

Sprachlich Verfestigtes analysieren. Theoretische und methodische Überlegungen

Sarah Brommer (Zürich)

Abstract

This article combines theoretical and methodological considerations. It addresses the question how formulaic language manifests itself from a pragmatic point of view and how it can be analyzed with corpuslinguistic methods. First, it is shown that formulaic language – on the performance level – appears and is visible in the form of language patterns. Furthermore, an overview of the conceptual diversity prevailing in this context is provided. Subsequently, the question is how to make formulaic language empirically tangible, and the strengths and weaknesses of various corpuslinguistic approaches are discussed. In the third section, some methodological consequences, which result from a (largely) inductive corpuslinguistic analysis, are highlighted. Firstly, it is argued that even an inductive approach in practice necessarily includes deductive operations involving, on one hand, the corpus structure, and on the other, the analysis itself. Secondly, it is described how the study design affects the kind of formulaic language that can be identified. Thirdly, the possibilities and limitations of establishing formulaic language at the level of word forms and at the morphosyntactic level are discussed.

1 Einleitung

Sprachliche Verfestigungen (als Prozess) und sprachlich Verfestigtes (als Produkt) sind in den letzten Jahren verstärkt in das Blickfeld wissenschaftlicher Auseinandersetzung geraten. Die Aufmerksamkeit ist nicht zuletzt auf neue korpuslinguistische Methoden zurückzuführen, die es erlauben, sprachlich Verfestigtes quantitativ zu analysieren, und sie zeigt sich in einer beträchtlichen Zahl empirischer Arbeiten zu dieser Thematik (vgl. u. a. Bubenhofer 2009; Filatkina 2009; Steyer 2013). Der vorliegende Beitrag vereint theoretische und methodische Überlegungen zur Frage, wie sich sprachlich Verfestigtes aus pragmatischer Sicht äußert und wie es sich korpuslinguistisch analysieren lässt.

Zunächst geht es in Kap. 2 darum, wie sich sprachlich Verfestigtes auf Performanz-Ebene manifestiert, nämlich in Form von musterhaftem Sprachgebrauch. Hierbei wird auch ein Überblick über die begriffliche Diversität gegeben, die in diesem Zusammenhang vorherrscht. Der Schwerpunkt der Ausführungen liegt dabei auf denjenigen Bezeichnungen, die im Rahmen korpuslinguistischer Arbeiten Verwendung finden, und im Folgenden referiert der Begriff *sprachlich Verfestigtes* auf Sprachgebrauchsmuster an der sprachlichen Oberfläche. Kap. 3 ist der Frage gewidmet, wie sich sprachlich Verfestigtes empirisch greifbar machen lässt, und es werden die Vor- und Nachteile verschiedener korpuslinguistischer Herangehensweisen diskutiert. Mit einem deduktiven Vorgehen lässt sich sprachlich Verfestigtes in

Texten gezielt in den Blick nehmen. Das auf Vorannahmen beruhende Verfahren ist jedoch zwangsläufig auf die Analyse desjenigen beschränkt, was zu erwarten bzw. zu vermuten ist. Demgegenüber ermöglicht es ein induktives Vorgehen, auch sprachlich Verfestigtes zu entdecken, das aufgrund seiner Alltäglichkeit unauffällig ist und möglicherweise verborgen bliebe. Kap. 4 beleuchtet einige methodische Konsequenzen, die sich aus einer (weitgehend) induktiven korpuslinguistischen Analyse ergeben. Erstens wird dargelegt, dass auch eine induktive Herangehensweise in der Praxis notwendigerweise deduktive Arbeitsschritte beinhaltet, die einerseits die Korpuszusammenstellung, andererseits die Analyse selbst betreffen. Zweitens wird erörtert, wie sich das Untersuchungsdesign darauf auswirkt, was sich an sprachlich Verfestigtem ermitteln lässt. Drittens werden die Möglichkeiten und Grenzen diskutiert, sprachlich Verfestigtes auf Ebene der Wortformen und auf morphosyntaktischer Ebene zu ermitteln. Ein kurzes Fazit rundet den Beitrag ab.

2 Perspektiven auf sprachlich Verfestigtes

2.1 Sprachlich Verfestigtes aus pragmatischer Sicht

Sprachlich Verfestigtes wird sichtbar im Sprachgebrauch, in konkreten Kommunikaten. Mit dem Ausdruck *Kommunikat* soll deutlich gemacht werden, dass sprachlich Verfestigtes in jeglicher Form von Kommunikation vorkommt und sich beispielsweise nicht auf schriftliche Texte beschränkt. Sprachlich Verfestigtes lässt sich beschreiben als der für einen bestimmten Sprachausschnitt typische Sprachgebrauch. Dieser lässt sich sowohl diachron als auch synchron betrachten, aber auch hinsichtlich seiner diatopischen, diastratischen oder diaphasischen Variation. So kann sprachlich Verfestigtes typisch sein für eine bestimmte Zeit; auf diese Weise ließe sich historische Variation analysieren. Es kann auch – mit Blick auf die synchrone Variation – typisch sein für regionalen Sprachgebrauch, für den Sprachgebrauch sozialer Gruppen oder für situativ bedingten Sprachgebrauch. Und sprachlich Verfestigtes kann typisch sein für eine bestimmte kommunikative Form, beispielsweise eine bestimmte Textsorte. Solch textsortentypischer Sprachgebrauch wird im Folgenden behandelt.

Hierfür ist es zunächst notwendig, auf die Begriffe *Textsorte* und *Textmuster* einzugehen, bei deren Definitionen es mitunter zu Überschneidungen kommt (vgl. z. B. Brinker/Cölfen/Pappert 2014: 133, 139; Sandig 2000: 102). Sowohl *Textsorte* als auch *Textmuster* erfassen den „Sachverhalt, dass wir aus unserer Alltagserfahrung heraus Wissen über Textkonventionen haben und Merkmale kennen, die Gruppen von Texten eigen sind“ (Fix 2008: 10). Während mit dem Begriff *Textsorte* auf konkrete Texte mit gemeinsamen Eigenschaften referiert wird, bezeichnet *Textmuster* die dahinterstehende mentale Größe, die als Richtschnur wirkt, wie Texte einer bestimmten Textsorte prototypisch beschaffen sind. Einzelne Textexemplare lassen sich aufgrund ihrer gemeinsamen Merkmale zu Textsorten zusammenfassen, und sie repräsentieren das ihnen zugrunde liegende Textmuster oder anders gesagt: Die Textexemplare einer Textsorte folgen einem gemeinsamen Textmuster (vgl. auch Heine mann 2000: 517). Das Textmuster ist dabei nicht als strikte Vorgabe zu verstehen, sondern als (gesellschaftlich akzeptierte) Orientierung mit prototypischen Elementen und Freiräumen, gewissermaßen als eine Idealvorstellung (vgl. Fix 2000; Sandig 2000).¹

¹ Der in der Textlinguistik ebenfalls gebräuchliche Begriff *Textklasse*, der teilweise als Oberbegriff zu *Textsorte*, teilweise auch synonym verwendet wird, bleibt unberücksichtigt, ebenso der Begriff *Texttyp*, der ebenfalls als

Zwischen sprachlich Verfestigtem, Textsorte und Textmuster besteht ein wechselseitiger Zusammenhang. Das Sprachgebrauchswissen erlaubt den Teilnehmern einer Kommunikationsgemeinschaft, sprachlich Verfestigtes, also konventionalisierte sprachliche Mittel, intuitiv kontextangemessen anzuwenden. Textsorten wiederum sind kontextgebunden und vorrangig kommunikativ-funktional bestimmt (vgl. u. a. Brinker/Cölfen/Pappert 2014: 138–147). Dies hat zur Folge, dass der Sprachgebrauch in Textsorten nicht beliebig ist, sondern dass – je nach Textsorte – bestimmte sprachliche Mittel Eingang in die Texte finden und dass jeder einzelne Text als Vertreter einer bestimmten Textsorte unterschiedliches sprachlich Verfestigtes aufweist. Der alltägliche Umgang mit Textsorten und ihrem typischen, verfestigten Sprachgebrauch führt dazu, dass sich Textmuster herausbilden, die wiederum als Textmusterwissen in das Sprachgebrauchswissen eingehen und auf die weitere Produktion und Rezeption von Texten wirken.

