

Alexandra N. Lenz (Wien)

Korpora zur deutschen Sprache in Österreich

System- und soziolinguistische Perspektiven

Abstract: Der Beitrag liefert einen Einblick in korpuslinguistische Projekte und Aktivitäten aus dem österreichischen Sprachraum. Der Fokus liegt auf zwei primär auf die Analyse gesprochener Sprache ausgerichteten Korpora (*DiÖ-Korpus* und *WBÖ-Korpus*) sowie auf dem medial wie konzeptionell schriftlich angelegten *Austrian Media Corpus*. Institutionell eingebettet sind die Korpora in den Spezialforschungsbereich „Deutsch in Österreich (DiÖ)“ sowie in die Aktivitäten des *Austrian Centre for Digital Humanities and Cultural Heritage* (ACDH-CH) an der Österreichischen Akademie der Wissenschaften. Die theoretisch-methodologische Perspektive der Diskussion ist eine variationslinguistische, wobei sozio- und systemlinguistische Aspekte im Beitrag Berücksichtigung finden.

1 Zu diesem Beitrag und seiner variationslinguistischen Ausrichtung

Der vorliegende Beitrag verfolgt das Ziel, einen Einblick in korpuslinguistische Projekte und Aktivitäten aus dem österreichischen Sprachraum zu geben, der zweifelsohne ein ideales Forschungslabor für die Untersuchung von Sprachdynamik im Allgemeinen darstellt (siehe Lenz 2018, 2019a). Dazu tragen zum einen die sehr komplexen Varietätenkonstellationen der deutschen Sprache in Österreich bei, die sich nicht nur durch eine Fülle lebendiger nonstandardsprachlicher Varietäten (insbesondere Dialekte, Regiolekte oder nonstandardsprachliche Soziolekte) auszeichnen, sondern auch durch standardsprachliche Register im pluri-zentrischen Spannungsfeld des gesamten deutschsprachigen Raums. Diese sprachinternen Varietätenkonstellationen gehen zudem einher mit einer spezifischen, historisch gewachsenen Mehrsprachigkeit. Die Reflexe dieser internen und äußeren Mehrsprachigkeit haben hohe sozio-symbolische und attitudinal-perzeptive Implikationen, die sich auch im öffentlichen Diskurs niederschlagen (vgl. Lenz 2019a, S. 335).

Die für diesen Beitrag ausgewählten Korpora sind institutionell eingebettet, einerseits in den Spezialforschungsbereich „Deutsch in Österreich (DiÖ)“ (siehe

<https://doi.org/10.1515/9783111085708-004>

dazu Kap. 2) und andererseits in die Aktivitäten des „Österreichischen Zentrums für Digitale Geisteswissenschaften und Kulturelles Erbe“ (*Austrian Centre for Digital Humanities and Cultural Heritage*, ACDH-CH) an der Österreichischen Akademie der Wissenschaften. Die theoretisch-methodologische Perspektive, die der Diskussion zugrunde liegt, ist eine variationslinguistische. *Variationslinguistik* wird hier gefasst als eine Perspektivierung von Sprache mit Fokus auf ihrer Variabilität, die mit Mattheier (1984, S. 769) auf die allgemeine Eigenschaft von Sprache abzielt,

daß die Zuordnung zwischen Ausdrücken und Inhalten in beiden Richtungen nicht immer eindeutig ist, daß Inhalte von mehreren Ausdrucksseiten repräsentiert werden können und daß dieselben Ausdrucksseiten mehrere Inhalte widerspiegeln können. Der Begriff ‚Variation‘ kann dann die Realisierung von Variabilität innerhalb einer historischen Sprache bzw. bei einem Sprecher bezeichnen.

Eine umfassende Analyse sprachlicher Variation muss sowohl sprachinterne als auch sprachexterne Steuerungsfaktoren berücksichtigen. Damit steht Variationslinguistik an der Schnittstelle von Sozio- und Systemlinguistik, indem sie nämlich einerseits Sprache im sozialen Kontext und andererseits Sprache in ihrem Systemkontext analysiert. Sprachexterne Faktoren, mit denen Sprachvariation korrelieren kann, sind dann etwa soziodemographische Aspekte wie Raum, Alter oder Gender. Sprachinterne Faktoren sind linguistische Steuerungsfaktoren, zu denen etwa die Lautumgebung einer Variante oder die syntaktische Einbettung einer Konstruktion gehören. Digitale Variationslinguistik, wie sie gerade im Hinblick auf Sprachkorpora eine besondere Rolle spielt, kann dann gefasst werden als Variationslinguistik, die in ihren Forschungsprozess digitale Methoden und Tools einbindet, sei es bei der

Generierung und Erschließung von sprachwissenschaftlich relevanten Daten, [...] [der] Aufbereitung und Anreicherung der Daten (z. B. in Form von Transkriptionen und Annotationen) [...] [der] Analyse und Interpretation von Forschungsfragen [...] [oder aber der] digitale[n] Bereitstellung der erhobenen, aufbereiteten und analysierten Daten (Lenz 2019b, S. 5).

Der Computerlinguistik nähert sich die Digitale Linguistik insbesondere dann an, wenn sie digitale Methoden, Tools und Korpora nicht nur anwendet und nutzt, sondern selbst entwickelt bzw. ausbaut.

2 Ausgewählte Korpora und ihre Analysemöglichkeiten

Aus der Fülle von Korpora zur österreichischen Sprachlandschaft werden im Folgenden drei Korpora ausführlicher vorgestellt, unter denen zwei primär auf die Analyse gesprochener Sprache abzielen (DiÖ- und WBÖ-Korpus) und ein drittes medial und konzeptionell schriftlich ausgerichtet ist (*Austrian Media Corpus: amc*). Der gebotenen Kürze wegen finden hier lediglich ausgewählte und primär anwendungsbezogene Aspekte Berücksichtigung, während insbesondere technische/texttechnologische Details nur am Rande erwähnt werden können.

2.1 Das Korpus des SFB DiÖ

Das DiÖ-Korpus wird seit 2016 im Rahmen des vom FWF (Fonds zur Förderung der wissenschaftlichen Forschung) finanzierten Spezialforschungsbereichs „Deutsch in Österreich. Variation – Wandel – Kontakt“ (FWF F060) aufgebaut. Der SFB wird an den Universitäten Wien, Salzburg, Graz und der Österreichischen Akademie der Wissenschaften ÖAW durchgeführt (siehe <https://dioe.at> (Stand: 20.7.2022); Lenz 2018,2019a; Budin et al. 2019; Koppensteiner/Lenz 2017). Wie sich im Titel des SFB andeutet, liegen seine Forschungsschwerpunkte auf erstens sprachinterner Variation, zweitens Sprachkontakt und drittens attitudinal-perzeptiven Aspekten. Das DiÖ-Korpus setzt sich aus neu erhobenen Daten einerseits und andererseits aus bereits existenten, dann im Rahmen des SFB aufbereiteten (v. a. historischen) Daten zusammen. Im Rahmen der komplexen Erhebungen mündlicher Daten werden seit 2016 verschiedene Methoden in variierenden Settings eingesetzt, um – neben attitudinal-perzeptiven Daten – unterschiedliche Registerausschnitte der individuellen Sprachrepertoires zu gewinnen: Zum einen werden Konversationsdaten in Interviews, Gruppendiskussionen (gerade im schulischen Kontext) sowie in Gesprächen unter Freunden erhoben. Zum anderen werden kontrollierte Sprachdaten mit Methoden wie Übersetzungs- und Leseaufgaben (siehe etwa Lanwermeier et al. 2019) oder Sprachproduktionsexperimente (siehe etwa Lenz et al. 2019) erhoben. Zusätzlich kommen mündliche und schriftliche Fragebögen zum Einsatz, teilweise kombiniert mit Sprachproben, die im Rahmen von Hörerurteilstests zu bewerten sind (siehe etwa Koppensteiner/Lenz 2020).

