

Christa Dürscheid und Stephan Elspaß

23 Variantengrammatik des Standarddeutschen

1 Gegenstand

1.1 Organisationsrahmen

Gegenstand des Forschungsprojekts *Variantengrammatik des Standarddeutschen* ist die grammatische Variation in der deutschen Standardsprache. Das Projekt wurde zunächst im Rahmen der D-A-CH-Vereinbarung der Forschungsförderungsinstitutionen der Schweiz, Österreichs und Deutschlands seit dem 1. August 2011 für einen Zeitraum von 36 Monaten gefördert; ein Verlängerungsantrag für weitere drei Jahre wurde bewilligt, so dass die Arbeiten nun bis ins Jahr 2018 fortgesetzt werden können.¹ Projektstandorte sind die Universitäten Zürich, Graz und – seit dem Jahr 2012 – Salzburg (bis 2012: Augsburg); die Projektleitung haben Christa Dürscheid, Stephan Elspaß und Arne Ziegler; das Team besteht aus sechs Doktorandinnen und Doktoranden (je zwei an jedem Projektstandort), studentischen und wissenschaftlichen Mitarbeitenden, einem Computerlinguisten und dem Projektkoordinator.

Das Projekt hat sich zum Ziel gesetzt, auf der Grundlage eines großen Korpus von standardsprachlichen Texten aus allen Regionen des zusammenhängenden deutschen Sprachgebiets (unterteilt in 15 Sektoren) die Variation in der Grammatik der geschriebenen deutschen Standardsprache (genauer: in der Grammatik des Gebrauchsstandards) zu erfassen, zu dokumentieren und für Laien wie für Wissenschaftler zugänglich zu machen. Unter ‚Standardsprache‘ wird dabei explizit der *geschriebene* Standard verstanden, d. h. eine Varietät, die „im allgemeinen nicht mit der Muttermilch, sondern in einem jahrelangen, explizit gesteuerten Prozeß als Sprache der Distanz erworben“ wird und die ein Schreiber „unter bestimmten Umständen verwendet“ (Eisenberg 2007: 226). Dabei verstehen wir unter ‚Gebrauchsstandard‘ im Anschluss an Ammon (1995) und Berend (2005a) das, was im Gebrauch der Standardsprache (z. B. in Preetexten) regelhaft vorkommt und mit dem kodifizierten Standard übereinstimmen kann, aber nicht übereinstimmen muss.

Eine umfassende und systematische Darstellung der arealen Variation in der Grammatik der deutschen Standardsprache ist bislang ein Forschungsdesiderat. Dass es eine solche Darstellung noch nicht gibt, hat nicht zuletzt mit historischen, auch

¹ Die Förderung erfolgte für die erste Projektphase über die DFG (EL 500/3-1), den FWF (I 716-G18) und – als Lead-Agency – den SNF (100015L_134895). Die zweite Projektphase wird über den FWF (I 2067-G23) und – wiederum als Lead-Agency – den SNF (100015L_156613) gefördert.

ideologisch aufgeladenen Konzepten von ‚Standardsprache‘ zu tun (vgl. Durrell 2000; Elspaß 2005). So sind viele Grammatiken des Deutschen noch vom Idealbild einer variationsfreien Standardsprache, einer eher sprachökonomischen als sprachökologischen Sichtweise auf Sprache geleitet (vgl. Schmidlin 2011) sowie von dem – auch praktisch-didaktischen – Interesse, ein möglichst homogenes grammatisches System zu beschreiben. Abweichungen davon werden, wie bereits ein Blick in die wichtigsten Grammatiken des Gegenwartsdeutschen zeigt, mit vagen Etiketten wie „umgangssprachlich“, „regional“ oder „landschaftlich“ behaftet (vgl. Elspaß 2010). In der Regel bleibt den Grammatikschreibern auch nichts anderes übrig, als solche Etiketten zu verwenden, da umfassende, auf großen Korpora basierende empirische Studien zur diatopischen Variation in der deutschen Grammatik fehlen und spezifischere Markierungen deshalb kaum möglich sind. Dieses Forschungsdesiderat und der daraus resultierende ‚tote Winkel‘ in der Grammatikographie wie auch in der Grammatikforschung sind die Hauptmotivation für das Projekt. Das Projekt reiht sich damit in neuere Forschungsentwicklungen ein – einschlägig sind in diesem Kontext v. a. die Arbeiten von Mattheier (1997), Christen (2004), Ammon (1995), Eichinger & Kallmeyer (2005), Schmidlin (2011) sowie Lenz & Glauninger (2015); aus sprachdidaktischer Perspektive sind die Publikationen von Eva Neuland zu nennen (z. B. Neuland 2006).

Das Konzept der „nationalen Varietäten“, das in der Forschung seit den 1990er Jahren immer mehr an Bedeutung gewann (vgl. Ammon 1995; von Polenz 1999: 412 ff.), stellt auch für unser Projekt einen Referenzpunkt dar. Doch übernehmen wir die Bezeichnung „nationale Varietäten“ nicht; auch schließen wir uns nicht einem pluri-zentrischen Konzept an, das auf ‚nationale Zentren‘ abhebt (vgl. zur Kritik daran u. a. Reiffenstein 2001; Auer 2013). Denn zum einen machen die Merkmale einer Varietät nicht an der Landesgrenze Halt, zum andern gilt nicht in allen Fällen, dass eine Standardvarietät im ganzen Land im Gebrauch ist. Ersteres ist etwa im Sprachgebrauch in Bayern und Österreich zu beobachten, Letzteres sieht man an der Schweiz, wo Deutsch auf Kantons- und Gemeindeebene nur in 17 der 26 Kantone gesprochen wird. In den Projektpublikationen ziehen wir es daher vor, von ‚arealen‘, nicht von ‚nationalen‘ Varietäten zu sprechen.

Was die bisherige Forschung zur diatopischen Standardvariation des Deutschen betrifft, lag der Fokus v. a. auf lexikalischen, nicht auf grammatischen Phänomenen. So fallen in Ammon (1995) die Passagen, die dem Wortschatz gewidmet sind, erheblich ausführlicher aus als die entsprechenden Abschnitte zur Grammatik. Zum Wortschatz des österreichischen Deutsch etwa gibt es in diesem Übersichtswerk 20 Seiten, zur Grammatik nur knapp zwei Seiten. In den wenigen Abschnitten zur Grammatik werden fast ausschließlich lexemgesteuerte Phänomene dargestellt (v. a. zum Substantivgenus und -numerus), strukturell-syntaktische Aspekte dagegen werden kaum thematisiert. Entsprechendes lässt sich für die Arbeiten von Rash (2002) (zum Schweizer Standarddeutsch) und von Zibrowa (1995) (zum Standarddeutschen in Österreich und der Schweiz) konstatieren. In einem wichtigen Grundlagenwerk für unser Projekt, im *Variantenwörterbuch* (VWB, vgl. Bickel, Hofer & Suter in diesem Band), sind neben

lexikalischen Varianten zwar auch eine Reihe von Phänomenen erfasst, die sich auf der Ebene der Morphosyntax verorten lassen, doch sind diese Angaben nicht korpuslinguistisch abgestützt und bieten nur eine sehr schmale Basis für die grammatische Beschreibung der jeweiligen Varietät. Das sei hier nur kurz am Beispiel der Konstruktion [*am* + Infinitiv] verdeutlicht:

Das VWB (2004: 34) ordnet diese Konstruktion als „Grenzfall des Standards“ ein und kennzeichnet sie als national in „CH D“ verbreitet. Hierzu ergeben sich kritische Fragen: So ist nach Angaben der Duden-Grammatik (2009: 594) die Progressivkonstruktion [*sein* + *am* + Infinitiv] „in der gesprochenen Sprache weiter verbreitet als in der Standardschriftsprache“.² Offenbar ist sie also schon, wenn auch noch nicht sehr verbreitet, in der „Standardschriftsprache“ angekommen. Diesen Schluss muss man auch ziehen, wenn man grammatische Konstruktionen dann als Varianten des Standarddeutschen ansieht, wenn sie mit einer gewissen Frequenz in der Presse belegt sind. Die im *Atlas zur deutschen Alltagsprache* (AdA, vgl. Möller & Elspaß in diesem Band) zu findenden Angaben zur *am*-Konstruktion (vgl. AdA, „Verlaufsformen“) werfen zudem die Frage auf, ob diese grammatische Variante nicht inzwischen in *allen* deutschsprachigen Ländern üblich ist, aber Erweiterungen – etwa durch Objekte des Infinitivverbs (vgl. *ich bin die Äpfel am schälen*) – ein anderes regionales Profil aufweisen als die ‚einfachen‘ Konstruktionen. Auch stellt sich die Frage, mit welchen anderen Funktionsverben (z. B. *halten*, *haben*, *bleiben*) sich der *am*-Infinitiv verbinden lässt (vgl. hierzu Van Pottelberge 2004: 189 ff.) und ob diese Verbindungen wiederum regional beschränkt sind.

Solche Fragen sollen mit der *Variantengrammatik* beantwortet werden. Ihr Ziel ist, standardsprachliche, aber areal eingeschränkt auftretende Phänomene in der Grammatik zu dokumentieren und auf diese Weise die standardsprachliche Variation im Deutschen ebenbürtig zu erfassen – wenn sie auch in ihrem Vorkommen nicht ebenbürtig ist. Denn dass hinsichtlich der Verbreitung von Varianten eine Asymmetrie vorliegt und dass insbesondere großräumig in Deutschland verbreitete Varianten häufig dominieren, ist unbestritten. Dies freilich steht nicht im Widerspruch dazu, dass das Deutsche eine pluriareale Sprache ist und der grammatischen Variation sowohl länderübergreifend als auch innerhalb eines Landes Rechnung zu tragen ist.

