine Unterdisziplin der Wortschatzdidaktik im Regelunterricht ist bisher kaum beforscht. In wenigen Sammelbänden, die sich Wortschatz und Wortschatzarbeit widmen, werden auch phraseodidaktische Perspektiven beleuchtet (siehe dazu Pohl & Ulrich 2011; Merten & Kuhs 2012; Kilian & Eckhoff 2015; Kuhs & Merten 2015; Kuhs & Merten 2019 u. a.)<sup>2</sup>. Neben kleineren empirischen Studien, u. a. in den genannten Sammelbänden, fehlt es bislang an einer Basis von größer angelegten Querschnittstudien, die gezielt Wortschatz- und Phrasemkompetenz von Schülerinnen und Schülern im deutschen Bildungssystem überprüfen. Die Wortschatzdidaktik wird in keinem eigenen Kompetenzbereich dargestellt, sondern gehört gleichzeitig zu allen definierten Kompetenzbereichen.<sup>3</sup> Bislang bekannt ist nur die DESI-Studie, eine größer angelegte Querschnittstudie, bei der die sprachlichen Leistungen im Jahr 2001 von etwa 11.000 Schülerinnen und Schülern der 9. Klasse, also ungefähr im Alter von 15 Jahren, an deutschen Schulen aller Schularten in den Fächern Deutsch und Englisch erhoben wurden. Hier wurde ein deutliches Defizit in Bezug auf die Wortschatzkenntnisse der Lernenden sichtbar (vgl. Willenberg 2008).

### 2.2 Grundwortschätze und Phraseminventare

In den 1970er Jahren gab es eine große Diskussion um Grundwortschätze, also der Sammlung von Einzellexemen (ein Überblick dazu bei Schnörch 2002). Gegenwärtig werden Grundwortschätze in deutschen Schulen vorrangig in den Grundschulen für die Klassenstufen 1–4 bzw. 1–6, also für das Alter von 6–11 bzw. 6–14, eingesetzt (für den Rechtschreibunterricht z. B. für das Bundesland Brandenburg siehe Grünack, Mühlens & Waldmann 2011, aus genuin wortschatzdidaktischer Perspektive siehe Decker-Ernst & Oomen-Welke 2019). Im Bereich des Unterrichts Deutsch als Fremdsprache werden Grundwortschätze aktuell eingesetzt, nicht nur in Bezug auf Einzellexeme (siehe z. B. Tschirner 2008), sondern auch für feste Wortverbindungen (siehe z. B. Hallsteinsdóttir, Šajánková & Quasthoff 2006 und Hessky & Ettinger 1997).

<sup>2</sup> Einen vielversprechenden Sammelband für die Phraseodidaktik stellt ein Sammelband von Mückel (i. E.) dar.

<sup>3</sup> Zum Stellenwert von Wortschatz in der Deutschdidaktik und einer möglichen bzw. gewünschten Verortung siehe Steinhoff (2009).

### 3 Explorative Datenanalyse

Ettinger (2007) sieht in der Überprüfung von Frequenzen von Phrasemen eine Leerstelle der Phraseodidaktik. Mit einer frequenzbasierten Forschung könne man verhindern, dass im Unterricht mit nur selten im Sprachgebrauch auftretenden Phrasemen gearbeitet wird. Er hält daher die Überprüfung von schon vorhandenen Inventaren für sinnvoll (vgl. Ettinger 2007, 902f).<sup>4</sup> Zunächst scheint jedoch die explorative Untersuchung eines dafür geeigneten Korpus sinnvoll. Im Folgenden wird daher eine qualitative Datenanalyse beschrieben, die ein vorliegendes frequenzbasiertes Phraseminventar aus der Fremdsprachendidaktik in einem primärsprachlich-didaktischen Kontext überprüft.

#### 3.1 Datengrundlage

Zuerst wird das Korpus, das die Grundlage für die explorative Datenanalyse bildet, und anschließend das Phraseminventar, das für den Vergleich der in der Analyse identifizierten Phraseme als Referenzliste hinzugezogen wird, erläutert.

##### 3.1.1 Das Korpus *Ref10*

Das zugrundeliegende Korpus *Ref10* wurde im Rahmen des Projektes *Wortschatzwissen. Ein Referenzwortschatz für die Sekundarstufe I<sup>5</sup>* aufgebaut. Das Korpus umfasst derzeit ungefähr 6,5 Millionen Token (Stand: 02.04.2020) und enthält Texte, die entweder Jugendliche ungefähr im Alter von 11–16 Jahren (Sekundarstufe I) adressieren oder von ihnen selbst geschrieben sind. Das Korpus *Ref10* teilt sich insgesamt in fünf Subkorpora auf (siehe Abb. 1).