Ein wesentlicher soziolinguistischer Parameter, der die Datenerhebungen des SFB steuert, ist die Arealität. Die Erhebungen berücksichtigen sowohl politische Einheiten (alle neun Bundesländer) als auch sprachgeographische Räume

(zu diesen siehe Wiesinger 1983, Kt. 474; Lenz 2019c; siehe unten Abb. 1). Während umfangreiche Dialekterhebungen bislang an 56 dörflichen Ortspunkten durchgeführt wurden, wird das vertikale Gesamtspektrum der Dialekt-Standard-Achse an 15 ausgewählten dörflichen Ortspunkten erfasst. In Ergänzung zu den ländlichen Ortspunkten finden ebenso groß- und kleinstädtische Zentren Berücksichtigung, insbesondere die Städte Wien und Graz sowie Schulorte im Westen Österreichs. Im gesprochen sprachlichen SFB-Korpus vertreten sind bislang (Stand: Mai 2022) ca. 470 Sprecher/-innen mit variierenden sozio-demographischen Hintergründen, wobei in Phase I (2016–2019) v. a. L1-Sprechende von DiÖ und in Phase II (2020f.) auch L2-Sprechende und diese insbesondere in urbanen Kontexten Berücksichtigung finden. Aktuell umfasst das DiÖ-Korpus insgesamt ca. 1057 h Sprachaufnahmen. Hinzu kommen verschiedenste Fragebogenerhebungen (von aktuell bereits 4.700 Proband/-innen, Stand: Mai 2022), die sich über ganz Österreich verteilen und teils sprachgebrauchsorientiert, teils mehr attitudinal-perzeptiv ausgerichtet sind.

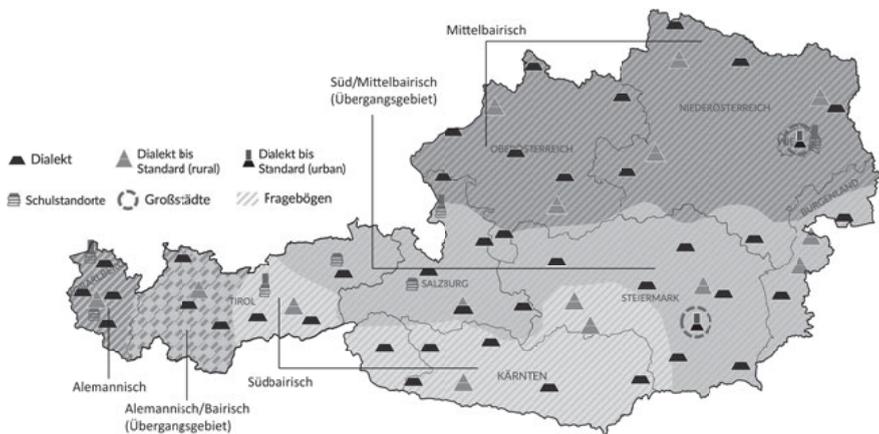


Abb. 1: SFB DiÖ – Datenerhebungen im „(sprach)geographischen“ Raum (Stand: Mai 2022)

Um das historische Datenspektrum zu vervollständigen, werden zudem schriftliche historische Quellen in das SFB-Korpus integriert, die insbesondere im Kontext der Mehrsprachigkeit in Österreich stehen, wie etwa Daten aus Volkszählungen, Schulstatistiken, Gesetzestexten oder Klassenbüchern (siehe Kim 2021; Newerkla 2020; Schinko 2021).

Die verschiedenen Datentypen des DiÖ-Korpus erfordern unterschiedliche Formen der Datenverarbeitung, die den individuellen und integrierten Anforderungen der jeweiligen quantitativen und qualitativen Datenanalysen gerecht wer-

den. Daher bilden verschiedene Arten von Transkriptionen, Annotationen und Klassifizierungen der Korpusdaten einen zentralen Teil unseres Arbeitsprogramms (siehe Pluschkovits/Kranawetter 2021; Korecky-Kröll et al. im Dr.).

Um einen Einblick in die variationslinguistischen Analysemöglichkeiten des DiÖ-Korpus zu vermitteln, wird im Folgenden das Subkorpus des Teilprojekts (project part) PP03 herangezogen, das Sprachrepertoires und Varietätenspektren in Österreich erhebt und analysiert (siehe Lenz 2018). Mit Fokus auf ländlichen Räumen geht PP03 der Frage nach, wer in Österreich welche/s/n Varietät/Register/Stil von DiÖ mit wem wann und wozu spricht. Das Subkorpus von PP03 wird seit 2016 an 15 Ortspunkten Österreichs bei bis zu 230 autochthonen Sprecher/-innen mit variierenden soziodemographischen Hintergründen erhoben. Die ländlichen Ortspunkte verteilen sich auf die verschiedenen Dialekträume Österreichs (siehe die hellgrauen Dreiecke in Abb. 1). Die Gewährspersonen haben bislang bis zu acht verschiedene Erhebungssettings durchlaufen, wobei es das Ziel der verschiedenen Settings ist, auch verschiedene Ausschnitte/Register/Sprechlagen der individuellen Sprachrepertoires zu evozieren, die auch interindividuelle, sprecherübergreifende Vergleiche ermöglichen sollen (siehe Lenz 2019a). Zwei der Erhebungssettings konzentrieren sich auf gesprochene Daten in freie(re)n Gesprächskontexten: erstens ein leitfadengesteuertes Interview, das Sprachdaten in einer eher formellen Gesprächssituation evoziert, und zweitens ein Gespräch unter Freunden aus demselben lokalen Netzwerk, um Sprachdaten in einer eher informellen Situation zu erheben.¹ Diese zwei Konversationssettings wurden ergänzt durch sechs (mehr oder weniger) standardisiert-kontrollierte Erhebungssettings. Über die Analyse intraindividuelle(r) Sprachrepertoires hinaus hatte die komplexe Datenerhebung zum Ziel, verschiedene Systemebenen zu berücksichtigen. Um niedrige Gebrauchsfrequenzen bestimmter Phänomene auszugleichen, kamen speziell entwickelte Sprachproduktionsexperimente (insbesondere im Hinblick auf Syntax, siehe Lenz et al. 2019), Übersetzungsaufgaben bzw. Leseaufgaben (insbesondere im Hinblick auf Phonetik/Phonologie) zum Tragen. Die Übersetzungsaufgaben und Experimente fanden dabei in zwei verschiedenen Durchgängen statt, um einerseits (eher) nonstandardsprachliche versus andererseits (eher) standardsprachliche Sprachverhaltensmuster zu evozieren. Insgesamt umfasst das Subkorpus von PP03/PP08 bislang rund 525 Stunden gesprochene Sprache (Stand: Mai 2022).

