

ABEL, ANDREA / GLAZNIEKS, AIVARS (2015): *Wo Sprachkompetenzforschung auf Varietätenlinguistik trifft: Empirische Befunde aus dem Varietäten-Lernerkorpus „KoKo“*. In: LENZ, ALEXANDRA / AHLERS, TIMO / GLAUNINGER, MANFRED (Hrsgg.): *Dimensionen des Deutschen in Österreich. Variation und Varietäten im sozialen Kontext*. Frankfurt am Main: Peter Lang. 257-282.
[Pre-publication draft. Not for quotation or copying]

Wo Sprachkompetenzforschung auf Varietätenlinguistik trifft: Empirische Befunde aus dem Varietäten-Lernerkorpus „KoKo“

Andrea Abel / Aivars Glaznieks

1 Einleitung

Im Zuge der Analyse von Schülertexten werden häufig mehr oder weniger gravierende Abweichungen vom standardsprachlichen konventionellen Sprachgebrauch beschrieben (z. B. Wodak & Rheindorf 2011). Die Abweichungen können unterschiedlicher Art sein und reichen von lernerspezifischen Phänomenen zu allgemesprachlichen Erscheinungen, die in den Bereich des Sprachwandels einzuordnen sind. Lernerspezifische Phänomene sind nicht einheitlich und umfassen neben sprachlichen „Systemfehlern“ (Frage nach der Korrektheit sprachlicher Formen, z. B. grammatische Fehler) auch „Normfehler“ (Frage nach der Angemessenheit sprachlicher Formen, z. B. stilistische Varianten) auf unterschiedlichen Ebenen der linguistischen Beschreibung (Eisenberg 2007: 212, Schneider 2013: 30). Im L1-Lernerkorpus „KoKo“, das im Verlauf des Projekts „Bildungssprache im Vergleich“ seit 2010 aufgebaut wird,¹ können beispielsweise folgende von OberschülerInnen verfasste Sätze gefunden werden:²

(1) *Wir sind jeder ein Individuum für sich. (ID1571)*

(2) *Natürlich kann man sich nicht alles von Jugendlichen erwarten, sie befinden sich ja erst in einer Phase, in der sie solche Werte entwickeln sollen. (ID2798)*

1 am Institut für Fachkommunikation und Mehrsprachigkeit der EURAC (Bozen/Italien), vgl. http://www.korpus-suedtirol.it/bildungssprache_de.htm (Februar 2014)

2 Die Schreibweisen der Beispiele entsprechen dem Original; es wurden keine Fehler verbessert. Selbstkorrekturen, Tilgungen o. Ä. werden in den transkribierten Beispielen aus Gründen der Lesbarkeit allerdings nicht wiedergegeben.

Ausgehend von einer standardsprachlichen Norm ist der Satz in (1) grammatisch nicht korrekt gebildet. Das Personalpronomen im Plural kann keine Verbindung mit dem im Singular stehenden Indefinitpronomen eingehen, sie müssen im Numerus kongruent sein (vgl. Duden 2005: 325, 1003–1004). Der Satz in Beispiel (2) hingegen wird wahrscheinlich nur von einem Teil der Sprechergemeinschaft als ungrammatisch beurteilt werden. Die Verwendung des Verbs *erwarten* in Verbindung mit einem Reflexivpronomen mag bundesdeutschen und Schweizer SprecherInnen ins Auge fallen, kann aber mit Ziegler (2010) für das österreichische Deutsch als „Tendenz zum vermehrten Gebrauch des Reflexivpronomens“ (67) beschrieben werden, die sich in der österreichischen Standardvarietät manifestiert hat. Beispiel (2) illustriert folglich ein Sprachwandelphänomen, das die österreichische Standardsprache betrifft, Beispiel (1) hingegen einen grammatischen Fehler, der sowohl die bundesdeutsche als auch die österreichische Standardsprache betreffen dürfte.

Neben grammatischen Besonderheiten finden sich in „KoKo“ auch Auffälligkeiten auf anderen sprachlichen Ebenen:

- (3) *Man kann nicht alle Jugendlichen in einem Topf werfen. (ID1587)*
- (4) *Mann kann nur sehr schlecht alle Jugendlichen in den gleichen Topf schmeißen. (ID2603)*
- (5) *Jedoch stellt sich die Frage, kann man wirklich alle Jugendlichen in einen Topf stecken? (ID2658)*
- (6) *Der Aussage des Deutschen kann ich nicht ganz zu stimmen, da er alle Jugendlichen unter einen Hut steckt. (ID1762)*
- (7) *Aber oft ist es schwierig etwas durchzusetzen, denn häufig wird die Jugend von älteren Menschen nicht ernst genommen oder es werden alle jungen Menschen in eine Schublade gesteckt. (ID2935)*

Richtet man den Blick ausschließlich auf den Bereich des Wortschatzes, ist festzustellen, dass in den Beispielsätzen (3) bis (5) eine formelhafte Sequenz, und zwar ein metaphorisches Idiom (Burger 2010: 63), in unterschiedlichen Varianten realisiert wird. Es handelt sich jeweils um lexikalische Variation, die das verwendete Verb betrifft. Nimmt man eine Standard-Grundform *alles in einen Topf werfen* an (vgl. Duden 2008: 777, ÖWB 2006: 660), stellen die Sätze (4) und (5) Abweichungen davon dar, die allein aufgrund der semantischen Ähnlichkeit der Verben *werfen*, *stecken* und *schmeißen* (‘etwas von A nach B bewegen’) durchaus in vielen Kontexten akzeptiert werden könnten. In Beispiel (6) weist die verwendete formelhafte Sequenz neben ihrer formalen Abweichung zur Standard-Grundform *alles unter einen Hut bringen* eine semantische Auffälligkeit auf: Die ursprüngliche Bedeutung (‘in Übereinstimmung, in Einklang bringen’, Duden 2008: 381, ‘koordinieren’, ÖWB 2006: 324) kann in dieser Variante nicht mehr angenommen werden; anscheinend verwendet der oder die SchreiberIn diese Form in der Bedeutung ‘alle(s) gleich beurteilen, ohne bestehende Unterschiede zu berücksichtigen’, also ähnlich der Bedeutung von *alles in einen Topf werfen*. Beispiel (7) schließlich weist keinerlei formale Auffälligkeit auf, hier sind die Abweichungen rein semantischer Natur. *Schublade* wird als

Metapher in der Bedeutung ‘Kategorie’ verwendet, die entweder bereits definiert wurde oder in irgendeiner Weise definit erscheint (i. S. einer bestimmten Schublade). Nur dann entsteht die Bedeutung ‘zu Unrecht in eine bestimmte Kategorie einordnen’. In (7) drängt sich jedoch eine ähnliche Lesart wie in (3) bis (6) auf.

In Schülertexten ist es oftmals nicht sofort ersichtlich, wie solche und ähnliche Abweichungen vom Standard eingestuft werden müssen und welche Ursachen sie haben. Die Abweichungen können einerseits auf mangelnde Kenntnisse der SchreiberInnen über die Zielsprache (hier: Standardsprache, bildungssprachliches Register) zurückzuführen sein („Kompetenzfehler“, vgl. Schneider 2013: 30) und wären damit lernerspezifische Phänomene. Sie können andererseits auch allgemeine Veränderungen im Gebrauch der Sprache widerspiegeln und müssten gegebenenfalls als Reproduktionen eines allgemeinen Sprachwandelphänomens beschrieben werden. Zusätzlich muss beachtet werden, dass die deutsche Sprache eine plurizentrische Sprache ist, die mehrere Voll- und Halbzentren aufweist (vgl. Ammon et al. 2004). In jedem Zentrum sind mehr oder weniger unterschiedliche Standards akzeptiert. Bei kontrastiven Untersuchungen verschiedener Voll- und Halbzentren können systematische Abweichungen von einem Standard (z. B. dem bundesdeutschen) prinzipiell auch Hinweise auf Eigenschaften eines anderen Standards (z. B. den österreichischen) liefern und folglich diatopische Varianz anzeigen. Schließlich muss berücksichtigt werden, dass SchülerInnen bei der Textproduktion auch z. B. Flüchtigkeitsfehler („Performanzfehler“, vgl. Schneider 2013: 30) unterlaufen, die im Text zurückbleiben und nicht eindeutig von Kompetenzfehlern unterschieden werden können.³

Im Folgenden wird ein Verfahren vorgestellt, mit dem das Varietäten-Lernerkorpus „KoKo“ systematisch nach Auffälligkeiten durchsucht und diese u. a. im Hinblick auf Lernerspezifik, Diatopik und Sprachwandel analysiert werden. Das Verfahren wird – im Anschluss an eine Einführung in den Bereich der formelhaften Sequenzen (Kapitel 2) – anhand einer Beispielanalyse der Vorkommnisse des Phrasems *alles in einen Topf werfen* sowie formal und semantisch ähnlicher Phraseme erläutert (Kapitel 3). In Kapitel 4 werden die Vorkommnisse der untersuchten Phraseme hinsichtlich lernerspezifischer, diatopischer und sprachwandelbezogener Auffälligkeiten diskutiert. Es zeigt sich, dass für alle Erklärungsmodelle Kandidaten im Lerner-Varietätenkorpus „KoKo“ gefunden werden können. Die Ergebnisse der Korpusanalyse werden abschließend (Kapitel 5) zusammengefasst.

2 Formelhafte Sequenzen

Unter der Bezeichnung *formelhafte Sequenzen* werden „mehrmorphemische Sequenzen“ subsumiert, die aus psycholinguistischer Perspektive „nicht mittels Regeln konstruiert werden, sondern – wie ein einzelnes Lexem – als Ganzes abgerufen werden“ (Aguado 2002: 30). Im Allgemeinen manifestieren sich formelhafte Sequenzen in

3 vgl. zur Problematik der Unterscheidung von Produktionsfehlern und regionalen Besonderheiten in der Sprachproduktion im Erstspracherwerb u. a. Katerbow et al. 2011

„strings of linguistic items where the relation of each item to the rest is relatively fixed, and where the substitutability of one item by another of the same category is relatively constrained” (Perkins 1999). Die Eigenschaft der „Festigkeit“ lässt, obwohl die Sequenz als Ganzes wahrgenommen wird, diskontinuierliche Formen sowie grammatische, lexikalische u. a. Varianten einer Sequenz zu (u. a. Wray & Perkins 2000, Heid 2008: 344–345, Burger 2010: 25–27). *Formelhafte Sequenz* wird in dieser Arbeit als Oberbegriff verwendet, mit welchem alle Formen fester polylexikalischer Einheiten bezeichnet werden. Dazu gehören etwa Kollokationen, Routineformeln, metaphorische Idiome u. a. (vgl. Burger 2010).

