Cordula Meißner und Franziska Wallner

GeSIG – Das gemeinsame sprachliche Inventar der Geisteswissenschaften

Ein korpuslinguistisches Projekt zur Erschließung
von fachübergreifendem Wortschatz

Sprache ist nicht nur ein Instrument, um Sachverhalte zu vermitteln, sondern spielt für das wissenschaftliche Denken eine konstitutive Rolle. Diese für alle Wissenschaften geltende Tatsache wird im Bereich der Geisteswissenschaften besonders deutlich, da hier selbst die Gegenstände der Forschung größtenteils sprachlich verfasst sind (vgl. Kretzenbacher 2010). Sprache fungiert hierbei als Medium der Aneignung, Verarbeitung sowie Weiterentwicklung wissenschaftlicher Erkenntnis. Entscheidend für die sprachliche Realisierung dieser Funktionen sind die Ausdrucksmittel der allgemeinen bzw. alltäglichen Wissenschaftssprache, mit denen die Zusammenhänge zwischen den fachterminologisch ausgedrückten Sachverhalten hergestellt und ausgestaltet werden (vgl. Ehlich 1999: 8–9). Doch gerade ihr adäquater Gebrauch stellt für Novizen des Wissenschaftsbetriebs eine große Herausforderung dar. So ist die Aneignung wissenschaftssprachlicher Kompetenz eine komplexe Lernaufgabe. Studierende finden hierbei kaum Unterstützung, denn während die Terminologie Gegenstand der Fachausbildung ist, wird vorausgesetzt, dass die sprachlichen Mittel der allgemeinen Wissenschaftssprache von vornherein beherrscht werden.

Obwohl die Erforschung der Wissenschaftssprache auch in Bezug auf das Deutsche in den letzten Jahren deutliche Fortschritte gemacht hat, stehen bislang keine Wortschatzsammlungen zur Verfügung, welche die allgemeine fachübergreifende Wissenschaftssprache bedarfsgerecht repräsentieren. Dies wäre jedoch die Voraussetzung für eine adäquate Unterstützung von Aneignung und Vermittlung wissenschaftssprachlicher Kompetenz.

Vor diesem Hintergrund setzt sich das im vorliegenden Beitrag vorgestellte Projekt GeSIG zum Ziel, den fachübergreifend gebrauchten Wortschatz – das gemeinsame sprachliche Inventar – der Geisteswissenschaften korpuslinguistisch zu erschließen. Der Beitrag stellt das methodische Vorgehen zur Ermitt-

Cordula Meißner, Universität Leipzig, Herder-Institut, 04081 Leipzig,
E-Mail: cordula.meissner@uni-leipzig.de

Franziska Wallner, Universität Leipzig, Herder-Institut, 04081 Leipzig,
E-Mail: f.wallner@rz.uni-leipzig.de

Open Access. © 2018 Cordula Meißner und Franziska Wallner, publiziert von De Gruyter. Dieses Werk ist lizenziert unter der Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Lizenz.

<https://doi.org/10.1515/9783110476958-019>

lung des Inventars vor, gibt einen ersten Einblick in seine Zusammensetzung und umreißt Forschungsbereiche, in denen das Inventar den Ausgangspunkt weitergehender Untersuchungen bilden kann.

1 Zum Stand der Lexikografie der Wissenschaftssprache

Zur Erforschung der Wissenschaftssprache hat sich seit den 1980er Jahren eine eigenständige Disziplin herausgebildet, die ihren Gegenstand theoretisch nicht mehr als Teil der traditionellen Fachsprachenforschung verortet: Anknüpfend an Weinrichs Forderung, dass „[d]ie Dyas Wissenschaftstheorie – Wissenschaftsgeschichte [...] zu einer Trias Wissenschaftstheorie – Wissenschaftsgeschichte – Wissenschaftssprachforschung zu erweitern [ist]“ (Weinrich 1989: 154), wird die Wissenschaftssprachforschung nunmehr den wissenschaftsreflektierenden Disziplinen zugeordnet (vgl. Kretzenbacher 1998).

Während sich traditionelle Arbeiten der Fach- und Wissenschaftssprachforschung nahezu ausschließlich auf den terminologisch fixierten Fachwortschatz konzentrierten, rückten durch Konzepte wie das der ‚allgemeinen Wissenschaftssprache‘ (Schepping 1976) und das der ‚alltäglichen Wissenschaftssprache‘ (Ehlich 1993) diejenigen nicht-terminologischen sprachlichen Mittel ins Zentrum der Aufmerksamkeit, die pragmatisch-methodische Inhalte ausdrücken und so disziplinenübergreifend Verwendung finden. Zum anderen wird die mit der Polysemie bzw. Vagheit der Ausdrücke verbundene inhaltliche Flexibilität in verschiedenen fachlichen und textuellen Kontexten als eine charakteristische Eigenschaft dieser Ausdrucksmittel in den Blick genommen (vgl. Ehlich 2007: 104–105). Die alltägliche Wissenschaftssprache zeichnet sich durch die „spezifische Nutzung von Teilen der Alltagssprache für die Zwecke der Wissenschaft“ (Ehlich 2000: 52) aus. Diesen disziplinenübergreifend verwendeten sprachlichen Mitteln kommt eine besondere wissenschaftsmethodologische Bedeutung zu, stellen sie doch Ausdrucksressourcen bereit etwa für Formen des Voraussetzens, des Begründens, des Folgerns, des Übertragens und des Ableitens. Daneben finden Formen der metakommunikativen Bezugnahme (wie bspw. zur Textstrukturierung und Lenkung der Rezeptionserwartung) und der intertextuellen Bezugnahme (etwa durch Zitate, Verweise und Reformulierungen) ihren Ausdruck mit Hilfe allgemeinwissenschaftssprachlicher Mittel. Die sprachliche Ausgestaltung dieser Handlungen ist geprägt durch den kontroversenorientierten Charakter von Wissenschaftskommunikation (vgl. Ehlich 1993, Steinhoff 2008, Feilke 2010). Eine Analyse der sprachlichen Mittel der