¹ Interviews und Freundesgespräche wurden in Kooperation mit PP08 erhoben, in dem es um attitudinal-perzeptive Aspekte mit Fokus auf Standardsprachlichkeit in Österreich geht (siehe Koppensteiner/Lenz 2017).

Um die intra- und interindividuelle Komplexität der Variationsmuster im PP03-Korpus zu illustrieren, gibt Abbildung 2 zunächst ausgewählte Analyseergebnisse mit Fokus auf lautliche Variation wieder. Konkret geht es um die individuellen Gebrauchsfrequenzen der Reflexe zu mhd. /ei/ bei insgesamt 28 Sprechenden an sechs Ortspunkten Österreichs, die sich auf sechs Dialekt-räume verteilen (von links nach rechts in Abb. 2): Raggal im Höchstalemannischen, Tarrenz im bairisch-alemannischen Übergangsgebiet, Weißbriach im Südbairischen, Taufkirchen im Westmittelbairischen, Neumarkt/Ybbs im Ostmittelbairischen sowie Neckenmarkt im südmittelbairischen Übergangsgebiet. Verschiedene Schraffuren und Farbanteile der Balkendiagrammanteile visualisieren verschiedene Realisierungsvarianten pro Person in jeweils sechs verschiedenen Erhebungssettings. Personenübergreifend ist eine höhere Nonstandardsprachlichkeit in Dialektübersetzung, Freundesgespräch und Interview zu erkennen, die sich in schraffierten und gepunkteten Balkenanteilen nieder-

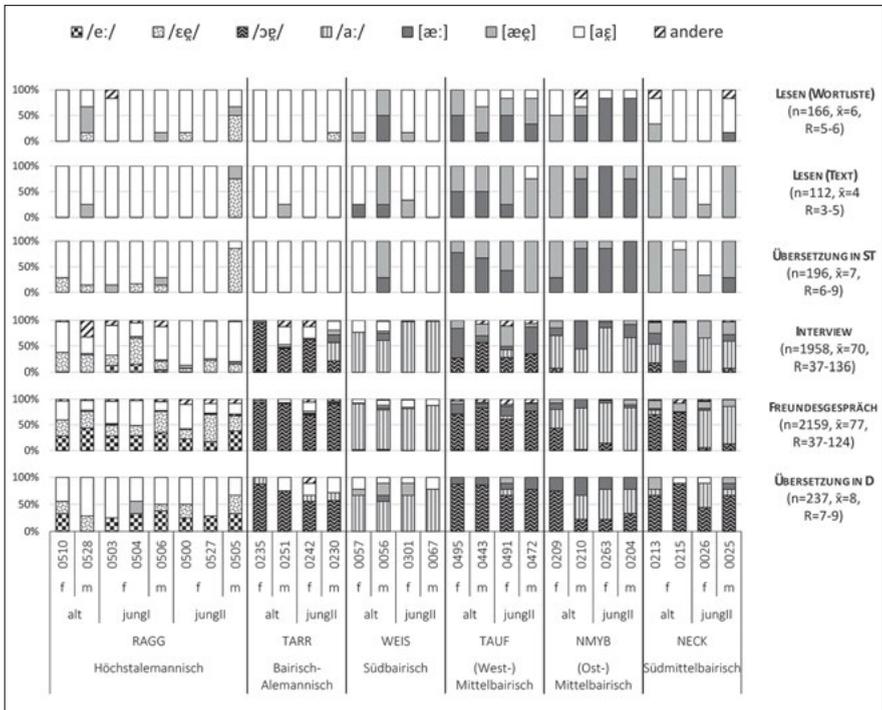


Abb. 2: Individuelle Frequenzen der Reflexe zu mhd. /ei/. Datenbasis: 4.828 Belege bei 28 Sprecher/-innen aus sechs Erhebungssettings an sechs Ortspunkten des SFB DiÖ – Subkorpus von PP03 (Abbildung aus Fanta-Jende in Vorb.)

schlägt. Schwarz-/Grau- und Weißtöne stehen hingegen für standardsprachliche bzw. standardsprachnähere Varianten, die gerade in der Übersetzung in den Standard (ST) und bei Leseaufgaben (einmal Wortliste, einmal Nordwind- und-Sonne-Text) auftreten. Neben intersituativen Differenzen sind auch offensichtliche Unterschiede zwischen den Spracharealen zu erkennen, insbesondere zwischen alemannischen und bairisch-österreichischen Sprechenden. Insgesamt ermöglicht das methodische Inventar von PPO3 einen sehr detaillierten Einblick in die Repertoires der Sprechenden, ihre intraindividuellen Variationsmuster sowie interindividuelle Differenzen, und dies von dialektale(re)n bis hin zu standardsprachliche(re)n Registern inklusive des mittleren Bereichs auf der Dialekt-Standard-Achse. Zu einer ausführlichen Interpretation des Variationsphänomens sei verwiesen auf Fanta-Jende (2020, 2021).

Ähnliche Variationstendenzen zeigen sich auch mit Blick auf grammatische Phänomene, die in Abbildung 3 am Beispiel von Konjunktiv-II-Realisierungen exemplarisch skizziert werden. Die Datenbasis liefern nun 32 Sprecher/-innen an vier ausgewählten Ortspunkten in Österreich in ihren beiden Experimentrunden (einmal mit Fokus auf standardsprachliche(re)m und einmal mit Fokus auf nonstandardsprachliche(re)m bzw. dialektale(re)m Sprachverhalten) ergänzt durch Interview und Freundesgespräch. Differenziert werden sechs Konstruktionsvarianten, die sich in analytische, periphrastische und synthetische Varianten einteilen lassen.

Offensichtlich nimmt vom Dialektexperiment über das Freundesgespräch hin zum Interview und schließlich zum standardsprachlichen Experimentdurchgang nicht nur insgesamt die Variationsvielfalt ab, sondern auch die Frequenzen nonstandardsprachlicher Konjunktivvarianten werden reduziert. Wie auch beim lautlichen Variationsbeispiel (siehe oben Abb. 2) geschieht die intersituative Anpassung der grammatischen Varianten graduell; die Sprachverhaltensmuster werden nicht abrupt, sondern fließend angepasst. Die intersituative Abnahme an dunklen Grautönen „von unten nach oben“ reflektiert dabei die Abnahme an periphrastischen *täte*-Konjunktiven (z. B. *täte singen*). Während der basisdialektale *täte+at*-Konjunktiv (z. B. *darat singen*) ohnehin nur selten im Korpus vertreten ist, werden synthetische *at*-Konjunktive von mehreren (bairischen) Sprecher/-innen bis ins Interview hinein realisiert. *Würde*-Konjunktive sind in unterschiedlichem Maße bei allen Sprecher/-innen vorzufinden, ihre Frequenzen nehmen gerade in „oberen“ Bereichen des Spektrums und mit Abnahme dialektaler Alternativen zu. Zu einer ausführlichen Interpretation des Variationsphänomens – unter Berücksichtigung weiterer (teils standardsprachlicher, teils nonstandardsprachlicher) synthetischer Varianten (siehe Grünanteile in Abb. 3) – sei verwiesen auf Breuer/Wittibschlager (2020) und Stöckle/Wittibschlager (2022).