Auf den Umstand, dass eine solche Variantengrammatik ein Desiderat darstellt, hat früh schon Ursula Götz (1995) hingewiesen. Sie wertete ausgewählte Grammatiken des Deutschen aus und stellte fest, dass zwar einzelne Fälle von Variation (z. B. unter dem Label „schweizerisch“) erfasst würden, dies jedoch sehr unsystematisch geschehe. Außerdem fänden sich ungenaue Angaben zur arealen Verbreitung. Zu diesem Ergebnis gelangen auch Christa Dürscheid und Patrizia Sutter in ihrem Aufsatz

² So auch in der siebten Auflage der Duden-Grammatik von 2005; und schon der Bearbeiter der sechsten Auflage von 1998 (Rd.nr. 147, Anm. 1) hält fest: „auch standardsprachlich, besonders in der gesprochenen Sprache, möglich“.

„Grammatische Helvetismen im Wörterbuch“ (vgl. Dürscheid & Sutter 2014a, siehe auch Dürscheid & Sutter 2014b). So weisen sie nach, dass im Zweifelsfälle-Duden (vgl. Duden 2011) einige Variantenphänomene, auf die in der Forschungsliteratur hingewiesen wird, gar nicht erfasst sind (z. B. die Reduktion von *dass*-Sätzen auf Verberstsätze in Konstruktionen wie *Es ist gut, kommst du*, vgl. Dürscheid & Hefti 2006). Zudem sind die Ausführungen zu einschlägigen grammatischen Phänomenen (z. B. zur Pluralbildung, Verbflexion, Perfektbildung, Aufspaltung von Pronominaladverbien) oft am norddeutschen Gebrauchsstandard orientiert. Schon die Zusammenstellung von Götz (1995: 226–227) hinterließ den Eindruck, dass in den Grammatiken Markierungen vom Typus „süddeutsch“, „oberdeutsch“, „im Süden des deutschen Sprachgebiets“, „im Schweizerdeutschen“, „süddeutsch/österreichisch“ oder „in der Schweiz und in Österreich“ gegenüber Markierungen wie „in Norddeutschland“, „rheinisch“, „in Berlin und Niedersachsen“ oder „ostmitteldeutsch“ überwiegen, dass die in diesen Regionen vorkommenden standardsprachlichen Varianten demnach also einen markierten Fall darstellen. Eine systematische Auszählung der Markierungen in der CD-Rom-Version des Zweifelsfälle-Dudens erbrachte dazu das folgende Ergebnis (Tab. 1, übernommen aus Dürscheid & Sutter 2014b: 114):

Tab. 1: Areale Markierungen im Zweifelsfälle-Duden.

| „norddt.“ | „süddt.“ | „deutschl.“ | „schweiz.“ | „österr.“ | Gesamt |
|-----------|----------|-------------|------------|-----------|--------|
| 29 | 63 | 10 | 97 | 149 | 348 |

Diese Auszählungen lassen vermuten, dass der Sprachgebrauch in Deutschland resp. in Norddeutschland eher als Normalfall angesehen wird (da hier die wenigsten Markierungen vorliegen). Eine solch deutschlandzentrierte Sichtweise war lange Zeit auch im Deutsch als Fremdsprache-Unterricht vorherrschend. Doch auch hier scheint ein Umdenken eingesetzt zu haben: Im Zusammenhang mit der in den letzten zwei Jahrzehnten immer wieder aufgeworfenen Frage, welches Deutsch man im Ausland lehren soll, wurde wiederholt auf die Realität der arealen Variation im Deutschen hingewiesen (vgl. z. B. Hensel 2000; Berend 2005b sowie Durrell & Langer 2005). So hat nun auch die zweite Auflage des Handbuchs „Deutsch als Fremd- und Zweitsprache“ (vgl. Krumm et al. 2010) das Thema aufgegriffen und enthält ein Kapitel zur diatopischen Variation im Deutschen, das Beiträge zur Dialekt- und Standardsprachsituation im deutschsprachigen Raum umfasst.

1.2 Korpus und Methode

Wie bereits erwähnt, schließt unser Projekt auf theoretischer Ebene an die Arbeiten von Ammon (1995) und an das VWB (2004) an, erweitert diesen Ansatz aber um die

pluriareale Perspektive. Dem trägt der Korpusaufbau Rechnung, der – soweit dies möglich ist – die gesamte areale Bandbreite der Variation im geschriebenen Standard berücksichtigt (und nicht nur relativ wenige, sog. ‚Qualitätszeitungen‘). Mit einer solch breiten Textauswahl sollen nicht zuletzt auch das Leseverhalten und die Bedürfnisse der Sprachteilhaber ernst genommen werden – in dem Sinne, dass ihnen eine verlässliche Antwort auf die Frage gegeben werden soll, in welchen Gebieten des deutschsprachigen Raums man sich mit welchen Varianten „im geschriebenen Standard unauffällig bewegen kann“ (Eisenberg 2007: 226). Für das Projekt sind drei Arbeitsphasen vorgesehen, wobei die ersten beiden inzwischen weitgehend abgeschlossen sind und die dritte Phase nun im Zentrum steht. Die drei Phasen werden im Folgenden aufgelistet und anschließend erläutert:

- (1) Erstellung eines digitalen Korpus des Gebrauchsstandards in Zeitungstexten der deutschsprachigen Länder sowie einer Datenbank grammatischer Varianten,
- (2) vollständige Korpusanalyse und linguistische Beschreibung der ermittelten grammatischen Phänomene,
- (3) Aufbereitung der Forschungsergebnisse und ihre Darstellung in einer Grammatik, die nach Ende der Projektlaufzeit als digitale Publikation erscheinen wird.

zu (1): Das Korpus, das in der ersten Arbeitsphase erstellt wurde und auf den Lokal- und Regionalteilen von 69 Zeitungen aus dem zusammenhängenden deutschsprachigen Raum basiert, umfasst fast 600 Mio. Wörter. Es ist tokenisiert, lemmatisiert, POS-annotiert (TreeTagger und RFTagger) und enthält zusätzlich Annotationen auf syntaktischer Ebene. Die Qualität der automatischen Annotationen wurde durch eine stichprobenartige Kontrolle von 898 Tokens evaluiert. Berücksichtigt wurden dabei Lemma-, Encoding- und Tagging-Merkmale (POS und RF): Die Lemma-Erkennung weist eine Fehleranfälligkeit von 4,55 % auf, wobei hier die Schwachstellen bei Eigennamen und bei fremdsprachigem Material liegen, also in einem Bereich, der bei automatisierter Annotation immer durch eine relativ hohe Fehlerquote charakterisiert ist. Die Encoding-Fehler sind statistisch nicht relevant; in der Stichprobe gab es nur drei Annotationsfehler, die auf das Encoding zurückgeführt werden konnten. Das POS-Tagging führt im nominalen Bereich und bei der Erkennung von Verbalformen zu Fehlern, doch liegt die Quote hier jeweils unter 5 % und ist damit im akzeptablen Bereich. Bessere Ergebnisse liefert der RF-Tagger, der insgesamt eine Fehlerhäufigkeit von nur 3 % aufweist. Davon betreffen die meisten Annotationsfehler Verbalformen, während die Attributzuweisung im nominalen Bereich sehr gute Ergebnisse zeigt. Außerdem ist kein Zusammenhang bei den fehlerhaften Zuweisungen zwischen POS- und RF-Tagger zu erkennen. Da das Abfragesystem eine Suche mit beiden Taggern gleichzeitig ermöglicht, kann die Fehlerquote so auf ein Minimum reduziert werden, was die Reliabilität der Korpusabfragen sicherstellt.

Korpusaufbau und Annotation der Korpusdaten sind mittlerweile zu einem Abschluss gekommen. Ergänzt wurde das Großkorpus um die Daten aus einem Pilotkorpus, das bereits im Vorfeld des Projekts erstellt wurde, knapp 6,4 Mio. Wörter umfasst

und es dem Projektteam erlaubte, schon vor Fertigstellung des Großkorpus Hypothesen zu Variationsphänomenen zu überprüfen und auf dieser Basis erste Forschungsergebnisse zu publizieren (z. B. in der Arbeit von Dürscheid, Elspaß & Ziegler 2015). Dieses Pilotkorpus beschränkt sich nur auf Quellen aus Österreich, Deutschland und der Schweiz und ist – anders als das Großkorpus – nicht weiter in Sektoren untergliedert.