Der so definierte Untersuchungsgegenstand nimmt den Großteil sprachlicher Äußerungen ein. Schätzungen zufolge sind etwa 70% der normalerweise von Muttersprachlern verwendeten Sprache formelhaft (Wray & Perkins 2000: 1–2). Formelhafte Sequenzen sind daher ein äußerst relevanter Bereich für viele linguistische Teildisziplinen. Ihre Relevanz sowohl für die Varietäten- und Korpuslinguistik als auch für die Untersuchung von Lernersprachen wird im Folgenden dargestellt.

Varietätenlinguistik und formelhafte Sequenzen

Formelhafte Sequenzen werden im Rahmen der europäischen Phraseologieforschung auch aus varietätenlinguistischer Perspektive untersucht (z. B. Korhonen et al. 2010). Dabei stehen meist idiomatisierte, formelhafte Sequenzen im Mittelpunkt des Interesses (u. a. Földes 1996, Ernst 2007), sogenannte „Idiome“ (vgl. Burger 2010: 15).

Prinzipiell kann diatopische Varianz im Bereich der Phraseologie auf allen Ebenen der linguistischen Beschreibung auftreten. Földes (1996) hat dies ausführlich am Beispiel des österreichischen Deutsch⁴ gezeigt, das er in seiner Untersuchung mit der bundesdeutschen Varietät kontrastiert. Er findet Unterschiede auf folgenden Ebenen:

1. Phonetik / Prosodie: *etw. springt jmdm. ins Aug‘, für etw. g‘rade stehen*
2. Wortbildung: *(genau sein) bis aufs I-Tüpfel, im Büchel stehen*
3. Morphosyntax:
 - a. Genus: *zerrinnen wie der Butter in der Sonne, in ein Eck stellen*
 - b. Präpositionen: *am Prüfstand stehen, am Zahnfleisch gehen*
 - c. Valenz: *über den Durst trinken (bdt. *etw. über den Durst trinken*)*
 - d. Verbflexion: *sein-Perfekt bei *stehen, liegen, sitzen**
 - e. Kasus- und Deklinationsunterschiede: *jmdm. ins Knie ficken (bdt. *jmdn. ins Knie ficken*)*
4. Lexikon:
 - a. territoriale Doubletten: *einen Knödel im Hals haben (ndt. *einen Kloß im Hals haben*)*

4 Földes räumt ein, dass in seinem Untersuchungskorpus nicht immer eindeutig zu entscheiden war, „ob ein Phraseologismus in der österreichischen nationalen Standardvarietät anzusiedeln ist, oder diatopischen, diastratischen bzw. diaphasischen Restriktionen unterliegt.“ (1996: 33, Fußnote 26). Die folgenden Beispiele nähren den Verdacht, dass es sich nicht ausschließlich um die österreichische Standardvarietät handelt.

- b. Varianz innerhalb von Wortfeldern: *jmdm. ins Gäu kommen* (bdt. *jmdm. ins Gehege kommen*)
- c. völlig verschiedene Konstituenten: *keinen Tau von etw. haben* (bdt. *keinen Schimmer von etw. haben*)
- d. ohne Entsprechung: *(etw.) in Evidenz halten*
- e. Regionale Bedeutungsvarianz eines Phraseologismus: *deutsches Eck*

5. Orthographie: sein Placet (bdt. Plazet) geben

Die von Földes angeführten Ebenen der Varianz entsprechen im Wesentlichen den Ebenen, die Schmidlin (2011: 72) allgemein für die intralinguale Varianz der plurizentrischen Sprache Deutsch aufführt. Die diatopische Varianz, die man im Bereich der Phraseologie nachweisen kann, reflektiert letztlich die allgemein für die jeweilige Varietät festzustellenden regionalen Besonderheiten (vgl. Burger 2010: 211).

Für die schweizerdeutsche Phraseologie müssen besonders die Arbeiten von Burger (1996) und Schmidlin (2007) beachtet werden. Sie geben einen guten Überblick über die Besonderheiten der schweizerdeutschen Varietät. Für das Bundesdeutsche gibt es keine verlässliche Sammlung, die einen Überblick über solche Phraseologismen gibt, die ausschließlich in Deutschland verwendet werden. Piirainen (2006) weist darauf hin, dass die Sammlung des Duden-Verlags (Duden 2008) lediglich regionale Besonderheiten für Österreich und die Schweiz markiert. Im Umkehrschluss allerdings gelten die unmarkierten Phraseologismen des Werks häufig nicht für den gesamten deutschen Sprachraum, sondern geben teilweise rein bundesdeutsche Wendungen wieder (vgl. auch Polenz 1999: 123).

Eine problematische Fragestellung, die eng mit der Beurteilung von Variation bei formelhaften Sequenzen in der gesprochenen und geschriebenen Sprache zusammenhängt, ist: Wie erkennt man diatopische Varianz und wie kann man sie von Fehlern unterscheiden? In idiomatischen Wörterbüchern ist häufig nur eine Variante aufgeführt, sodass jede Abweichung davon eine nicht-lexikalisierte Variante derselben darstellen würde. Burger (2000: 42) vermutet, dass in der Sprachverwendung sehr viel mehr Varianz verbreitet ist und akzeptiert wird, als es die Wörterbücher suggerieren (vgl. Gehweiler et al. 2007). Um diesen Umstand zu berücksichtigen, muss anhand von spezifischen Varietätenkorpora datengeleitet gearbeitet und die Verwendung von Phraseologismen an authentischem Material untersucht werden.⁵

Korpuslinguistik und formelhafte Sequenzen

Für die Untersuchung der Verwendung formelhafter Sequenzen aller Art sind linguistisch aufbereitete Korpora heutzutage eine unersetzliche Datenquelle. Korpora können auf unterschiedliche Weise nach formelhaften Sequenzen untersucht werden. Sie können einerseits nur auf das Vorkommen einer bestimmten Sequenz, etwa eines Phraseologismus durchsucht werden. Die gängigen Zusatzannotationen nach Lemma und Wortart ermöglichen es dabei, potenzielle lexikalische oder morphosyntaktische Varianten des Phraseologismus bei der Suche nach diesem zu berücksichtigen. Je

5 Aspekte bewusst als Stilmittel eingesetzter Abweichungen werden im Rahmen dieses Beitrags nicht berücksichtigt.

stärker bei der Formulierung der Suchanfrage abstrahiert wird, desto eher kann andererseits mit Hilfe von Sequenzmustern nach formelhaften Sequenzen gesucht werden, die diesem Muster folgen, ohne die einzelnen Stellen des Musters mit einzelnen Wörtern und Lemmata füllen zu müssen (z. B.: V PP, i. e. Verbalphraseme mit Präpositionalphrase).

Bei der direkten Suche nach den Vorkommnissen einer bestimmten formelhaften Sequenz können mit Hilfe geeigneter Suchanfragen alle Textstellen gefunden werden, in denen die gesuchte Sequenz in ihrer Grundform oder einer Variante davon vorkommt (z. B. Fellbaum 2007). Bei der Suche nach formelhaften Sequenzen mit Hilfe von Sequenzmustern hingegen werden auf der Basis statistischer Häufigkeiten Listen von möglichen formelhaften Sequenzen bezogen auf *ein* Ziel-Item erstellt. Dazu wird kontrolliert, wie häufig das Ziel-Item in einem vordefinierten Kontext mit einem anderen Item gemeinsam auftritt. Die Ergebnislisten mit möglichen Kandidaten für eine formelhafte Sequenz müssen anschließend manuell evaluiert werden, wobei jeweils entschieden werden muss, ob es sich tatsächlich um eine formelhafte Sequenz handelt oder nur um die Kookurrenz zweier Items, die aus anderen Gründen (z. B. thematischen) zustande gekommen ist (Sailer 2007: 1065–1068). Spezifische Programme unterstützen zudem das automatische Auffinden von formelhaften Sequenzen in einem Korpus (z. B. Weller & Heid 2010, Vatvedt Fjeld et al. 2010).⁶ In einigen Untersuchungen werden die eben beschriebenen Methoden nicht alternativ, sondern ergänzend angewendet (z. B. Quasthoff et al. 2010). Ein kombiniertes Verfahren eignet sich sehr gut dazu, ein vollständiges Bild von der Verwendung formelhafter Sprache in einem bestimmten Korpus zu erhalten. Problematische Aspekte bei der korpuslinguistischen Untersuchung von formelhaften Sequenzen betreffen vor allem die Abhängigkeit der Suchergebnisse von der Korpusgröße sowie von der absoluten und relativen Häufigkeit bestimmter formelhafter Sequenzen (vgl. Moon 2007, Sailer 2007: 1062–1063). In Textsammlungen sind Idiome und Sprichwörter beispielsweise weit weniger häufig aufzufinden als Kollokationen (Moon 2007: 1050). Daher ist für eine Untersuchung des Gebrauchs weniger häufiger formelhafter Sequenzen wie Idiome eine Mindestkorpusgröße entscheidend. Moons Analyse der Vorkommen von Idiomen und Sprichwörtern in englischsprachigen Korpora bringt sie zu dem Fazit, „that at least 100 million words is an appropriate size of corpus for work on idioms and proverbs, with 50 million words being a minimum“ (Moon 2007: 1050).

Abfragbare und annotierte Lernerkorpora sind normalerweise nicht in dieser Größe vorhanden. Das TIGER-Korpus beispielsweise umfasst ca. 900.000 Wörter. Sailer gibt die recht pessimistische Einschätzung, dass solche Korpora kaum für die Phraseologieforschung verwendbar seien (Sailer 2007: 1062). Bevor dieser Punkt erneut aufgegriffen wird, werden Besonderheiten der Verwendung formelhafter Sequenzen in Lernaltersprachen vorgestellt.