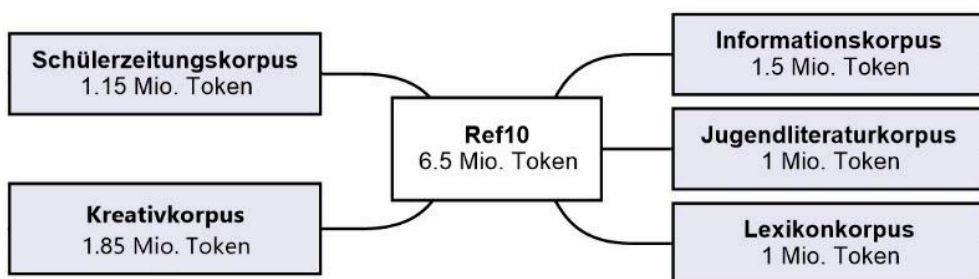


Abbildung 1. Subkorpora des Korpus *Ref10* mit Token-Angaben

<sup>4</sup> Einen Überblick über die Einsatzmöglichkeiten und Limitationen korpuslinguistischer und computergestützter Forschung für die Lexikologie und Phraseologie zeichnen Bubenhofer & Ptashnyk (2010) nach.

<sup>5</sup> Nähere Informationen zum Projekt sind abrufbar unter <https://wortschatzwissen.de/> (06.12.2020).

a) Subkorpora mit Texten **für** Jugendliche: insg. ungefähr 3,5 Millionen Token

Das Lexikonkorpus umfasst verschiedene Jugend- und Kinderlexika, das Jugendliteraturkorpus beinhaltet Werke aus Kinder- und Jugendliteratur und das Informationskorpus ist eine Sammlung aus Informationstexten wie z. B. Infobroschüren oder Texte auf Homepages über Mobilität, Aufklärung und Jugendreisen und weiteren Themen, die potenziell für Jugendliche relevant sind.

b) Subkorpora mit Texten **von** Jugendlichen: insg. ungefähr 3 Millionen Token

Das Schülerzeitungskorpus umfasst Schülerzeitungen verschiedener Schulen und das Kreativkorpus stellt eine Sammlung von Texten wie Forenbeiträgen, Buchrezensionen oder anderen von Jugendlichen selbst verfassten Texten dar.

### **3.1.2 Das phraseologische Optimum für Deutsch als Fremdsprache von Hallsteinsdóttir, Šajánková & Quasthoff (2006)**

Hallsteinsdóttir, Šajánková & Quasthoff (2006) entwickelten mit dem phraseologischen Optimum für Deutsch als Fremdsprache einen Vorschlag für einen Phrasem-Grundwortschatz für den Fremdsprachenunterricht, der nicht nur frequenzbasiert erstellt wurde, sondern erstmals auch Geläufigkeitsaspekte mit einbezieht. Sie haben für diese didaktische Phrasemliste insgesamt 1112 Phraseme<sup>6</sup> untersucht, die durch die Sichtung von verschiedenen phraseologischen Wörterbüchern für Deutsch als Fremdsprache gewonnen wurden. Diese wurden in einer Frequenzuntersuchung auf ihr Vorkommen im Korpus des Wortschatzlexikons *Deutscher Wortschatz*<sup>7</sup> der Universität Leipzig geprüft und in einer Fragebogenstudie in Bezug auf die Geläufigkeit bei Muttersprachlern der Deutschen Sprache untersucht (vgl. Hallsteinsdóttir, Šajánková & Quasthoff 2006, 119–123).

Die Ergebnisse der beiden Untersuchungen mündeten in die Einteilung der untersuchten Phraseme in neun verschiedene Gruppen. Es ergeben sich zunächst zwei Dimensionen: Frequenz im Korpus und Geläufigkeit bei Muttersprachlerinnen und Muttersprachlern. Diese werden in drei Grade aufgeteilt: niedrig (weniger als 100 Belegstellen im Korpus bzw. von 0–50 Teilnehmende der Geläufigkeitsstudie aktiv verwendet), mittel (100–500 Belegstellen bzw. von 51–75 T. aktiv verwendet) und hoch (über 500 Belegstellen bzw. von 76–101 T. aktiv verwendet) (vgl. Hallsteinsdóttir, Šajánková & Quasthoff 2006, 126) (s. Tab. 1).

---

<sup>6</sup> Die hier untersuchten Phraseme sind zum allergrößten Teil idiomatische Phraseme (Anm. der Autorin).

<sup>7</sup> Online-Ressource abrufbar unter <https://wortschatz.uni-leipzig.de/de> (06.04.2020).