Parallel zum Korpusaufbau haben wir in der ersten Arbeitsphase eine Variantendatenbank erstellt, die auf einer systematischen Durchsicht zahlreicher Referenzwerke (Wörterbücher, Grammatiken und wissenschaftlichen Darstellungen) basiert, sukzessive aber auch um solche Phänomene ergänzt wurde, von denen die Projektmitglieder aufgrund der Arbeit mit dem Pilotkorpus und eigener Beobachtungen (z. B. Belege aus der Presse) annahmen, dass es sich um diatopische Varianten handeln könnte (z. B. [*genug* + Adjektiv], vgl. *Das Eis ist genug dick*). Ob dies tatsächlich der Fall ist, werden die Suchabfragen im Großkorpus erweisen; ergibt sich kein signifikanter Wert, werden diese Phänomene nicht als Varianten in die *Varietengrammatik* aufgenommen. Dank der Kooperation mit dem Zentrum für Informationsmodellierung der Universität Graz steht die Variantendatenbank nun seit Dezember 2012 als webbasiertes, komfortables Arbeitsinstrument zur Verfügung. Eine Datenmaske erlaubt die Speicherung und Verknüpfung einer Vielzahl von Informationen (so u. a. Angaben zu den Referenzwerken, die das Phänomen nennen). Durch den Web-Zugriff ist zudem sichergestellt, dass alle Mitarbeitenden die Fortschritte in den diversen Arbeitsbereichen wahrnehmen und sich gegenseitig Feedback geben können, was eine wesentliche Grundlage für die Qualitätssicherung darstellt.

zu (2): Das Korpus wird derzeit an den drei Projektstandorten, getrennt nach verschiedenen grammatischen Phänomenbereichen (z. B. Wortstellung, Kasusreaktion, Nominal- und Verbalflexion), systematisch bearbeitet. Im Folgenden werden die wichtigsten Schritte in diesem Arbeitsprozess dargestellt, um einen Eindruck von der Vorgehensweise im Projekt zu geben: Bei den Suchabfragen muss auf eine Optimierung der Belege geachtet werden, d. h. es ist sicherzustellen, dass eine möglichst geringe Anzahl Fehltreffer in die Auswertung eingeht. Im Anschluss daran wird die Signifikanz der Ergebnisse berechnet. Ist die Anzahl gefundener Belege zu gering und somit absehbar, dass sie keine signifikanten Aussagen liefern, werden sie nicht weiter bearbeitet. Hierzu wurde festgelegt, dass im gesamten Korpus wie auch in einzelnen Sektoren (in denen das gesuchte Phänomen überhaupt belegbar ist) drei verschiedene Schwellenwerte (10, 25, 50) getestet werden, um abschätzen zu können, welche Phänomene je nach Schwellenwert unberücksichtigt bleiben. Finden sich 50 bis 500 Belege, erfolgt eine manuelle Durchsicht der Treffer. Sind die Ergebniszahlen zu hoch (> 500), um händisch bearbeitbar zu sein, werden 100 Belege zufällig ausgewählt, um die Quote der Fehltreffer zu ermitteln. Ist die Anzahl der Fehler über 15 %, muss die Suchanfrage verbessert oder das entsprechende Phänomen als nicht auswertbar qualifiziert werden. Ist die Fehlerquote unter 15 %, wird ein Sample von Fehlerquote mal 100 Belegen ausgewählt, händisch ausgewertet und auf die Gesamt-

menge hochgerechnet. So können auch Phänomene mit relativ hohen Belegzahlen zuverlässig bearbeitet werden. Was die statistische Auswertung der Treffer betrifft, ist grundsätzlich zu unterscheiden, ob Frequenzen ohne Vergleichszahl ausgewertet werden (z. B. [Funktionsverb + *am/beim* + Infinitiv]-Konstruktionen) oder ob zwei (oder mehr) Varianten miteinander verglichen werden (z. B. Genusvarianten). Über die Suchanfragen kann also zum einen die Frequenzverteilung einer Variante im Korpus ermittelt werden, zum andern, wo möglich, auch das Verhältnis zwischen zwei (oder mehr) Varianten. In beiden Fällen kann eine Chi-Quadrat-Auswertung eine Aussage zur Signifikanz liefern.

zu (3): Die dritte Arbeitsphase umfasst die distributionell-quantitativen und grammatischen Analysen der als statistisch relevant ermittelten Variantenphänomene und die schriftliche Aufbereitung der Ergebnisse für die *Variantengrammatik*. Diese wird sowohl Grundlagen-, Definitions-, Überblicks- und Einzelartikel umfassen als auch Karten, die nach den im Projekt bearbeiteten 15 Sektoren gegliedert sind und das areale Vorkommen der jeweiligen Variante veranschaulichen sollen. Die Artikeltexte werden auf einer kollaborativen Schreibplattform über Eingabemasken erstellt, die für die verschiedenen Artikeltypen einen jeweils spezifischen Aufbau haben. Dieses Wiki mit seinen verschiedenen Templates dient aber nicht nur als Grundlage für das Schreiben der Artikel, es wird auch die Basis für die digitale Publikation der *Variantengrammatik* sein. Im Vergleich zu der ursprünglich geplanten Printversion bietet eine solch digitale Ausgabe weitaus mehr Möglichkeiten der Informationsaufbereitung (z. B. in der Verlinkung), und es können neben näherungsweise Frequenzmarkierungen (z. B. ‚sporadisch‘, ‚gebräuchlich‘, ‚üblich‘ u. a.) auch präzisere Angaben (z. B. Prozentangaben zur Verwendung von *nachdem* als temporaler und als kausaler Subjunktion in einzelnen Regionen) gemacht werden. Außerdem sind in der digitalen Version umfassende kartographische Darstellungen zur arealen Variation möglich, was in einem einbändigen Druckwerk nur exemplarisch Platz finden könnte.

1.3 Umsetzung

Die Erstellung eines regional gewichteten Zeitungskorpus wurde im Frühsommer 2013 abgeschlossen. Mit fast 600 Mio. Wörtern ist das Korpus weitaus größer als ursprünglich geplant. Hinzu kommen die Daten aus dem Pilotkorpus (s. o.). Die Variantendatenbank umfasst derzeit ca. 3 500 Einträge, von denen ca. 2 000 auf Einzelphänomene aus dem Bereich der Wortbildung entfallen. Die Datenbank ist nach den folgenden sieben Kategorien untergliedert, die ihrerseits die Basis für die Struktur des Wiki darstellen: (1) Flexion, (2) Gebrauch des Artikels, (3) Genus, (4) Phraseologismen, (5) Valenz/Rektion, (6) Wortbildung, (7) Wortstellung/Satzgliedstellung. Die distributionell-quantitative Korpusanalyse sowie die linguistische Beschreibung der ermittelten grammatischen Phänomene sind ebenfalls zu großen Teilen abgeschlossen. Erste Beispielpartikel aus der *Variantengrammatik* wurden bereits veröffentlicht.

licht; sie sind in einem Themenheft des *Sprachspiegel* (vgl. Businger & Rimensberger 2014) zu finden. Einer dieser Artikel, der die Stellung des finiten Verbs in zusammengesetzten Tempusformen betrifft, wird zur Illustration auch im vorliegenden Beitrag abgedruckt (s. u. Kap. 2.3). Weitere projektbezogene Publikationen, in denen Teilergebnisse aus der Korpusanalyse vorgestellt werden, sind Elspaß & Niehaus (2014), Niehaus & Scherr (2013), Dürscheid & Sutter (2014a, b) sowie Dürscheid, Elspaß & Ziegler (2015). Eine erste Dissertation, die aus dem Projekt hervorgegangen ist und sich historisch ausgreifend mit der Wortstellungsvariation im Neuhochdeutschen am Beispiel von Genitivattributen, Verbclustern und Ausklammerung befasst, ist Niehaus (2014). Fünf weitere Dissertationsprojekte werden derzeit zum Abschluss gebracht. Es sind dies die Arbeiten von Patrizia Sutter zur diatopischen Variation im Wörterbuch, von Susanne Oberholzer zum Dialekt- und Standarddeutschgebrauch in Deutschschweizer Kirchen, von Anna Thurner zu Strukturpräferenzen im Spannungsfeld von Gradienz und Frequenz, von Elisabeth Scherr zu den epistemischen Modalitätsstrategien im Gebrauchsstandard und von Julia Engel zur regional unterschiedlichen Verb-, Substantiv- und Adjektivrektion.

Neben der Arbeit an den Dissertationen und der laufenden Publikation von Einzelveröffentlichungen liegt der Schwerpunkt der Arbeit nun auf dem Schreiben der Artikel. Dies geschieht in dem oben bereits erwähnten Wiki. Dieses soll am Ende der zweiten Projektlaufzeit als Open-Access-Publikation der wissenschaftlichen Gemeinschaft zur Verfügung gestellt werden. Im Folgenden wird ein Ausschnitt aus dem Projekt-Wiki wiedergegeben, um einen Eindruck davon zu vermitteln, welche Möglichkeiten für das Artikelschreiben auf dieser Plattform zur Verfügung stehen:

Hilfeseite Diskussion

Hilfe:Wikisyntax und -struktur

Dieses Papier enthält Hilfestellungen und allgemeine Hinweise zum Art dazu genutzt werden, um sich über die hier thematisierten Aspekte aus:

| Inhaltsverzeichnis [Verbergen] | |
|--------------------------------|----------------------------------------------------------------------|
| 1 | Erstes Artikelschreiben / Bearbeiten von Seiten / Löschen von Seiten |
| 2 | Überschriften von Artikeln/Seiten |
| 3 | Templates / fertige Artikel als Vorlage |
| 4 | Interne Links und Registereinträge |
| 5 | Kategorienstruktur / Kategorien |
| 6 | Sprache im Wiki |
| 7 | Hochladen von Dateien |
| 8 | Zusammenfassung Wiki-Syntaxbefehle |
| 9 | Hilfestellungen |

Abb. 1: Screenshot aus dem Projekt-Wiki.