6 Vgl. Bubenhofer & Ptashnyk (2010: 13) für einen Überblick über Software zum Auffinden von formelhaften Sequenzen in Textkorpora sowie für einen Überblick über weitere Methoden der (halb)automatischen Phrasemextraktion.

Lernersprachen und formelhafte Sequenzen

Formelhafte Sequenzen sind ein wesentlicher Bestandteil einer Sprache und tragen zur ökonomischen Bewältigung von Kommunikationssituationen bei. Gleichzeitig signalisieren sie, bewusst oder unbewusst, auch Zugehörigkeit zu einer bestimmten Sprechergemeinschaft, was unter anderem im Kontext der Spracherwerbsforschung und der Beschreibung von Lernersprachen beachtet und rege diskutiert wird (vgl. Aguado 2002, Wray 2000). Der Begriff *Lernersprache* wird meist für nicht-muttersprachliche Sprecher- und SchreiberInnen verwendet, so auch in der Korpuslinguistik (Granger 2008). Wenn im Folgenden von *LernerInnen* bzw. *Lernersprache* gesprochen wird, sind damit allerdings L1-LernerInnen bzw. deren Erstsprache(n) gemeint. L1-LernerInnen sind solche, die ihre Erstsprache(n) oder wesentliche Teile davon, wie etwa Schreib- und Textkompetenzen, erlernen.

Im Erstspracherwerb werden formelhafte Sequenzen ähnlich wie monolexikalische Elemente des Lexikons erlernt; dabei entspricht die Bedeutung der formelhafte Sequenzen zunächst oft nicht der konventionellen Bedeutung der Sprachgemeinschaft. Bedeutungen sind bei der Begegnung mit neuen Wörtern und formelhafte Sequenzen häufig noch nicht fest und können noch über Jahre einer Wandlung unterliegen (vgl. Häcki Buhofer 2007, Glaznieks 2011, McGregor et al. 2002). Das Verstehen und Lernen phraseologischer Verbindungen betrifft generell alle sprachlichen Entwicklungsstufen (s. Häcki Buhofer 2007: 867). In schriftlichen Texten können formelhafte Sequenzen für Lernende eine Fehlerquelle darstellen. Die Fehler zeigen sich in Abweichungen sowohl von der konventionellen Form als auch von der konventionellen Bedeutung. Diese Beobachtung wurde nicht nur bei SchülerInnen gemacht (u. a. Eichler 2004, Margewitsch 2006, Wodak & Rheindorf 2011), sondern gilt auch für das studentische Schreiben (z. B. Ortner 2006). Typische Fehler entstehen einerseits durch grammatische und syntaktische Normverletzungen, andererseits durch Missachtung von stilistischen bzw. semantischen Selektionsbeschränkungen (z. B. durch formale oder konzeptuelle Kontamination mehrerer formelhafte Sequenzen), durch Missachtung kontextueller Gebrauchsbeschränkungen (z. B. durch Registerwechsel) oder durch fehlende inhaltlich-funktionale Angemessenheit (bzgl. Informativität, Präzision z. B. mittels Klischees) (vgl. Margewitsch 2006). Ausgelöst werden sie häufig durch hyperkorrekte Formulierungsversuche in der Bildungssprache (Ortner 2006).

Im schulischen Kontext werden formelhafte Sequenzen meist in die linguistische Bewertung von geschriebenen Texten einbezogen, sei dies nun als Fehler, als wenig spezifische, „auffällige“ Verwendungsweise in Texten, die weder positiv noch negativ ist (vgl. Dürscheid et al. 2010), oder als Bewertungskriterium im diachronen Vergleich von Schülertexten (vgl. Wodak & Rheindorf 2011). Im studentischen Kontext werden formelhafte Sequenzen gerne unter dem Aspekt eines zu erlernenden wissenschaftlichen Stils berücksichtigt, dessen fester Bestandteil sie zweifellos sind und daher von den wissenschaftlichen Schreibnovizen erlernt werden müssen (vgl. Steinhoff 2007, Wallner 2010).

Die Untersuchung von Lernaltersprachen hat das Ziel, lerner-typische Phänomene sichtbar und auf diese aufmerksam zu machen. Die Erkenntnisse solcher Untersuchungen können anschließend in die didaktische Aus- und Fortbildung fließen, so dass die Ergebnisse letztlich den Lernenden zugutekommen.

Lernerkorpora und sprachliche Varietäten

Anhand abfragbarer und linguistisch annotierter Lernerkorpora können lerner-typische Phänomene auf unterschiedlichen sprachlichen Ebenen mit Unterstützung von Abfragetools systematisch untersucht werden (z. B. Schröder-Lenzen & Henn 2009). Generell dient die Analyse von Lernertexten, im spezifischen Fall von Schüler-texten, jedoch nicht nur dem Auffinden lerner-typischer Phänomene. Oftmals zeigen sich im diachronen Vergleich Veränderungen der sprachlichen Realität (z. B. Sieber 1998). Wodak & Rheindorf (2011) vermuten generell, dass Schülertexte ein Fundus für Sprachwandelphänomene sind. Zwar müssen sich die SchülerInnen an standardsprachliche Normen halten, wodurch ein „massive[r] Druck gegen Veränderungen“ (2011: 13) aufgebaut wird, andererseits kann eine Analyse des Gelingens bzw. Nicht-Gelingens dieser Anforderungen auch auf Phänomene des sprachlichen Wandels hindeuten.

Zur Beschreibung sprachlicher Varietäten jenseits lerner-typischer Auffälligkeiten, z. B. von diachroner, aber auch diatopischer o. ä. Varianz insbesondere auf der Ebene der Syntagmatik wurde das Potenzial von L1-Lernerkorpora und korpuslinguistischer Ansätze unseres Wissens bisher unzureichend ausgeschöpft.⁷ Die im Folgenden vorgestellten Analysen sollen aufzeigen, inwiefern spezifische Lernerkorpora diesbezüglich nützlich sein und einen Beitrag zur varietätenlinguistischen Forschung leisten können.

3 Korpusanalyse

Das L1-Lernerkorpus „KoKo“: Zahlen und Fakten

Das L1-Lernerkorpus „KoKo“ ist aus dem Projekt „Bildungssprache im Vergleich: korpusunterstützte Analyse der Sprachkompetenz bei Lernenden im deutschen Sprachraum (KoKo)“ entstanden, das seit 2010 am Institut für Fachkommunikation und Mehrsprachigkeit der Europäischen Akademie Bozen gemeinsam mit Partnern an der Bildungswissenschaftlichen Fakultät der Freien Universität Bozen und dem Institut für Germanistik der Universität Innsbruck durchgeführt wird. Das Projekt ist Teil der Initiative „Korpus Südtirol“, im Zuge welcher kontinuierlich Südtiroler Texte gesammelt und ein abfragbares Korpus aufgebaut werden (siehe z. B. Anstein et al. 2011, Abel & Anstein 2011).⁸

Das gesamte Lernerkorpus „KoKo“ umfasst 1503 Texte und hat eine Größe von 811.330 Tokens (ohne Satzzeichen, Stand: Dezember 2012). Davon stammen 1319

7 Zur gesprochenen Sprache vgl. u. a. Katerbow (2013).

8 Siehe <http://www.korpus-suedtirol.it>. (Februar 2014)

Texte von SchülerInnen mit der Erstsprache Deutsch. Die Rahmenbedingungen für die Textproduktion waren für alle teilnehmenden SchülerInnen gleich. Die standardisierte Vorgabe beinhaltete, dass die Aufgabe als schriftliche Arbeit während des Unterrichts bearbeitet werden musste. Als Mittel der Motivation sollten die Lehrkräfte gelungene Texte positiv in die Leistungsbewertung der SchülerInnen einfließen lassen. Alle Texte wurden zu ein und demselben Thema verfasst. Die Aufgabe bestand darin, zu einem Zitat von Hans Magnus Enzensberger über die Jugend einen Erörterungsaufsatz zu verfassen, in dem auch zum Thema persönlich Stellung genommen werden sollte. Die Schreibzeit betrug 120 Minuten, als Hilfsmittel waren Wörterbücher zugelassen.

Die Texte wurden von OberschülerInnen ein Jahr vor der Hochschulreife geschrieben. 89% davon waren zum Erhebungszeitpunkt zwischen 17 und 19 Jahre alt, 7% gaben an, älter als 19 Jahre zu sein, von etwa 4% fehlt diese Angabe. Die Erhebung fand an allgemeinbildenden und berufsqualifizierenden Oberschulen in drei Regionen des deutschen Sprachraums, und zwar in Nordtirol, Südtirol und Thüringen, statt. Die Teilnehmerauswahl erfolgte pro Region über eine stratifizierte Zufallsstichprobe mit den Schichtungsmerkmalen *Schultyp* und *Gemeindegröße*. Folgende regionale Zusammensetzung hat sich unter den genannten Vorgaben ergeben:

Tabelle 1: Zusammensetzung des Lernerkorpus „KoKo“

<i>Subkorpus (nach Region)</i>	<i>Gesamtanzahl</i>	<i>L1 Deutsch</i>
Nordtirol (NT):	233.098 Tokens (457 Texte)	206.439 Tokens (404 Texte)
Südtirol (ST):	222.209 Tokens (520 Texte)	192.891 Tokens (451 Texte)
Thüringen (TH):	353.674 Tokens (521 Texte)	317.075 Tokens (464 Texte)
ohne Angabe	2.349 Tokens (5 Texte)	---
gesamt	811.330 Tokens (1503 Texte)	716.405 Tokens (1319 Texte)

Zusätzlich zu den Texten ist eine Reihe von Informationen über die SchreiberInnen bekannt, die über einen Fragebogen erfasst und als Metadaten in das Korpus aufgenommen wurden. Im Korpus (Stand: Dezember 2012) befinden sich Informationen zur Muttersprache, Geschlecht, Angaben zum Schultyp, zur Deutschnote des vergangenen Schuljahrs bezogen auf den Erhebungszeitpunkt und zur Gemeindegröße. Zukünftig werden Angaben zum sozio-ökonomischen Hintergrund, Angaben zur Sprachbiographie (z. B. zum Sprach- und Dialektgebrauch in der Familie) sowie zu Gewohnheiten der Mediennutzung (u. a. Lese- und Schreibgewohnheiten, Medienkonsum) der SchreiberInnen hinzukommen.