allgemeinen Wissenschaftssprache gestattet somit Einblicke in zentrale Prozesse wissenschaftlichen Handelns. Sie ermöglicht es, die Funktionsweise von Wissenschaftssprache als facettenreiches, differenziertes Erkenntnisinstrument näher zu beleuchten, welches insbesondere für die geisteswissenschaftliche Forschung von grundlegender Bedeutung ist.

Die allgemeine Wissenschaftssprache der Geisteswissenschaften war bislang jedoch kaum Gegenstand der Forschung und wurde bis dato in keiner Weise lexikografisch erfasst und erschlossen. Vorliegende Arbeiten nehmen in der Regel das gesamte Spektrum der allgemeinen Wissenschaftssprache in den Blick, ohne spezifische Disziplinengruppen zu fokussieren. Sie widmen sich exemplarisch, z. T. auch unter kontrastiver Perspektive, einzelnen sprachlichen Aspekten (z. B. Graefen 1999; Jasny 2001; Fandrych 2004, 2005, 2006; Thielmann 2009; Meißner 2014; Redder, Heller & Thielmann 2014; Wallner 2014). Den bislang einzigen Versuch einer lexikografischen Erfassung der allgemeinen Wissenschaftssprache bilden die Wortschatzsammlungen von Heinrich Erk (1972, 1975, 1982, 1985). Auch sie fokussieren jedoch nicht die Sprache der Geisteswissenschaften, sondern berücksichtigen sämtliche akademische Fachdisziplinen und erlauben daher in Bezug auf die geisteswissenschaftlichen Fächer eine nur geringe Beschreibungsdetailliertheit. Die Arbeiten von Erk basieren auf insgesamt 102 Texten aus 34 Fachdisziplinen, „die traditionsgemäß zum Fächerkanon deutschsprachiger Universitäten gehören“ (Erk 1982: 9). Die Datenbasis umfasst pro Fach jeweils drei Texte verschiedener Textsorten (Lehrbuchtexte, fachwissenschaftliche Zeitschriftenartikel und populärwissenschaftliche Zeitschriftenartikel). Die Texte wurden in ihrer Länge genormt auf je 300 Zeilen mit durchschnittlich 63 Anschlägen pro Zeile (Erk 1972: 27). Insgesamt umfasst das Korpus 250.000 laufende Wörter (Token). Alle Wörter, die mindestens zehnmal vorkamen, fanden Eingang in Erks Sammlung. Die disziplinenübergreifende Verwendung ist dabei kein Selektionskriterium, es werden jedoch Angaben zur Verbreitung der Lexeme in einzelnen Texten gemacht.

Für ihre Zeit sind die Arbeiten von Heinrich Erk sehr fortschrittlich. Aus aktueller korpusmethodologischer Perspektive erscheint die verwendete Datengrundlage bei der Vielzahl einfließender fach- und textsortenbezogener Parameter jedoch recht schmal. Allerdings existiert bis heute keine adäquate empirische Forschungsgrundlage in Form von repräsentativen, ausreichend umfangreichen und ausgewogenen Korpora zur aktuellen deutschen Wissenschaftssprache. Dies gilt im Hinblick auf wissenschaftliche Texte, denn selbst die beiden größten schriftsprachlichen Korpus-sammlungen zum Deutschen (DWDS, IDS-Korpora) stellen aufgrund ihrer Zusammensetzung, ihres Umfangs und ihrer Zugänglichkeit keine geeignete Datengrundlage für eine systematische Beschreibung der Wissenschaftssprache dar. Hinsichtlich der gesproche-

nen Wissenschaftssprache liegt mit dem GeWiss-Korpus¹ eine erste öffentlich zugängliche Datengrundlage an wissenschaftlichen Vorträgen, Seminarreferaten und Prüfungsgesprächen vor. Diese kann jedoch aufgrund ihres geringen Umfangs und der Konzentration auf den Fachbereich der Philologie für das hier vorgestellte Projekt keine geeignete Grundlage bilden. Das Projekt GeSIG konnte somit nicht auf ein vorliegendes Korpus zurückgreifen. Es musste eine für seine Fragestellung adäquate Datengrundlage zusammenstellen, auf die im folgenden Abschnitt näher eingegangen wird.

2 GeSIG – ein korpuslinguistisches Projekt zur Erschließung des fachübergreifenden Wortschatzes der Geisteswissenschaften

Mit dem am Herder-Institut der Universität Leipzig angesiedelten Projekt GeSIG (Das gemeinsame sprachliche Inventar der Geisteswissenschaften)² wurde erstmals das Inventar der allgemeinen Wissenschaftssprache der Geisteswissenschaften auf empirischer Grundlage bestimmt. Hierzu wurden aktuelle korpusmethodologische Werkzeuge und Erschließungsverfahren eingesetzt, um datengeleitet einen vollständigen, systematischen Zugriff auf den allgemeinwissenschaftlichen Wortschatzbestand der Geisteswissenschaften zu ermöglichen. Im Folgenden soll das Vorgehen zur Ermittlung dieses Lemmabestandes vorgestellt werden.