<b>Frequenz und Geläufigkeit</b>	<b>Frequenz im Korpus</b>	<b>Geläufigkeit bei Muttersprachlern</b>
Hoch	<b>A</b> über 500 Belegstellen	<b>B</b> 76–101 aktiv verwendet
Mittel	<b>C</b> 101–500 Belegstellen	<b>D</b> 51–75 aktiv verwendet
Niedrig	<b>E</b> 0–100 Belegstellen	<b>F</b> 0–50 aktiv verwendet

Tabelle 1. Erweiterte Tabelle für die Auswertung der Geläufigkeit und der Frequenz von Phraseologismen (Hallsteinsdóttir, Šajánková & Quasthoff 2006, 127)

Die verschiedenen Grade der Dimensionen werden dann kombiniert: Gruppe AB enthält zum Beispiel alle Phraseme mit einer hohen Frequenz im Korpus und einer hohen Geläufigkeit bei Muttersprachlerinnen und Muttersprachlern. Dies sind 143 Phraseme und werden von Hallsteinsdóttir, Šajánková & Quasthoff (2006) u. a. als Kernbereich eines phraseologischen Optimums für Deutsch als Fremdsprache ausgemacht, dessen Phraseme im Grundwortschatz enthalten sein sollten, genauso wie die Phraseme der Gruppe CB (mittlere Frequenz, hohe Geläufigkeit: 258 P.) und Gruppe EB (niedrige Frequenz, aber hohe Geläufigkeit: 223 P.). Insgesamt gehören also etwa 600 Phraseme zum hier ausgemachten Kernbereich des phraseologischen Optimums für Deutsch als Fremdsprache. Sie machen außerdem einen Peripheriebereich aus (Gruppe AD, CD, ED, eventuell AF) und schließen die restlichen Phraseme von dem phraseologischen Optimum aus (vgl. Hallsteinsdóttir, Šajánková & Quasthoff 2006, 127f).<sup>8</sup>

### 3.2 Untersuchungsziel

Die Untersuchung hat zum Ziel, einen Einblick über die Phrasenausgewogenheit – d. h. die Verteilung von Phrasemen in Bezug auf Häufigkeit und Kategorie in Texten, Stichproben oder Korpora etc. – im Korpus *Ref10* zu erlangen. Gibt es Phrasemarten, die besonders häufig vorkommen, und zeichnet sich ein Unterschied zwischen den Subkorpora ab? Und welches Ergebnis ergibt ein Vergleich der im *Ref10* identifizierten Phraseme mit den Phrasemen des phraseologischen Optimums von Hallsteinsdóttir, Šajánková & Quasthoff (2006), das in dieser Untersuchung als Referenzwortschatz dienen soll, da es für eine allgemeine Zielgruppe entwickelt wurde?

Dafür ist eine explorative Datenanalyse einer kleinen Stichprobe des Korpus geeignet, die vor allem folgende Teilfragen beantworten soll:

<sup>8</sup> Man kann Genaueres zur Argumentation zum Ein- und Ausschluss in den Kernbereich bei Hallsteinsdóttir, Šajánková & Quasthoff (2006) nachlesen. Für die Bereitstellung der kompletten Auflistung aller Phraseme der Studie mitsamt der Gruppenzugehörigkeit möchte sich die Autorin bei Erla Hallsteinsdóttir herzlich bedanken.

1. Wie verteilen sich die identifizierten Phraseme auf verschiedene Phrasemkategorien?
2. Gibt es Phrasemkategorien, die besonders häufig/besonders selten in einem Subkorpus oder in der darüber liegenden Struktur (von/für Jugendliche(n)) vorkommen?
3. Wie viele Phraseme kommen auch im phraseologischen Optimum von Hallsteinsdóttir, Šajánková & Quasthoff (2006) vor und wenn ja, in welchem Bereich?
4. Kommen Phraseme vor, die nicht im phraseologischen Optimum aufgeführt sind?

### 3.3 Methodik

Zur Beantwortung der oben formulierten Forschungsfragen wurde zuerst deduktiv, auf Basis von theoretischer Forschung zur Klassifikation von Phrasemen, ein Kategoriensystem entwickelt, das die Grundlage der explorativen Datenanalyse einer Stichprobe des Korpus *Ref10* darstellt.

#### 3.3.1 Kategoriensystem

Das Kategoriensystem für die explorative Analyse umfasst neun Kategorien (s. Tab. 2). Die Phrasemtypen des Kategoriensystems orientieren sich vor allem an Farø (2015). Die Phrasemdefinition, die hier zugrunde gelegt wird, ist daher eine der Phraseme im weiteren Sinn, wonach damit jegliche feste Wortverbindung ohne weitere Einschränkungen gemeint ist (Farø 2015, 231). Ein *Idiom* wird als ein Phrasem im engeren Sinne aufgefasst. Unter *Kollokationen* werden Phraseme verstanden, die sich durch usuellen Gebrauch verfestigt haben, sich aber im Gegensatz zu Idiomen dekompositionell erschließen lassen (vgl. auch Belica & Perkuhn 2015, 213). Von Farø (2015) übernommen wurden die Kategorien *Funktionsverbgefüge* (Mehrwortverbindungen aus einem deverbalen Substantiv und einem desemantisierten Verb), *Phraseme mit unikalen Komponenten* (mit einer nur in dieser Wortverbindung vorkommenden lexikalischen Komponente), *pragmatische Phraseme* (Routineformeln) und *Geflügelte Worte* (mit prägnanter historischer, politischer etc. Verbindung zur Wirklichkeit). Weiterhin wurden einige Kategorien für diesen Zweck zusammengefasst, darunter *Phraseotermini* (phraseologische Fachbegriffe) und *Phraseonyme* (phraseologische Namen), weil sie ähnliche Anforderungen und Funktionen erfüllen. Gleiches gilt für *Sprichwort* (fertig ausgebildeter Allsatz ohne relevanten Urheber oder Ursprungskontext) und *Xenophrasem* (aus einer Fremdsprache stammend). *Paarformel* und *Phraseoduplikation* wurden zu einer Kategorie zusammengefasst, weil Zweiteres ein Spezialfall von Ersterem ist (vgl. Farø 2015, 231–236)<sup>9</sup>. Die Kategorie der Kollokation umfasst Mehrwortverbindungen, die nicht idiomatisch sind