Im Rahmen des Projekts werden die Texte in einer Kombination aus quantitativen, (halb-)automatischen korpuslinguistischen und qualitativen linguistischen Methoden nach orthographischen, grammatischen, lexikalischen und textuellen Merkmalen analysiert und miteinander verglichen. Die Metadaten werden in die Interpretation einbezogen und ermöglichen Aussagen darüber, welche Faktoren die Verteilung der Schreibkompetenz (mit-)bestimmen.

Die forschungsleitenden Fragen für die Analyse der Texte lauten:

- a) Welche Merkmale weisen Schülertexte (im bildungssprachlichen Kontext) auf?
- b) Welche Arten von Merkmalen können beschrieben werden?
- c) Welche Faktoren (Schultyp, Region, Geschlecht, sozio-ökonomischer Hintergrund, sprachliche Biographie etc.) bestimmen die Verteilung der Merkmale?

Methoden der Korpusanalyse

In Kapitel 2b wurden bereits mehrere Methoden beschrieben, die in der Korpuslinguistik zur Untersuchung von formelhaften Sequenzen in Textkorpora verwendet werden. Obwohl die Nützlichkeit kleiner Korpora für die Phraseologieforschung mitunter kritisch betrachtet wird (vgl. Sailer 2007: 1062), wird für die folgende Analyse dennoch ein relativ kleines, aber dafür inhaltlich sehr homogenes Korpus herangezogen. Dies erhöht die Vorkommenswahrscheinlichkeit thematisch naher Phraseologismen. Außerdem wurde ein kombiniertes Verfahren aus automatischen Suchanfragen nach struktureller Vorgabe („Konsultation“, vgl. Sailer 2007) und manueller Analyse einzelner formelhafter Sequenzen („Analyse“, vgl. Sailer 2007) angewandt. Dies hat den Vorteil, dass durch die automatische Extraktion möglicher Sequenzkandidaten ein umfassendes Bild der verwendeten formelhaften Sequenzen entstehen kann. Dabei können bereits Frequenzanalysen einbezogen werden, wodurch besonders häufig verwendete Sequenzen sehr schnell erkennbar werden. Da diese Methode nicht dazu geeignet ist, alle Varianten einer Sequenz in einem Korpus zu erfassen, muss sie durch eine anschließende Analyse der Vorkommnisse eines oder mehrerer bestimmter Sequenzen ergänzt werden.

Folgende Schritte sind bei dieser Vorgehensweise notwendig:

1. automatische Extraktion möglicher Kandidaten (mit Strukturvorgabe) für eine formelhafte Sequenz;
2. manuelle Analyse und Auswahl formelhafter Sequenzen;
3. Sortierung der Kandidatenliste im Hinblick auf die getroffene Auswahl, um alle Varianten einer formelhaften Sequenz in der Liste einer Grundform zuzuweisen (diese Sortierung richtet sich nach dem inhaltstragenden Kern des Phrasems);
4. Reformulierung der Suchanfrage, um weitere Varianten, die nicht in der Kandidatenliste sind, zu finden;
5. Verifizierung der formelhaften Verwendung an der jeweiligen Korpusfundstelle.

Varianz ist ein Merkmal, anhand dessen die Verwendung formelhafter Sequenzen beschrieben werden kann (vgl. forschungsleitende Frage b). Bei der Interpretation der Varianz müssen sowohl lernerspezifische Ursachen (z. B. Fehler) als auch die Metadateninformationen berücksichtigt werden (vgl. forschungsleitende Frage c). Aufgrund der Zusammensetzung des Korpus können sowohl lernerspezifische Merkmale als auch diatopische, diaphasische u. a. Merkmale erwartet werden. Außerdem kann das Korpus nach Indizien für mögliche Sprachwandelphänomene durchsucht werden.

Um jede sprachliche Auffälligkeit in eine bestimmte Kategorie einordnen zu können, müssen mögliche alternative Erklärungen ausgeschlossen werden. Damit nicht für jedes Phänomen eine eigene Erklärungshypothese formuliert werden muss, wird zunächst für alle Auffälligkeiten angenommen, sie seien lernertypische Phänomene (H_0), die gegen Alternativannahmen getestet werden müssen (H_1, H_2, \dots, H_n). Die Grundannahme lautet demzufolge:

- Das sprachlich auffällige Merkmal x im Lernerkorpus „KoKo“ ist ein Lernerphänomen ($= H_0$).

Mögliche Alternativannahmen lauten:

- H_1 : Das sprachlich auffällige Merkmal x im Lernerkorpus „KoKo“ ist ein diatopisches Phänomen.
- H_2 : Das sprachlich auffällige Merkmal x im Lernerkorpus „KoKo“ ist ein Sprachwandelphänomen.
- (...)
- H_n : Das sprachlich auffällige Merkmal x im Lernerkorpus „KoKo“ ist ein Y -Phänomen.

Sprachlich auffällige Merkmale werden mit Bezug auf normsetzende Instanzen (vgl. Ammon 2005) als solche ausgezeichnet. Als Sprachkodizes werden das Deutsche Universalwörterbuch (Duden 2001), der Duden der Redewendungen (Duden 2008) und das Österreichische Wörterbuch (ÖWB 2006) herangezogen, außerdem als Modelltexte die Korpusressourcen des Digitalen Wörterbuchs der deutschen Sprache (DWDS, Wortprofile 3.0) sowie Subkorpora aus dem Deutschen Referenzkorpus (DeReKo) und dem Korpus Südtirol. Schließlich wird auf die Kompetenz linguistischer Fachexperten aus dem Projektteam zurückgegriffen.

H_1 wird zunächst nur am „KoKo“-Korpus getestet und, falls sie von den Daten bestätigt wird, anhand von geeigneten Referenzkorpora auf ihre Allgemeingültigkeit überprüft. Dieses Vorgehen gilt für alle Hypothesen, die eine Einschränkung von H_0 erlauben (z. B. $x =$ ein diatopisch markiertes Lernerphänomen). Die Bestätigung einer Hypothese wie H_2 schließt die Gültigkeit der Grundannahme aus. H_2 muss direkt an geeigneten Referenzkorpora getestet werden.

Ergebnisse

Im ersten Schritt wurden alle 1503 Texte des „KoKo“-Korpus mithilfe eines Kollokationsextraktors (vgl. Weller & Heid 2010) automatisch analysiert („Konsultation“). Die Extraktion unterscheidet nicht zwischen Kollokationen und Idiomen, sondern sucht allgemein nach Mehrworteinheiten, die einem vorher festgelegten syntaktischen Muster entsprechen. Tabelle 2 gibt die häufigsten Muster unter Berücksichtigung aller potenziell auftretenden Stellen in den jeweiligen Phrasen des Suchmusters *Präpositionalphrase + Verbalkomplex (PP+VV)* wieder.

Tabelle 2: Liste potenzieller formelhafter Sequenzen mit der Struktur Präpositionalphrase + Verbkomplex im Lernerkorpus „KoKo“

Freq.	Präp.	Det.	attr. Adj.	Nomen	Verb	Num.	fS
95	in	indef	-	Interview	sagen	Sg	?
85	aus	-	-	Fehler	lernen	Pl	X
57	in	-	voll	Zug	genießen	Pl	X
49	in	indef	-	Interview	äußern	Sg	?
48	mit	-	-	Freund	treffen	Pl	X
45	in	def	-	Diskotheke	gehen	Sg	?
43	zu	-	-	Schule	gehen	Sg	X
43	in	-	-	Kopf	haben	Sg	X
43	aus	poss	-	Fehler	lernen	Pl	X
40	als	-	-	Jugendlich	haben	PlSg	--
37	in	indef	-	Topf	werfen	Sg	X
35	in	def	-	Jugend	machen	Sg	--
34	in	def	-	Jugend	haben	Sg	--
32	auf	def	-	Suche	sein	Sg	X
31	mit	def	-	Wochenzeitung	sagen	Sg	--
0	in	dem	-	Zeit	machen	Sg	--
30	in	def	-	Diskotheke	verbringen	Sg	--
29	in	def	-	Schule	gehen	Sg	X
29	in	def	-	Disko	gehen	Sg	?
28	nach	-	-	Haus	kommen	Sg	?
28	in	-	-	Leben	stehen	Sg	?
25	in	def	-	Lage	sein	Sg	X
24	in	-	-	Diskotheke	gehen	Pl	?
24	auf	-	eigen	Bein	stehen	Pl	X
23	in	-	-	Vordergrund	stehen	Sg	X
23	in	-	-	Leben	haben	Sg	--
22	über	def	-	Jugend	sprechen	Sg	--
22	in	poss	-	Jugend	machen	Sg	--
22	aus	poss	-	Leben	machen	Sg	?
21	in	indef	-	Interview	behaupten	Sg	?
20	zu	-	-	Thema	äußern	Sg	?

20	in	def	-	Jugend	geben	Sg	--
20	als	quant	benei- denW	Phase	bezeich- nen	Sg	--
(...)							

Im nächsten Schritt muss nun diese Liste möglicher Kandidaten für formelhafte Sequenzen manuell auf tatsächliche formelhafte Sequenzen überprüft werden. Die Spalte *fS* in Tabelle 2 zeigt das Ergebnis dieser manuellen Analyse (Schritt 2). Einige vom Programm vorgeschlagene Kandidaten können mit Hilfe der o. g. Norminstanzen als formelhafte Sequenz markiert werden (X), andere nicht eindeutig (?) und müssen weiter überprüft werden, zum Beispiel dann, wenn sich die Norminstanzen widersprechen. Wieder andere werden nicht als formelhafte Sequenz gekennzeichnet (--), da sie weder als solche nachweisbar sind noch als solche wahrgenommen werden. Die eindeutigen Fälle umfassen in dieser Liste Kollokationen (*im Vordergrund stehen*) und Idiome (*in vollen Zügen genießen*). Im Weiteren wird das Vorgehen am Beispiel des metaphorischen Idioms *alles in einen Topf werfen* verdeutlicht.

