IDS | LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

62. Jahrestagung
des Leibniz-Instituts für Deutsche Sprache

DEUTSCH IM EUROPÄISCHEN SPRACHRAUM

Stand und Perspektiven

10.–12. März 2026

Congress Center Rosengarten Mannheim

ABSTRACTS

Leibniz Gemeinschaft

Stand: 17. Februar 2026

Termine und Inhalte der Vorträge können sich kurzfristig ändern.
Bitte beachten Sie auch die entsprechenden Hinweise während der Jahrestagung.

Gestaltung: Norbert Cußler-Volz (IDS / Öffentlichkeitsarbeit)

DOI: 10.14618/6de0-3c41

DEUTSCH IM EUROPÄISCHEN SPRACHRAUM

Stand und Perspektiven

Tagungskonzeption

Die Stellung und Funktion des Deutschen im europäischen Sprachraum lassen sich auf unterschiedlichen Ebenen untersuchen: in strukturellen und lexikalischen Eigenschaften, in interaktionalen und diskursiven Mustern sowie in seiner gesellschaftlichen und institutionellen Verankerung. Die Analyse des Deutschen im europäischen Kontext eröffnet damit sowohl vergleichende als auch kontextualisierende Perspektiven und macht die Wechselwirkungen zwischen sprachlichen Systemen, Gebrauchspraktiken und soziopolitischen Rahmenbedingungen sichtbar. Mehrere Arbeiten, die thematisch in diese Fragestellungen hineinreichen, wurden im Programmbereich *Grammatisches Wissen und Sprachgebrauch* unter der Leitung von Dr. Marek Konopka erarbeitet. Sie stehen exemplarisch für eine empirisch fundierte Auseinandersetzung mit grammatischen Strukturen der deutschen Gegenwartssprache in geschriebener und gesprochener Form sowie für deren Einordnung in größere systematische und historische Zusammenhänge. Wir möchten an dieser Stelle an Dr. Marek Konopka erinnern, der im Dezember 2025 plötzlich und unerwartet verstorben ist.[1]

Im weiteren Fokus stehen diskursanalytische, soziolinguistische und sprachpolitische Perspektiven auf das Deutsche in Europa. Berücksichtigt werden dabei seine plurizentrische Ausprägung ebenso wie seine Rolle als Minderheiten-, Amts- und Wissenschaftssprache sowie als kommunikative Ressource in mehrsprachigen gesellschaftlichen Kontexten. Besondere Aufmerksamkeit gilt den Wechselwirkungen zwischen Sprache, institutionellen Rahmenbedingungen und gesellschaftlichem Wandel. Ein weiterer inhaltlicher Schwerpunkt liegt auf der vergleichenden Sprachforschung. Im Zentrum stehen Fragen nach geeigneten *tertia comparationis* für kontrastive Analysen, nach typologischen Schnittstellen zwischen europäischen Sprachen sowie nach methodischen Zugängen zur empirischen Erhebung und Auswertung sprachlicher Daten. Die vergleichende Perspektive ermöglicht differenzierte Einsichten in konvergente und divergente Entwicklungen innerhalb europäischer Sprachräume und trägt zur Systematisierung und Modellierung sprachlicher Variation mit besonderem Blick auf das Deutsche bei. Dabei werden unter anderem die folgenden Themenbereiche diskutiert, die Eigenschaften, Vergleichsperspektiven und gesellschaftliche Rahmenbedingungen des Deutschen im europäischen Sprachraum betreffen:

- Strukturelle, lexikalische, interaktionale und diskursive Eigenschaften des Deutschen im europäischen Kontext
- Plurizentrik des Deutschen und seine Funktionen in mehrsprachigen Gesellschaften
- Deutsch als Minderheiten-, Amts- und Wissenschaftssprache
- Diskursanalytische, soziolinguistische und sprachpolitische Zugänge
- Vergleichende Sprachforschung: typologische Schnittstellen und Methodik

[1] https://www.ids-mannheim.de/aktuell/personalia/nachruf-konopka/

DIE VORTRÄGE DER JAHRESTAGUNG:

DIENSTAG, 10. MÄRZ 2026, 10.00 UHR

Deutsch in Europa – Deutsch im D-A-CH-Raum. Soziolinguistische und didaktische Perspektiven

Christa Dürscheid (Universität Zürich (Schweiz))

Der Vortrag gibt zunächst einen kurzen Überblick über die Stellung der deutschen Sprache in Europa. Im Fokus stehen hier sowohl die sieben Länder, in denen Deutsch nationale oder regionale Amtssprache ist, als auch einige der Länder, in denen die deutsche Sprache als Minderheitensprache anerkannt wird. Dabei stütze ich mich auf die Ausführungen in dem Buch „Deutsch in Europa", das im Jahr 2025 unter der Projektleitung von Rita Franceschini und mir von der Deutschen Akademie für Sprache und Dichtung und der Akademienunion herausgegeben wurde.