<sup>9</sup> Die phraseologische Typologie von Farø ist weder als abgeschlossen zu sehen, noch trennscharf. Kategorien können sich überschneiden (vgl. Farø 2015, 236).



und unter keine der anderen Kategorien subsummiert werden konnten, wie z. B. (1) *Sorgen machen*.<sup>10</sup> Außerdem wurde die allgemeinere Kategorie der Idiome ergänzt, da es sich neben den genannten Strukturen auch um ein Idiom handeln kann, ohne in einer solchen speziellen Form wie z. B. als Paarformel vorzukommen.

Bez.	Phrasemkategorie		
1	Kollokation (Ko)	6	Phrasem mit unikalenen Komponenten (PuK)
2	Funktionsverbgefüge (FVG)	7	Paarformel oder Phraseoduplikation (PF)
3	Phraseoterminus oder Phraseonym (PT)	8	Pragmatisches Phrasem (PP)
4	Idiom (Id)	9	Geflügelte Worte (GW)
5	Sprichwort oder Xenophrasem (Spr/Xe)		

Tabelle 2. Kategoriensystem für die explorative Datenanalyse

### 3.3.2 Vorgehen

Die explorative Datenanalyse hat eine Stichprobe aus dem *Ref10*-Korpus zur Grundlage. Pro Subkorpus wurden Texte bzw. Textausschnitte im Umfang von 3.000 Token entnommen. Da das Korpus fünf Subkorpora umfasst, ist die Stichprobe insgesamt ungefähr 15.000 Token groß, was etwa 0,23 Prozent der Gesamtkorpusgröße entspricht.

Diese Stichprobe wurde qualitativ untersucht. Dabei wurde jeder Satz der Stichprobe daraufhin untersucht, ob ein Phrasem vorliegt und welcher Kategorie es zugeordnet werden konnte. Ein Phrasem durfte jeweils nur einer Kategorie zugeordnet werden. Wenn ein Phrasem der Kategorie Idiom zugeordnet werden konnte, wird zusätzlich überprüft, ob es auch Teil des phraseologischen Optimums von Hallsteinsdóttir, Šajánková & Quasthoff (2006) ist und ob es dort zum Kern- oder Peripheriebereich gehört.

### 3.4 Ergebnisse

Das erste Unterkapitel bezieht sich auf die ersten beiden Teilfragen, die die Verteilung der Phraseme auf die Kategorien in Bezug auf die verschiedenen Subkorpora behandeln. Danach werden die identifizierten Phraseme mit dem phraseologischen Optimum von Hallsteinsdóttir, Šajánková & Quasthoff (2006) verglichen, womit die letzten beiden Teilfragen beantwortet werden.

<sup>10</sup> Es handelt sich hierbei nicht um eine Definition, die Kollokationen als eine statistische Größe beschreiben, da die Analyse in dieser Untersuchung nur qualitativer Natur ist.



### 3.4.1 Verteilung auf Phrasemkategorien

Es konnten insgesamt 137 Phraseme identifiziert werden, die den neun Kategorien zugeordnet werden können (siehe Abb. 2). Die Betrachtung bezieht sich auf alle identifizierten Phraseme, auch wenn sie mehrmals vorgekommen sind (Token).