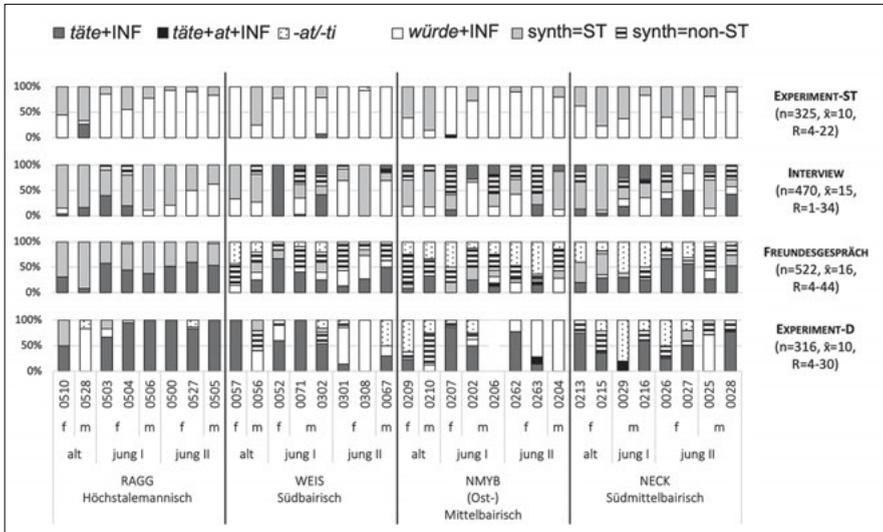


Abb. 3: Individuelle Frequenzen von Konstruktionsvarianten zum Konjunktiv II. Datenbasis: 1.633 Belege bei 32 Sprecher/-innen aus vier Erhebungssettings an vier Ortspunkten des SFB DiÖ – Subkorpus von PP03 (Abbildung aus Wittibschlager in Vorb.)

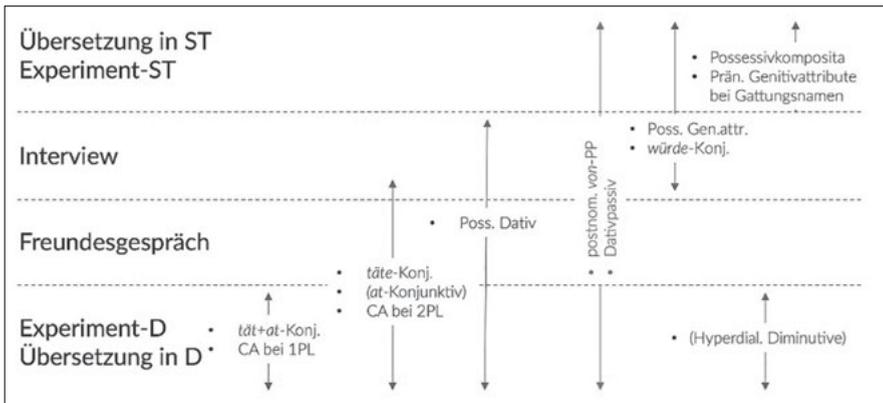


Abb. 4: Typen vertikaler Variation auf Basis des DiÖ-Subkorpus von PP03 mit Fokus auf grammatischen Varianten und ihren intra- wie intersituativen Auftretenshäufigkeiten (schemenhafte Darstellung)

Phänomenübergreifend ermöglicht die multidimensionale Anlage der Datenerhebungen und der Datenanalysen es, verschiedene Variantentypen vertikaler

Variation zu differenzieren, wie sie sich – etwa für die Grammatik – im Vergleich von bis zu sechs verschiedenen Situationen und Erhebungssettings abzeichnen (siehe Abb. 4).

So gibt es erstens grammatische Varianten, die v. a. in kontrolliert-dialektalen Settings (Dialektexperiment, Übersetzung in Dialekt) vorkommen, hingegen deutlich weniger in den Freundesgesprächen: Dazu gehören etwa *täte+at*-Konjunktive oder Complementizer Agreement („flektierte Konjunktionen“) in Nebensätzen mit 1PL (z. B. *wenn-ma mir singen* ‚wenn wir singen‘; siehe Fingerhuth/Lenz 2020). Es gibt zweitens Varianten, die in den Nonstandard-Settings bis ins Freundesgespräch hinein relativ hohe Frequenzen zeigen, die dann aber zum Interview und anderen standardsprachorientiert(er)en Settings hin frequenziell abnehmen. Als Beispiel dienen der *täte*-Konjunktiv, synthetische *at*-Konjunktive (z. B. *singat* ‚würde singen‘) oder Complementizer Agreement bei 2PL (z. B. *ob-s es morgen arbeitets* ‚ob ihr morgen arbeitet‘). Drittens gibt es Varianten, die bis in die Interviews – mit abnehmender Frequenz – gebraucht werden und erst in den kontrollierte(re)n Standard-Settings eindeutig vermieden werden. Dazu gehören etwa Possessive Dative (z. B. *dem Peter sein Freund*; siehe Goryczka et al. im Dr.). Viertens zeigen sich Varianten, die generell die (kontrollierten wie freie(re)n) standardsprachorientierten Settings (Interview, Experiment (ST), Übersetzung in ST) kennzeichnen. Beispielshaft zu nennen sind possessive Genitivattribute (z. B. *Peters Hund* oder *der Hund des Mannes*). Besonders interessant sind auch Varianten, die vor allem bis ausschließlich in den stark kontrollierten Standard-Settings auftreten und potenziell als Hyperkorrekturen interpretiert werden können. Dazu gehören etwa Possessivkomposita wie *Hundeball* oder pränominale Genitivattribute bei Gattungsnamen (*des Hundes Ball*), denen auf der anderen Seite der D-ST-Achse dialektale Hyperformen (*Hyperdialektalismen*) gegenüberstehen, die gerade in den stark kontrollierten Dialektsettings evoziert werden (siehe Lenz 2004). Schließlich finden sich auch grammatische Varianten, die kaum intersituative Variation zeigen und damit vertikal „unmarkiert“ erscheinen. Beispiele hierfür sind postnominale possessive *von*-Konstruktionen (z. B. *der Hund von Peter*) oder Dativpassive (z. B. *er kriegt/bekommt eine Brille auf die Nase gesetzt*, siehe Lenz et al. 2019).