Die ideale Datengrundlage für eine vollständige Beschreibung der allgemeinen Wissenschaftssprache der Geisteswissenschaften würde für alle zugehörigen Fachbereiche umfangreiche Korpora sowohl verschiedener Textsorten als auch Diskursarten umfassen. Wie aus der oben beschriebenen unzureichenden Lage der Korpuslandschaft zur Wissenschaftssprache des Deutschen hervorgeht, hatte das Projekt GeSIG nicht die Möglichkeit, hierbei auf vorhandene Ressourcen zurückzugreifen, sondern musste eine neue Datengrundlage aufbauen. Da die Erstellung von Korpora der gesprochenen Sprache sehr zeitintensiv und aus diesem Grund kaum für größere Korpusumfänge zu realisieren ist, wurde die Entscheidung getroffen, sich bei der Zusammenstellung der Datengrundlage auf schriftliche wissenschaftliche Texte zu konzentrieren. Als

¹ Vgl. Fandrych, Meißner & Slavcheva (2014) sowie unter <https://gewiss.uni-leipzig.de> (28. 11. 2017).

² Vgl. <http://research.uni-leipzig.de/gesig/> (28. 11. 2017).

Textsorte wurde für den Korpusaufbau die Dissertation gewählt, da sie – anders als bspw. die Rezension oder der Zeitschriftenaufsatz – das gesamte Spektrum des in Textform niedergelegten wissenschaftlichen Erkenntnisprozesses in besonderer Breite und Vollständigkeit abbildet. Sie enthält Auseinandersetzungen zum Forschungsstand und die eigene Positionierung, die Begründung und Entfaltung von Forschungsfragen, deren Operationalisierung und Bearbeitung, die Sicherung und Einordnung von Ergebnissen sowie die Formulierung von Schlussfolgerungen und Desideraten. Sie bietet damit einen repräsentativen Zugang zur Untersuchung der Vielfalt wissenschaftssprachlicher Handlungen in den Geisteswissenschaften und ermöglicht so die Ermittlung der allgemeinwissenschaftssprachlichen Ausdrucksmittel, die diesen Handlungen zugrunde liegen. Zudem erlaubt die Textsorte Dissertation aufgrund ihres Umfangs den Aufbau einer ausreichend großen Datengrundlage.

Als Analysegrundlage wurde folglich ein Korpus geisteswissenschaftlicher Dissertationen zusammengestellt. Zur Operationalisierung des Gebietes der Geisteswissenschaften wurde die Umfangsbestimmung des Wissenschaftsrates (2010) zugrunde gelegt. Diese ist an die Systematik des statistischen Bundesamtes angelehnt und umfasst die dort unterschiedenen Fächergruppen Sprach- und Kulturwissenschaften (ohne Psychologie, Erziehungswissenschaften und Sonderpädagogik) sowie Kunst und Kunstwissenschaften.³ So operationalisiert umfassen die Geisteswissenschaften 19 Fachbereiche. Eingeschlossen sind Fächer wie Philosophie, Sprach- und Literaturwissenschaften, Geschichtswissenschaften, Regionalstudien, religionsbezogene Wissenschaften, die bekenntnisgebundenen Theologien, die Ethnologien sowie die Kunst-, Theater- und Musikwissenschaften (vgl. Statistisches Bundesamt 2013). Abbildung 1 zeigt die 19 Fachbereiche und die ihnen zugeordneten Fächer im Einzelnen.

³ Die Einteilung des Statistischen Bundesamtes basiert auf der institutionellen Organisation des Bereiches sowie auf Stellungnahmen von Fachverbänden. Sie wird fortlaufend überprüft und weiterentwickelt (vgl. Projektgruppe „Fächerklassifikation und Thesauri“ 2014).