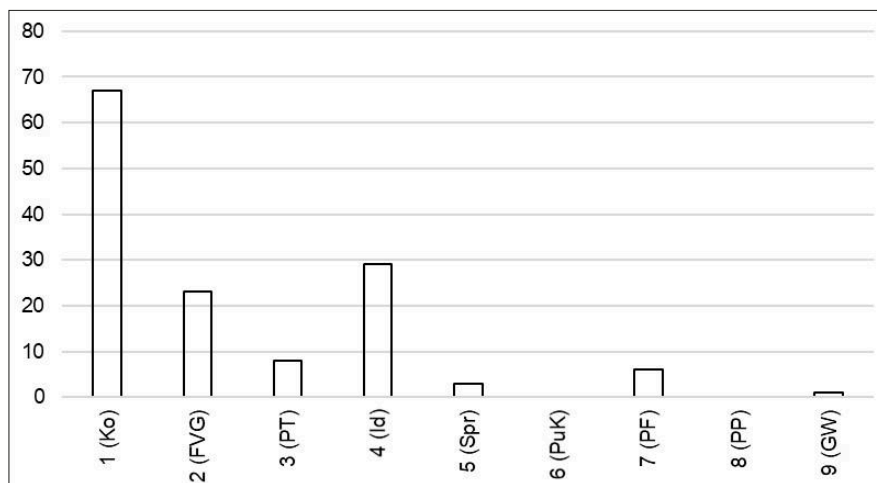


Abbildung 2. Anzahl der Phraseme pro Kategorie

Die meisten Phraseme konnten der Kategorie 1 (Kollokationen) zugeordnet werden, dies waren insgesamt 67 Phraseme. 29 Phraseme wurden der Kategorie 4 (Idiome) und 23 der Kategorie 2 (Funktionsverbgefüge) zugeordnet.

Man sieht außerdem schnell, dass einige Kategorien bei der genaueren Darlegung der Ergebnisse vernachlässigt werden können, weil nur wenige oder keine Phraseme diesen Kategorien zugeordnet werden konnten (dazu zählen die Kategorien 5–9). Auch wenn in Kategorie 3 acht Phraseme zuzuordnen waren, wird diese Kategorie in der weiteren Auswertung aus inhaltlichen Gründen nicht beachtet. Phraseotermini und Phraseonyme haben einen speziellen thematischen Bezug, müssten daher im Unterricht zum Gegenstand gemacht werden und können in einer allgemeinen Betrachtung für diesen Zweck vernachlässigt werden. Im Folgenden wird daher nur auf die 119 Phraseme der Stichprobe eingegangen, die den Kategorien mit den meisten Vorkommen zugeordnet werden können, also Kollokationen, Funktionsverbgefüge und Idiome (siehe Tab. 3).

Kategorie	Info-Korpus <sup>E</sup>	JL-Korpus <sup>E</sup>	Lexikonkorpus <sup>E</sup>	SZ-Korpus <sup>J</sup>	Kreativkorpus <sup>J</sup>	Insg.
<b>1 Kollokationen</b>	15	9	16	16	11	67
<b>2 FVG</b>	11	2	6	4	0	23
<b>4 Idiome</b>	10	7	2	4	6	29
<b>Insg.</b>	36	18	24	24	17	119

Tabelle 3. Anzahl an Phrasemen in den Kategorien 1, 2 und 4 pro Subkorpus<sup>11</sup>

Im Informationskorpus kommen die meisten Phraseme vor, wobei die Kategorien ausgeglichen sind, die Kollokationen sich hier aber besonders zu häufen scheinen. Die anderen Subkorpora sind nicht auf diese Weise ausgeglichen, hier ist der Überhang zu Kollokationen auffällig. Die Ausnahmen stellen hier das Kreativkorpus und das Jugendliteraturkorpus dar. Mit elf Kollokationen, aber sechs Idiomen im Kreativkorpus und neun Kollokationen im JL-Korpus zu sieben Idiomen sind hier relativ ausgewogene Verhältnisse vorzufinden.

Man kann beobachten, dass die Verteilung der Funktionsverbgefüge in den verschiedenen Subkorpora variiert. In den zwei Subkorpora mit einem ausgewogenen Kollokation-Idiom-Verhältnis konnten entweder keine (Kreativkorpus) oder nur wenige (JL-Korpus) Funktionsverbgefüge identifiziert werden. Bei der Stichprobe des Kreativkorpus handelt es sich um einen Blogbeitrag und einen Reisebericht, beide wurden von Jugendlichen geschrieben. Die Stichprobe des JL-Korpus umfasst kurze Ausschnitte zweier Jugendromane, die im Schulunterricht behandelt werden. In der Stichprobe des Informationskorpus, bei dem die Stichprobe einen Ausschnitt eines Magazins für Berufsanfänger und Ausbildungsinteressierte umfasst, und des Lexikonkorpus konnten hingegen besonders viele Funktionsverbgefüge, in Zahlen elf und sechs, identifiziert werden.

Des Weiteren scheint die relative Häufigkeit der Phrasemkategorien eine nähere Betrachtung zu rechtfertigen. Wenn man sich das relative Vorkommen der drei Kategorien, aufsummiert nach den Subkorpora mit Texten von und für Jugendliche(n), (siehe hierfür Abb. 3) anschaut, erkennt man, dass Idiome jeweils mit einer relativen Häufigkeit von ungefähr einem Fünftel vorkommen.

<sup>11</sup> Ein hochgestelltes E zeigt an, dass es sich um ein Subkorpus mit Texten von Erwachsenen für Jugendliche handelt, das hochgestellte J weist Subkorpora aus, in denen Texten versammelt sind, die Jugendliche geschrieben haben.