Da der vorliegende Publikationskontext nur einen eingeschränkten Blick auf die Vielfalt an Analysemöglichkeiten des DiÖ-Korpus zulässt, sei insbesondere auf die bereits erschienenen Publikationen aus dem SFB-Kontext verwiesen, die verschiedenste sozio- und systemlinguistisch ausgerichtete Fragestellungen zu Sprachgebrauch, Sprachkontakt sowie Spracheinstellungen und -perzeption fokussieren (siehe www.dioe.at/aktuelles/publikationen, Stand: 20.7.2022). Zusammengefasst zeichnet sich das DiÖ-Korpus durch folgende Besonderheiten aus:

- Der Korpuszusammenstellung liegt eine multidimensionale Perspektivierung von Sprachvariation zugrunde, die sowohl sozio- als auch systemlinguistische Aspekte berücksichtigt.
- Das Korpus erhebt den Anspruch, Sprachrepertoires verschiedenster L1- und L2-Typen abzubilden, wobei der Schwerpunkt auf ersteren liegt.
- Extralinguistisch finden durch die Auswahl und Zusammenstellung der Sprecher/-innensamples verschiedene sozio-situative Parameter Berücksichtigung, z. B. „Situation“ (bis zu acht Settings pro Individuum), „Soziodemographie“ (Alter, Geschlecht, Bildungsgrade, Berufsart u. a.), „Raum“ (z. B. Stadt, Kleinstadt, Dorf), „Medialität“ (gesprochensprachlich, schriftsprachlich).
- Der systemübergreifende Ansatz erlaubt es schließlich, das Korpus für die Analysen von Sprachvariation auf verschiedenen Systemebenen heranzuziehen (insbesondere durch die Kombination freier(er) Konversationsdaten einerseits und kontrollierter(er) Experiment-/Testaufgaben andererseits).

2.2 Das Korpus des Wörterbuchs der bairischen Mundarten in Österreich (WBÖ)

Das Korpus des „Wörterbuchs der bairischen Mundarten in Österreich“ (WBÖ) – siehe ausführlich dazu Stöckle (2021) – basiert maßgeblich auf dem WBÖ-Hauptkatalog. Dieser umfasst ca. 3,6 Mio. handschriftliche Belegzettel, die v. a. auf flächendeckenden Befragungen in der ersten Hälfte des 20. Jahrhunderts basieren. Die Belegzettel werden aktuell als hochauflösende Bilddigitalisate im Rahmen des WBÖ-Projekts am ACDH-CH erschlossen und langzeitarchiviert. Zumindest ein Großteil der Handzettel, nämlich alle ab der Buchstabenkombination „Di/Ti“, wurde bereits volltexterfasst. Seit 2018 ist die WBÖ-Datenbank auch als XML-TEI-Datenbank online frei zugänglich, und zwar über das „Lexikalische Informationssystem Österreich“ (LIÖ) (siehe <https://lio.dioe.at>, Stand: 20.7.2022), das auch als Publikationsplattform für die fortlaufend geschriebenen Lexikonartikel des WBÖ dient. Die LIÖ-Benutzeroberfläche ermöglicht verschiedene Such- und Filtermöglichkeiten in der WBÖ-Datenbank, die über Videotutorials erläutert werden.

Mit dem WBÖ-Korpus sind einige forschungspraktische Herausforderungen verbunden: Der Datensammlung liegt ein primär lexikographischer Zugang zugrunde, der darin zum Ausdruck kommt, dass das Korpus primär über Lemmata/Einzelwörter strukturiert ist, wobei diese Lemmata erst mit den Buchstaben „Di (bzw. Ti)“ beginnen. Die linguistischen Annotationen der Daten sind nur rudimentär und unsystematisch, was die Suchmöglichkeiten im Korpus

erschwert. Darüber hinaus handelt es sich um (mündlichkeitsorientierte) Dialektdaten, die medial-schriftlich vorliegen. Mit den genannten und vielen anderen Aspekten hängt auch zusammen, dass soziolinguistische Tiefenanalysen, die über den Raumparameter hinausgehen, erschwert sind. Dass die WBÖ-Datenbank trotz dieser Herausforderungen dennoch weit über lexikographische Belange hinaus genutzt werden kann, zeigen derweil eine Fülle von systemlinguistischen Analysen, die auf dem Material aufbauen. Als Beispiel dient hier eine Auswertung der knapp 800 Belege zur Subjunktion *wenn* in einem Nebensatz mit 2SG (z. B. *wenn-st morgen kommst* ‚wenn du morgen kommst‘) (siehe Abb. 5). Die ersichtlichen Arealstrukturen mit Konzentration des Phänomens im Mittel- und Nordbairischen (siehe Abb. 5) und angrenzenden südmitteleuropäischen Übergangsgebiet decken sich mit Analysen auf Basis anderer Datentypen (siehe Fingerhuth/Lenz 2020; Lenz/Ahlers/Werner 2014).

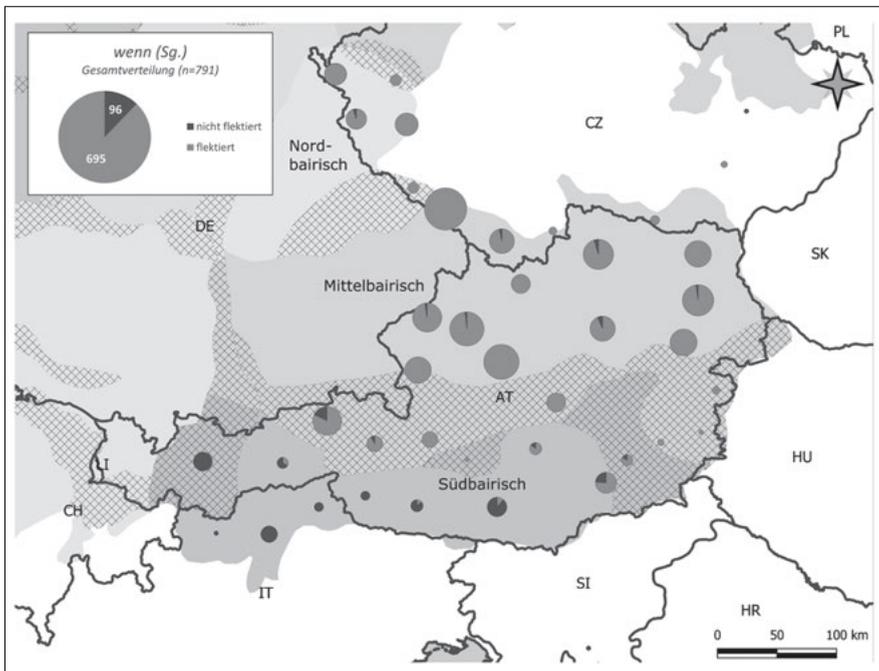


Abb. 5: Gesamtverteilung und sprachgeographische Verteilung (absolute Häufigkeiten) flektierter und nicht-flektierter Varianten des Komplementierers *wenn* für die 2SG auf Basis des WBÖ-Korpus. Datenbasis: 791 Belege. Grau hinterlegt ist der bairische Sprachraum nach Wiesinger (1983, Karte 47.4). (Abbildung aus Stöckle/Hemetsberger/Stütz 2021, Abb. 3)

Die Ergiebigkeit der WBÖ-DB konnte in den letzten Jahren durch vielfältige weitere Analysen etwa auf der Lautebene, der Grammatik oder der Wortbildung belegt werden.² Nicht nur bezogen auf das WBÖ illustrieren diese Analysen, welche Schätze die Daten der germanistischen Dialektlexikographie bereitstellen.