Aus Tabelle 2 können zunächst alle Vorkommnisse mit dem Bestandteil *Topf* sortiert und extrahiert werden (Schritt 3). Da im weiteren Verlauf der Analysen nur Daten von OberschülerInnen mit der L1 Deutsch berücksichtigt werden, werden während dieses Schrittes alle Vorkommnisse, die nicht von SchreiberInnen mit L1 Deutsch stammen, aussortiert. In Tabelle 3 wird das Ergebnis dieses Sortier- und Extraktionsprozesses wiedergegeben.

Tabelle 3: Liste potenzieller formelhafter Sequenzen mit der Struktur Präpositionalphrase + Verbkomplex und dem nominalen Bestandteil Topf im Lernerkorpus „KoKo“ (Filter: L1 Deutsch)

Freq.	Präp.	Det.	attr. Adj.	Nomen	Verb	Num.
32	in	indef	-	Topf	werfen	Sg
10	in	indef	-	Topf	stecken	Sg
5	in	indef	-	Topf	schmeißen	Sg
2	in	indef	-	Topf	hauen	Sg
2	in	dem	-	Topf	werfen	Sg
2	in	def	gleich	Topf	schmeißen	Sg
1	in	dem	-	Topf	stecken	Sg
1	in	dem	-	Topf	schmeißen	Sg
1	in	def	gleich	Topf	werfen	Sg
1	in	def	gleich	Topf	stecken	Sg

Durch dieses Verfahren wurden 57 Vorkommnisse der Struktur *PP+VV* mit dem nominalen Bestandteil *Topf* gefunden. In Tabelle 3 zeigt sich bereits die erste Möglichkeit von Varianz: Die Stellen des Determinierers sowie des attributiven Adjektivs können häufig unbesetzt bleiben, die lexikalische Varianz des Verbs umfasst die Verben *werfen*, *stecken*, *schmeißen* und *hauen*.

Um tatsächlich alle Vorkommnisse des Idioms *in einen Topf werfen* im Korpus zu bestimmen, muss das methodische Vorgehen an dieser Stelle gewechselt werden („Analyse“). Mit geeigneten Suchanfragen können weitere Varianten des Idioms entdeckt werden (Schritt 4). Eine solche Reformulierung der Suchanfrage erhöht die Anzahl der Vorkommnisse des Idioms bei SchreiberInnen mit L1 Deutsch auf 67.

Im abschließenden Schritt 5 müssen alle 67 extrahierten Vorkommnisse über die jeweilige Korpusfundstelle auch als formelhafte Sequenz verifiziert werden. Im vorliegenden Beispiel ist das besonders wichtig, da *in einen Topf werfen* und mögliche Varianten davon oftmals auch in freier (vgl. *das Gemüse in einen Topf werfen / geben* u. a. in Kochrezepten) oder metaphorischer Verwendung (vgl. *Geld in einen gemeinsamen Topf werfen / einzahlen*) vorkommen. Folgende Korpusfundstellen illustrieren die idiomatische Verwendungsweise der extrahierten Sequenz:

- ID1041 (NT): „Ein junger Mensch ist womöglich noch unsicher und leichter beeinflussbar, als eine Person, die schon etwas erfahrener ist, aber genau aus diesem Grund ist es wichtig, die Jugend nicht zu kritisieren, sondern ihnen zu helfen, die richtigen Entscheidungen zu treffen und sie auf den richtigen Weg zu bringen. Meiner Meinung nach ist es falsch *<alle Jugendlichen in einen Topf zu werfen>*. Menschen machen Fehler, egal wie alt sie sind und man sollte sich in seiner Jugend auch austoben können und nicht nach einem strengen Schema leben müssen.“
- ID2658 (NT): „Wie bereits erwähnt, ist laut Hans Magnus Enzensberger die Jugend keine beneidenswerte Phase, denn viele Leiden beispielsweise an einer Klammottensucht und wollen immer das haben, was andere auch haben, damit sie "cool" sind und dazu gehören. Jedoch stellt sich die Frage, kann man wirklich *<alle Jugendlichen in einen Topf stecken>*? Meiner Meinung nach, gibt es genug Jugendliche, die ihre Freizeit nicht nur mit Diskothekbesuche, Alkohol und Tabak verbringen.“

An dieser Stelle soll nicht weiter auf die lexikalische Varianz der verbalen Komponente des metaphorischen Idioms *in einen Topf werfen* eingegangen werden, sondern das Augenmerk auf ein anderes Phänomen gelegt werden: Bei der manuellen Analyse der in Tabelle 2 aufgeführten Kandidaten für formelhafte Sequenzen, besonders bei der Verifizierung der formelhaften Verwendung (Schritt 5), sind einige metaphorische Idiome aufgefallen, die zumindest teilweise mit derselben Bedeutung verwendet werden wie das Ausgangsidiom *in einen Topf werfen*. Es handelt sich dabei um die Idiome *über einen Kamm scheren*, *auf eine Stufe stellen*, *unter einen Hut bringen*, *in einen Sack stecken*, *in eine Schublade stecken* und Varianten davon.⁹ Die folgenden Beispiele illustrieren diese Beobachtung:

9 Der Begriff *Varianten* wird hier allgemein als Abweichung von einer im Wörterbuch verzeichneten

- ID2149 (TH): „Doch das trifft nur für manche Jugendlichen zu, nicht alle haben einen schwachen Charakter und lassen sich zu reden. Man kann nicht *<alle Jugendlichen über einen Kamm schären>* auch nicht Hans Magnus Enzensberger. Auch nicht jeder Jugendlicher lässt sich auf den Gruppenzwang ein, es gibt viele die ein Individuum sein wollen und gerade deswegen ihr eigenes Ding durchziehen.“
- ID1451 (TH): „Trotzdem sollte er bedenken, dass die Jugendzeit vor allem, dazu dient Erfahrungen zu sammeln, auch wenn dabei Fehler entstehen. Aber ich finde es falsch von ihm, dass er *<alle Jugendlichen auf eine Stufe stellt>*, denn es gibt verschiedene Persönlichkeiten. Das Zitat von dem deutschen Schriftsteller und Essayist Hans Magnus Enzensberger kritisiert die jugendliche Phase eines Lebens, die nur aus Unsicherheiten, Dummheiten, Gruppenzwang und Unsouveränität besteht.“
- ID1780 (ST): „Trotz der vielen Aspekte, die im ersten Blick alle negativ erscheinen, muss man die Meinung Enzensbergers nicht teilen. Vor allem die Tatsache, dass er *<alle Jugendlichen in einem Sack gibt>*, lässt mit sich diskutieren. Es gibt zahlreiche Kiffer und Trinker, aber mindestens, genauso viele welche wenig konsumieren oder gar nichts.“
- ID1762 (NT): „Der Aussage des Deutschen kann ich nicht ganz zustimmen, da er *<alle Jugendlichen unter einen Hut steckt>*. Es kommt stark auf die Persönlichkeit an, ob man ohne große Probleme zum gestandenen Mann heranwächst oder einen die Pubertät, wie ein Schlag trifft.“
- ID2944 (NT): „Zu allererst muss man erwähnen, dass diese Aussage als kaum ernstzunehmend angesehen werden kann. Herr Enzensberger *<steckt alle jungen Menschen in eine Schublade>*. Jeder Mensch ist in seinem Charakter verschieden und zeichnet sich dadurch aus.“

Tabelle 4 fasst alle metaphorischen Idiome zusammen, die quasi-synonym im „KoKo“-Korpus verwendet werden.¹⁰

Tabelle 4: Übersicht über metaphorische Idiome, die mit der Bedeutung 'Verschiedenes gleich behandeln' im Lernerkorpus „KoKo“ (Filter L1 Deutsch) verwendet werden.

Grundform verstanden.

- 10 vgl. auch folgende Fälle, bei denen kein zugrunde liegender Phraseologismus ausgemacht werden konnte:
- ID1142: „Sicherlich ist die Jugend, mitten im zweiten Weltkrieg wie jene von Enzensberger nicht sonderlich berauschend, aber deshalb darf man nicht so stark Kritik über junge Menschen ausüben und *<alle in dieselbe Kiste stecken>*.“
- ID2184: „Meiner Meinung nach *<werden die Jugendlichen alle in eine Ecke geschoben>* und als Säufer und als unnütz bezeichnet.“
- ID1344: „Genau aus diesen Gründen soll man seine Jugend ausleben, herausfinden, was man will und *<sich nicht von anderen in eine Schiene stecken lassen>*, die einem vorgiebt, wie man zu sein hat.“
- ID1730: „Ich finde so eine Aussage darf man nicht verallgemeinern. Denn man kann nicht *<alle jungen Menschen auf einen "Haufen kerren">*, schließlich hat jeder Jugendliche andere positive und negative Eigenschaften. Keiner von ihnen ist perfekt!“

		‘Verschiedenes gleich behandeln’
KoKo: Filter L1 Deutsch		
(1)	in DET (ADJA) Topf + VV	67
(2)	über DET (ADJA) Kamm + VV	14
(3)	auf DET (ADJA) Stufe + VV	3
(4)	unter DET (ADJA) Hut + VV	12
(5)	in DET (ADJA) Sack + VV	2
(6)	in DET (ADJA) Schublade + VV	14
AKKO ADV(Dir) + VV		113

Auffällig in Tabelle 4 sind die vier letztgenannten Idiome (3)-(6) in ihrer Verwendung mit der Bedeutung ‘Verschiedenes gleich behandeln’, da diese von den Wörterbuchbedeutungen teilweise erheblich abweicht. Für die jeweiligen Idiome verzeichnet der Duden 2008 folgende Beschreibungen¹¹:

alle[s] in einen Topf werfen *umgangssprachlich*: ‘alles, alle gleich [schlecht] beurteilen, ohne die bestehenden Unterschiede zu berücksichtigen’

alle[s] über einen Kamm scheren: ‘alle[s] gleich behandeln und dabei wichtige Unterschiede nicht beachten’

jemanden, etwas auf eine Stufe [mit jemandem, etwas] / auf die gleiche Stufe [wie jemand, etwas] stellen: ‘jemanden, etwas als im Rang gleichwertig beurteilen, darstellen, im Niveau, im Rang einander gleichstellen’

unter einen Hut bringen *umgangssprachlich*: ‘einigen; in Einklang, Übereinstimmung bringen’

jemanden in den Sack stecken: ‘jemandem überlegen sein’

jemanden in eine Schublade stecken: ‘jemanden (leichtfertig) in eine Kategorie einordnen’

Gemeinsam ist allen aufgeführten Idiomen ihre Struktur PREP DET (ADJA) NN VV. Laut der Einträge in Duden (2008) und im ÖWB (2006) kann man *alles in einen Topf werfen* und *alles über einen Kamm scheren* als quasi-synonym betrachten. Beide beschreiben Umstände, in denen unterschiedliche Dinge / Personen so behandelt werden, als gäbe es keine Unterschiede zwischen diesen Dingen / Personen. Die

11 Der Duden (2008) und das Österreichische Wörterbuch (ÖWB 2006) verzeichnen weder unterschiedliche Grundformen noch Bedeutungen der genannten metaphorischen Idiome, auch wenn der genaue Wortlaut der Bedeutungserklärung bisweilen abweichen mag. Das Idiom *alles über einen Kamm scheren* ist allerdings nicht im Österreichischen Wörterbuch verzeichnet. Das Idiom *jemanden in eine Schublade stecken* ist in keinem der Werke aufgeführt, findet sich aber im Deutschen Universalwörterbuch (Duden 2001) sowie im Deutschen Wörterbuch von Brockhaus und Wahrig (1983) unter dem Lemma *Schublade* als Wendung („von diesen Leuten wirst du immer gleich in eine bestimmte Schublade gesteckt“).