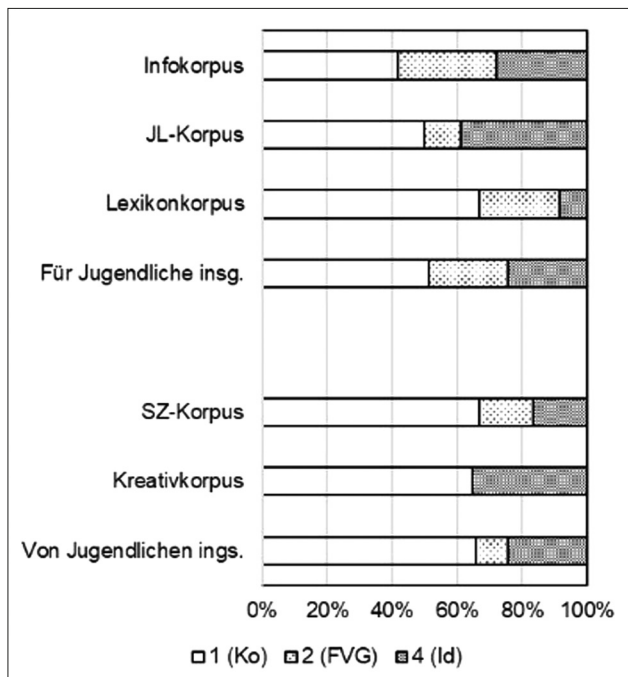


Abbildung 3. Anteil der Phraseme der Kategorien 1, 2 und 4 pro Subkorpus

### 3.4.2 Vergleich mit dem phraseologischen Optimum von Hallsteinsdóttir, Šajánková & Quasthoff (2006)

Nun zur Frage, ob die identifizierten Phraseme der Kategorie Idiome auch im phraseologischen Optimum von Hallsteinsdóttir, Šajánková & Quasthoff (2006) vorkommen. In der Untersuchung wurden 29 Idiome identifiziert. Davon kommen 17 nicht, neun genau in dieser Nennform und drei nur teilweise wie im phraseologischen Optimum vor.

Teilweise bedeutet, dass diejenigen Idiome den Phrasemen des Optimums nur teilweise entsprechen. Zum Beispiel kam in der Stichprobe aus dem Informationskorpus (2) *Steine liegen im Weg* vor, das der Redewendung *jemandem Steine aus dem Weg räumen/jemandem Steine in den Weg legen* ähnlich ist, die im phraseologischen Optimum der Gruppe CB und damit ebenfalls dem Kernbereich angehört, aber nicht voll übereinstimmt.

Es wurden mit wenigen Ausnahmen nur Phraseme gefunden, die sich im Kernbereich des Optimums befinden, also von Hallsteinsdóttir, Šajánková & Quasthoff (2006) als relevant eingestuft wurden und daher aus deren Sicht zu einem Grundwortschatz gehören sollten (vgl. Kap. 3.1.2). Zu den Idiomen, die nicht im phraseologischen Optimum aufgeführt wurden, aber in der Untersuchung in der Stichprobe identifiziert wurden, gehö-

ren zum Beispiel (3) *sich durchs Leben boxen*<sup>12</sup> und (4) *sich ein Bild über etwas machen*, obwohl es korrekterweise wohl *sich ein Bild von etwas machen*<sup>13</sup> heißen müsste.

#### 4 Zusammenfassung und Diskussion

Mithilfe der explorativen Datenanalyse einer Stichprobe des Korpus *Ref10*, bei der zunächst deduktiv ein Kategorienraster erstellt wurde, konnte ein Blick auf die Struktur des Korpus in Bezug auf die Verteilung der Phraseme auf verschiedene Phrasemkategorien gewonnen werden. Es konnte gezeigt werden, dass der Anteil an Kollokationen in allen Subkorpora besonders hoch ist und außerdem vor allem Idiome sowie Funktionsverbgefüge vorkommen.<sup>14</sup> In Bezug auf die Kategorie der Funktionsverbgefüge scheint sich eine Textsortenspezifität anzudeuten, da sie in Jugendromanen und von Jugendlichen selbst verfassten, kreativen Texten nicht oder nur selten vorkamen. In anderen Subkorpora waren sie aber doch im Verhältnis zu den anderen Kategorien gehäuft zu finden. Es konnte außerdem gezeigt werden, dass ein Großteil der in der Stichprobe des *Ref10*-Korpus identifizierten Idiome nicht oder nur teilweise, d. h. in einer nicht aufgeführten Variante, im phraseologischen Optimum von Hallsteinsdóttir, Šajánková & Quasthoff (2006) vorkommen. Diejenigen, die auch im Optimum aufgeführt sind, waren größtenteils dem von Hallsteinsdóttir, Šajánková & Quasthoff (2006) bezeichneten Kernbereich zuzuordnen. Der Vergleich mit dem phraseologischen Optimum von Hallsteinsdóttir, Šajánková & Quasthoff (2006) wirft außerdem methodische Überlegungen und Konflikte auf. Von einem didaktischen Gesichtspunkt müsste man sich überlegen, wie man mit Phrasemvarianz umgehen soll. Kann man bei abweichender Verwendung von einer Phrasemkenntnis sprechen, und ist dies kompetenter Phrasemgebrauch? Usuelle Varianzen wurden zumindest im phraseologischen Optimum aufgenommen, aber bei der Analyse wurden auch starke Abweichungen gefunden, die zumindest eine Kenntnis des Phrasems nahelegen, aber eine Unsicherheit im Umgang mit Phrasemen zeigen. Des Weiteren hat der Abgleich gezeigt, dass es eine Dunkelziffer an Phrasemen gibt, die mit einem computergestützten Verfahren nicht zu finden wären, wenn man die Nennformen wortgetreu im Korpus suchen würde. Dies müsste man bei einer computergestützten Analyse im Blick haben und mit einbeziehen oder gut begründet ausklammern, denn unter den Idiomen, die in dieser Untersuchung gefunden wurden, sind Phraseme, die nicht ganz irrelevant für die Lebenswirklichkeit von Jugendlichen sein dürften und in anderen Phrasemwörterbüchern zumindest vorkommen.