01 Sprach- und Kulturwissenschaften	08 Altphilologie (klass. Philologie), Neugriechisch	09 Kunst, Kunstwissenschaft
01 Sprach- und Kulturwissenschaften allgemein Interdisziplinäre Studien (Schwerpunkt Sprach- und Kulturwissenschaften), Lernbereich Sprach- und Kulturwissenschaften, Medienwissenschaft	Byzantinistik, Griechisch, Klassische Philologie, Latein, Neugriechisch 09 Germanistik (Deutsch, germanische Sprachen ohne Anglistik) Dänisch, Deutsch für Ausländer, Germanistik/Deutsch, Niederdeutsch, Niederländisch, Nordistik/Skandinavistik (Nordische Philologie)	74 Kunst, Kunstwissenschaft allgemein Interdisziplinäre Studien (Schwerpunkt Kunst, Kunstwissenschaften), Kunsterziehung, Kunstgeschichte, Kunstwissenschaft, Restaurierungskunde
02 Evang. Theologie, Religionslehre Diakoniewissenschaft, Ev. Religionspädagogik, Ev. Theologie, Religionslehre	10 Anglistik, Amerikanistik Amerikanistik/Amerikakunde, Anglistik/Englisch 11 Romanistik Französisch, Italienisch, Portugiesisch, Romanistik (Roman. Philologie), Spanisch	75 Bildende Kunst Bildende Kunst/Graphik, Bildhauerei/Plastik, Malerei, Neue Medien
03 Kath. Theologie, Religionslehre Caritaswissenschaft, Kath. Religionspädagogik, Kath. Theologie, Religionslehre		76 Gestaltung Angewandte Kunst, Edelstein- und Schmuckdesign, Graphikdesign/Kommunikationsgestaltung, Industriedesign/Produktgestaltung, Textilgestaltung, Werkerziehung
04 Philosophie Ethik, Philosophie, Religionswissenschaft		
05 Geschichte Alte Geschichte, Archäologie, Geschichte, Mittlere und neuere Geschichte, Ur- und Frühgeschichte, Wirtschafts-/Sozialgeschichte	12 Slawistik, Baltistik, Finno-Ugristik Baltistik, Finno-Ugristik, Polnisch, Russisch, Slawistik (Slaw. Philologie), Sorbisch, Südslawisch (Bulgarisch, Serbokroatisch, Slowenisch usw.), Tschechisch, Westslawisch	77 Darstellende Kunst, Film und Fernsehen, Theaterwissenschaft Darstellende Kunst/Bühnenkunst/Regie, Film und Fernsehen, Schauspiel, Tanzpädagogik, Theaterwissenschaft
06 Bibliothekswissenschaft, Dokumentation Bibliothekswissenschaft/-wesen, Dokumentationswissenschaft	13 Außereuropäische Sprach- und Kulturwissenschaften Ägyptologie, Afrikanistik, Arabisch/Arabistik, Asiatische Sprachen und Kulturen/Asienwissenschaften, Außereuropäische	78 Musik, Musikwissenschaft Dirigieren, Gesang, Instrumentalmusik, Jazz und Populärmusik, Kirchenmusik, Komposition, Musikerziehung, Musikwissenschaft/-
07 Allgemeine und vergleichende Literatur- und Sprachwissenschaft		

Allgemeine Literaturwissenschaft, Allgemeine Sprachwissenschaft/Indogermanistik, Angewandte Sprachwissenschaft, Berufsbezogene Fremdsprachenausbildung, Computerlinguistik	Sprachen und Kulturen in Ozeanien und Amerika, Hebräisch/Judaistik, Indologie, Iranistik, Islamwissenschaft, Japanologie, Kaukasistik, Orientalistik/Altorientalistik, Sinologie/Koreanistik, Turkologie	geschichte, Orchestermusik, Rhythmik, Tonmeister
--	--	--

14 Kulturwissenschaften i. e. S.

Europäische Ethnologie und Kulturwissenschaft, Ethnologie, Volkskunde

Abb. 1: Die 19 den Geisteswissenschaften zugeordneten Fachbereiche nach der Umfangsbestimmung des Wissenschaftsrates (2010: 15), die auf der Einteilung des Statistischen Bundesamtes basiert (Basis hier: Statistisches Bundesamt 2013).

Die 19 geisteswissenschaftlichen Fachbereiche bildeten die Grundlage für den Aufbau von 19 entsprechenden Fachbereichskorpora. Dabei wurde für jeden Bereich ein Korpus aus mindestens 10 Dissertationen und mit einer Mindestgröße von 1 Mio. Token zusammengestellt. Es wurde hierbei auf eine möglichst breite Abdeckung des Disziplinenpektrums der Fachbereiche geachtet. Die Zuordnung der Dissertationen zu Fachbereichen erfolgte anhand des Instituts bzw. Lehrstuhls, an dem diese eingereicht wurden. Insgesamt umfasst die so erhobene Datengrundlage 197 Dissertationen und rund 22,8 Mio. Token.

Die Sprachdaten wurden anschließend für die korpuslinguistische Analyse bereinigt und aufbereitet. Sie wurden mit Hilfe des TreeTaggers (Schmid 1995) nach Wortarten annotiert und lemmatisiert. Dabei lag das Stuttgart-Tübingen-Tagset (STTS) (Schiller et al. 1999) zugrunde – ein Wortartenkategorien-set, welches sich als Quasi-Standard für die Annotation von Wortarten im Deutschen etabliert hat (vgl. Zinsmeister 2015: 104).

Um das Konzept der allgemeinen Wissenschaftssprache zu operationalisieren, wurde das Charakteristikum ihrer disziplinenübergreifenden Verwendung herangezogen. Die sprachlichen Mittel der allgemeinen Wissenschaftssprache der Geisteswissenschaften wurden demnach empirisch bestimmt als Schnittmenge der Wortschätze einzelner geisteswissenschaftlicher Fachbereiche. Hierzu wurde für jedes der 19 Fachbereichskorpora mit Hilfe der Korpusanalyse-Software AntConc (Anthony 2014) eine Lemmaliste generiert. Aus den so erstellten 19 Fachbereichs-Lemmalisten wurde anschließend mit Hilfe der Software R⁴ eine Schnittmenge gebildet. Der in dieser Schnittmenge enthaltene Wortschatz umfasst jene sprachlichen Mittel, die der Form nach in den 19 geisteswissenschaftlichen Fachbereichen übergreifend gebraucht werden und repräsentiert damit die sprachlichen Mittel der allgemeinen Wissenschaftssprache der Geisteswissenschaften. Im folgenden Abschnitt wird die Zusammensetzung des auf diese Weise ermittelten Inventars näher beschrieben.