Was die Textproduktion und -rezeption betrifft, so besitzt der verfestigte Sprachgebrauch ein „Kontextualisierungspotential“ (Feilke 2003: 213) in zweifacher Hinsicht: Aus Produzentensicht macht sprachlich Verfestigtes den zum Text zugehörigen Kontext kenntlich. Der Produzent orientiert sich in seinem sprachlichen Handeln teils bewusst, teils unbewusst (s.o.) an seinem Musterwissen: seinem Wissen darüber, wie der Sprachgebrauch in einer jeweiligen Kommunikationssituation typischerweise beschaffen ist. Indem der Produzent auf sprachlich Verfestigtes zurückgreift, signalisiert er die Zugehörigkeit des Texts zu einem bestimmten Kontext oder einem spezifischen Sprachausschnitt. Die Kontextualisierung findet jedoch auch in umgekehrter Richtung statt, aus rezeptiver Perspektive. Rezipienten eines Texts können mithilfe ihres Textmusterwissens aufgrund des sprachlich Verfestigten die Zugehörigkeit des Texts erkennen und ihn kontextualisieren.² Weiterführend lassen sich anhand des verfestigten Sprachgebrauchs Rückschlüsse auf zugrundeliegende Handlungsmuster ziehen. Sprachlich Verfestigtes ist letztlich „das sprachlich fassbare Produkt von sozialem Handeln“ (Bubenhof 2009: 4). Solch eine Perspektive, die von der sprachlichen Oberfläche auf zugrundeliegende Kontexte und soziale Handlungsmuster schließt, steht konträr zur Perspektive, aufgrund eines zugrundeliegenden Kontextes und damit verbundenen sozialen Handlungen auf erwartbare sprachliche Muster zu schließen (so bspw. Steinhoff 2007). Ebendiese Überlegung bildet die Grundlage für deduktive Analysen des Sprachgebrauchs, bei denen für feststehende Thesen passende Belege gesucht werden und sprachlich Verfestigtes verifikationsorientiert untersucht wird (vgl. Kap. 3.1).

Neben der Kontextualisierung dient sprachlich Verfestigtes dazu, den Grad der Typizität eines Texts zu erkennen. Je nachdem, wie viel sprachlich Verfestigtes sich in einem Text findet, wird der Text als mehr oder weniger typisch wahrgenommen. Die Rezeption erfolgt immer vor dem Hintergrund des vorhandenen Textmusterwissens und der Kenntnis darüber, dass in einer bestimmten Kommunikationssituation etwas typischerweise so ausgedrückt wird, ob-

Oberbegriff zu *Textsorte*, aber auch synonym zu *Textmuster* gebraucht wird. Denn mit Blick auf das sprachlich Verfestigte ist keine weitere Ausdifferenzierung notwendig. Relevant ist nur, dass sich einerseits konkrete Textexemplare aufgrund ihrer Merkmale zusammenfassen lassen (= *Textsorte*) und dass es andererseits eine dahinterstehende mentale Größe gibt, die als Referenz dient (= *Textmuster*).

² Feilke (1996: 268) spricht von einer „Kenntnis des Kontextualisierungspotentials“; bei Feilke (2003: 217–219) finden sich diese Überlegungen bezogen auf die Kontextualisierung von Texten und Textsorten; s. a. Bubenhof/Scharloth (2010: 87).

wohl auch alternative Ausdrucksmöglichkeiten zur Verfügung stehen würden (vgl. Bubenhofer 2009: 5; Köller 2009: 1222).

Die Typizität eines Texts zeigt sich in verschiedener Hinsicht, beispielsweise in der Wahl des Mediums, der Art der Textstrukturierung, der Adressatenorientierung und nicht zuletzt auf der Textoberfläche, im Gebrauch verfestigter Wörter und Wortverbindungen. Dabei gibt es sprachlich Verfestigtes, das thematisch bedingt ist (hierzu zählt z. B. der Fachwortschatz), und es gibt solches, das nicht inhaltlich gefüllt ist und deshalb themenunabhängiger gebraucht werden kann (z. B. grammatische Konstruktionen). All diese verschiedenen Arten von sprachlich Verfestigtem sind aus pragmatischer, sprachgebrauchsanalytischer Sicht relevant, wenn es darum geht, sprachlich Verfestigtes als Teil der Performanz zu beschreiben.

2.2 Sprachlich Verfestigtes aus korpuslinguistischer Sicht

Auch die korpuslinguistische Perspektive ist eine pragmatische, sprachgebrauchsanalytische, da sie sprachlich Verfestigtes an der sprachlichen Oberfläche verortet und sichtbar macht. Zwar gehen die Auffassungen darüber, was unter Korpusbasiertheit als empirischem Prinzip genau zu verstehen ist, auseinander (vgl. Belica/Steyer 2008: 10). Einigkeit besteht aber darin, dass sich verfestigter Sprachgebrauch in der für einen bestimmten Sprachausschnitt typischen Verwendung von einzelnen Wörtern und Wortverbindungen zeigt und dass sich diese Ausdruckstypik „als statistisch messbare Kookkurrenz operationalisier[en]“ (Feilke 2012: 24) lässt (vgl. Scharloth/Bubenhofer 2011: 207). Auf diese Weise soll das Konventionelle, Regelhafte im Sprachgebrauch offengelegt werden. In diesem Zusammenhang ist die Rede von Sprachgebrauchsmustern (Bubenhofer 2009), Prozeduren- bzw. Routineausdrücken (Feilke 2012, 2014) oder auch usuellen Wortverbindungen (Steyer 2013).³ Die Begriffe überschneiden sich – ungeachtet der verschiedenen Betrachtungsweisen (siehe unten) – weitgehend: Alle genannten Formen des sprachlich Verfestigten

- sind an einen spezifischen Verwendungskontext gebunden und für diesen Kontext typisch, gleichzeitig dienen sie aufgrund ihrer Typizität der Kontextualisierung (vgl. Kap. 2.1),
- weisen aufgrund ihrer Kontextgebundenheit eine idiomatische Prägung auf (idiomatisch i. S. Feilkes und nicht im traditionellen phraseologischen Verständnis, vgl. Feilke 1993: 15–17),
- sind konventionalisiert,
- zeichnen sich durch wiederkehrenden Gebrauch aus (sind also rekurrent).