2.3 Das *Austrian Media Corpus* (amc)

Als drittes exemplarisches Korpus zur und aus der österreichischen Forschungslandschaft wird schließlich ein standardschriftsprachliches Korpus herangezogen, das *Austrian Media Corpus* (amc) (siehe Dorn et al. im Dr.). Beim amc-Projekt handelt es sich um eine Kooperation der Austrian Presse Agentur (APA) als Datengeber und dem ACDH-CH als Host und Pfleger des Korpus. Das amc basiert maßgeblich auf den Textinhalten österreichischer Tageszeitungen, Magazine und APA-Meldungen (journalistische Prosa). Es umfasst aktuell (in der Version 4.1) 47 Mio. Zeitungsartikel, die aus ca. 11 Mrd. Tokens bestehen. Da die Medienbranche erst im Laufe der Zeit in die Welt der volldigitalen Produktion eingestiegen ist, beginnt das amc in den späten 1980er Jahren recht bescheiden mit Agenturmeldungen der APA. Ab 1990 steigt der Datenzuwachs merklich, sodass seit 2000 eine recht stabile Korpuszunahme von jährlich zwischen 1,5 und 2 Millionen Artikeln zu verzeichnen ist. Die Korpusmanager- und Textanalyse-Software NoSketch Engine bietet verschiedene Such- und Filteroptionen, die auf die annotierten Metadaten zurückgreifen, wie etwa Medientyp, Publikationsorgan, Erscheinungsdatum oder Regionen- und Ressortzuordnungen (siehe Ziegler 2021). Gerade die Regionalunterteilung des amc (siehe Ziegler 2021, Abb. 1) wurde etwa genutzt, um die nationalen bzw. regionalen Angaben in der Neuauflage des Variantenwörterbuchs (Ammon/Bickel/Lenz (Hg.) 2016) empirisch zu fundieren. Weitere exemplarische Analysen auf Basis des amc sind auf der amc-Homepage zu finden (siehe <https://amc.acdh.oeaw.ac.at/publikationen/>, Stand: 20.7.2022). Jüngst wurde das Korpus bspw. genutzt, um Austriazismen, Wortbildungsvariation, expressive Adjektivkomposita oder Klimadiskurse zu analysieren.

Eine Beispielsauswertung mit Blick auf Fugenelemente in Komposita ist in Abbildung 6 einzusehen (siehe Ziegler 2021). Die jeweils nebeneinander angezeigten beiden Karten zu *Geschenk(s)korb* bzw. *Geschenk(s)papier* visualisieren die komplementären Häufigkeiten der Variante ohne Fugen-s (jeweils links) bzw.

² Siehe www.oeaw.ac.at/acdh/sprachwissenschaft/projekte/wboe/aktivitaeten/wboe-relevante-publikationen (Stand: 20.7.2022).

mit Verfügung (rechts) auf Basis der amc-Regionen „awest“, „amitte“, „asuedost“ und „aost“. Wie die Beispiel(s)fälle illustrieren, ist die für Österreich postulierte Bevorzugung der s-Fuge in Komposita mit *Geschenk-* als Erstglied (siehe Ammon/Bickel/Lenz (Hg.) 2016, Lemma *Geschenks-*) vor dem Hintergrund einzelwortspezifischer und regionaler Ausprägungen zu interpretieren.

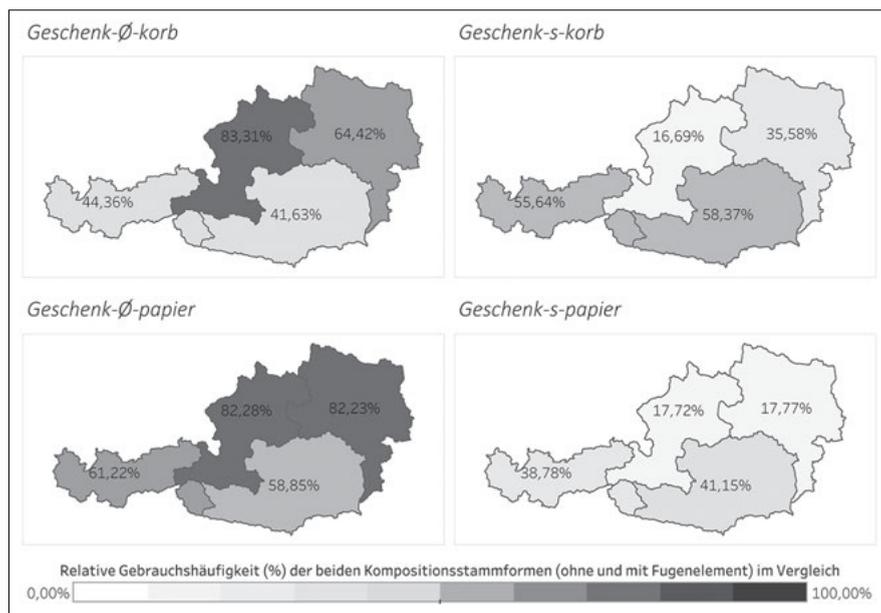


Abb. 6: Relative Häufigkeiten (in %) der Verfügungsschwankungsfälle *Geschenkkorb* (n = 5873) vs. *Geschenkskorb* (n = 3048) bzw. *Geschenkpapier* (n = 1842) vs. *Geschenkspapier* (n = 448) in den amc-Regionen (aus Ziegler 2021, Abb. 3)

3 Zusammenfassung und Ausblick

Die im Vorangehenden skizzierten drei Korpora zum österreichischen Sprachraum bringen im Hinblick auf soziolinguistische und systemlinguistische Analysemöglichkeiten sehr unterschiedliche Potenziale und Herausforderungen mit sich:

Da das Korpus des SFB DiÖ gerade auf variationslinguistische Fragestellungen hin ausgerichtet ist, finden sozio- wie systemlinguistische Aspekte bei der Erstellung des Korpus eine starke Berücksichtigung. Die gesprochensprachlichen Daten, die die Variationsrepertoires von (soziodemographisch variierenden) Indi-

viduen in (bislang) bis zu acht Settings widerspiegeln, erlauben quantitative wie qualitative Tiefenbohrungen zur intra- wie interindividuellen Variation und dies im Vergleich verschiedener Regionen und Netzwerke. Die Fülle von soziodemographischen und sprachbiographischen Metadaten sowie Spracheinstellungs- und -perceptionsdaten pro Individuum erlaubt zudem multidimensionale Analysen zu extralinguistischen Parametern, die mit den Sprachgebrauchsmustern in Beziehung gesetzt werden können.

Das amc ist aufgrund seiner arealen Flächendeckung und seiner (stetig wachsenden) Größe zweifelsohne für standardschriftsprachliche (insbesondere quantitativ ausgerichtete) Korpusanalysen höchst attraktiv. Es eignet sich besonders für vielfältigste systemlinguistische Analysen, die auch zumindest mit dem extralinguistischen Faktor „Raum“ korreliert werden können.