3 Das gemeinsame sprachliche Inventar der Geisteswissenschaften: Erste Ergebnisse

Die aus den 19 Fachbereichs-Lemmalisten gebildete Schnittmenge umfasste nach Ausschluss von Elementen, die automatisch als Eigennamen (NE),⁵ Kardinalzahlen (CARD), Partikelverbzusätze (PTKVZ) und fremdsprachliches Material (FM) getaggt wurden, zunächst 4.668 Lemmata. Darüber hinaus wurden

⁴ Vgl. <https://www.rstudio.com/> (28. 11. 2017).

⁵ Die hier aufgeführten Kurzformen entsprechen den im STTS verwendeten Kürzeln für Wortartenkategorien (Schiller et al. 1999).

nach einer genaueren Prüfung der Schnittmenge einzelne durch den TreeTagger fehlerhaft ausgewiesene Elemente getilgt. Dazu gehörten u. a.

- falsch lemmatisierte Elemente (bspw. **anmessen* für *angemessen*, welches fälschlich als Partizip erkannt und entsprechend als Verb lemmatisiert wurde, als Adjektiv aber bereits im Inventar enthalten war),
- falsch getaggte Elemente (bspw. **Kühn*, welches fälschlich als Nomen (NN) getaggt wurde, obwohl es sich um einen Eigennamen (NE) handelt)
- sowie fälschlich nicht als fremdsprachlich erkannte Wörter (bspw. *Most* (NN) für engl. *most*).

Nach diesem Bereinigungsschritt enthält das Inventar 4.490 Lemmata. Den größten Teil bilden Inhaltswörter: Nomen sind mit 1.681 Lemmata am häufigsten vertreten, gefolgt von Adjektiven mit 1.171 Lemmata, Verben mit 1.108 Lemmata und Adverbien⁶ mit 261 Lemmata. Abbildung 2 zeigt im Detail, wie sich die Lemmata des Inventars auf die einzelnen Wortarten verteilen.

Abb. 2: Zusammensetzung des Inventars nach Wortarten.

Nomen, Adjektive, Verben und Adverbien machen somit 94 % aller Lemmata des Inventars aus. Doch um welche Inhaltswörter handelt es sich dabei? Tabelle 1 zeigt die jeweils 20 häufigsten Vertreter dieser Wortarten.⁷

⁶ Nach den Regeln der Wortartenannotation des Stuttgart-Tübingen-Tag-Sets (STTS) zählen nur die nicht von Adjektiven abgeleiteten, nicht-flektierbaren Modifizierer zu den Adverbien (Schiller et al. 1999: 55–56).

⁷ Eine umfassende Dokumentation und Beschreibung des GeSIG-Inventars wird in Meißner & Wallner (2019) gegeben. Daneben ist das GeSIG-Inventar elektronisch verfügbar unter <http://GeSIG-Inventar.ESV.info> (27. 09. 2018).

Tab. 1: Die 20 häufigsten Lemmata des GeSIG-Inventars für Nomen, Verben, Adjektive und Adverbien.

	Nomen	Verben	Adjektive	Adverbien
1	<i>Jahr</i>	<i>sein</i>	<i>neu</i>	<i>auch</i>
2	<i>Form</i>	<i>werden</i>	<i>erst</i>	<i>so</i>
3	<i>Mensch</i>	<i>haben</i>	<i>groß</i>	<i>nur</i>
4	<i>Frage</i>	<i>können</i>	<i>weit</i>	<i>hier</i>
5	<i>Arbeit</i>	<i>sollen</i>	<i>deutsch</i>	<i>noch</i>
6	<i>Zeit</i>	<i>lassen</i>	<i>eigen</i>	<i>aber</i>
7	<i>Text</i>	<i>müssen</i>	<i>gut</i>	<i>jedoch</i>
8	<i>Sprache</i>	<i>geben</i>	<i>verschieden</i>	<i>selbst</i>
9	<i>Bild</i>	<i>zeigen</i>	<i>deutlich</i>	<i>also</i>
10	<i>Gott</i>	<i>finden</i>	<i>folgend</i>	<i>bereits</i>
11	<i>Beispiel</i>	<i>machen</i>	<i>unterschiedlich</i>	<i>mehr</i>
12	<i>Teil</i>	<i>sehen</i>	<i>möglich</i>	<i>dann</i>
13	<i>Bedeutung</i>	<i>stehen</i>	<i>hoch</i>	<i>immer</i>
14	<i>Begriff</i>	<i>kommen</i>	<i>stark</i>	<i>sehr</i>
15	<i>Kapitel</i>	<i>darstellen</i>	<i>wichtig</i>	<i>wieder</i>
16	<i>Wort</i>	<i>wollen</i>	<i>zweit</i>	<i>schon</i>
17	<i>Kunst</i>	<i>bleiben</i>	<i>einzel</i>	<i>allerdings</i>
18	<i>Sinn</i>	<i>gehen</i>	<i>spät</i>	<i>zwar</i>
19	<i>Werk</i>	<i>handeln</i>	<i>bestimmt</i>	<i>da</i>
20	<i>Fall</i>	<i>führen</i>	<i>sozial</i>	<i>erst</i>

Diese Lemmata des GeSIG-Inventars weisen die typischen Charakteristika der allgemeinen Wissenschaftssprache auf: Auf den ersten Blick wirken sie unauffällig und scheinen sich nicht von allgemeinsprachlicher Lexik zu unterscheiden. Mit ihnen ist jedoch ein breites und vielfältiges Spektrum an handlungsspezifischen Gebrauchsmustern verbunden. Meißner & Wallner (2016) belegen dies in einer ersten Korpusanalyse zum GeSIG-Inventar am Beispiel der Kollokation *Bild zeichnen*. Die Basis *Bild* zählt zu den 10 häufigsten Nomen des Inventars. Das Verb *zeichnen* ist wiederum dessen meist gebrauchter Vollverb-Kookkurrenzpartner. Die Analyse von *Bild zeichnen* zeigt, dass diese Kollokation in wissenschaftlichen Texten zur Realisierung verschiedener wissenschaftssprachlicher Handlungen gebraucht wird, wie u. a.