Bedeutungen der anderen Idiome divergieren hingegen teilweise beträchtlich davon. Die Verwendung dieser Idiome in der Bedeutung ‘alle(s) gleich beurteilen, ohne die bestehenden Unterschiede zu berücksichtigen’ bedarf daher einer Erklärung.

4 Interpretation

Auffälligkeiten als lernerspezifisches Phänomen

H₀ folgend, soll bei der Erklärung der Verwendung der o. g. Idiome in der von den Einträgen der angegebenen Wörterbücher abweichenden Bedeutung (i. e. das sprachlich auffällige Merkmal x) zunächst von einem lernerspezifischen Phänomen ausgegangen werden. H₀ soll an dieser Stelle wiederholt werden:

- Das sprachlich auffällige Merkmal x im Lernerkorpus „KoKo“ ist ein Lernerphänomen (= H₀).

Die Verwendungen der metaphorischen Idiome (3)–(6) in Tabelle 4 wären demnach fehlerhaft; die Fehler bei der Produktion betreffen jedes Mal die Bedeutung (*konzeptuelle Kontamination*, vgl. Margewitsch 2006), oftmals treten auch Varianten auf, die außerdem als Fehler in der Wiedergabe der richtigen Form des Idioms (*formale Kontamination*, vgl. Margewitsch 2006) beschrieben werden können. Dies ist vor allem bei den kontaminierten Formen *alle unter einen Hut stecken* und *alle auf eine Stufe stellen* der Fall. Die strukturelle Ähnlichkeit der vorliegenden Idiome mit den Idiomen *alles in einen Topf werfen* bzw. *alles über einen Kamm scheren* könnte der Grund für die formale und konzeptuelle Kontamination sein.

Abbildung 1 gibt alle Verwendungsweisen unterschiedlicher Idiome mit der Bedeutung ‘alle(s) gleich beurteilen ohne die bestehenden Unterschiede zu berücksichtigen’ schematisch wieder. Sie stellt die quasi-synonymen Idiome *alles über einen Kamm scheren* und *alles in einen Topf werfen* im Zentrum der Abbildung dar. Neben diesen Idiomen finden sich im Korpus mehr oder weniger strukturell ähnliche Idiome, die in Abbildung 1 an der Peripherie außerhalb der gestrichelten Linie aufgeführt sind. Ein Großteil dieser strukturell ähnlichen Idiome wird im „KoKo“-Korpus in derselben Bedeutung verwendet wie *alles über einen Kamm scheren* und *alles in einen Topf werfen*. Die inhaltliche Kontamination ist teilweise auch mit einer formalen Kontamination verbunden. Die fehlerhaften Verwendungsweisen sind in Abbildung 1 an der Peripherie innerhalb der gestrichelten Linie dargestellt.

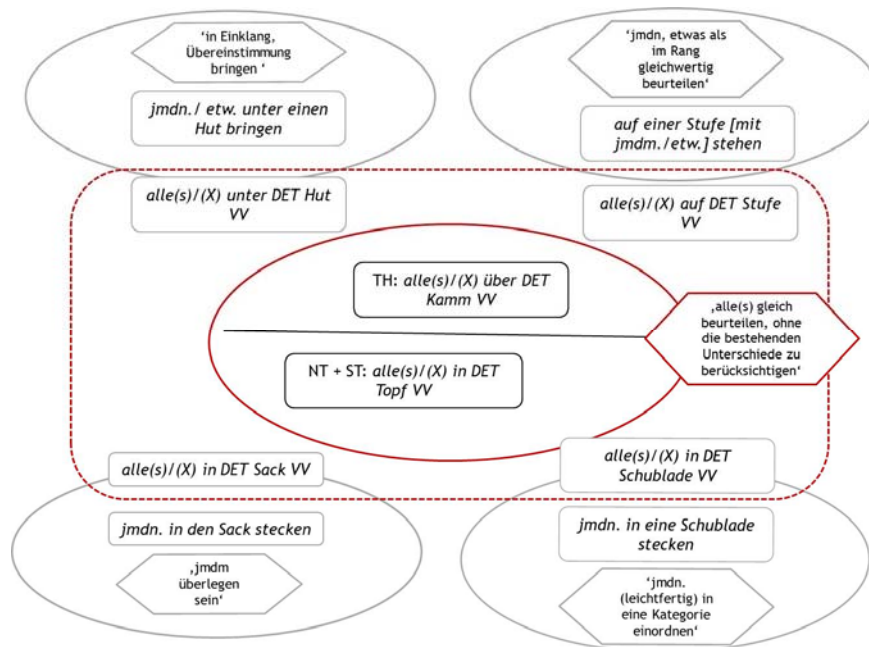


Abb. 1: Verwendungsweisen unterschiedlicher Idiome mit der Bedeutung ‚alle(s) gleich beurteilen ohne die bestehenden Unterschiede zu berücksichtigen‘ im „KoKo“-Korpus (Filter L1 Deutsch)

Diatopische Varianz in einem Varietäten-Lernerkorpus

Wie oben beschrieben muss H_0 gegen mögliche Alternativerklärungen getestet werden. Eine mögliche Alternativhypothese ist H_1 , die besagt, dass es sich bei den zu beschreibenden Phänomenen um diatopische Varianten handelt, d. h. dass zum Beispiel die Verwendung von *alle unter einen Hut stecken* mit der Bedeutung ‚alles, alle gleich [schlecht] beurteilen, ohne die bestehenden Unterschiede zu berücksichtigen‘ eine regionale Variante darstellt. Wie bereits in Kapitel 3.b angedeutet, würde die Bestätigung von H_1 nicht zwangsläufig bedeuten, dass H_0 zurückgewiesen werden muss; H_1 kann H_0 auch lediglich einschränken, da Fehler spezifisch für eine Region sein könnten.

Zunächst wird daher überprüft, ob die oben beschriebenen Beobachtungen im Lernerkorpus „KoKo“ (Filter L1 Deutsch) einer regionalen Verteilung folgen. Betrachtet man nur diejenigen Fälle, in denen die Bedeutung ‚Verschiedenes gleich behandeln‘ für die extrahierten Idiome vorliegt, ergibt sich folgende regionale Verteilung:

Tabelle 5: Regionale Verteilung der Verwendung von metaphorischen Idiomen mit der Bedeutung ‚Verschiedenes gleich behandeln‘ im Lernerkorpus „KoKo“. Zeichenerklärung: * signifikanter Unterschied (log-likelihood) mit $p < 0.5$; Werte,

die aufgrund geringer Frequenzen (i. e. wenn der Erwartungswert < 1) mit Vorsicht betrachtet werden sollten, werden mit † gekennzeichnet.¹²

	‘Verschiedenes gleich behandeln’				log likelihood			vs.
	gesamt	NT	ST	TH	NT vs. ST	NT vs. TH	ST TH	
KoKo: Filter L1								
auf DET (ADJA) Stufe + VV	3	0	0	3	0,00†	-3,01	-2,85	
unter DET (ADJA) Hut + VV	12	5	1	6	2,65	0,16	-1,91	
in DET (ADJA) Sack + VV	2	0	2	0	-2,91†	0,00†	3,89†*	
in DET (ADJA) Schublade + VV	14	6	4	4	0,28	1,72	0,49	
AKKO ADV(Dir) + VV	113	44	41	28				

Tabelle 5 gibt kaum Hinweise auf eine regionale Verteilung. Das Idiom *in DET (ADJA) SACK + VV*, das in der auffälligen Bedeutung nur in den Texten aus Südtirol vorkommt, wird zu selten verwendet, als dass über seine Verteilung Aussagen getroffen werden könnten. Dasselbe gilt für das Idiom *auf DET (ADJA) Stufe + VV*. Auch hier kann mangels ausreichender Belege keine Aussage über die regionale Verteilung gemacht werden. Für die restlichen beiden Idiome gilt, dass die Alternativhypothese H_1 , nämlich dass es sich beim vom Wörterbucheintrag abweichenden Gebrauch um ein diatopisches Phänomen handelt, nicht bestätigt wird, da sie über die drei Regionen ohne signifikante Unterschiede verteilt sind. Die Fallzahlen im Korpus sind allerdings insgesamt zu gering, um zu beurteilen, ob H_1 ein alternatives Erklärungsmodell für H_0 darstellt.

Obwohl nicht unmittelbar im Fokus von H_1 , soll an dieser Stelle auf eine Beobachtung aufmerksam gemacht werden: Beim Gebrauch der quasi-synonymen Idiome *in einen Topf werfen* und *über einen Kamm scheren* zeigt sich eine regionale Verteilung. Während ersteres Idiom signifikant häufiger in den Texten aus Nordtirol und Südtirol vorkommt als in den Texten aus Thüringen, wird dort signifikant häufiger letzteres Idiom verwendet.¹³ Jedoch muss auch diese Verteilung zumindest im Falle von *über einen Kamm scheren* aufgrund der geringen Vorkommnisse mit Vorsicht genossen werden. In diesem Zusammenhang sollte berücksichtigt werden, dass die Häufigkeit des Gebrauchs bestimmter Konstruktionen nicht als einziger Indikator für deren Geltung infrage kommt (vgl. Eichinger 2005: 369).