Das WBÖ-Korpus ist – natürlich in Abhängigkeit von der konkreten Forschungsfrage – das Korpus, das die größten Herausforderungen mit sich bringt. Zu den primär konzeptionell-mündlichen Dialektdaten, die überwiegend per Fragebogen schriftlich erhoben wurden, gibt es zumindest einige Hintergrundinformationen zu den „Sammler/-innen“ und ihrer Soziodemographie. Die linguistischen Annotationen des Materials sind rudimentär und bislang nicht systematisch angelegt. Trotz dieser und anderer korpuspezifischer Probleme belegen die bisher vorliegenden Analysen auf Basis der WBÖ-Datenbank dennoch die Ergiebigkeit des Materials, insbesondere für systemlinguistisch ausgerichtete Dialektfragestellungen.

Die für diesen Beitrag exemplarisch ausgewählten drei Korpora stellen nur einen Ausschnitt all der Text- und Sprachressourcen dar, die (nicht nur) für die germanistische Sprachwissenschaft von Interesse sein könnten. Im weiteren Kontext des ACDH-CH an der ÖAW werden weitere Korpora auf- und ausgebaut bzw. gehostet. Zu diesen gehören etwa – um zumindest zwei weitere Beispiele explizit zu erwähnen – das Korpus „Österreichische Dialektaufnahmen im 20. Jahrhundert“ (siehe Lenz et al. 2020) sowie das „Wienerische Digitalium“, der digitalen Ausgabe der historischen Zeitung „Wien[n]erisches Diarium“ (heute: „Wiener Zeitung“) (siehe Resch/Kampkaspar 2019).

Allen in diesem Beitrag skizzierten oder zumindest erwähnten Korpora ist – wie Sprach- und Textkorpora grundsätzlich – gemein, dass sie im Hinblick auf variationslinguistische Fragestellungen, seien sie mehr soziolinguistisch oder systemlinguistisch ausgerichtet, unterschiedliche Stärken und Herausforderungen mit sich bringen. Wie ein Blick in kontrastiv-synoptische Analysen über verschiedene Korpora hinweg zeigt (siehe z. B. Lenz 2013), lohnt es sich generell, die Grenzen von Einzelkorpora zu verlassen und ihre individuellen Schwächen durch die Kombination mit „Nachbarkorpora“ auszugleichen.

Literatur

- Ammon, Ulrich/Bickel, Hans/Lenz, Alexandra N. (Hg.) (2016): Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz, Deutschland, Liechtenstein, Luxemburg, Ostbelgien und Südtirol sowie Rumänien, Namibia und Mennonitensiedlungen. 2., völl. neu bearb. u. erw. Aufl. Berlin/Boston: De Gruyter.
- Breuer, Ludwig M./Wittibschlager, Anja (2020): The variation of the subjunctive II in Austria. Evidence from urban and rural analyses. In: *Linguistic Variation* 20, 1, S. 136–171.
- Budin, Gerhard/Elspaß, Stephan/Lenz, Alexandra N./Newerkla, Stefan M./Ziegler, Arne (2019): The Research Project (SFB) ‘German in Austria’. *Variation – Contact – Perception*. In: Bülow, Lars/Herbert, Kristina/Fischer, Ann Kathrin (Hg.): *Dimensions of linguistic space: Variation – Multilingualism – Conceptualisations*. (= *Schriften zur deutschen Sprache in Österreich* 45). Frankfurt a. M. u. a.: Lang, S. 7–35.
- Dorn, Amelie/Höll, Jan/Ziegler, Theresa/Koppensteiner, Wolfgang/Pirker, Hannes (im Dr.): Die österreichische Presselandschaft digital: Das Austrian Media Corpus (AMC) – Aufbau, Bedienung und Möglichkeiten. In: Kupietz/Schmidt (Hg.).
- Fanta-Jende, Johanna (2020): Varieties in contact. Horizontal and vertical dimensions of phonological variation in Austria. In: Lenz, Alexandra N./Maselko, Mateusz (Hg.): *Variationist Linguistics meets Contact Linguistics*. (= *Wiener Arbeiten zur Linguistik* 6). Göttingen: Vienna University Press, S. 203–240.
- Fanta-Jende, Johanna (2021): Situational effects on intra-individual variation in German – Reflexes of Middle High German *ei* in Austrian speech repertoires. In: Werth, Alexander/Bülow, Lars/Pfenninger, Simone E./Schiegg, Markus (Hg.): *Intra-individual variation in language*. (= *Trends in Linguistics. Studies and Monographs [TiLSM]* 363). Berlin/Boston: De Gruyter, S. 87–125.
- Fanta-Jende, Johanna (im Dr.): Intra- und interindividuelle Variation in Österreich. Phonetisch-phonologische Analysen. In: Kupietz/Schmidt (Hg.).
- Fingerhuth, Matthias/Lenz, Alexandra N. (2020): Variation and dynamics of “Complementizer Agreement” in German. *Analyses from the Austrian language area*. In: *Linguistic Variation* 21, 2, S. 322–369.
- Goryczka, Pamela/Wittibschlager, Anja/Korecky-Kröll, Katharina/Lenz, Alexandra N. (im Dr.): Variation und Wandel adnominaler Possessivkonstruktionen im Deutschen. Horizontal-areale und vertikal-soziale Analysen zum österreichischen Sprachraum. In: *Zeitschrift für Dialektologie und Linguistik (ZDL)*.
- Kim, Agnes (2021): The melting pot revisited: Historical sociolinguistic perspectives on migration and language contact in Vienna. In: Auer, Anita/Thorburn, Jennifer (Hg.): *Approaches to migration, language and identity*. (= *Language, Migration and Identity* 4). Oxford u. a.: Lang, S. 11–40.
- Koppensteiner, Wolfgang/Lenz, Alexandra N. (2017): Theoretische und methodische Herausforderungen einer perzeptiv-attitudinalen Standardsprachforschung. *Perspektiven aus und auf Österreich*. In: Sieburg, Heinz/Solms, Hans-Werner (Hg.): *Das Deutsche als plurizentrische Sprache. Ansprüche – Ergebnisse – Perspektiven*. (= *Sonderheft Zeitschrift für deutsche Philologie* 136). Berlin: ESV, S. 43–68.
- Koppensteiner, Wolfgang/Lenz, Alexandra N. (2020): Tracing a standard language in Austria using methodological microvariations of verbal and matched guise technique. In: *Linguistik Online* 102, 2, S. 47–82.