- in Metakomentierungen⁸ (*Aufgrund der Korrelation der Daten [...] lässt sich folgendes **Bild zeichnen**. (09_DAF_3)*),

⁸ Bei Metakomentierungen bzw. Textkomentierungen handelt es sich um Äußerungen bzw. Textsequenzen, die der Gliederung und rezipientenbezogenen Erwartungslenkung dienen (vgl. näher zu diesem Konzept Fandrych & Graefen 2002, Fandrych 2014).

- in Verweisen auf andere Forschungspositionen und Primärquellen (*Hough/Richards, The Battle of Britain, 90f. zeichnet ebenfalls ein etwas differenzierteres Bild ...* (10_ANG_8)),
- in evaluierenden Charakterisierungen von Forschungsdaten (*Diese Ergebnisse zeichnen ein klares und zugleich kontrastreiches Bild.* (11_FRZ_2))

Das Beispiel der Kollokation *Bild zeichnen* macht zudem deutlich, wie sich der kontroversenorientierte Charakter von Wissenschaftskommunikation sprachlich manifestiert. So zeigt die Analyse von Belegstellen, die diese Kollokation enthalten, dass mittels qualifizierender Adjektivattribute zu *Bild* auf einen wissenschaftlichen Erkenntniszuwachs Bezug genommen und auf diese Weise die jeweils realisierten Handlungen kontroversenorientiert persuasiv gestaltet werden können. Dies zeigt sich auch in den oben genannten Belegen am Gebrauch der Adjektive *differenziert*, *klar* und *kontrastreich*. Auch die in Tabelle 1 mit aufgeführten Lemmata *neu*, *deutlich* und *wichtig* zählen zu den Adjektiven, mit denen in dieser Form auf einen wissenschaftlichen Erkenntniszuwachs Bezug genommen werden kann. Allerdings ist die Auswahl der für diese Funktion in Frage kommenden Adjektive konventionalisiert und nicht beliebig erweiterbar. Nur eine Beschreibung der Ausdrucksmittel der allgemeinen Wissenschaftssprache in ihrem Gebrauchskontext ermöglicht es, das vielfältige wissenschaftssprachspezifische Funktionenspektrum sowie die entsprechenden Realisierungskonventionen zu erfassen. Das GeSIG-Inventar bietet einen fundierten Ausgangspunkt für solche Analysen und damit zu einer umfangreichen Erschließung der allgemeinen Wissenschaftssprache der Geisteswissenschaften.

4 Ausblick

Das GeSIG-Inventar kann den Ausgangspunkt für Untersuchungen in verschiedenen Forschungsbereichen bieten. Für die Erforschung des Deutschen als Wissenschaftssprache eröffnet es die Möglichkeit, ausgehend von der daten geleiteten Analyse der Verwendungsspektren der Lemmata deren multiple Funktionalitäten aufzudecken, zu systematisieren und auf diese Weise den gesamten durch die Mittel der allgemeinen Wissenschaftssprache für die Geisteswissenschaften eröffneten wissenschaftssprachlichen Handlungsraum zu erschließen. Diese Untersuchungsperspektive ließe sich auch in die diachrone Dimension fortsetzen. So wäre für die Lemmata des Inventars nachzuzeichnen, wann sich aus deren ursprünglich gemeinsprachlichem Gebrauch eine wissenschaftssprachliche Funktion herausgebildet hat und wie die Entfaltung des

wissenschaftssprachlichen Funktionenspektrums im historischen Verlauf erfolgt ist.

Daneben kann das Inventar für sprachkontrastive Analysen nutzbar gemacht werden. Im Konzept der alltäglichen Wissenschaftssprache ist die Sichtweise angelegt, dass für jede Sprachgemeinschaft mit einer ausgebildeten Wissenschaftssprache eine alltägliche Wissenschaftssprache anzunehmen ist. Im Sinne einer „Wissenschaftssprachkomparatistik“ (Ehlich 2006) wäre dann der jeweils sprachspezifischen Nutzung gemeinsprachlicher Ressourcen für die einzelnen alltäglichen Wissenschaftssprachen nachzugehen. Das für das Deutsche ermittelte gemeinsame sprachliche Inventar der Geisteswissenschaften kann hierbei als Grundlage des Vergleichs dienen.

Neben diesen Fragen der sprachwissenschaftlichen Grundlagenforschung stellt das GeSIG-Inventar eine Vorarbeit für die lexikografische Aufarbeitung der allgemeinen Wissenschaftssprache der Geisteswissenschaften dar. Als fachübergreifend gebrauchter Lemmabestand bildet es eine empirisch fundierte Basis für eine entsprechende Lemmaselektion.