Die einzelnen Kriterien – Kontextgebundenheit, Typizität, Idiomatizität, Konventionalität und Rekurrenz – stehen in einem wechselseitigen Abhängigkeitsverhältnis: Die kontextgebundene Verwendung eines sprachlichen Ausdrucks ist Voraussetzung dafür, dass dieser Ausdruck als idiomatisch (= gekoppelt an ein spezifisches Gebrauchsschema) bezeichnet werden kann. Der Prozess der idiomatischen Prägung lässt sich als Konventionalisierungsprozess beschreiben. Damit es hierzu kommt, ist es nötig, dass ein sprachlicher Ausdruck in einem spezifischen Verwendungskontext wiederholt gebraucht wird und aufgrund seiner Rekurrenz für diesen typisch ist. Zum Kriterium der Rekurrenz ist einzuwenden, dass nicht die absolute Verwen-

³ Der Fokus liegt hier auf jüngeren korpuslinguistischen Arbeiten, ungeachtet dessen, dass sprachlich Verfestigtes auch schon früher Gegenstand wissenschaftlicher Betrachtung war (vgl. bspw. Coulmas 1981; Gülich 1997; Stein 1995), insbesondere auch in der Phraseologie (für einen Überblick vgl. Burger et al. 2007).

dungshäufigkeit eines sprachlichen Ausdrucks entscheidend ist, sondern seine „Einschlägigkeit [...] für bestimmte Gebrauchsschemata und Genrekontexte“ (Feilke 2012: 16). Diese lässt sich anhand eines Korpusvergleichs offenlegen bzw. ermitteln (siehe unten). Das Ziel besteht darin, sprachlich Verfestigtes weitestgehend automatisiert zu entdecken, indem der Computer möglichst viele vorstrukturierende Arbeitsschritte abnimmt. Die Korpuslinguistik bedient sich hierfür computerlinguistischer Methoden und Werkzeuge. Die Ergebnisse der automatischen Analyse lassen sich dann durch den Menschen weiterführend interpretieren (vgl. Belica/Steyer 2008: 12; Oakes 2014).

Worin unterscheiden sich nun die Bezeichnungen *Prozedurenausdruck*, *usuelle Wortverbindung* und *Muster*? In erster Linie weisen sie auf unterschiedliche Betrachtungsperspektiven hin: So steht bei dem Begriff *Prozedurenausdruck* (bzw. *Routineausdruck*)⁴ der Aspekt des Handelns im Vordergrund (vgl. Feilke 2012: 7), während der Begriff *usuelle Wortverbindung* vor allem das rekurrente Vorkommen fokussiert, den häufigen, gewohnheitsmäßigen Gebrauch (vgl. Steyer 2013: 14). Demgegenüber betont *Muster* bzw. *Musterhaftigkeit* den Vorbildcharakter (vgl. Bubenhofer 2009: 24). Ein Muster gibt Prototypisches vor (vgl. Fix 2009: 1302f.) und wirkt als Orientierung. Genau darin liegt m. E. die Stärke des Begriffs *Sprachgebrauchsmuster*, den ich im Folgenden übernehme und als Synonym zu *sprachlich Verfestigtem* verstehe.

Abschließend noch eine Anmerkung: Alle genannten Bezeichnungen beziehen sich nur auf mehrgliedrige Einheiten.⁵ Deshalb ist explizit darauf hinzuweisen, dass ich den Begriff *Sprachgebrauchsmuster* sowohl auf eine Wortform als auch auf eine Verbindung von Wortformen beziehe. Denn auch eine einzelne, singuläre Wortform kann die oben genannten Kriterien erfüllen: Die Verwendung einer Wortform kann an einen bestimmten Kontext gebunden und für diesen typisch und im oben beschriebenen Sinne idiomatisch sein, sie kann konventionalisiert erfolgen und die Wortform selbst rekurrent auftreten. – Im Folgenden geht es nun um die Frage, mit welchen methodischen Herangehensweisen sich sprachlich Verfestigtes empirisch greifbar machen lässt.

3 Sprachlich Verfestigtes an der Textoberfläche empirisch greifbar machen

Modernes korpuslinguistisches Arbeiten zeichnet sich dadurch aus, große Mengen verfügbarer Daten mittels computerlinguistischer Methoden zu analysieren. Das Vorgehen setzt umfangreiche Korpora voraus, einerseits aufgrund der ungleichen Häufigkeitsverteilung von Wörtern (vgl. Sinclair 1991: 18, 143) und andererseits aufgrund des Erkenntnisinteresses korpuslinguistischer Analysen: Zumeist stehen keine Vorkommenshäufigkeiten einzelner Wortformen im Fokus, sondern das Ziel besteht im Aufdecken mehrgliedriger musterhafter Struk-

⁴ Der Begriff *Prozedurenausdruck* (Feilke 2014) ersetzt den bedeutungsgleichen Begriff *Routineausdruck*, der in früheren Arbeiten Feilkes Verwendung findet (noch Feilke 2012). Während dieser eher das Gewohnheitsmäßige betont, rückt mit *Prozedurenausdruck* der Prozess des Handelns in den Vordergrund. Der Begriff ist Teil eines umfassenden Konzeptes, das den Zusammenhang zwischen Schreibprozess, Handlungsmustern und Produkt (Text) darstellt.

⁵ Es besteht in der Forschung ganz grundsätzlich eine Konzentration auf mehrgliedrige Ausdrücke (neben den genannten Arbeiten s. a. Sanderson 2008; Tognini-Bonelli 2001). Eine Ausnahme bildet Biber (2006), dessen Studie zum wissenschaftlichen Sprachgebrauch auch Wortschatzuntersuchungen beinhaltet, wobei er sich in späteren Arbeiten ebenfalls auf Mehrworteinheiten konzentriert (vgl. Biber 2009).

turen, denen aufgrund ihres viel selteneren Vorkommens nur auf Grundlage großer Textmengen nachgegangen werden kann. Bei der Arbeit mit Korpora lassen sich zwei Herangehensweisen (induktiv und deduktiv) unterscheiden, die in ihrer jeweiligen zugrundeliegenden Methodologie grundverschieden sind.

3.1 Deduktive Herangehensweise

Bei einem deduktiven Vorgehen erfolgt die Analyse des Sprachgebrauchs interessensgeleitet und zielgerichtet untersucht anhand bereits vorher formulierter Fragestellungen und Hypothesen. Dass dabei Unerwartetes oder Überraschendes offengelegt wird, ist weder zu erwarten noch ist es vordringliches Ziel dieser Vorgehensweise. Vielmehr besteht das Anliegen darin, die formulierten Fragestellungen und Hypothesen empirisch breit abgestützt zu beantworten bzw. zu überprüfen. Korpusdaten stellen dann in erster Linie das „empirische Gewissen oder auch Ruhekränchen“ (Steyer 2013: 14) dar.