Wie Abb. 1 zeigt, ist das Wiki u. a. nach „Kategorien“ und „Artikeltypen“ untergliedert. Die Kategorien entsprechen den in der Variantendatenbank erfassten Phänomenbereichen, unter den Artikeltypen sind Grundlagenartikel, Überblicksartikel und Einzelartikel erfasst; die Definitionsartikel werden zu einem späteren Zeitpunkt ergänzt. Jedes Projektmitglied ist für einen Phänomenbereich zuständig, sichtet dazu die Angaben in der Variantendatenbank, führt die Suchabfragen im Korpus durch, verfasst die Einzelartikel (z. B. zu *ändern/sich ändern*; *rentieren/sich rentieren*) und – falls erforderlich – auch die Überblicksartikel, auf die von den Einzelartikeln aus verwiesen wird (und vice versa). Im Sinne des Vier-Augen-Prinzips werden im Anschluss daran alle diese Arbeiten von einem zweiten Projektmitglied überprüft und abschließend von der Projektleitung einer letzten Durchsicht unterzogen. Parallel dazu arbeitet die Projektleitung derzeit an den Grundlagenartikeln zu den sieben Phänomenbereichen, die mehrere Unterpunkte umfassen (z. B. Wortbildung > Substantiv > Derivation; Wortbildung > Substantiv > Komposition > Fugenelemente). Exemplarische Auszüge aus diesen Grundlagenartikeln sind ebenfalls im *Sprachspiegel* einsehbar (vgl. Businger & Rimensberger 2014). Im weiteren Projektverlauf wird die Bearbeitung aller in der Datenbank erfassten Varianten in der eben geschilderten Weise fortgesetzt; zudem wird weiter am Aufbau des Portals für die *Variantengrammatik* (inkl. aller Verlinkungen, Karten, möglicher Suchabfragen) gearbeitet. Die Komplettierung aller Arbeiten ist für das Jahr 2018 vorgesehen.

2 Inhaltliche Vertiefung

2.1 Vorbemerkungen

Die *Variantengrammatik* wird das erste Handbuch sein, in dem die diatopische Variation auf grammatischer Ebene systematisch dargestellt ist und die Möglichkeiten ausgeschöpft sind, welche die Digitalisierung bietet (z. B. bzgl. Verlinkung, Multimodalität, Belegeinsicht, kartographischer Aufbereitung). Die grammatische Beschreibung der Phänomene in den Grundlagenartikeln soll den Ansprüchen an eine wissenschaftliche Grammatik genügen, sie soll aber – ähnlich der Duden-Grammatik – auch für linguistische Laien verständlich sein. Die distributionell-quantitativen Ergebnisse aus den Korpusanalysen werden Eingang in die Einzel- und Überblicksartikel finden (z. B. die statistischen Berechnungen zur Verwendung von Partikel- resp. Präfixverben), und in den Definitionsartikeln werden ergänzende Erläuterungen zu wichtigen grammatischen Termini gegeben (z. B. Morphem). Im Folgenden werden vier Beispiele für solche Artikel angeführt, aber auch die Arbeitsweise im Projekt genauer dargestellt. Dabei wird zum einen auf das Pilotkorpus Bezug genommen (in Kap. 2.2), das bereits zu Beginn der Projektarbeit zur Verfügung stand, zum anderen auf das Groß-

korpus (in Kap. 2.3), auf dessen Basis seit Abschluss der Korpusarbeiten alle quantitativen Analysen erfolgen.

2.2 Statistische Berechnungen am Beispiel der Verbrektion

Zur Verbrektion wird es einen Grundlagenartikel geben, in dem die wichtigsten grammatischen Informationen zusammenfassend dargestellt sind. Daneben ist ein Überblicksartikel geplant, der die diatopischen Unterschiede in der Verbrektion beschreibt. Hinzu kommen Einzelartikel zu den Verben, die eine variable Rektion aufweisen (z. B. *rufen* + Dativ/*rufen* + Akkusativ). Einige dieser Verben sind in der Literatur bereits erfasst und haben deshalb Eingang in unsere Variantendatenbank gefunden. Dazu gehören z. B. die Verben *vergessen* und *kündigen*. Im *Variantenwörterbuch* heißt es unter dem Eintrag zu *vergessen*, dass die Variante [*vergessen* + *auf*] in Österreich standardsprachlich sei (in der Bedeutung von ‚an etwas nicht rechtzeitig denken; etwas vergessen‘). Man erfährt aber nicht, auf welcher quantitativen Basis die Autoren zu diesen Angaben gelangen und wie das Phänomen auf grammatischer Ebene einzuordnen ist. Hier wird ein Mehrwert der *Variantengrammatik* deutlich. Über die Korpusanalyse wird ermittelt, wie häufig die Variante [*auf etwas vergessen*] im Korpus vorkommt, und über die statistische Berechnung mittels Chi-Quadrat-Test angegeben, ob es sich um eine Variante handelt, deren Auftreten in einem (oder mehreren) der 15 Sektoren statistisch signifikant ist.

Zur Illustration seien nun zwei Beispiele aus dem Pilotkorpus angeführt, im Anschluss daran wird am Beispiel des Verbs *vergessen* gezeigt, wie die Signifikanz von Varianten berechnet wird. Beide Belege stammen aus österreichischen Zeitungen (d. h. aus dem A-Subkorpus):

- (1) Aufgrund seiner unauffälligen Erscheinung *vergisst* man zu leicht *auf* die 67 PS (49 kW), die sich im Motorraum verstecken.
- (2) Stilistisch an aktuellen wie gleichsam klassischen Sounds aus England orientiert, entfaltet das neue Album ein mannigfaltiges Kaleidoskop unterschiedlichster Genres, ohne dabei *auf* die eigene Handschrift zu *vergessen*.

In den 69 resp. 55 Fällen, in denen *vergessen* im CH- resp. D-Subkorpus vorkommt, regiert dieses Verb den Akkusativ. Im A-Subkorpus dagegen tritt das Verb *vergessen* mit zwei Rektionsvarianten auf: 116 Mal regiert es den Akkusativ, 16 Mal die Präposition *auf* (vgl. die beiden obigen Beispiele). Das sind immerhin 12,6 % aller Belege von *vergessen* aus dem Pilotkorpus. Führt man mit diesen Zahlen nun den Chi-Quadrat-Test durch, dann zeigt sich, dass das Resultat im Ländervergleich mit größter Wahrscheinlichkeit nicht zufällig sein kann. Das Ergebnis, das sich über diesen Test berechnen lässt, ist $\chi^2 = 16,686$, $df = 2$, $p < 0,001$. Wie diese Berechnung zeigt, unterscheidet

sich der Gebrauch des Verbs *vergessen* im A-Subkorpus in hoch signifikantem Maß vom Gebrauch dieses Verbs im CH- und D-Subkorpus. Das gilt auch für das Verb *kündigen*, zu dem in der *Variantengrammatik* ebenfalls ein Einzelartikel vorgesehen ist. Die Abfrage im Pilotkorpus ergibt dazu folgendes Bild:

Im A-Subkorpus stehen alle 13 Treffer zu *kündigen* mit Akkusativ. Ein Beispiel hierfür stammt aus dem *Echo Salzburg*: „Er sagte zum anderen Gesprächsteilnehmer, dass dieser als Koch anfangen könne und seinen Lehrling werde er einfach kündigen.“ Führt man den Chi-Quadrat-Test durch, ergibt sich im Ländervergleich ein Wert von $\chi^2 = 18,054$, $df = 2$, $p < 0,001$, d. h. es kann mit 99,9 % Wahrscheinlichkeit davon ausgegangen werden, dass es sich bei diesem Resultat nicht um einen Zufallswert handelt, sondern die Akkusativreaktion bei *kündigen* in den österreichischen Daten signifikant ist.

Zum Schluss sei noch ein Verb genannt, dessen Rektion in keinem Nachschlagewerk als diatopisch variabel angegeben wird. Es ist dies das Verb *beantragen*, das, so unsere Vermutung, in Schweizer Standardtexten unter bestimmten Bedingungen mit einem Dativobjekt auftreten kann (vgl. als konstruiertes Beispiel: *Die Kommission beantragt der Fakultät, das Verfahren fortzuführen.*). Hier stellt sich die Frage, ob dies nur eine formelhafte Verwendung im akademischen Kontext ist oder die Dativreaktion auch in anderen Kontexten vorkommen kann. Konsultiert man dazu das Pilotkorpus, dann zeigt sich folgendes Bild: Das Verb *beantragen* regiert in 76 % aller Fälle, in denen es im CH-Subkorpus verwendet wird, ein Dativobjekt. Die Belege lassen darauf schließen, dass diese Konstruktion v. a. dann vorkommt, wenn es um Beantragungen mit offiziell-amtlichem Charakter geht. Alternativ dazu finden sich im Korpus aber auch Belege für die Verwendung von *beantragen* mit der Präposition *bei*, *zuhanden* oder *an*. Im Folgenden werden hierfür vier Beispiele aus dem CH-Subkorpus gegeben:

- (3) Der Regierungsrat des Kantons Bern beantragt dem Grossen Rat, die Auflösung der Burgergemeinde Meiringen zu genehmigen. (*beantragen* + Dativ)
- (4) Der Pächter kann im Falle einer ordentlichen Pachtauflösung beim Richter eine einmalige Erstreckung der Pacht beantragen. (*beantragen* + *bei*)
- (5) An die Erneuerung der Traubachbrücke und der Bohlbachbrücke in Habkern mit einer Gesamtkostensumme von 1,4 Millionen Franken wird zuhanden des Kantons ein zinsloses Bundesdarlehen von 280'000 Franken beantragt. (*beantragen* + *zuhanden*)
- (6) Für folgende Baugesuche wurde die Bewilligung an das Regierungsstatthalteramt beantragt oder erteilt. (*beantragen* + *an*)

Im D-Subkorpus wie auch im A-Subkorpus liegt der Prozentsatz von *beantragen* + Dativ dagegen bei 0 %, hier gibt es also keinen einzigen Beleg zur Dativreaktion

von *beantragen*. Der Chi-Quadrat-Test ergibt im Ländervergleich den Signifikanzwert $p < 0,001$, d. h. das Ergebnis ist wiederum hoch signifikant. Es steht somit fest, dass es zu diesem Verb einen Einzelartikel geben wird, in dem die Rektionsunterschiede erläutert und die Distribution der Rektionsvarianten anhand einer Karte dargestellt werden.