- Korecky-Kröll, Katharina/Wittibschlager, Anja/Pluschkovits, Markus/Tavernier, Florian/Fanta-Jende, Johanna/Stiglbauer, Rita/Bal, Jakob/Kranawetter, Katharina/Stocker, Rebecca (im Dr.): Erhebung, Aufbereitung und (kollaborative) Nutzung des Korpus „Deutsch in Österreich. Variation – Kontakt – Perzeption“. In: Kupietz/Schmidt (Hg.).
- Kupietz, Marc/Schmidt, Thomas (Hg.) (im Dr.): Neue Entwicklungen in der Korpuslandschaft der Germanistik: Beiträge zur IDS-Methodenmesse 2022. (= Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache (CLIP) 11). Tübingen: Narr.
- Lanwermyer, Manuela/Fanta-Jende, Johanna/Lenz, Alexandra N./Korecky-Kröll, Katharina (2019): Competing norms of standard pronunciation – Evidence from <-ig->-variation in Austria. In: *Dialectologia et Geolinguistica* 27, 1, S. 143–175.
- Lenz, Alexandra N. (2004): Hyperforms and variety barriers. In: Gunnarsson, Britt-Louise/Bergström, Lena/Eklund, Gerd/Fridell, Staffan/Hansen, Lise H./Karstadt, Angela/Nordberg, Bengt/Sundgrenand Mats Thelander, Eva (Hg.): *Language variation in Europe. Papers from the Second International Conference on Language Variation in Europe, ICLaVE 2, Uppsala University, Sweden, June 12–14, 2003*. Uppsala: Universitetsstryckeriet, S. 281–294.
- Lenz, Alexandra N. (2013): Vom >kriegen< und >bekommen<. Kognitiv-semantische, variationslinguistische und sprachgeschichtliche Perspektiven. (= *Linguistik– Impulse & Tendenzen* 53). Berlin/Boston: De Gruyter.
- Lenz, Alexandra N. (2018): The Special Research Programme: German in Austria: Variation – Contact – Perception. In: Ammon, Ulrich/Costa, Marcella (Hg.): *Sprachwahl im Tourismus – mit Schwerpunkt Europa. Language choice in tourism – Focus on Europe. Choix de langues dans le tourisme – focus sur l’Europe*. (= *Sociolinguistica* 32). Berlin/Boston: De Gruyter, S. 269–277.
- Lenz, Alexandra N. (2019a): Der SFB „Deutsch in Österreich. Variation – Kontakt – Perzeption“. In: Eichinger, Ludwig M./Plewnia, Albrecht (Hg.): *Neues vom heutigen Deutsch. Empirisch – methodisch – theoretisch*. (= *Jahrbuch des Instituts für Deutsche Sprache* 2018). Berlin/Boston: De Gruyter, S. 335–338.
- Lenz, Alexandra N. (2019b): Digitale Sprachwissenschaft. Herausforderungen und Perspektiven. In: *Akademie im Dialog* 15, S. 5–17.
- Lenz, Alexandra N. (2019c): Bairisch und Alemannisch in Österreich. In: Herrgen, Joachim/Schmidt, Jürgen Erich (Hg.): *Sprache und Raum. Ein internationales Handbuch der Sprachvariation*. Bd. 4: Deutsch. Unter Mitarbeit von Hanna Fischer und Brigitte Ganswindt. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 30.4). Berlin/Boston: De Gruyter, S. 318–363.
- Lenz, Alexandra N./Breuer, Ludwig Maximilian/Huber, Christian/Fischer, Benjamin/Graf, Bernhard (2020): „Österreichische Dialektaufnahmen im 20. Jahrhundert“ – Zur Genese, Aufbereitung und wissenschaftlichen Nutzung eines einmaligen Sprachkorpus. In: *Jahrbuch des Phonogrammarchivs* 10, S. 128–140.
- Lenz, Alexandra N./Ahlers, Timo/Werner, Martina (2014): Zur Dynamik bairischer Dialektsyntax – eine Pilotstudie. In: *Zeitschrift für Dialektologie und Linguistik* LXXXI 81, 1, S. 1–33.
- Lenz, Alexandra N./Breuer, Ludwig Maximilian/Fingerhuth, Matthias/Wittibschlager, Anja/Seltmann, Melanie (2019): Exploring syntactic variation by means of “Language Production Experiments” – Methods from and analyses on German in Austria. In: *Journal of Linguistic Geography* 7, 2, S. 63–81.

- Mattheier, Klaus J. (1984): Sprachwandel und Sprachvariation. In: Besch, Werner/Reichmann, Oskar/Sonderegger, Stefan (Hg.): Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. 1. Teilbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft 2.1). Berlin: De Gruyter, S. 768–779.
- Newerla, Stefan Michael (2020): Mehrsprachigkeit und moderne Fremdsprachenausbildung in der ausgehenden Habsburgermonarchie am Beispiel der Schulbücher eines Wladimir Hanaček. In: Schörg, Christine/Sippl, Carmen (Hg.): Die Verführung zur Güte. Beiträge zur Pädagogik im 21. Jahrhundert. Festschrift für Erwin Rauscher. (= Pädagogik für Niederösterreich 8). Innsbruck/Wien: StudienVerlag, S. 217–232.
- Pluschkovits, Markus/Kranawetter, Katharina (2021): Annotation von Sprachdaten eines variationslinguistischen Großprojekts am Beispiel des Spezialforschungsbereichs ›Deutsch in Österreich‹. In: Wiener Linguistische Gazette 89, S. 167–189.
- Resch, Claudia/Kampkaspar, Dario (2019): DIGITARIUM – Unlocking the treasure trove of 18th-century newspapers for digital times. In: Wallnig, Thomas/Romberg, Marion/Weis, Joelle (Hg.): Digital Eighteenth century: Central European perspectives. Wien u. a.: Böhlau, S. 49–64.
- Schinko, Maria (2021): »Böhmisch erwünscht«?! Das Tschechische und Slowakische im Gerichtsbezirk Stockerau um 1900. In: Wiener Linguistische Gazette 89, S. 423–462.
- Stöckle, Philipp (2021): Wörterbuch der bairischen Mundarten in Österreich (WBÖ). In: Lenz, Alexandra N./Stöckle, Philipp (Hg.): Germanistische Dialektlexikographie zu Beginn des 21. Jahrhunderts. (= Zeitschrift Für Dialektologie und Linguistik. Beihefte 181). Stuttgart: Steiner, S. 11–46.
- Stöckle, Philipp/Wittibschlager, Anja (2022): Zur Sprachdynamik des Konjunktivs im Bairischen in Österreich. In: Linguistik online 114, 2, S. 43–65
- Stöckle, Philipp/Hemetsberger, Christina/Stütz, Manuela (2021): Die WBÖ-Belegdatenbank als Quelle für syntaktische Analysen – Möglichkeiten, Grenzen, Perspektiven. In: Wiener Linguistische Gazette 89, S. 579–626.
- Wiesinger, Peter (1983): Die Einteilung der deutschen Dialekte. In: Besch, Werner/Kopp, Ulrich/Putschke, Wolfgang/Wiegand, Herbert Ernst (Hg.): Dialektologie: Ein Handbuch zur deutschen und allgemeinen Dialektforschung. 2. Halbbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft 1.2). Berlin/New York: De Gruyter, S. 807–900.
- Wittibschlager, Anja (in Vorb.): Verbalgrammatische Dynamik – Variationslinguistische Analysen im österreichischen Sprachraum. Dissertation. Wien: Universität Wien.
- Ziegler, Theresa (2021): Über Geschenk-s-körbe und Schokolade-n-torten (zu runden Geburtstagen). Sneak Peek auf eine Abschlussarbeit über areal-horizontale Verfügungstendenzen bei NN-Komposita mit schwankenden Fugenelementen in der österreichischen Standard(schrift)sprache. In: Wiener Linguistische Gazette 89, S. 55–83.