Beispielhaft für dieses Vorgehen ist die Analyse von Texten unter Berücksichtigung ihrer Funktion: Aufbauend auf der Frage, welche kommunikative Funktion Texte erfüllen, werden die erwartbaren sprachlichen Mittel untersucht. Dieses Vorgehen sucht nach der „Korrelation [...] zwischen Außersprachlichem (Tätigkeitsbereiche, Kommunikationssituationen) und für dieses Außersprachliche typische[n] sprachliche[n] Gebrauchsweisen“ (Fix 2009: 1310f.; vgl. auch Hoffmann 2009: 1326f.). Indem die funktionalen Anforderungen an Texte als Ausgangspunkt für Untersuchungen genommen werden, konzentriert sich die Analyse auf sprachlich Verfestigtes, das sich funktional erklären lässt. Dieses Vorgehen hat jedoch erstens zur Folge, dass sprachlich Verfestigtes nur dort aufgedeckt wird, wo ein Text typische Funktionen erfüllt. Zweitens bleibt auf diese Weise das von den Rezipienten (und Forschenden) nicht explizit erwartbare sprachlich Verfestigte bei der Analyse unentdeckt.

Der Fokus einer deduktiven Herangehensweise liegt also auf dem Bekannten bzw. Erwartbaren, und die Untersuchung beschränkt sich zwangsläufig auf die Elemente, die sich theoretisch, beispielsweise funktional, begründen lassen. Es ist daher eine naheliegende Frage, ob sich im Sprachgebrauch nicht auch jenseits des Bekannten und Erwartbaren sprachlich Verfestigtes, Musterhaftes findet und wie sich dieses aufdecken lässt.⁶ Genau hier setzt das induktive Vorgehen an, das nicht von vorneherein interessengeleitet auf bestimmte Muster beschränkt ist.

3.2 Induktive Herangehensweise

Geht man wie oben beschrieben davon aus, dass beim Schreiben sprachlich Verfestigtes vielfach unbewusst Eingang in die Texte findet (vgl. Kap. 2.1), dann ist sprachlich Verfestigtes erwartbar, das sich deduktiv nicht erschließen lässt, da es sich auch dem Blick von Forschenden entzieht. Dieses sprachlich Verfestigte dient gleichsam als „Gerüst des Sprachgebrauchs“ (Steyer 2013: 13), auch wenn es in seiner „unauffälligen Häufigkeit verborgen“ (ebd.) bleibt, gerade weil es alltäglich ist (ebenso Linke 2011: 39; Sinclair 1991: 4). Soll der Sprachgebrauch hinsichtlich des Verfestigten, Musterhaften analysiert werden, gilt es, auch das Unauffällige offenzulegen und sich dem Sprachgebrauch möglichst unvoreingenommen zu nähern.

⁶ Vgl. die Untersuchungen von Wray und Perkins, wonach ca. 70 Prozent der Sprache als musterhaft angesehen werden können (zitiert nach Wallner 2014: 59).

Kupietz und Keibel ist daher zuzustimmen, wenn sie es zum „programmatischen Leitprinzip“ erheben, „dem Phänomen *Sprache* empirisch gerecht zu werden“ (Kupietz/Keibel 2009: 49, Kursiv. i. O.). Dies ist das Anliegen einer induktiven korpuslinguistischen Herangehensweise.⁷

Korpuslinguistisches Arbeiten in diesem engeren Sinne (der Induktion verpflichtet) ist mehr als ein bestimmtes methodisches Vorgehen: Es ist vielmehr eine spezifische Sichtweise auf Sprache, und der Sprachgebrauch selbst stellt den Erkenntnisgegenstand dar. Voraussetzung ist eine umfangreiche Datengrundlage, die es erlaubt, Sprache unvoreingenommen hinsichtlich ihrer Struktur und ihrer Verwendung zu untersuchen. Sprache wird als Datensammlung aufgefasst, und Strukturen, die statistisch auffällig sind, werden sichtbar gemacht. Diese Auffälligkeiten müssen nicht in absoluten Zahlen hochfrequent, sondern im untersuchten Sprachausschnitt nur überzufällig häufig auftreten. Auf diese Weise wird auch offengelegt, was weder den Sprachbenutzern noch den Forschenden bewusst auffällt, was aber trotzdem – da statistisch auffällig – signifikant ist (vgl. Bubenhofer 2009: 16).

Dieses induktive Vorgehen wird von Tognini-Bonelli als „corpus-driven“-Vorgehen explizit gemacht und streng vom oben beschriebenen deduktiven, „corpus-based“-Vorgehen abgegrenzt (vgl. Tognini-Bonelli 2001: 84). Beim induktiven Vorgehen dienen die Daten als Ausgangspunkt für die Analyse, und jegliche Theorie baut auf diesen Daten auf. Dadurch können auch „Evidenzen in den Fokus [geraten], die entweder quer zu den vorher existierenden Erwartungen stehen und die Grundlage für neue Hypothesen sind, oder im besten Fall sogar solche Evidenzen, die die Bildung neuer interpretativer linguistischer Analysekategorien nahelegen“ (Scharloth/Bubenhofer 2011: 197f.).

Für diese Art der Sprachgebrauchsanalyse ist es notwendig, sprachlich Verfestigtes bzw. Musterhaftigkeit auf die Textoberfläche zu beziehen und an statistische Signifikanz zu knüpfen. Denn nur so lässt sich der für den untersuchten Sprachausschnitt verfestigte, musterhafte Sprachgebrauch korpuslinguistisch offenlegen.⁸ Sprachlich Verfestigtes, musterhafte Wörter und Wortverbindungen, werden also auf Basis ihrer statistischen Signifikanz ermittelt. Welche Konsequenzen dieses Vorgehen hat und dass das induktive Vorgehen seine Grenzen hat, wird im folgenden Kapitel dargelegt.

⁷ Um das Profil der Korpuslinguistik zu schärfen, sprechen sich Korpuslinguisten dafür aus, nur von korpuslinguistischem Vorgehen i. e. S. zu sprechen, wenn eine induktive Herangehensweise gewahrt ist (vgl. z. B. Bubenhofer 2009; Steyer 2013; Tognini-Bonelli 2001).

⁸ Das methodische Vorgehen hat zur Folge, dass dieser Musterbegriff im engeren Sinne kleinräumig angelegt ist und nicht die Text-Ebene (bspw. keine Strukturierungsmuster oder Argumentationsmuster) erreicht.

4 Methodische Konsequenzen

4.1 Weitestgehende Induktion als Ziel

In der Praxis lässt sich die oben beschriebene, von Tognini-Bonelli theoretisch vorgenommene programmatische Gegenüberstellung von induktivem vs. deduktivem Vorgehen in dieser Radikalität nicht beibehalten. Denn keine Analyse kann vorbehaltlos durchgeführt werden, und auch in einem der Induktion verpflichteten Vorgehen sind notwendigerweise deduktive Arbeitsschritte enthalten. Diese betreffen einerseits das Korpus und damit die Grundlage einer jeden Analyse und andererseits die Analyse selbst.