12 Für die Festlegung von sinnvollen Schwellen für die Reliabilität statistischer Tests nach der Cochran rule, vgl. Rayson et al. (2004) und Anstein (2013).

13 Vgl. das Vorkommen von *in einen Topf werfen* in NT (31 Belege) und ST (34 Belege) gegenüber TH (2 Belege) (log likelihood (NT vs. ST) = -0,42; (NT vs. TH) = 44,61*; (ST vs. TH) = 52,56*) mit dem Vorkommen von *über einen Kamm scheren* in NT (2 Belege) und ST (0 Belege) gegenüber TH (12 Belege) (log likelihood (NT vs. ST) = 2,46†; (NT vs. TH) = -4,27*; (ST vs. TH) = -11,40*).

Vergleiche zu Referenzkorpora

Da diatopische Varianz keine alternative Erklärung für die von den Wörterbucheinträgen abweichenden Bedeutungen im „KoKo“-Korpus liefert, soll abschließend eine weitere Alternativhypothese getestet werden. H₂ lautet angepasst: Die von den Wörterbucheinträgen abweichenden Bedeutungen der o. g. Idiome sind ein Sprachwandelphänomen.

H₂ sollte mithilfe von Korpora getestet werden, die keine Lernertexte enthalten. Hierfür wurde das Deutsche Referenzkorpus (DeReKo) herangezogen. Das DeReKo beinhaltet ein österreichisches Zeitungskorpus (OZK), in dem einige Jahrgänge von 1991 bis einschließlich 2000 unterschiedlicher österreichischer Tageszeitungen (*Oberösterreichische Nachrichten, Kleine Zeitung, Neue Kronen-Zeitung, Die Presse, Salzburger Nachrichten, Tiroler Tageszeitung, Vorarlberger Nachrichten*) zu finden sind. Das OZK hat eine Größe von 469.356.467 Tokens. Zusätzlich zum OZK wurde ein Teilkorpus mit bundesdeutschen Tageszeitungen erstellt (DZK: *Berliner Morgenpost, Frankfurter Rundschau, Mannheimer Morgen, Nürnberger Nachrichten, Rhein-Zeitung*), das sich auf denselben Zeitraum bezieht und in etwa dieselbe Größe (477.924.157 Tokens) aufweist. Da keine Südtiroler Tageszeitung im DeReKo verfügbar ist, wurde zusätzlich das Dolomiten-Korpus (DOL) des Korpus Südtirol verwendet. Dieses besteht aus 66.348.067 Tokens und umfasst ausgewählte Jahrgänge von 1991 bis 2006. Um H₂ zu testen, wird sowohl in den Referenzkorpora als auch im „KoKo“-Korpus die Verwendung der semantisch und formal abweichend verwendeten Idiome mit der Häufigkeit der Verwendung in konventioneller Bedeutung verglichen. Tabelle 6 gibt den Anteil der abweichenden Bedeutung im Verhältnis zur Verwendung in der konventionellen Bedeutung an.

Tabelle 6: Vergleich der Verwendung von metaphorischen Idiomen mit der Bedeutung ‘Verschiedenes gleich behandeln’ im Lernerkorpus „KoKo“ (Filter LI Deutsch) mit Referenzkorpora.

	‘Verschiedenes gleich behandeln’			
	KoKo	OZK	DOL	DZK
	absolut	absolut	absolut	absolut
auf DET (ADJA) Stufe + VV	3 / 5	0 / 525	0 / 67	0 / 399
unter DET (ADJA) Hut + VV	12 / 16	3 / 1985	0 / 263	1 / 2002
in DET (ADJA) Sack + VV	2 / 2	4 / 20	0 / 2	12 / 24
in DET (ADJA) Schublade + VV	14 / 27	1 / 150	0 / 16	5 / 398

Generell zeigt sich, dass im Lernerkorpus die abweichenden Verwendungen einen größeren Anteil bei der Verwendung der untersuchten Phraseologismen ausmachen als in den Referenzkorpora, in denen sie nur vereinzelt auftreten und als marginale Erscheinungen betrachtet werden können. Dieses Ergebnis deutet darauf hin, dass die abweichenden Verwendungsweisen weiterhin als Lernerphänomene eingestuft werden können. Eine Ausnahme stellt die Verwendung von *in DET (ADJA)*

Sack + *VV* in der Bedeutung ‘Verschiedenes gleich behandeln’ dar, das in einem Fünftel der Fälle im OZK und in der Hälfte der Fälle im DZK in dieser Bedeutung vorkommt. Hier scheint also die Verwendung mit einer Bedeutung verbreitet zu sein, die von der konventionellen Bedeutung abweicht, was als Hinweis auf ein Sprachwandelphänomen und die Entstehung eines neuen Gebrauchsstandards (vgl. Eisenberg 2007) interpretiert werden kann. Zwei Beispiele aus dem OZK und dem DZK sollen diese Verwendungsweise abschließend illustrieren:

Die Presse, 09.05.1998: „Das erste ist der Wahlsieg des holländischen Sozialdemokraten Wim Kok. Oberflächlich betrachtet reiht er sich ein in eine Serie von sozialdemokratischen Wahlsiegen, von England über Frankreich bis Dänemark, und scheint die triviale Auffassung zu bestätigen, die Geschichte sei ein Uhrwerk, in dem das Pendel einmal nach rechts und einmal nach links ausschlägt. Bei ausreichender Distanz zu den Fakten könnte man sich vielleicht dazu verleiten lassen, *<Rasmussen und Jospin mit Kok und Blair in einen roten Sack zu stecken>* und dann den Schluß zu ziehen, daß das Herz des Kontinents wieder links schlägt. Aber die Dinge sind halt sehr kompliziert, wie auch Sozialdemokraten gelegentlich zugeben.“ (Originalressort: Seite Zwei; Geht Europa nach links?)

Rhein-Zeitung, 13.01.1998: „Man sollte endlich damit aufhören, jeden, der etwas anders aussieht, als Herr Panthel es für gut befindet, zu diskriminieren. Piercing, dort wo Herr Panthel es nicht vermutet, schrille Farben in den Haaren und ein "skurriles" Aussehen darf nicht dazu führen, *<alle (in irgendeiner Weise) kreativen Jugendlichen in einen Sack mit denen zu stecken, die nächstens anderer Leute Eigentum beschmieren>*. Dieses Menschenbild halte ich im Gestern, Heute und Morgen für Untragbar für unsere Gesellschaft.“ (Gelbachtal einst ein Paradies)

5. Fazit

Die vorgestellte Analyse der Verwendung von formelhaften Sequenzen im L1-Varietäten-Lernerkorpus „KoKo“ hat anhand der Verwendung des metaphorischen Idioms *alles in einen Topf werfen* sowie konzeptionell und strukturell verwandter Idiome wie *alles über einen Kamm scheren*, *alles unter einen Hut bringen* u. a. gezeigt, wie lernerspezifische bzw. fehlerhafte Gebrauchsweisen von regionalen und Sprachwandelphänomenen unterschieden werden können. Mithilfe eines hypothesentestenden Verfahrens wurden verschiedene Erklärungsmöglichkeiten bei der Interpretation der Ergebnisse berücksichtigt.

Im Bereich der Phraseologismen kann regionale Varianz auftreten, wobei ein Phrasem in verschiedenen Regionen unterschiedliche Bedeutungen oder unterschiedliche formale Merkmale haben kann. Daher muss neben der Grundannahme, dass sich lernerspezifische Auffälligkeiten zeigen, in einem Varietäten-Lernerkorpus immer auch diatopische Varianz als Erklärungsmöglichkeit für Phänomene semantischer oder struktureller Unterschiede beim Gebrauch einer formelhaften Sequenz in Betracht gezogen werden. Des Weiteren können Abweichungen von einem Standard

in einem Schülerkorpus auch Hinweise auf Sprachwandelphänomene geben; auch diese Erklärungsmöglichkeit für unterschiedliche Verwendungsweisen muss daher berücksichtigt werden. Freilich kann es noch weitere Ursachen für Variation geben, diese mussten in dieser Arbeit allerdings unberücksichtigt bleiben.

Die Ergebnisse der Untersuchung der Verwendung des metaphorischen Idioms *alles in einen Topf werfen* sowie konzeptionell und strukturell verwandter Idiome hat eine regionale Verteilung der Quasi-Synonyme *alles in einen Topf werfen* und *alles über einen Kamm scheren* ergeben (die zumindest für ersteres auch im OZK und DZK bestätigt werden kann). Eine regionale Verteilung der anderen metaphorischen Idiome konnte nicht bestätigt werden – hierfür sind die Belege im Korpus teilweise zu selten.

Ein Vergleich zu österreichischen, Südtiroler und bundesdeutschen Referenzkorpora hat gezeigt, dass einige der im „KoKo“-Korpus vorkommenden auffälligen Verwendungen, vor allem von *alles in einen Sack stecken* in der Bedeutung ‚Verschiedenes gleich behandeln‘, in österreichischen und bundesdeutschen Tageszeitungen relativ häufig vorkommen. Dies zeigt, dass es sich nicht nur um ein schüler- oder lernerspezifisches Phänomen handelt, sondern auch darüber hinaus in großen Teilen des deutschsprachigen Raumes zu finden ist. Hier müssten weitere Untersuchungen verifizieren, ob es sich tatsächlich um eine „neue“ Verwendung des Idioms, also um ein Sprachwandelphänomen handelt, das im gesamten deutschen Sprachraum vorkommt oder von bestimmten regionalen Zentren ausgeht.

Die vorgestellte Analyse hat ein ausgewähltes Phänomen im L1-Varietäten-Lernerkorpus „KoKo“ herausgegriffen. Es ist zu erwarten, dass künftige Analysen auf unterschiedlichen sprachlichen Ebenen ebenso relevante Hinweise sowohl auf Lernerpezifika als auch auf varietätenlinguistische bzw. Sprachwandelphänomene liefern können.