<sup>12</sup> Als Varianz von *sich durch Leben schlagen* zu finden bei ettinger-phraseologie.de (<https://www.ettinger-phraseologie.de/pages/buchinhalt/f/f.15.php#17>, 21.04.20) und in *Duden-Redewendungen* (2013, 461).

<sup>13</sup> *Sich ein Bild von jmd. /etwas machen* ist bei ettinger-phraseologie.de zu finden (<https://www.ettinger-phraseologie.de/pages/buchinhalt/e/e.13.php#7>, 21.04.20).

<sup>14</sup> Turgay (2019) stellt bei ihrer Studie, in der sie Phraseme in Schüleraufsätzen untersucht und die Häufigkeit der einzelnen Phrasemkategorien bestimmt, ebenfalls fest, dass Kollokationen (sowie feste Phrasen) dominieren.

Dieser explorative Ansatz kann als ein Einstieg in das Phraseminventar des Korpus *Ref10* dienen und Orientierung für eine computergestützte Korpusanalyse dieses Korpus bieten<sup>15</sup>. Die Untersuchung in diesem Beitrag konnte zeigen, dass eine Überprüfung des phraseologischen Optimums von Hallsteinsdóttir, Šajánková & Quasthoff (2006) am Korpus *Ref10* Potential hat, um ein idiomatisches Phraseminventar an einem Korpus für eine bestimmte Zielgruppe – Jugendliche im Alter von 11–16 Jahren – zu überprüfen. Allerdings wurden auch Probleme offengelegt, die eine solche Datenanalyse mit sich bringen würde. Zum einen lässt man die zahlreichen Kollokationen außer Acht, die in der Stichprobe identifiziert werden konnten. Bei einem modernen Phraseologie-Ansatz sollten diese nicht unhinterfragt ausgeklammert werden. Außerdem besteht auch das pragmatische Problem der Phrasemvarianz sowie der nachvollziehbaren, aber unter solchen Gesichtspunkten kritischen Unvollständigkeit des Inventars: Wenn man sich auf ein Phraseminventar blind verlässt, können unter Umständen für die Zielgruppe relevante Phraseme übersehen werden, wie in der Untersuchung an einer Stichprobe gezeigt werden konnte. Eine Lösung könnte ein rein datengeleiteter Ansatz bieten, allerdings sind hier zum derzeitigen Standpunkt auch Limitationen in Bezug auf das Identifizieren von Phrasemen gesetzt. Vermutlich wäre hier ein Ergänzen der beiden Verfahren sinnvoll. Die Entwicklung eines Phraseminventars für Schülerinnen und Schüler der Sekundarstufe I kann aber nur ein Ansatzpunkt sein, um die Phraseodidaktik für den Deutschunterricht weiterzuentwickeln. Es fehlen aktuelle Studien über vorhandene Phrasemkompetenzen für die Sekundarstufe I, die korpuslinguistische Ansätze sinnvoll ergänzen würden.

## Quellenverzeichnis

- Duden-Redewendungen. 2013. *Redewendungen. Wörterbuch der deutschen Idiomatik*. Werner Scholze-Stubenrecht, Wolfgang Worsch. Berlin: Bibliographisches Institut GmbH.
- ettinger-phraseologie. Ettinger, Stefan. Verfügbar unter <https://www.ettinger-phraseologie.de/>.
- Wortschatzwissen.de. *Ein Referenzwortschatz für die Sekundarstufe I*. Universität Duisburg-Essen. Verfügbar unter <https://www.wortschatzwissen.de/>.

## Literaturverzeichnis

- Belica, Cyril & Rainer Perkuhn. 2015. Feste Wortgruppen/Phraseologie I. Kollokationen und syntagmatische Muster. In *Handbuch Wort und Wortschatz*. Ulrike Haß & Petra Storjohann, Hg. Berlin, Boston: De Gruyter. 201–225.