Korpuslinguistik analysiert und beschreibt Sprache im Gebrauch, zugrunde liegen reale Sprachdaten. Ein Korpus ist also eine Sammlung konkreter sprachlicher Äußerungen. Ob es sich dabei um Alltagssprache, literarische Texte oder Gebrauchstexte, um medial geschriebene oder gesprochene Äußerungen oder beispielsweise nur um Äußerungen einer spezifischen Produzentengruppe (mögliche Faktoren sind etwa Alter, Geschlecht, Herkunft) handelt, hängt von der Fragestellung ab. Dass Sprache im Gebrauch untersucht werden soll, impliziert also nicht, dass jede Art von Gebrauch Eingang in das Korpus findet. Vielmehr geschieht schon die Entscheidung, welcher Sprachausschnitt induktiv korpuslinguistisch untersucht werden soll, auf der Basis von Vorüberlegungen. Mit der Zusammenstellung des Korpus wird bereits eine bestimmte Auswahl getroffen. Je nachdem, welche Art von Gebrauch im Erkenntnisinteresse liegt, ist das Korpus anders zusammenzustellen. Diese Vorauswahl, die sich notwendigerweise auf die möglichen Analyseergebnisse auswirkt, widerspricht nicht dem Anspruch einer induktiven Datenanalyse. Entscheidend ist, dass die Daten induktiv, frei von Vorannahmen analysiert werden. Die Daten selbst jedoch sind bewusst, in Abhängigkeit von der Fragestellung und nach spezifischen Kriterien ausgewählt.⁹

Das Offenlegen des sprachlich Verfestigten erfolgt dann streng induktiv: Weitgehend automatisiert (mit den damit verbundenen Vor- und Nachteilen, siehe unten) wird alles ermittelt, was für den Sprachgebrauch im Untersuchungskorpus signifikant ist im Vergleich zum Sprachgebrauch im Referenzkorpus. Im weiteren Verlauf der Analyse ergänzen jedoch deduktive Arbeitsschritte das Vorgehen. Denn das korpuslinguistische Potential erschöpft sich nicht allein in induktiven, quantitativen Analysen, die für sich genommen wenig Aussagekraft haben. Vielmehr müssen die induktiv gewonnenen Daten in weiteren Analyseschritten selektiert, klassifiziert und interpretiert werden. Eine Selektion ist notwendig, weil sich automatisiert nur ermitteln lässt, was überhaupt signifikant für das Untersuchungskorpus ist. Jedoch ist nicht alles, was der Computer objektiv als statistisch signifikant erkannt hat, interessant für die jeweilige, einer jeden Analyse zugrundeliegende Fragestellung. Die Schwierigkeit besteht darin, aus der großen Menge an ermitteltem sprachlich Verfestigtem dasjenige zu identifizieren, das in pragmatischer Hinsicht relevant ist (vgl. Scharloth/Bubenhofner 2011: 197). Diese Beurteilung der pragmatischen Relevanz wie auch die sich anschließende qualitative Analyse des sprachlich Verfestigten lassen sich nicht rein induktiv und automatisiert vornehmen. In der

⁹ Dabei ist zu bedenken, dass sich die Verfügbarkeit von Daten darauf auswirkt, was überhaupt korpuslinguistisch analysiert wird: Ob, in welchem Umfang und in welcher Qualität Daten verfügbar sind, ist je nach Sprachausschnitt sehr unterschiedlich und hat zur Folge, dass sich einzelne Sprachausschnitte eher (z. B. journalistische Texte), andere Sprachausschnitte weniger (z. B. medial mündliche Texte) für eine korpuslinguistische Analyse anbieten.

Praxis sind also die induktive und die deduktive Perspektive sinnvoll miteinander zu verknüpfen (vgl. Bubenhofer 2009; Steyer 2013). Entscheidend ist, dass jedes Vorgehen streng von den empirischen Korpusdaten geleitet ist und ein manuelles Eingreifen wie die Selektion der Daten oder ihre linguistische Klassifikation und Interpretation immer nachgelagert erfolgt.

4.2 „Nebenwirkungen“ des Untersuchungsdesigns

Wie bereits erwähnt wird sprachlich Verfestigtes, werden musterhafte Wörter und Wortverbindungen auf Basis ihrer statistischen Signifikanz ermittelt. Eine induktive korpuslinguistische Analyse verlangt daher neben einem Untersuchungskorpus immer auch ein Referenzkorpus. Denn statistische Signifikanz lässt sich nur im Vergleich beurteilen: Sprachlich Verfestigtes ist signifikant für einen Sprachausschnitt im Vergleich zu einem anderen Sprachausschnitt. Der notwendige Korpusvergleich wirkt sich folgendermaßen auf das Ergebnis der induktiven Analyse aus: Erstens wird mit der vergleichenden Korpusanalyse als statistisch signifikant und typisch für den Sprachgebrauch im Untersuchungskorpus nur erkannt, was nicht gleichzeitig auch in den Texten des Referenzkorpus auftritt. Sprachlich Verfestigtes, das sowohl typisch für das Untersuchungskorpus als auch typisch für das Referenzkorpus und demzufolge in beiden Korpora vertreten ist, kann mit dem Untersuchungsdesign nicht erfasst werden.

Wenn etwas für das Untersuchungskorpus als signifikant ermittelt wird, bedeutet dies zweitens nichts anderes, als dass es in diesem Korpus häufiger als erwartet vorkommt – bzw. dass es im Referenzkorpus seltener auftritt, als zu erwarten wäre. Es kann also theoretisch der Fall sein, dass etwas als signifikant ermittelt wird, was im Referenzkorpus unterdurchschnittlich häufig auftritt, dessen Vorkommen im Untersuchungskorpus aber lediglich einer Normalverteilung entspricht. Dies allerdings wäre ein Zeichen für ein schlecht gewähltes Referenzkorpus, das aufgrund seiner Zusammensetzung nicht geeignet ist, den durchschnittlichen Sprachgebrauch zu repräsentieren.

Die genannten „Nebenwirkungen“ der korpusvergleichenden Analyse müssen bedacht werden, denn sie haben Einfluss darauf, was die Analyse an sprachlich Verfestigtem offenlegt. Wenn Aussagen zur Signifikanz und Typizität für den untersuchten Sprachausschnitt (für das Untersuchungskorpus) gemacht werden, ist die Zusammensetzung des Referenzkorpus daher immer zu berücksichtigen, ebenso bei der Interpretation der Ergebnisse. Im Folgenden werden nun methodische Konsequenzen thematisiert, die sich aus dem weitgehend automatisierten Vorgehen ergeben und die das Ermitteln von sprachlich Verfestigtem auf Ebene der Wortformen und auf morphosyntaktischer Ebene betreffen.

4.3 Sprachlich Verfestigtes auf Wortformen- und morphosyntaktischer Ebene ermitteln

Das automatisierte Ermitteln dessen, was im Untersuchungskorpus überzufällig häufig auftritt, kann auf Ebene der Wortformen, aber auch auf Lexem- oder – noch abstrakter – auf morphosyntaktischer Ebene vorgenommen werden, wenn die Korpusdaten entsprechend morphosyntaktisch annotiert sind. Im Folgenden werden erstens die Vor- und Nachteile annotierter Daten diskutiert und es wird dafür argumentiert, sprachlich Verfestigtes grundsätzlich auf Wortformen-Ebene zu ermitteln. Zweitens wird – als eine Konsequenz, die sich aus der Ermittlung auf Wortformen-Ebene ergibt – auf den Umgang mit Mustervarianten eingegangen.