2.3 Beispiele für Musterartikel³

Im Folgenden wollen wir anhand von vier Fallbeispielen zeigen, wie die distributionell-quantitativen Befunde in den Überblicks- und Einzelartikeln unserer *Varietätsgrammatik* dargestellt werden. Weiters wird zur Illustration unserer Arbeit am Beispiel von „(der) Kosovo“ der Weg von den Rohdaten zum entsprechenden Einzelartikel nachgezeichnet. Vorweg ist aber noch eine Bemerkung zu den quantitativen Angaben erforderlich: In den Artikeln finden sich – der besseren Lesbarkeit halber – Frequenzaussagen zur diatopischen Distribution in den einzelnen Sektoren, wie z. B. „sporadisch“ oder „gebräuchlich“. Diesen Angaben liegen Zahlenwerte zugrunde, die vom Projektteam wie folgt festgelegt wurden:

- „sporadisch“: die Variante kommt zu weniger als 5 % vor,
- „kommt (selten) vor“: die Variante kommt zu 5–20 % vor,
- „gebräuchlich“: die Variante kommt zu 21–50 % vor,
- „üblich“/„mehrheitlich verwendet“: die Variante kommt zu 51–94 % vor,
- „gilt (fast) ausnahmslos“: die Variante kommt zu 95 % oder mehr vor.

Weiter werden Vorkommen mit weniger als zehn absoluten Belegen in einem Sektor für den betreffenden Sektor mit einem Asterisk (*) gekennzeichnet. In Anbetracht der niedrigen Belegzahlen mag es verwundern, dass solche Angaben (die insbesondere die kleinen Länder bzw. die kleinen Sektoren betreffen) überhaupt aufgenommen werden, doch kann es durchaus sein, dass diese für die Dokumentation diatopischer Variation relevant sind. So erscheint es plausibel, bei einem Verhältnis von 8 : 2 für Variante A gegenüber Variante B keine zufällige Mehrheit von Variante A anzunehmen, wenn sich in einer angrenzenden Region eine solche Mehrheit statistisch signifikant belegen lässt.

(a) Überblicksartikel: *Artikelgebrauch bei Länder- und Landschaftsnamen*

Als erstes sei ein Beispiel für einen Überblicksartikel in der *Varietätsgrammatik* gegeben. Der Artikel bezieht sich auf die Verwendung des Artikels bei Länder- und Land-

³ Für die Hilfe bei den Auswertungen danken wir Julia Engel, Konstantin Niehaus und Gerard Adarve.

schaftsnamen. Vorweg sei noch darauf hingewiesen, dass wir uns bei den Siglen für die jeweils genannten Länder und Regionen an der „Liste der arealen Abkürzungen“ im VWB (2004: XVIII) orientieren, also z. B. das Kürzel STIR für ‚Südtirol‘ verwenden, LIE für ‚Liechtenstein‘ und A-MITTE für ‚Oberösterreich und Salzburg‘.

Artikelgebrauch bei Länder- und Landschaftsnamen

Die Mehrzahl der Ländernamen sowie viele Landschaftsnamen werden standardsprachlich ohne Artikel gebraucht (*Ungarn, Argentinien, Anatolien*). Dabei handelt es sich meist um Länder- und Landschaftsnamen mit neutralem Genus. Verwendet wird der definite Artikel bei femininen und pluralischen sowie bei attribuierten Raumnamen (= Länder- und Landschaftsnamen): *die Türkei, die USA, das schöne Italien, das kalte Sibirien*. Zudem ist bei einigen Landschaftsnamen im Neutrum der Artikelgebrauch lexikalisiert, sodass auch diese immer mit definitivem Artikel stehen (*das Elsass, das Allgäu*). Handelt es sich um eine Konstruktion mit Präposition, so verschmelzen im Falle der Verwendung des Artikels häufig Präposition und Artikel: *im Breisgau*. Bei Aufzählungen oder Listen geographischer Namen kann der definite Artikel grundsätzlich aber auch entfallen. Die Verwendung des Artikels ist in bestimmten Konstruktionen in Verbindung mit allen Länder- und Landschaftsnamen möglich (*Ein Serbien, wie es niemand kennt. ... das Serbien seiner Jugend...*). Variation besteht bei einigen Ländernamen, die mit oder ohne definiten Artikel verwendet werden und auch in der Deklination schwanken können, hier handelt es sich häufig um jüngere Ländernamen: *Iran, Irak, Jemen, Kongo, Libanon, Niger, Oman, Senegal, Sudan, Tschad, Kosovo*. Im Sprachgebrauch herrscht hier zumeist die Verwendung des Artikels vor. Lediglich bei *Niger* und *Oman* ist gemeindeutsch neben dem definiten auch der Nullartikel gebräuchlich, auch *Senegal* wird ohne Artikel verwendet, dies jedoch nur selten. Regionale Variation in der Standardsprache zeigt sich bei den Ländernamen *Iran* und *Kosovo*, bei dem bei Verwendung des definiten Artikels zudem das Genus variieren kann. Die artikellosen Varianten der Ländernamen sind neben der mehrheitlich verwendeten Variante mit Artikel nur in CH gebräuchlich, sonst treten sie selten auf. Bei den Landschaftsnamen variiert der Artikelgebrauch bei *Vorarlberg* und *Tirol*: Die Standardsprache in CH tendiert hier im Gegensatz zum restlichen deutschen Sprachgebiet eher zu einer Verwendung des Artikels.

(b) Einzelartikel: Kosovo

Möchte der Benutzer der *Variantengrammatik* nun wissen, wie sich der Artikelgebrauch bei bestimmten Ländern darstellt, kann er den Einzelartikel konsultieren, von dem aus mit einem Link auf den Überblicksartikel verwiesen wird (s. den markierten Text am Ende des Artikels):

Kosovo – Artikelgebrauch/Genus: – CH, LIE, LUX; **das** D; **der** ÜBERALL; *Kosovo* wird in D und A fast ausnahmslos mit Artikel verwendet (auch in *BELG und *STIR). In CH, LIE und LUX ist neben der mehrheitlichen Verwendung des Artikels auch die artikellose Variante in Gebrauch, in LIE und LUX allerdings nur selten. *Nebenan gibt eine Roma-Frau aus Kosovo ein Fernsehinterview* (Bernerzeitung, CH); *Für sie war es der erste Besuch in Kosovo* (St. Galler Tagblatt, CH); *Kosovo gehört nicht unbedingt zu den Topnationen im europäischen Handball* (Luxemburger Tageblatt, LUX). Wird der definite Artikel verwendet, so ist nur in D neben dem maskulinen auch neutrales Genus gebräuchlich. *Voraussichtliche Empfängerländer sind in diesem Jahr Bulgarien, Georgien, Kasachstan, das Kosovo, Moldau, die Mongolei, Polen, Rumänien, Serbien, die Slowakei, Weißrussland und das Westjordanland* (Osna-

brücker Zeitung, D-NORDWEST); *Das Kosovo bleibt Schwerpunkt der Hilfeleistung, da es das ärmste Land Europas ist* (Oberbayerische Volkszeitung, D-SÜDOST). In A hingegen wird ausnahmslos maskulines Genus verwendet. Dies scheint auch im restlichen Gebiet üblich zu sein. *Der 17. Februar 2008 ist für den Kosovo bekanntlich ein historisches Datum* (Der Vinschger, STIR). Die Variante mit neutralem Artikel (wie sie in D vorkommt) kann im Übrigen als ein Sonderfall bezeichnet werden. Die Verwendung des Artikels ist hier in Zusammenhang mit der historischen und politischen Entwicklung des Kosovos von der Region zum (von den meisten Mitgliedsstaaten der UN) anerkannten Staat zu sehen. Denn während Ländernamen im Neutrum im Normalfall ohne Artikel gebraucht werden, ist die Verwendung des neutralen Artikels bei Landschaftsnamen in vielen Fällen der Standard (vgl. Artikelgebrauch bei Länder- und Landschaftsnamen).

An dieser Stelle sei noch kurz erläutert, auf welche quantitativen Daten sich Aussagen wie „ausnahmslos“, „üblich“, „gebräuchlich“ stützen; es soll also der Weg von der Korpusanalyse zum Einzelartikel dargestellt werden. Exemplarisch wird die Vorgehensweise anhand des Artikelgebrauchs beim Substantiv *Kosovo* gezeigt (analog dazu verläuft auch die Analyse der Genusvariation bei *Kosovo*). Die Resultate der Auszählungen (vgl. Tab. 2) werden hier allerdings nur auf Länderebene angeführt, da es in diesem Fall zwischen den Regionen keine signifikanten Unterschiede gibt.