Wichtige weitergehende Forschungsaufgaben, die auf Basis des GeSIG-Inventars ermöglicht werden, liegen im Bereich der Studienvorbereitung und Wissenschaftspropädeutik. Erst wenn für die sprachlichen Mittel der allgemeinen Wissenschaftssprache die im Sprachgebrauch zu Tage tretenden Verwendungsbereiche und die daraus resultierenden Polysemiebeziehungen erforscht sind, ist es möglich, adäquate Hilfsmittel bzw. Informationsressourcen zur Unterstützung der Aneignung, der Vermittlung und des Gebrauchs der deutschen Wissenschaftssprache zu entwickeln. Hierin liegt für die Vermittlung des Deutschen als (fremder) Wissenschaftssprache ein besonderes Desiderat, stellt doch die bis in gemeinsprachliche Verwendungen hineinreichende Polysemie allgemeinwissenschaftlicher Ausdrucksmittel eine erhebliche Lernschwierigkeit dar.

Literaturverzeichnis

- Anthony, Laurence (2014): *AntConc* (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. <http://www.laurenceanthony.net/> (28. 11. 2017).
- Ehlich, Konrad (1993): Deutsch als fremde Wissenschaftssprache. In: *Jahrbuch Deutsch als Fremdsprache* 19, 13–42.
- Ehlich, Konrad (1999): Alltägliche Wissenschaftssprache. In: *Info DaF* 26 (1), 3–24.
- Ehlich, Konrad (2000): Deutsch als Wissenschaftssprache für das 21. Jahrhundert. In: *German as a Foreign Language* 2000 (1), 47–63.
- Ehlich, Konrad (2006): Mehrsprachigkeit in der Wissenschaftskommunikation – Illusion oder Notwendigkeit. In: Konrad Ehlich & Dorothee Heller (Hrsg.), *Die Wissenschaft und ihre Sprachen*. Frankfurt am Main: Lang, 17–38.

- Ehlich, Konrad (2007): *Sprache und sprachliches Handeln* (Bd. 1). Berlin, New York: de Gruyter.
- Erk, Heinrich (1972): *Zur Lexik wissenschaftlicher Fachtexte: Verben, Frequenz und Verwendungsweise. Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts*. München: Hueber.
- Erk, Heinrich (1975): *Zur Lexik wissenschaftlicher Fachtexte: Substantive, Frequenz und Verwendungsweise. Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts*. München: Hueber.
- Erk, Heinrich (1982): *Zur Lexik wissenschaftlicher Fachtexte: Adjektive, Adverbien und andere Wortarten, Frequenz und Verwendungsweise. Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts*. München: Hueber.
- Erk, Heinrich (1985): *Wortfamilien in wissenschaftlichen Texten. Ein Häufigkeitsindex. Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts*. München: Hueber.
- Fandrych, Christian (2004): Bilder vom wissenschaftlichen Schreiben. Sprechhandlungs-ausdrücke im Wissenschaftsdeutschen: Linguistische und didaktische Überlegungen. In: *Materialien Deutsch als Fremdsprache* 73, 269–291.
- Fandrych, Christian (2005): Räume und Wege in der Wissenschaft. Einige zentrale Konzeptualisierungen von wissenschaftlichem Schreiben im Deutschen und Englischen. In: Ulla Fix, Gotthard Lerchner, Marianne Schröder & Hans Wellmann (Hrsg.), *Zwischen Lexikon und Text. Lexikalische, stilistische und textlinguistische Aspekte*. Leipzig, Stuttgart: Sächsische Akademie der Wissenschaft, 20–33.
- Fandrych, Christian (2006): Bildhaftigkeit und Formelhaftigkeit in der allgemeinen Wissenschaftssprache als Herausforderung für Deutsch als Fremdsprache. In: Konrad Ehlich & Dorothee Heller (Hrsg.), *Die Wissenschaft und ihre Sprachen*. Frankfurt am Main: Lang, 39–61.
- Fandrych, Christian (2014): Metakomentierungen in wissenschaftlichen Vorträgen. In: Christian Fandrych, Cordula Meißner & Adriana Slavcheva (Hrsg.), *Gesprochene Wissenschaftssprache: Korpusmethodische Fragen und empirische Analysen*. Heidelberg: Synchron, 95–111.
- Fandrych, Christian, Cordula Meißner & Adriana Slavcheva (Hrsg.) (2014): *Gesprochene Wissenschaftssprache: Korpusmethodische Fragen und empirische Analysen*. Heidelberg: Synchron.
- Fandrych, Christian & Gabriele Graefen (2002): Text-commenting devices in German and English academic articles. In: *Multilingua* 21, 17–43.
- Feilke Helmuth (2010): Schriftliches Argumentieren zwischen Nähe und Distanz – am Beispiel wissenschaftlichen Schreibens. In: Vilmos Ägel & Mathilde Hennig (Hrsg.): *Nähe und Distanz im Kontext variationslinguistischer Forschung*. Berlin, New York: de Gruyter, 209–231.
- Graefen, Gabriele (1999): Wie formuliert man wissenschaftlich? In: *Materialien Deutsch als Fremdsprache* 52, 222–239.
- Jasny, Sabine (2001): *Trennbare Verben in der gesprochenen Wissenschaftssprache und die Konsequenzen für ihre Behandlung im Unterricht für Deutsch als fremde Wissenschaftssprache*. Regensburg: Fachverband Deutsch als Fremdsprache.
- Kretzenbacher, Heinz (1998): Fachsprache als Wissenschaftssprache. In: Lothar Hoffmann, Hartwig Kalverkämper & Herbert Ernst Wiegand (Hrsg.), *Fachsprachen. Handbücher zur Sprach- und Kommunikationswissenschaft* (HSK) (Bd. 14.1). Berlin, New York: de Gruyter, 133–142.