Zunächst einmal bedeutet eine Annotation einen Informationsgewinn, da den Primärdaten zusätzliche, in diesem Fall morphosyntaktische Informationen hinzugefügt werden. Dadurch wird es möglich, ähnliche Wortgruppen (wie *nachgewiesen werden kann*, *nachgewiesen werden können*, *nachgewiesen werden konnte* usw.) zusammenzufassen, was sich auf die Vorkommenshäufigkeit und statistische Signifikanz auswirkt.¹⁰ Außerdem lassen sich auf Basis annotierter Daten musterhafte Wortgruppen ermitteln, die sich aus Wortformen, Lemmata und Wortartenklassen zusammensetzen (sog. „komplexe n-Gramme“, Bubenhofer 2015; „Collostructions“, Stefanowitsch/Gries 2003; oder „Wortverbindungsmuster“, Steyer 2013). Es handelt sich hierbei um sprachlich Verfestigtes auf einer bereits abstrakteren Ebene, bestehend aus festen lexikalischen Elementen und variablen Leerstellen.¹¹

Die erweiterten Analysemöglichkeiten sprechen dafür, Korpusdaten morphosyntaktisch zu annotieren. Doch es würde einen erheblichen Nachteil mit sich bringen, sprachlich Verfestigtes allein auf Grundlage der annotierten Daten und damit der Lemmata zu ermitteln. Denn Lemmata sind bereits Abstraktionen von den einzelnen Wortformen. Mit der Lemmatisierung gehen „Differenzierungsmerkmale, die sich in der morphosyntaktischen Ausprägung der Wörter zeigen“ (Bubenhofer 2009: 126), verloren; es kommt zu einem Informationsverlust (ähnlich argumentieren auch Steyer 2013: 342; Tognini-Bonelli 2001: 73f.). Untersuchungen haben gezeigt, dass einzelne Wortformen eines Lexems unterschiedlich gebraucht werden und dass sich Listen von musterhaften Mehrworteinheiten, die auf Basis der Lemmata ermittelt werden, stark von solchen unterscheiden, die auf Basis der Wortformen entstanden sind (vgl. Bubenhofer 2009: 128; vgl. Brommer z. Publ. eing.). Verfestigt ist der Sprachgebrauch demzufolge auf Ebene der Wortformen und nicht auf Lexem-Ebene; Sprachgebrauchsmuster sind (nicht nur, aber auch) Wortformen-Muster. Die Analyse des musterhaften Sprachgebrauchs in wissenschaftlichen Aufsätzen (vgl. Brommer z. Publ. eing.) hat beispielsweise offengelegt, dass das Partizip *ausgehend (von)* anders verwendet wird und aus pragmatischer Sicht zu unterscheiden ist von der Verwendung des Verbs *ausgehen (von)* als Prädikat (z. B. *lässt sich davon ausgehen*). Verfestigt sind in diesem Fall beide Wortformen, aber unterschiedlich. Um dem gerecht zu werden und differenzierte Gebrauchsweisen erfassen zu können, ist eine induktive korpuslinguistische Analyse sinnvollerweise auf Wortformen-Ebene vorzunehmen.¹² Annotierte Daten bieten wiederum die Möglichkeit, in nachgeordneten Analyseschritten das

¹⁰ Manche Wortgruppen sind so selten, dass sie überhaupt erst mit einem solchen Zusammenfassen als statistisch signifikant ermittelt werden.

¹¹ Das Muster *im Rahmen der* ließe sich dann abstrahieren zu der Collostruction *im Rahmen der N*, wobei *N* ein Substantiv repräsentiert, das in diesem Fall im Genitiv steht und entweder als Singular (Genus: fem.) oder als Plural (Genus: mask./fem./neutr.) realisiert und ggf. attributiv erweitert werden kann. Unter der Collostruction *im Rahmen der N* würden alle Muster, die die vorgegebene Konstruktion erfüllen, subsumiert, so z. B. *im Rahmen der Arbeit* und *im Rahmen der Untersuchungen*. Anhand der Beispiele wird deutlich, dass die Leerstellen bzw. die Leerstelle (in diesem Fall das *N*) nicht beliebig besetzt werden können. Neben den syntaktischen Beschränkungen unterliegt die Besetzung auch semantisch-pragmatischen Restriktionen, die durch den situativen Kontext bestimmt sind. – Ein weiterer Vorteil der morphosyntaktischen Annotation zeigt sich nicht zuletzt, wenn deduktiv etwas im Korpus untersucht werden soll. In diesem Fall lässt sich zielgerichtet nach Wortformen oder Lemmata suchen, ohne Rücksicht auf die morphosyntaktische Ausprägung im Einzelnen.

¹² Was bei der Analyse auf Grundlage von Lemmata zu beachten ist, gilt in noch größerem Maß für Analysen auf höheren Abstraktionsebenen, beispielsweise wenn beim Ermitteln von Collostructions Wortartenklassen berücksichtigt werden. Jede Art von Abstraktion birgt die Gefahr eines Informationsverlusts.

zu betrachten, zu vergleichen, zu überprüfen, was zuvor an Erkenntnissen aus den (nicht annotierten) Daten gewonnen wurde. Es ist also anzuraten, die Annotation als eine zusätzliche Information hinzuzufügen und nicht die ursprünglich nicht annotierten Daten zu überschreiben. Annotierte Daten können dann nach wie vor ohne Berücksichtigung der Annotation analysiert werden. Entscheidend ist, dass dies tatsächlich gemacht wird, dass die zusätzlichen Informationen zunächst also ausgeblendet werden und nicht die Analyse beeinflussen.

Das auf Wortformen-Ebene automatisiert ermittelte sprachlich Verfestigte, die Wortformen-Muster sind jedoch aus mehreren Gründen nur mit Einschränkung zur weiteren Analyse geeignet: Erstens hat das Vorgehen zur Folge, dass die Signifikanz eines Musters für alle zugehörigen Schreibvarianten einzeln ermittelt wird.¹³ Dies wirkt sich auf die Frequenz eines Musters bzw. der einzelnen Mustervarianten und in der Folge auf deren Signifikanz aus. Da der Computer nicht unterscheidet, ob die orthographischen Varianten eines Musters einer veränderten Rechtschreibregelung oder der Satzstellung geschuldet sind (z. B. *in bezug auf*, *in Bezug auf*, *In Bezug auf*), ist in Abhängigkeit vom jeweiligen Untersuchungsinteresse zu überlegen, (ortho-)graphische Varianten ggf. nachträglich zusammenzufassen. Wird eine diachrone Fragestellung verfolgt, kann es sinnvoll sein, Varianten, die auf die veränderte Rechtschreibung zurückzuführen sind, zu erhalten; andernfalls sind sie grundsätzlich zusammenzufassen. Bei der satzstellungsbedingten Groß-/Kleinschreibung sind hingegen Vor- und Nachteile des Zusammenfassens abzuwägen unter Berücksichtigung der Fragestellung: Wenn sprachlich Verfestigtes mit Blick auf den Satzbau analysiert werden soll (z. B. musterhafte Vorfeldbesetzungen), ist es notwendig, dass die einzelnen Varianten berücksichtigt werden. Denn nur dann lässt sich untersuchen, welche Muster aufgrund der Großschreibung an den Satzanfang gestellt werden. Geht es allerdings darum, welche Muster allgemein signifikant für den Sprachgebrauch im Untersuchungskorpus sind – ungeachtet ihrer syntaktischen Einbettung –, ist es zweckmäßig, die Varianten zusammenzufassen.

Neben der (ortho-)graphischen Variation führt das automatisierte Ermitteln von sprachlich Verfestigtem auf Wortformen-Ebene zu morphosyntaktischer Variation (z. B. *im Bereich dieser*, *im Bereich der*, *im Bereich des*, *im Bereich von*).¹⁴ Auch in diesem Fall kann es sinnvoll sein, die einzelnen Varianten zusammenzufassen, da bestimmte Unterschiede in der morphosyntaktischen Struktur eines Musters (wie bspw. das variable Genus des angefügten Genitiv-Attributs) für die Typizität des Musters unerheblich sind. Durch ein Nivellieren dieser morphosyntaktischen Ausprägungen lassen sich musterhafte Verwendungsweisen systematisieren und die Signifikanz der einzelnen Muster besser beurteilen.