Wie in Tab. 2 zu sehen, sind in den beiden rechten Spalten die für die relativen Zahlen eingeführten Kodierungen (z. B. „sporadisch“) vermerkt. Dass die hier angeführten Resultate, die auch im Wiki zugänglich gemacht werden sollen, keine zufällige Verteilung widerspiegeln, zeigt die Durchführung des Chi-Quadrat-Tests, der im

Tab. 2: Ergebnis der Korpusanalyse für *Kosovo*.

| <i>Kosovo</i> | ohne Artikel | mit Artikel | gesamt | ohne Artikel | mit Artikel | ohne Artikel | mit Artikel |
|---------------|--------------|-------------|--------|--------------|-------------|---------------------|-----------------------|
| | (absolut) | | | (%) | | | |
| BELG | 0 | 4 | 4 | 0 | 100 | — | *gilt ausnahmslos |
| LIE | 1 | 6 | 7 | 14 | 86 | *kommt (selten) vor | *üblich/mehrheitlich |
| STIR | 1 | 20 | 21 | 5 | 95 | kommt (selten) vor | gilt fast ausnahmslos |
| LUX | 6 | 22 | 28 | 21 | 79 | gebräuchlich | üblich/mehrheitlich |
| D | 7 | 714 | 721 | 1 | 99 | sporadisch | gilt fast ausnahmslos |
| A | 15 | 440 | 455 | 3 | 97 | sporadisch | gilt fast ausnahmslos |
| CH | 38 | 67 | 105 | 36 | 64 | gebräuchlich | üblich/mehrheitlich |

vorliegenden Fall einen Wert von $\chi^2 = 256,4107$, $df = 6$, $p = < 0,001$ ergibt und damit die Signifikanz der Daten bestätigt.

(c) Einzelartikel: (E-)Mail

Das nächste Fallbeispiel betrifft den Genusgebrauch bei Substantiven. Entlehnte Substantive können das Genus aus der Herkunftssprache übernehmen (z. B. *die Saison* von frz. *la saison*), dies ist aber nicht notwendigerweise der Fall (z. B. *das Püree* von frz. *la purée*).⁴ Bei Entlehnungen aus Sprachen wie dem Englischen, in denen es keine Genusunterscheidung gibt, ist die Genuszuweisung zum Zeitpunkt der Entlehnung prinzipiell offen. Sie erfolgt erst in einem Prozess der Konventionalisierung, in dem formale und/oder semantische Prinzipien zusammenspielen. Im Ergebnis können Sprecher verschiedener Regionen durchaus verschiedene Genera präferieren. Das folgende Beispiel zeigt die Ergebnisse unserer Korpusanalyse wieder an einem Einzelfall auf, an der Genusvariation bei (E-)Mail. Tab. 3 stellt die Rohdaten dar.

Der Chi-Quadrat-Test ergibt hier im Regionenvergleich einen Wert von $\chi^2 = 1961,841$, $df = 14$, $p = < 0,001$, d. h. dass die Verteilung der Resultate nicht zufällig ist. Legt man diese Ergebnisse zugrunde, dann sind sowohl die Einträge im *Variantenwörterbuch* als auch in *Duden Online* zu präzisieren. Zwar kann die Angabe im VWB (2004: 215) bestätigt werden, dass *E-Mail* „in A meist Neutrum, in D meist Femininum“ ist. Allerdings weicht der Sprachgebrauch in A-West – wie so häufig – auch in diesem Fall vom Gebrauch in den anderen Regionen Österreichs deutlich ab, und auch die Aussage im VWB, dass das Genus „in CH schwankend“ sei, ist im Vergleich mit den Befunden für die österreichischen Regionen nicht berechtigt („schwankend“ wäre, wenn man diese Kategorie überhaupt übernehmen wollte, der Gebrauch in A-SÜDOST). *Duden Online* (1. Juli 2014) gibt – klar deutschlandzentriert – zunächst die Variante *die E-Mail* an und fügt sachlich ungenau an: „auch, besonders süddeutsch, österreichisch, schweizerisch: das E-Mail [...]“. Der Einzelartikel zum Genusgebrauch von (E-)Mail gestaltet sich auf der Basis dieses Korpusbefundes wie folgt:

(E-)Mail – Genus: In D, BE und LUX fast ausnahmslos *die E-Mail*. In A-WEST und STIR üblicherweise *die E-Mail*, daneben auch *das E-Mail* gebräuchlich. In CH und den anderen Gebieten von A mehrheitlich *das E-Mail*, aber auch *die E-Mail* gebräuchlich.

⁴ Ob in den Fällen, in denen das Genus der Herkunftssprache mit dem Genus im Deutschen übereinstimmt, tatsächlich eine Entlehnung anzunehmen ist oder andere Faktoren eine Rolle spielen, ist nicht immer eindeutig: Bei *Saison* ist beispielsweise auch eine semantische Motivierung (*Jahreszeit*) denkbar (vgl. Schulte-Beckhausen 2002: 35).

Tab. 3: Ergebnis der Korpusanalyse für *die/das (E-)Mail*.

| | <i>die (E-)Mail (absolut)</i> | <i>das (E-)Mail</i> | <i>gesamt</i> | <i>die (E-)Mail (%)</i> | <i>das (E-)Mail</i> | <i>die (E-)Mail</i> | <i>das (E-)Mail</i> |
|--------------|---------------------------------------|-------------------------|---------------|---------------------------------|-------------------------|-----------------------------|-------------------------|
| D-NORDWEST | 500 | 4 | 504 | 99 | 1 | gilt fast ausnahmslos | sporadisch |
| D-MITTELOST | 552 | 5 | 557 | 99 | 1 | gilt fast ausnahmslos | sporadisch |
| D-NORDOST | 593 | 7 | 600 | 99 | 1 | gilt fast ausnahmslos | sporadisch |
| D-MITTELWEST | 595 | 4 | 599 | 99 | 1 | gilt fast ausnahmslos | sporadisch |
| D-SÜDWEST | 472 | 13 | 485 | 97 | 3 | gilt fast ausnahmslos | sporadisch |
| D-SÜDOST | 884 | 19 | 903 | 98 | 2 | gilt fast ausnahmslos | sporadisch |
| LUX | 19 | 1 | 20 | 95 | 5 | gilt fast ausnahmslos | kommt (selten) vor |
| BELG | 46 | 1 | 47 | 98 | 2 | gilt fast ausnahmslos | sporadisch |
| A-WEST | 56 | 33 | 89 | 63 | 37 | üblich/ mehrheitlich | gebräuchlich |
| A-MITTE | 64 | 79 | 143 | 45 | 55 | gebräuchlich | üblich/ mehrheitlich |
| A-ost | 9 | 20 | 29 | 31 | 69 | gebräuchlich | üblich/ mehrheitlich |
| A-SÜDOST | 147 | 148 | 295 | 50 | 50 | gebräuchlich | gebräuchlich |
| CH | 91 | 155 | 246 | 37 | 63 | gebräuchlich | üblich/ mehrheitlich |
| LIE | 3 | 2 | 5 | 60 | 40 | (*) üblich/ mehrheitlich | (*) gebräuchlich |
| STIR | 19 | 3 | 22 | 86 | 14 | üblich/ mehrheitlich | kommt (selten) vor |

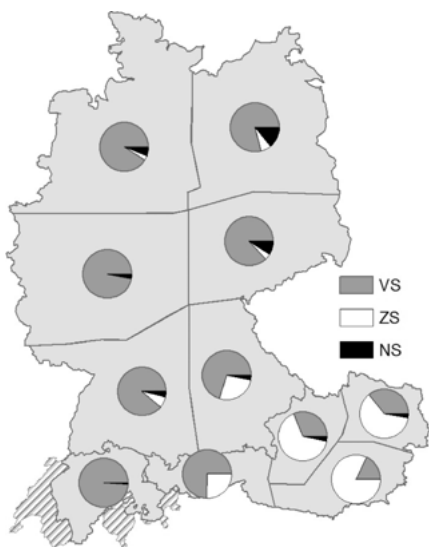
(d) Überblicksartikel: Stellung des finiten Verbs in zusammengesetzten Tempusformen mit Modalverb oder lassen

Das letzte Beispiel betrifft die Satzsyntax, genauer die Stellung des finiten Verbs in zusammengesetzten Tempusformen mit Modalverb oder mit dem Verb *lassen*. Der Überblicksartikel zu diesem Thema wird um eine Karte (s. u.) ergänzt, die über die Distribution der verschiedenen Stellungsvarianten informieren soll.