<sup>15</sup> Diese computergestützte Korpusanalyse wird derzeit im Dissertationsprojekt der Autorin umgesetzt.

- Bubenhof, Noah & Stefaniya Ptashnyk. 2010. Korpora, Datenbanken und das Web. State of the Art computergestützter Forschung in der Phraseologie und Lexikographie. In *Korpora, Web und Datenbanken. Corpora, Web and Databases. Computergestützte Methoden in der modernen Phraseologie und Lexikographie. Computer-Based Methods in Modern Phraseology and Lexicography*. Stefaniya Ptashnyk, Erla Hallsteinsdóttir & Noah Bubenhof, Hg. Baltmannsweiler: Schneider Hohengehren. 7–19.
- Burger, Harald, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick. 2007. Phraseologie: Objektbereich, Terminologie und Forschungsschwerpunkte. In *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung*. Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick, Hg. Berlin: Walter de Gruyter. 1–10.
- Decker-Ernst, Yvonne & Ingelore Oomen-Welke. 2019. *1000 Wörter Basiswortschatz Deutsch für die Grundschule. Wortschatzvermittlung in Erst- und Zweitsprache*. Stuttgart: Fillibach bei Klett.
- Ettinger, Stefan. 2007. Phraseme im Fremdsprachenunterricht. In *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung*. Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick, Hg. Berlin: Walter de Gruyter. 893–908.
- Grünack, Iris, Heike Mühlens & Elvira Waldmann. 2011. *Unterrichtsentwicklung. Grundwortschatz für die Grundschule in Brandenburg. Rechtschreiben*. Ludwigsfelde-Struveshof: Landesinstitut für Schule und Medien Berlin-Brandenburg.
- Farø, Ken. 2015. Feste Wortgruppen/Phraseologie II. Phraseme. In *Handbuch Wort und Wortschatz*. Ulrike Haß & Petra Storjohann, Hg. Berlin, Boston: De Gruyter. 226–247.
- Hallsteinsdóttir, Erla, Monika Šajánková & Uwe Quasthoff. 2006. Phraseologisches Optimum für Deutsch als Fremdsprache. Ein Vorschlag auf Basis von Frequenz- und Geläufigkeitsuntersuchungen. *Linguistik Online* 27, 117–136.
- Hessky, Regina & Stefan Ettinger. 1997. *Deutsche Redewendungen. Ein Wörter- und Übungsbuch für Fortgeschrittene*. Tübingen: Narr.
- Kilian, Jörg & Jan Eckhoff, Hg. 2015. *Deutscher Wortschatz – beschreiben, lernen, lehren. Beiträge zur Wortschatzarbeit in Wissenschaft, Sprachunterricht, Gesellschaft*. Frankfurt am Main u. a.: Lang.
- Kuhs, Katharina & Stephan Merten, Hg. 2015. *Arbeiten am Wortschatz. Sprache und Sprachgebrauch untersuchen*. Trier: WVT Wissenschaftlicher Verlag Trier.
- Kuhs, Katharina & Stephan Merten, Hg. 2019. *Arbeiten am Wortschatz. Schreiben*. Trier: WVT Wissenschaftlicher Verlag Trier.
- Merten, Stephan & Katharina Kuhs, Hg. 2012. *Perspektiven empirischer Sprachdidaktik*. Trier: WVT Wissenschaftlicher Verlag Trier.
- Mückel, Wenke. 2014. Primärsprachliche Phraseodidaktik – Skizze eines Aufgabenfeldes. *Neuphilologische Mitteilungen* 115 (4), 391–399.

- Mückel, Wenke Hg. i. E. *Didaktische Perspektiven der Phraseologie. Deutschsprachige Phraseodidaktik international*. De Gruyter.
- Pohl, Inge & Winfried Ulrich, Hg. 2011. *Wortschatzarbeit*. Baltmannsweiler: Schneider-Verl. Hohengehren.
- Schnörch, Ulrich. 2002. *Der zentrale Wortschatz des Deutschen. Strategien zu seiner Ermittlung, Analyse und lexikografischen Aufarbeitung*. Tübingen: Narr.
- Steinhoff, Torsten. 2009. Wortschatz – eine Schaltstelle für den schulischen Spracherwerb? *Siegener Papiere zur Aneignung sprachlicher Strukturformen* 17.
- Tschirner, Erwin. 2008. Das professionelle Wortschatzminimum im Deutschen als Fremdsprache. *Deutsch als Fremdsprache* 45 (4), 195–208.
- Turgay, Katharina. 2019. „Ein Klima zwischen Leser und Verfasser“. Ein Überblick zur Verwendung von Phrasemen in Schüleraufsätzen. In *Arbeiten am Wortschatz – Schreiben*. Katharina Kuhs & Stephan Merten, Hg. Trier: WVT Wissenschaftlicher Verlag Trier. 79–94.
- Willenberg, Heiner. 2008. Wortschatz Deutsch. In *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*. Eckhard Klieme, Hg. Weinheim: Beltz. 72–80.