Die angesprochenen Punkte machen deutlich, dass es sinnvoll ist, von den automatisiert ermittelten Wortformen-Mustern zu abstrahieren und zugrundeliegende Muster abzuleiten (z. B. *in/In bezug/Bezug auf*; *im Bereich* + *Genitiv-/präpositionales Attribut*). Auf dieselbe Art –

¹³ Diese (ortho-)graphische Variation ist zum einen der Reform der Rechtschreibregelung geschuldet. Sie betrifft die *ss/ß*-Schreibung ebenso wie Anpassungen im Bereich der Laut-Buchstaben-Zuordnung und Änderungen im Bereich der Getrennt-/Zusammenschreibung sowie der Groß-/Kleinschreibung. Zum anderen kommt es zu Variation in der Groß-/Kleinschreibung, die auf die Stellung des Musters im Satz zurückzuführen ist.

¹⁴ Der Kern dieser Muster besteht in der Präpositionalphrase *im Bereich*, die attributiv ergänzt wird durch eine Nominalphrase im Genitiv (erkennbar an den Artikeln *der*, *des* bzw. den entsprechenden Pronomen) oder durch eine Präpositionalphrase, eingeleitet mit der Präposition *von*. Die einzelnen Muster unterscheiden sich allein in der Realisierung ihrer attributiven Ergänzung, der Kern ist immer der gleiche.

datengeleitet und manuell – lassen sich von den Wortformen-Mustern morphosyntaktische Muster ableiten, deren Musterhaftigkeit in ihrer morphosyntaktischen Struktur liegt. Beispielsweise lassen die Wortformen-Muster *nachgewiesen werden kann, aufgezeigt werden können, veranschaulicht werden konnte* usw. auf die Musterhaftigkeit der Passivkonstruktion mit Modalverb *können* schließen. Die Signifikanz solcher morphosyntaktischen Muster lässt sich wiederum deduktiv anhand der annotierten Korpusdaten überprüfen. Es ist auf diese Weise möglich, von sprachlich Verfestigtem auf Wortformen-Ebene auf sprachlich Verfestigtes auf morphosyntaktischer Ebene zu schließen.

5 Fazit

Im vorliegenden Beitrag wurde sprachlich Verfestigtes als an der Textoberfläche beobachtbares Performanzphänomen beschrieben. Diese Sichtweise fügt sich in die in der Textlinguistik beobachtbare Entwicklung einer „Rehabilitierung der Textoberfläche“ (Feilke 2000: 78), wonach die sprachliche Gestaltung und damit auch das sprachlich Verfestigte als eine (notwendige) Dimension der Textsortenbeschreibung angesehen wird (vgl. z. B. Fandrych/Thurmair 2011; Krieg-Holz/Bülow 2016). Da das im Sprachgebrauch vorhandene sprachlich Verfestigte nicht zwangsläufig offensichtlich und auf Basis von Vorannahmen analysierbar ist, ist eine induktive korpuslinguistische Analyse die geeignete Methode, das sprachlich Verfestigte offenzulegen. Damit steht die Herangehensweise im Einklang mit dem Anliegen der quantitativen Textlinguistik und Stilistik, quantitative Eigenschaften von Texten – und hierzu ist auch das sprachlich Verfestigte zu zählen – mit Hilfe deskriptiver statistischer Verfahren zu bestimmen (vgl. Schmitz 2000: 196; Mehler 2008: 339). Zu bedenken ist, dass das Vorgehen um deduktive Arbeitsschritte zu ergänzen ist und dass bei der Analyse der Ergebnisse immer die Zusammensetzung des Untersuchungs- und Referenzkorpus berücksichtigt werden muss. Auch sollten die quantitativen Analysen die qualitativen nicht ersetzen, sondern objektivieren und ergänzen.

Wenn sich auch mit dieser Analysemethode, die Musterhaftigkeit an statistische Signifikanz bindet, nicht alles sprachlich Verfestigte entdecken lässt (vgl. Kap. 4.2), so ist das Offenlegen von signifikanten Sprachgebrauchsmustern doch ein vielversprechender Weg zur Identifikation von sprachlich Verfestigtem.

Literaturverzeichnis

- Belica, Cyril/Steyer, Kathrin (2008): Korpusanalytische Zugänge zu sprachlichem Usus. In: Vachková, Marie (ed.). *Beiträge zur bilingualen Lexikographie*. Prag, Univerzita Karlova v Praze: 7–24.
- Biber, Douglas (2006): *University Language. A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: Benjamins. (= *Studies in Corpus Linguistics* 23).
- Biber, Douglas (2009): A corpus-driven approach to formulaic language in English. Multiword patterns in speech and writing. *International Journal of Corpus Linguistics* 14/3: 275–311.
- Brinker, Klaus/Cölfen, Hermann/Pappert, Steffen (2014): *Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden*. 8., neu bearbeitete und erweiterte Auflage. Berlin: Schmidt. (= *Grundlagen der Germanistik* 29).
- Brommer, Sarah (z. Publ. eing.): *Sprachliche Muster in wissenschaftlichen Texten – eine induktive korpuslinguistische Analyse* (Dissertation).
- Bubenhof, Noah (2009): *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin/New York: de Gruyter. (= *Sprache und Wissen* 4).
- Bubenhof, Noah (2015): „Muster aus korpuslinguistischer Sicht“. In: Dürscheid, Christa/Schneider, Jan Georg (ed.): *Handbuch Satz – Äußerung – Schema*. Berlin/Boston, de Gruyter: 485–502. (= *Handbücher Sprachwissen* 4).
- Bubenhof, Noah/Scharloth, Joachim (2010): „Kontext korpuslinguistisch. Die induktive Berechnung von Sprachgebrauchsmustern in großen Textkorpora“. In: Klotz, Peter et al. (eds.): *Kontexte und Texte. Soziokulturelle Konstellationen literalen Handelns*. Tübingen, Narr: 85–107. (= *Europäische Studien zur Textlinguistik* 8).
- Burger, Harald et al. (2007): Phraseologie: Objektbereich, Terminologie und Forschungsschwerpunkte. In: Burger, Harald et al. (eds.): *Phraseologie. Ein internationales Handbuch zeitgenössischer Forschung*. Berlin/New York, de Gruyter: 1–10. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 28.1)
- Coulmas, Florian (1981): *Routine im Gespräch. Zur pragmatischen Fundierung der Idiomatik*. Wiesbaden: Akademische Verlagsgesellschaft Athenaion.
- Fandrych, Christian/Thurmair, Maria (2011): *Textsorten im Deutschen. Linguistische Analysen aus sprachdidaktischer Sicht*. Tübingen: Stauffenburg. (= *Linguistik* 57).
- Feilke, Helmuth (1993): Sprachlicher Common sense und Kommunikation. Über den „gesunden Menschenverstand“, die Prägung der Kompetenz und die idiomatische Ordnung des Verstehens. *Der Deutschunterricht* 45/6: 6–21.
- Feilke, Helmuth (1996): *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt am Main: Suhrkamp.
- Feilke, Helmuth (2000): „Die pragmatische Wende in der Textlinguistik“. In: Brinker, Klaus et al. (eds.): *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. Berlin/New York, de Gruyter: 64–82. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 16.1).
- Feilke, Helmuth (2003): Textroutine, Textsemantik und sprachliches Wissen. In: Linke, Angelika et al. (eds.): *Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis*. Tübingen, Niemeyer: 209–230. (= *Reihe Germanistische Linguistik* 245).
- Feilke, Helmuth (2012): Was sind Textroutinen? – Zur Theorie und Methodik des Forschungsfeldes. In: Feilke, Helmuth/Lehnen, Katrin (eds.): *Schreib- und Textroutinen. The-*