Stellung des finiten Verbs in zusammengesetzten Tempusformen mit Modalverb oder lassen

In zusammengesetzten (auch: analytischen) Tempusformen, die eine finite, also mit Personalendungen versehene Form eines Hilfsverbs (*haben, werden*), sowie ein Modalverb (*können, sollen, müssen, dürfen*) oder *lassen* enthalten, kann die Wortstellung in verschiedener Weise stilistisch und regional variieren. Regionale Variation besteht in der Standardsprache vorwiegend in dreigliedrigen Verbgruppen. Solche Verbgruppen treten nur in eingeleiteten Nebensätzen auf, wo alle verbalen Elemente direkt nacheinander stehen können (z. B. *Ich freue mich, dass er hat kommen können*). Dabei hängt von einer finiten Form von *haben* (z. B. *hätte*) bzw. *werden* (z. B. *würde*) eine infinite Form (‚Ersatzinfinitiv‘) eines Modalverbs oder *lassen* ab und davon wiederum der Infinitiv eines Vollverbs. Das finite Hilfsverb kann vor, zwischen oder hinter die infinite Gruppe treten; man spricht von Voranstellung (Kartenlegende: VS), Zwischenstellung (ZS) und Nachstellung (NS):

(a) Handelt es sich um eine finite Form von *haben*, geht diese überwiegend der Verbgruppe voran (... *dass er es hätte wissen müssen*). Dies gilt für CH und D-WEST fast ausnahmslos. Die Nachstellung des Finitums tritt selten (bzw. sporadisch) auf, vor allem in D-NORDOST, und auch dort fast nur nach infinitem *lassen* (... *dass er es mich wissen lassen hätte*). Die Nachstellung nach einem infiniten Modalverb, z. B. *müssen* (... *dass er es wissen müssen hätte*), kommt nur sporadisch vor. Dagegen ist die Verwendung des finiten Verbs zwischen den beiden infiniten Teilen (... *dass er es wissen hätte müssen*) in D-SÜDOST und A-WEST gebräuchlich, in allen anderen Gebieten von A wird die Zwischenstellung sogar mehrheitlich verwendet. Die Karte zeigt die Verteilung:



(b) Handelt es sich um eine finite Form von *werden*, ist zwischen Verbgruppen mit einem Modalverb und solchen mit *lassen* zu unterscheiden. In Verbgruppen mit *lassen* wird in allen Regionen fast ausnahmslos die Nachstellung verwendet (... *ob sie es machen lassen würde*). Ist ein Modalverb beteiligt, steht das Finitum – außer in A – überwiegend am Anfang der Verbgruppe (... *ob sie es würde schaffen können*). Daneben ist die Stellung am Ende der Verbgruppe in D gebräuchlich, in CH kommt sie sporadisch vor (... *ob sie es schaffen können würde*). In A ist neben Voran- und Nachstellung auch die Zwischenstellung gebräuchlich (... *ob sie es schaffen würde können*). Für die infinite Gruppe gilt standardsprachlich das Prinzip, dass das Modalverb oder *lassen* immer rechts von bzw. hinter dem von ihm abhängigen Infinitiv steht (*wissen müssen, schaffen können*). Die Voranstellung von Modalverben im infiniten Teil der Verbgruppe, wie sie aus der standardnahen gesprochenen Sprache und noch aus älteren Texten bekannt ist (z. B. ... *dass ich hätte können schreiben*), ist in der gegenwärtigen Standardsprache nicht mehr produktiv. Sie ist aber in einer erstarrten Wendung wie *es hat nicht sollen sein* erhalten, die im gesamten Sprachgebiet üblich ist.

2.4 Grundsätzliche Aspekte

Im vorangehenden Abschnitt wurde anhand einiger Beispiele gezeigt, wie die *Variante(n)grammatik* aufgebaut sein wird und welche Arbeit schon geleistet wurde. Dabei wurde aber auch deutlich, wie viele Analysen noch nötig sind, um auf der Basis systematischer Suchabfragen eine umfassende Darstellung diatopischer Variation präsentieren zu können. Bislang orientierten sich die Suchabfragen auch nur an solchen Phänomenen, von denen von vornherein angenommen wird, dass es sich um grammatisch-diatopische Varianten handeln könnte und die deshalb Eingang in unsere Variantendatenbank gefunden haben. Neu hinzu sollen nun in der Korpusrecherche noch automatisierte, induktive Suchverfahren kommen. Dazu gehören das Clustering von Korpustexten sowie n-Gramm-Analysen, die es allenfalls erlauben, neue Kandidaten für grammatische Variation zu finden. In einer n-Gramm-Analyse werden Frequenzen komplexer n-Gramme in den Teilkorpora miteinander verglichen, um so bisher nicht dokumentierte Varianten aufzudecken. Grundlage für das Clustering der Korpustexte anhand der in ihnen auftretenden grammatischen Varianten sind die durch die induktiven und deduktiven Analysen gewonnenen grammatischen Varianten. Dadurch können Teilkorpora mit ähnlichem grammatischem Profil gruppiert werden. Die durch diesen datengeleiteten Analyseschritt entstehenden diatopischen Einheiten – ausgehend von den Orten, in denen die Zeitungen erscheinen – können dann mit den Ergebnissen für die zu Projektbeginn definierten Sektoren verglichen werden.

Den Mehrwert unserer Arbeiten gegenüber bisherigen Studien zur diatopischen Variation sehen wir primär in den folgenden drei Bereichen: (1) Deutschlehrende und Deutschstudierende (ob im muttersprachlichen oder nichtmuttersprachlichen Kontext), aber auch interessierte Laien werden auf die Frage, ob bestimmte *grammatische* Varianten als standardsprachlich gelten können, verlässliche Antworten erhalten, die nicht auf subjektiven Normurteilen der Autoren beruhen, sondern auf Einordnungen, die sich empirisch, aus einheitlich angewandten und transparenten

korpuslinguistischen Verfahren ergeben. (2) Die Ergebnisse sind auch in normen-theoretischer Perspektive von erheblicher Relevanz: Es gibt bisher weder verlässliche noch sprachpolitisch zeitgemäße Angaben dazu, welche grammatischen Varianten, die nicht gemeindeutsch, sondern nur areal verbreitet sind, zum Standard gehören, noch wo sie verwendet werden. (3) Die Ergebnisse sollen nicht nur bestehende Lücken in den Grammatiken füllen, sondern auch solche Analysen korrigieren, die – mangels geeigneter Korpora – in die Grammatikwerke Eingang gefunden haben (vgl. z. B. die Kritik von Negele 2011: 223–225 an der Darstellung von Pronominaladverbien in gegenwärtigen Grammatiken des Deutschen).

3 Ausblick

Ein langfristiges Ziel unserer Projektarbeit ist, die *Variantengrammatik* in bereits bestehende Netzangebote (wie <<http://www.woerterbuch-portal.de/>> oder <<http://woerterbuchnetz.de/>> [31. Januar 2015]) zu integrieren und auf diese Weise eine direkte Verlinkung von unseren Artikeln zu anderen Wörterbüchern oder Websites, welche sich mit Variantenphänomenen beschäftigen, zu ermöglichen. Diese Vernetzung ist nicht nur aus Sicht der Benutzerinnen und Benutzer wünschenswert, sie würde auch die Bekanntheit und die Nutzung unserer *Variantengrammatik* steigern. Das ist uns wichtig, damit die *Variantengrammatik* tatsächlich den intendierten Nutzerkreis (neben anderen Grammatikerinnen und Grammatikern z. B. DaF-Lehrpersonen, aber auch sprachinteressierte Laien) erreicht, was wiederum eine Voraussetzung dafür ist, dass die angestrebten Ziele in sprachpolitischer und sprachdidaktischer Hinsicht erreicht werden.

Damit kommen wir zu einem letzten Aspekt, den wir hier erwähnen wollen. Bei aller methodischer Strenge, was die Korpusorientierung und die empirisch gestützten Analysen betrifft, werden in der täglichen Arbeit im Projekt auch Probleme offensichtlich, die nur langfristig zu lösen sind. Zum ersten ist das *Variantengrammatik*-Korpus trotz seiner über die ursprüngliche Planung weit hinausgehenden Größe für eine Abfrage mancher Phänomene immer noch zu klein bzw. für einzelne Regionen zu klein, als dass man belastbare Ergebnisse erzielen könnte. Das betrifft vor allem die kleineren Staaten bzw. Regionen Ostbelgien, Luxemburg, Liechtenstein und Südtirol. Diesbezüglich wird das Projekt dankbar für weitere Korpusinitiativen in einzelnen dieser Länder bzw. Regionen sein. Zum zweiten wird sich das Bild der grammatischen (wie auch der lexikalischen und lautlichen) Variation im Gebrauchsstandard des Deutschen stetig verändern. Es versteht sich also von selbst, dass das Projekt in künftigen Arbeiten eine Fortsetzung erfahren muss, wenn es nicht schon nach zehn Jahren veraltet sein soll. Allerdings müssen diese Arbeiten dann nicht ‚bei Null‘ beginnen, sondern können auf den Daten, Methoden und Ergebnissen des hier vorgestellten Projekts aufbauen.