- Kretzenbacher, Heinz (2010): Fach- und Wissenschaftssprachen in den Geistes- und Sozialwissenschaften. In: Hans-Jürgen Krumm, Christian Fandrych, Britta Hufeisen & Claudia Riemer (Hrsg.), *Deutsch als Fremd- und Zweitsprache. Handbücher zur Sprach- und Kommunikationswissenschaft* (HSK) (Bd. 35.1). Berlin, New York: de Gruyter, 493–501.
- Meißner, Cordula (2014): *Figurative Verben in der allgemeinen Wissenschaftssprache des Deutschen. Eine Korpusstudie*. Tübingen: Stauffenburg.
- Meißner, Cordula & Franziska Wallner (2016): Persuasives Handeln im wissenschaftlichen Diskurs und seine lexikografische Darstellung: das Beispiel der Kollokation Bild zeichnen. In: *Studia Linguistica* 35 (=Acta Universitatis Wratislaviensis No 3742), 235–252.
- Meißner, Cordula & Franziska Wallner (2019): *Das gemeinsame sprachliche Inventar der Geisteswissenschaften. Lexikalische Grundlagen für die wissenschaftspropädeutische Sprachvermittlung*. Berlin: Erich Schmidt Verlag.
- Projektgruppe „Fächerklassifikation und Thesauri“ (2014): Empfehlungen zur Revision der Fächersystematiken des Statistischen Bundesamtes (2014) Veröffentlichung im Rahmen des Projekts „Spezifikation Kerndatensatz Forschung“. Institut für Forschungs- information und Qualitätssicherung (iFQ) / Fraunhofer – Institut für Angewandte Informationstechnik (FIT) / Geschäftsstelle des Wissenschaftsrates.
http://www.kerndatensatz-forschung.de/version1/PGK_Empfehlungen_zur_Revision_der_Faechersystematiken_2014.pdf (28. 11. 2017).
- Redder, Angelika, Dorothee Heller & Winfried Thielmann (Hrsg.) (2014): *Eristische Strukturen in Vorlesungen und Seminaren deutscher und italienischer Universitäten. Analysen und Transkripte*. Heidelberg: Synchron.
- Schepping, Heinz (1976): Bemerkungen zur Didaktik der Fachsprache im Bereich des Deutschen als Fremdsprache. In: Dietrich Rall, Heinz Schepping & Walter Schleyer (Hrsg.), *Didaktik der Fachsprache. Beiträge zu einer Arbeitstagung der RWTH Aachen vom 30.9. bis 4. 10. 1974*. Bonn-Bad Godesberg: DAAD, 13–34.
- Schiller, Anne, Simone Teufel, Christine Stöckert, & Christine Thielen (1999): Guidelines für das Tagging deutscher Textcorpora mit STTS. (Kleines und großes Tagset). Universität Stuttgart, Institut für maschinelle Sprachverarbeitung; Universität Tübingen, Seminar für Sprachwissenschaft. <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf> (21. 12. 2016).
- Schmid, Helmut (1995): Improvements In Part-of-Speech Tagging With An Application To German. In: Institut für maschinelle Sprachverarbeitung (Stuttgart): *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger2.pdf> (28. 11. 2017).
- Statistisches Bundesamt (2013): *Bildung und Kultur. Studierende an Hochschulen – Fächersystematik*. <https://www.destatis.de/DE/Methoden/Klassifikationen/BildungKultur/StudentenPruefungsstatistik.pdf> (16. 10. 2014). [Das Dokument unterliegt einer regelmäßigen Aktualisierung. Unter dem angegebenen Link ist immer die neuste Fassung zugänglich. Auf die hier verwendete Fassung von 2013 kann nicht mehr zugegriffen werden.]
- Steinhoff, Torsten (2008): Kontroversen erkennen, darstellen, kommentieren. In: Ines Bons, Dennis Kaltwasser & Thomas Gloning (Hrsg.): *Fest-Platte für Gerd Fritz*. Gießen (13 Seiten). http://www.festschrift-gerd-fritz.de/files/steinhoff_2008_kontroversen_erkennen-darstellen-und-kommentieren.pdf (28. 11. 2017)

- Thielmann, Winfried (2009): *Deutsche und englische Wissenschaftssprache im Vergleich: Hinführen, Verknüpfen, Benennen*. Heidelberg: Synchron.
- Wallner, Franziska (2014): *Kollokationen in Wissenschaftssprachen. Zur lernerlexikographischen Relevanz ihrer wissenschaftssprachlichen Gebrauchsspezifika*. Tübingen: Stauffenburg.
- Weinrich, Harald (1989): Formen der Wissenschaftssprache. In: Akademie der Wissenschaften zu Berlin (Hrsg.), *Jahrbuch 1988 der Akademie der Wissenschaften*. Berlin: Akademie der Wissenschaften, 119–158.
- Wissenschaftsrat (2010): Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften. Drs. 10039–10. <http://www.wissenschaftsrat.de/download/archiv/10039-10.pdf> (28. 11. 2017).
- Zinsmeister, Heike (2015): Chancen und Grenzen von automatischer Annotation. In: *Zeitschrift für germanistische Linguistik* 43 (1), 84–111.