- orie, *Erwerb und didaktisch-mediale Modellierung*. Frankfurt am Main u. a., Peter Lang: 1–31. (= *Forum angewandte Linguistik* 52).
- Feilke, Helmuth (2014): Argumente für eine Didaktik der Textprozeduren. In: Feilke, Helmuth/Bachmann, Thomas (eds.): *Werkzeuge des Schreibens. Beiträge zu einer Didaktik der Textprozeduren*. Stuttgart, Fillibach bei Klett: 11–34.
- Felder, Ekkehard/Gardt, Andreas (2015): Sprache – Erkenntnis – Handeln. In: Felder, Ekkehard/Gardt, Andreas (eds.): *Handbuch Sprache und Wissen*. Berlin/Boston, de Gruyter: 3–33. (= *Handbücher Sprachwissen* 1)
- Filatkina, Natalia (2009): Historische formelhafte Sprache als „harte Nuss“ der Korpus- und Computerlinguistik. Ihre Annotation und Analyse im HiFoS-Projekt. *Linguistik online* 39/3: 75–95.
- Fix, Ulla (2000): Wie wir mit Textsorten umgehen und sie verändern – die Textsorte als ordnender Zugriff auf die Welt. *Der Deutschunterricht* 52/3: 54–65.
- Fix, Ulla (2008): Ansprüche an einen guten (?) Text. *Aptum. Zeitschrift für Sprachkritik und Sprachkultur* 4/1: 1–20.
- Fix, Ulla (2009): Muster und Abweichung in Rhetorik und Stilistik. In: Fix, Ulla et al. (eds.): *Rhetorik und Stilistik. Ein internationales Handbuch zeitgenössischer Forschung*. Berlin/New York, de Gruyter: 1300–1315. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 13.2).
- Gülich, Elisabeth (1997): Routineformeln und Formulierungsroutinen. Ein Beitrag zur Beschreibung formelhafter Texte. In: Wimmer, Rainer (ed.): *Wortbildung und Phraseologie*. Tübingen, Narr: 131–176. (= *Studien zur deutschen Sprache* 9).
- Heinemann, Wolfgang (2000): Textsorte – Textmuster – Texttyp. In: Brinker, Klaus et al. (eds.): *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. Berlin/New York, de Gruyter: 507–523. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 16.1).
- Hoffmann, Michael (2009): Situation als Kategorie von Rhetorik und Stilistik. In: Fix, Ulla et al. (eds.): *Rhetorik und Stilistik. Ein internationales Handbuch zeitgenössischer Forschung*. Berlin/New York, de Gruyter: 1316–1335. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 31.2).
- Köller, Wilhelm (2009): „Stil und Grammatik“. In: Fix, Ulla et al. (ed.): *Rhetorik und Stilistik. Ein internationales Handbuch zeitgenössischer Forschung*. Berlin/New York, de Gruyter: 1210–1229. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 31.2).
- Krieg-Holz, Ulrike/Bülow, Lars (2016): *Linguistische Text- und Stilanalyse. Eine Einführung*. Tübingen: Narr Francke Attempto. (= *narr Studienbücher*).
- Kupietz, Marc/Keibel, Holger (2009): Gebrauchsbasierte Grammatik: Statistische Regelmäßigkeit. In: Konopka, Marek/Strecker, Bruno (eds): *Deutsche Grammatik – Regeln, Normen, Sprachgebrauch*. Berlin/New York, de Gruyter: 33–52. (= *Jahrbuch des Instituts für Deutsche Sprache* 2008).
- Linke, Angelika (2011): Signifikante Muster – Perspektiven einer kulturanalytischen Linguistik. In: Wåghäll Nivre, Elisabeth et al. (eds.): *Begegnungen: Das VIII. Nordisch-Baltische Germanistentreffen in Sigtuna vom 11. bis zum 13.6.2009*. Stockholm, Acta Universitatis Stockholmiensis: 23–44. (= *Stockholmer Germanistische Forschungen* 74).
- Mehler, Alexander (2008): „Eigenschaften der textuellen Einheiten und Systeme“. In: Köhler, Reinhard et al. (ed.): *Quantitative Linguistik. Ein internationales Handbuch*

- zeitgenössischer Forschung. Berlin/New York, de Gruyter: 325–348. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 27).
- Oakes, Michael P. (2014): *Literary detective work on the computer*. Amsterdam: Benjamins. (= *Natural Language Processing* 12).
- Sanderson, Tamsin (2008): *Corpus, Culture, Discourse*. Tübingen: Narr. (= *Language in Performance* 39).
- Sandig, Barbara (2000): Text als prototypisches Konzept. In: Mangasser-Wahl, Martina (ed.). *Prototypentheorie in der Linguistik. Anwendungsbeispiele – Methodenreflexion – Perspektiven*. Tübingen, Stauffenburg: 93–112.
- Scharloth, Joachim/Bubenhof, Noah (2011): Datengeleitete Korpuspragmatik: Korpusvergleich als Methode der Stilanalyse. In: Felder, Ekkehard et al. (eds.): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen*. Berlin/New York, de Gruyter: 195–230. (= *Linguistik – Impulse & Tendenzen* 44).
- Schmitz, Ullrich (2000): „Statistische Methoden in der Textlinguistik“. In: Brinker, Klaus et al. (ed.): *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. Berlin/New York, de Gruyter: 196–201. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 16.1).
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford: University Press.
- Stefanowitsch, Anatol/Gries, Stefan (2003): „Collocations: Investigating the Interaction of Words and Constructions“. *International Journal of Corpus Linguistics* 2/8: 209–243.
- Stein, Stephan (1995): *Formelhafte Sprache. Untersuchungen zu ihren pragmatischen und kognitiven Funktionen im gegenwärtigen Deutsch*. Frankfurt am Main: Peter Lang. (= *Sprache in der Gesellschaft: Beiträge zur Sprachwissenschaft* 22).
- Steinhoff, Torsten (2007): *Wissenschaftliche Textkompetenz. Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten*. Tübingen: Niemeyer. (= *Reihe Germanistische Linguistik* 280).
- Steyer, Kathrin (2013): *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr. (*Studien zur deutschen Sprache* 65).
- Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam: Benjamins. (= *Studies in Corpus linguistics* 6).
- Wallner, Franziska (2014): *Kollokationen in Wissenschaftssprachen. Zur lernerlexikographischen Relevanz ihrer wissenschaftssprachlichen Gebrauchsspezifika*. Tübingen: Stauffenburg. (= *Deutsch als Fremd- und Zweitsprache. Schriften des Herder-Instituts* 5).