Literatur

- Ammon, Ulrich (1995): *Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten*. Berlin, New York: de Gruyter.
- Ammon, Ulrich, Hans Bickel, Jakob Ebner et al. (2004): *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol*. Berlin, New York: de Gruyter.
- Auer, Peter (2013): Enregistering pluricentric German. In Augusto Soares da Silva (Hrsg.), *Pluricentricity. Language Variation and Sociocognitive Dimensions*, 19–48. Berlin, Boston: de Gruyter Mouton.
- Berend, Nina (2005a): Regionale Gebrauchsstandards – Gibt es sie und wie kann man sie beschreiben? In Ludwig M. Eichinger & Werner Kallmeyer (Hrsg.), *Standardvariation. Wie viel Variation verträgt die deutsche Sprache?*, 143–170. Berlin, New York: de Gruyter.
- Berend, Nina (2005b): Variation ja, aber welche? Zur Frage der Vermittlung von sprachlichen Varianten im Unterricht Deutsch als Fremdsprache. In DAAD (Hrsg.), *Germanistentreffen Deutschland – Großbritannien, Irland 30.9.–3.10.2004*, 279–296. Bonn: DAAD.
- Bickel, Hans & Christoph Landolt (2012): *Schweizerhochdeutsch. Wörterbuch der Standardsprache in der deutschen Schweiz*. Hrsg. v. Schweizerischen Verein für deutsche Sprache. Mannheim, Zürich: Dudenverlag.
- Businger, Martin & Bettina Rimensberger (Hrsg.) (2014): Schwerpunkt: Variantengrammatik. *Sprachspiegel* 04/14, 98–118.
- Christen, Helen (Hrsg.) (2004): *Dialekt, Regiolekt und Standardsprache im sozialen und zeitlichen Raum*. Wien: Edition Praesens.
- Dudenredaktion (Hrsg.) (2009): *Die Grammatik*. 8., überarb. Aufl. Mannheim u. a.: Dudenverlag. [7. Aufl. 2005, 6. Aufl. 1998].
- Dudenredaktion (Hrsg.) (2011): *Richtiges und gutes Deutsch*. 7. Aufl., digitale Version. Mannheim u. a.: Dudenverlag.
- Durrell, Martin (2000): Standard Language and the Creation of National Myths in Nineteenth-Century Germany. In Jürgen Barkhoff, Gilbert Carr & Roger Paulin (Hrsg.), *Das schwierige neunzehnte Jahrhundert. Germanistische Tagung zum 65. Geburtstag von Eda Sagarra im August 1998*, 15–26. Tübingen: Niemeyer.
- Durrell, Martin & Nils Langer (2005): Gutes Deutsch und schlechtes Deutsch an britischen und irischen Hochschulen. Zur Akzeptanz von Variation im DaF-Unterricht. In DAAD (Hrsg.), *Germanistentreffen Deutschland – Großbritannien, Irland 30.9.–3.10.2004*, 297–314. Bonn: DAAD.
- Dürscheid, Christa & Inga Hefti (2006): Syntaktische Merkmale des Schweizer Standarddeutsch. Theoretische und empirische Aspekte. In Christa Dürscheid & Martin Businger (Hrsg.), *Schweizer Standarddeutsch. Beiträge zur Varietätenlinguistik*, 131–161. Tübingen: Narr.
- Dürscheid, Christa & Patrizia Sutter (2014a): Wie werden grammatische Helvetismen in Nachschlagewerken behandelt? Ratgeber geben nicht immer Rat – oder unterschiedlichen. *Sprachspiegel* 04/14, 111–118.
- Dürscheid, Christa & Patrizia Sutter (2014b): Grammatische Helvetismen im Wörterbuch. *Zeitschrift für Angewandte Linguistik* 60 (1), 37–65.
- Dürscheid, Christa, Stephan Elspaß & Arne Ziegler (2011): Grammatische Variabilität im Gebrauchsstandard – das Projekt „Variantengrammatik des Standarddeutschen“. In Marek Konopka, Jacqueline Kubczak, Christian Mair, František Štícha & Ulrich H. Waßner (Hrsg.), *Grammatik und Korpora 2009*, 123–140. Tübingen: Narr.
- Dürscheid, Christa, Stephan Elspaß & Arne Ziegler (2015): *Variantengrammatik des Standarddeutschen. Konzeption, methodische Fragen, Fallanalysen*. In Lenz & Glauning (Hrsg.), 207–235.

- Eichinger, Ludwig M. & Werner Kallmeyer (Hrsg.) (2005): *Standardvariation. Wie viel Variation ver­ trägt die deutsche Sprache?* Berlin, New York: de Gruyter.
- Eisenberg, Peter (2007): Sprachliches Wissen im Wörterbuch der Zweifelsfälle. Über die Rekonstruktion einer Gebrauchsnorm. *Aptum. Zeitschrift für Sprachkritik und Sprachkultur* 3/07, 209–228.
- Elsaß, Stephan (2005): Zum sprachpolitischen Umgang mit regionaler Variation in der Standard­ sprache. In Jörg Kilian (Hrsg.), *Sprache und Politik. Deutsch im demokratischen Staat*, 294–313. Mannheim u. a.: Dudenverlag.
- Elsaß, Stephan (2010): Regional Standard Variation in and out of Grammarians' Focus. In Alex­ andra N. Lenz & Albrecht Plewnia (Hrsg.), *Grammar between Norm and Variation*, 127–144. Frankfurt a. M.: Lang.
- Elsaß, Stephan (2012): Areal Variation in der Grammatik des Standarddeutschen. In Elena Kan (Hrsg.): *Wort – Text – Kultur. Beiträge zur Germanistik*, 219–240. Chabarovsk: Staatliche Fernöstliche Humanwissenschaftliche Universität.
- Elsaß, Stephan & Péter Maitz (Hrsg.) (2011): *Sprache und Diskriminierung*. Themenheft der Zeit­ schrift *Der Deutschunterricht* 63 (6).
- Elsaß, Stephan, Julia Engel & Konstantin Niehaus (2013): Areal Variation in der Grammatik des Standarddeutschen – Problem oder Aufgabe? *German as a Foreign Language* 2/13, 44–64.
- Elsaß, Stephan & Konstantin Niehaus (2014): Exploring the standardization of a modern pluriareal language. Concepts and corpus designs not only for German. *Orð og tunga* 16, 47–67. [Sonder­ heft zum Thema ‚Standardisierung‘, hrsg. v. Ásta Svavarsdóttir & Guðrún Þórhallsdóttir].
- Götz, Ursula (1995): Regionale grammatische Varianten des Standarddeutschen. *Sprachwissen­ schaft* 20, 222–238.
- Hensel, Sonja (2000): Welches Deutsch sollen wir lehren? Über den Umgang mit einer plurizentri­ schen Sprache im DaF-Unterricht. *Zielsprache Deutsch* 31, 31–39.
- Krumm, Hans-Jürgen, Christian Fandrych, Britta Hufeisen & Claudia Riemer (Hrsg.) (2010): *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch*. 2., neu bearb. Aufl. 2 Bde. Berlin, New York: de Gruyter.
- Lenz, Alexandra N. & Manfred M. Glauning (Hrsg.) (2015): *Standarddeutsch im 21. Jahrhundert. Theoretische und empirische Ansätze mit einem Fokus auf Österreich*. Wien: Vienna University Press.
- Maitz, Péter & Stephan Elsaß (2012): Pluralismus oder Assimilation? Zum Umgang mit Norm und arealer Variation in Deutschland und anderswo. In Susanne Günthner, Wolfgang Imo, Dorothee Meer & Jan Georg Schneider (Hrsg.), *Kommunikation und Öffentlichkeit: Sprachwissenschaftliche Potenziale zwischen Empirie und Norm*, 43–60. Berlin, Boston: de Gruyter.
- Mattheier, Klaus J. (Hrsg.) (1997): *Norm und Variation*. Frankfurt a. M. u. a.: Lang.
- Negele, Michaela (2011): *Varianten der Pronominaladverbien im Neuhochdeutschen. Grammatische und soziolinguistische Untersuchungen*. Berlin, New York: de Gruyter.
- Neuland, Eva (Hrsg.) (2006): *Variation im heutigen Deutsch: Perspektiven für den Sprachunterricht*. Frankfurt a. M. u. a.: Lang.
- Niehaus, Konstantin (2014): *Wortstellungsvarianten im Schriftdeutschen. Über Kontinuitäten und Diskontinuitäten in neuhochdeutscher Syntax*. Dissertation, Universität Augsburg.
- Polenz, Peter von (1999): *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart*. Band III: 19. und 20. Jahrhundert. Berlin, New York: de Gruyter.
- Rash, Felicity J. (2002): *Die deutsche Sprache in der Schweiz: Mehrsprachigkeit, Diglossie und Ver­ änderung*. Bern: Lang.
- Reiffenstein, Ingo (2001): Das Problem der nationalen Varietäten. Rezensionssatz zu Ulrich Ammon: Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten. *Zeitschrift für deutsche Philologie* 120, 78–89.

- Scherr, Elisabeth & Konstantin Niehaus (2013): „...weil man den Gebrauchsstandard erheben wird wollen.“ Variabilität und funktionale Äquivalenz in der Standardsyntax am Beispiel der ‚Zwischenstellung‘ in Verbalkomplexen. In Jörg Hagemann, Wolf Peter Klein & Sven Staffeldt (Hrsg.), *Pragmatischer Standard*, 75–84. Tübingen: Stauffenburg.
- Schmidlin, Regula (2011): *Die Vielfalt des Deutschen: Standard und Variation. Gebrauch, Einschätzung und Kodifizierung einer plurizentrischen Sprache*. Berlin, New York: de Gruyter.
- Schulte-Beckhausen, Marion (2002): *Genusschwankung bei englischen, französischen, italienischen und spanischen Lehnwörtern im Deutschen. Eine Untersuchung auf der Grundlage deutscher Wörterbücher seit 1945*. Frankfurt a. M.: Lang.
- Van Pottelberge, Jeroen (2004): *Der am-Progressiv. Struktur und parallele Entwicklungen in den kontinentalgermanischen Sprachen*. Tübingen: Narr.
- VWB = Ammon, Ulrich, Hans Bickel, Jakob Ebner et al. (2004).
- Zibrowa, Galina J. (1995): *Österreichisches und schweizerisches Deutsch*. Moskau: Vyshaja Shkola.
- Ziegler, Arne (2011): Standardsprachliche Variation als Ausgangspunkt grammatischer Reflexion. In Klaus-Michael Köpcke & Arne Ziegler (Hrsg.), *Grammatik – Lehren – Lernen – Verstehen. Zugänge zur Grammatik des Gegenwartsdeutschen*, 245–264. Berlin, New York: de Gruyter.