Literatur

- Abel, Andrea / Anstein, S.: Korpus Südtirol – Varietätenlinguistische Untersuchungen. In: A. Abel & R. Zanin [Hgg.]: Korpora in Lehre und Forschung, Bozen-Bolzano, 2011, 29–54.
- Anstein, Stefanie: Computational Approaches to the Comparison of Regional Variety Corpora - Prototyping a Semi-automatic System for German, Dissertation, Universität Stuttgart, 2013.
- Anstein, Stefanie et al.: Korpus Südtirol – Aufbau und Abfrage. In: A. Abel / R. Zanin [Hgg.]: Korpora in Lehre und Forschung, Bozen-Bolzano, 2011, 15–28.
- Aguado, Karin: Formelhafte Sequenzen und ihre Funktion für den L2-Erwerb. In: Zeitschrift für Angewandte Linguistik 37, 2002, 3–26.
- Ammon, Ulrich: Standard und Variation. Norm, Autorität, Legitimation. In: L. M. Eichinger / W. Kallmeyer [Hgg.]: Standardvariation, Wie viel Variation verträgt die deutsche Sprache? Berlin, 2005, 28–40.
- Ammon, Ulrich et al. [Hgg.]: Variantenwörterbuch des Deutschen, Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol, Berlin, 2004.
- Brockhaus/ Wahrig: Deutsches Wörterbuch in sechs Bänden, Bd. 5, Wiesbaden, Stuttgart, 1983.

- Bubenhofner, Noah / Ptashnyk S.: Korpora, Datenbanken und das Web. State of the Art computer-gestützter Forschung in der Phraseologie und Lexikographie. In: S.Ptashnyk et al. [Hgg.]: Korpora, Web und Datenbanken, Baltmannsweiler, 2010, 7–19.
- Burger, Harald: Zur Phraseologie des Schweizer Hochdeutschen. In: J. Korhonen [Hg.]: Studien zur Phraseologie des Deutschen und des Finnischen II, Bochum, 1996, 461–488.
- Burger, Harald: Konzepte von Variation in der Phraseologie. In: A. Häcki Buhofer [Hg.]: Vom Umgang mit sprachlicher Variation, Tübingen, 2000, 35–51.
- Burger, Harald: Phraseologie, Eine Einführung am Beispiel des Deutschen, Berlin, 2010.
- Duden: Deutsches Universalwörterbuch, Mannheim, 2001.
- Duden: Die Grammatik, 7. Auflage, Mannheim, 2005.
- Duden: Redewendungen, Wörterbuch der deutschen Idiomatik, Mannheim, 2008.
- Dürscheid, Christa et al.: Wie Jugendlich schreiben, Schreibkompetenz und neue Medien, Berlin, 2010.
- DWDS: Das Digitale Wörterbuch der deutschen Sprache (<http://www.dwds.de/>) (Februar 2014)
- Eichinger, Ludwig M.: Standardnorm, Sprachkultur und die Veränderung der normativen Erwartungen. In: L. M. Eichinger / W. Kallmeyer [Hgg.]: Standardvariation, Wie viel Variation trägt die deutsche Sprache? Berlin, 2005, 363–381.
- Eichler, Wolfgang: Sprachbewusstsein und grammatisches und stilistisches Formulieren, Falsche Kollokationen und verformelter Sprachgebrauch in Oberstufenaufsätzen. In: ELiSe 4 (1), 2004 155–163.
- Eisenberg, Peter: Sprachliches Wissen im Wörterbuch der Zweifelsfälle, Über die Rekonstruktion einer Gebrauchsnorm. In: Aptum, Zeitschrift für Sprachkritik und Sprachkultur 3 /2007, 209–228.
- Ernst, Peter: Redewendungen im österreichischen Deutsch. In: tribüne 4, 2007, 4–6.
- Fellbaum, Christiane. [Hg.]: Idioms and Collocations, London, 2007.
- Földes, Csaba: Deutsche Phraseologie kontrastiv, Heidelberg, 1996.
- Gehweiler, Elke et al.: Types of changes in idioms – some surprising results of corpus research. In: C. Fellbaum [Hg.]: Idioms and Collocations, London, 2007, 109–137.
- Glaznieks, Aivars: Emotionsbezeichnungen im kindlichen Wortschatz, Die Ausdifferenzierung der lexikalischen Felder „Angst“ und „Ärger“ bei Kindern im Vor- und Grundschulalter, Hamburg, 2011.
- Granger, Sylviane: Learner corpora. In: A. Lüdeling / M. Kytö [Hgg.]: Corpus Linguistics, An International Handbook, Vol. 1, Berlin, 2008, 259–275.
- Häcki Buhofer, Annelies: Phraseme im Erstspracherwerb. In: H. Burger et al. [Hgg.]: Phraseologie, Ein internationales Handbuch der zeitgenössischen Forschung, Bd. 2, Berlin, 2007, 854–869.
- Heid, Ulrich: Computational phraseology, An overview. In: S. Granger / F. Meunier [Hgg.]: Phraseology, An interdisciplinary perspective, Amsterdam, 2008, 337–360.
- Katerbow, Matthias et al.: Einstellung zur kindlichen Sprachvariation, Eine empirische Pilotstudie zur Fremdbewertung regionalsprachlicher und /oder störungsbedingter Variation bei Kindern, Linguistische Berichte 228, 2011, 467–484.
- Katerbow, Matthias: Spracherwerb und Sprachvariation, Eine phonetisch-phonologische Analyse zum regionalen Erstspracherwerb im Moselfränkischen, Berlin, 2013.
- Korhonen, Jarmo et al.: EUROPHRAS 2008. Beiträge zur internationalen Phraseologiekonferenz vom 13.–16.8.2008 in Helsinki, Helsinki, 2010.
- Margewitsch, Erika: Formelhafter Sprachgebrauch in Schülertexten, Oldenburg, 2006.
- McGregor, Karla K. et al.: Semantic representation and naming in young children; Journal of Speech, Language, and Hearing Research 45, 2002, 332–346.
- Moon, Rosamund: Corpus linguistic approaches with English corpora. In: H. Burger et al. [Hgg.]: Phraseologie. Vol. 2, Berlin, 2007, 1045–1059 (= HSK 28.2).
- ÖWB = Österreichisches Wörterbuch, Wien, 2006.
- Ortner, Hanspeter: Die Bildungssprache im Visier der Sprachkritik. In: tribüne 1, 2006, 4–11.

- Perkins, Mick R.: Productivity and formulaicity in language development. In: M. Garman et al. [Hgg.]: *Issues in Normal and Disordered Child Language, From Phonology to Narrative*, Special Issue of *The New Bulmershe Papers*, Reading, 1999, 51–67.
- Piirainen, Elisabeth: Phraseologie in arealen Bezügen, ein Problemaufriss. In: *Linguistik online* 27, 2, 2006.
- Polenz, Peter v.: Deutsch als plurinationale Sprache im postnationalistischen Zeitalter. A. Gardt et al. [Hgg.]: *Sprachgeschichte als Kulturgeschichte*, Berlin, 1999, 115–132.
- Quasthoff, Uwe et al.: Häufigkeit und Struktur von Phraseologismen am Beispiel verschiedener Web-Korpora. In: S. Ptashnyk et al. [Hgg.]: *Korpora, Web und Datenbanken*, Baltmannsweiler, 2010, 37–53.
- Rayson, Paul et al.: Extending the Cochran rule for the comparison of word frequencies between corpora. In: G. Purnelle et al. [Hgg.]: *Proceedings of the 7th International Conference on Statistical Analysis of Textual Data (JADT)*, Louvain-la-Neuve, 2004, 926–936.
- Sailer, Manfred: Corpus linguistic approaches with German corpora. In: H. Burger et al. [Hgg.]: *Phraseologie*. Vol. 2, Berlin, 2007, 1060–1071 (= HSK 28.2).
- Schmidlin, Regula: Phraseological expressions in German standard varieties. H. Burger et al. [Hgg.]: *Phraseologie*. Vol. 1, Berlin, 2007, 551–562 (= HSK 28.1).
- Schmidlin, Regula: *Die Vielfalt des Deutschen, Standard und Variation*, Berlin, 2011.
- Schneider, Jan G.: Sprachliche ‚Fehler‘ aus sprachwissenschaftlicher Sicht. In: *Sprachreport* 1–2/2013, 30–37.
- Schründer-Lenzen, Agi / Henn, D.: Entwicklung eines Computerprogramms zur Analyse der schriftlichen Erzählfähigkeit. In: D. Lengyel et al. [Hgg.]: *Von der Sprachdiagnose zur Sprachförderung*, Münster, 2009, 91–107.
- Sieber, Peter: *Parlando in Texten*. Tübingen, 1998.
- Steinhoff, Torsten: *Wissenschaftliche Textkompetenz*, Tübingen, 2007.
- Vatvedt Fjeld, Ruth et al.: Semi-Automatic Retrieval of Phraseological Units in a Corpus of Modern Norwegian. In: S. Ptashnyk et al. [Hgg.]: *Korpora, Web und Datenbanken*, Baltmannsweiler, 2010, 111–117.
- Wallner, Franziska: Kollokationen in Wissenschaftssprachen. In: S. Ptashnyk et al. [Hgg.]: *Korpora, Web und Datenbanken*, Baltmannsweiler, 2010, 197–214.
- Weller, Marion / Heid, U.: Extraction of German multiword expressions from parsed corpora using context features, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, 2010, 3195–3201.
- Wodak, Ruth / Rheindorf, M.: *Wandel der deutschen Sprache: eine textsortenbezogene Pilotstudie*, Wien, 2011.
- Wray, Alison: Formulaic sequences in second language teaching: principle and practice. In: *Applied Linguistics* 21 (4), 2000, 463–489.
- Wray, Alison / Perkins, M. R.: The functions of formulaic language: an integrated model. In: *Language & Communication* 20, 2000, 1–28.
- Ziegler, Arne: Er erwartet sich nur das Beste – Reflexivierungstendenz und Ausbau des Verbalparadigmas in der österreichischen Standardsprache. In: D. Bittner / L. Gaeta [Hgg.]: *Kodierungstechniken im Wandel, Das Zusammenspiel von Analytik und Synthese im Gegenwartsdeutschen*, Berlin, 2010, 67–81.