Im zweiten Teil des Vortrags liegt der Schwerpunkt auf den Ländern, die von Ulrich Ammon als die Vollzentren der deutschen Sprache bezeichnet wurden (in absteigender Sprecherzahl): Deutschland (D), Österreich (A), Schweiz (CH). Zunächst werden einige Hinweise und Beispiele zum Gebrauch des Standarddeutschen im D-A-CH-Raum gegeben – basierend auf den drei Nachschlagewerken, in denen die standardsprachlichen Charakteristika umfassend dokumentiert sind (dazu hier nur die Kürzel: VWB, VG und AADG). Dann werde ich zeigen, dass dahinter ein varietätenbezogener, quantifizierender Ansatz steht, der in neueren Arbeiten durch Untersuchungen ergänzt wird, in denen das sprachliche Handeln in konkreten Kommunikationssituationen im Fokus steht. Aktuelle Publikationen dazu situieren sich in der Variationspragmatik (vgl. Stumpf et al. 2025), schließen aber auch an die interaktionale Soziolinguistik an (vgl. Eckert 2000, Spitzmüller 2022).

Der dritte Teil des Vortrags widmet sich didaktischen Fragen. In der DaF-Didaktik wird dafür argumentiert, dass im Unterricht ein Bewusstsein für die kulturelle und sprachliche Vielfalt des Deutschen aufgebaut werden sollte. Das DACH-Prinzip, das von einer Arbeitsgruppe im Internationalen Deutschlehrerinnen- und Deutschlehrerverband (IDV) ausformuliert wurde, bringt diese Auffassung deutlich zum Ausdruck (vgl. dazu auch Shafer et al. 2020). Im Vortrag wird aber nicht nur auf DaF-didaktische Aspekte eingegangen, sondern dafür argumentiert, dass auch im Erstsprachunterricht eine Reflexion über den Gebrauch des Deutschen im D-A-CH-Raum angestrebt werden sollte – und mehr noch: eine Reflexion über die Stellung der deutschen Sprache in Europa. Wie dies umgesetzt werden kann (z.B. als gemeinsames Projekt der Schulfächer Deutsch, Erdkunde, Geschichte), soll abschließend gezeigt werden.

Referenzen:

Dürscheid, Christa/Franceschini, Rita (Projektleitung) (2025): Deutsch in Europa. Vielfalt, Sprachnormen und Sprachgebrauch. Vierter Bericht zur Lage der deutschen Sprache. Hrsg. von der

Deutschen Akademie für Sprache und Dichtung und der Union der deutschen Akademien der Wissenschaften. Tübingen: Narr.

Eckert, Penelope (2000): Linguistic Variation as Social Practice. The Linguistic Construction of Identity in Belten High. Malden, MA/Oxford: Wiley-Blackwell (= Language in Society 27).

Shafer, Naomi/Middeke, Annegret/Hägi-Mead, Sara/Schweiger, Hannes (Hrsg.) (2020): Weitergedacht. Das DACH-Prinzip in der Praxis. Göttingen: Universitätsverlag (= Materialien Deutsch als Fremd- und Zweitsprache 103).

Spitzmüller, Jürgen (2022): Soziolinguistik. Eine Einführung. Heidelberg: J. B. Metzler.

Stumpf, Sören/Merten, Marie-Luis/Kabatnik, Susanne/Zollner, Sebastian (Hrsg.) (2025): Variationspragmatik. Regionale Vielfalt und situative Unterschiede im Sprachgebrauch. Tübingen: Narr (= Studien zur Pragmatik 8).

DIENSTAG, 10. MÄRZ 2026, 11.15 UHR

Die deutsche Sprache in Europa: Historisches Erbe und aktuelle Tendenzen

Claudia Maria Riehl (Ludwig-Maximilians-Universität München)

In weiten Teilen Europas bildet die deutsche Sprache ein historisches Erbe: So gibt es seit dem Mittelalter Ansiedlungen deutschsprachiger Gruppen, die teilweise noch bis heute Dialekte oder Regionalsprachen, die sich aus deutschen Dialekten herausgebildet haben, bewahrt haben. Andere Siedlungen sind in der Neuzeit entstanden und entwickelten ihre eigenen auf verschiedenen deutschen Dialekten basierenden Mischdialekte (Riehl 2025; Riehl/Beyer 2021). Viele dieser Gruppen hatten auch lange Zeit Zugang zur deutschen Schriftsprache oder haben es noch bis heute. Andere deutschsprachige Gemeinschaften wurden erst nach dem Ersten Weltkrieg zu Sprachminderheiten in anderssprachigen Nationen, wie etwa die Deutschen in Südtirol oder in Dänemark (Beyer/Plewnia 2019; Riehl/Beyer 2021).

In diesem Vortrag wird zunächst ein historischer Überblick über die Entstehung der verschiedenen deutschsprachigen Gemeinschaften in Europa und die soziolinguistischen Rahmenbedingungen gegeben. Danach werden die wesentlichen Veränderungen nach dem Zweiten Weltkrieg und die verschiedenen Faktoren diskutiert, die dazu geführt haben, dass bestimmte deutsche Sprachminderheiten ihre Varietät bis heute bewahren konnten. Hier spielen die Institutionalisierung der Sprache (durch Schule, Religion) sowie Spracheinstellungen eine bedeutende Rolle. Weiter wird gezeigt, wie sich das Varietätenspektrum, über das die Sprecherinnen und Sprecher verfügen, in den verschiedenen Regionen unterscheidet: Wir finden eine Vielzahl von Mehrsprachigkeitskonstellationen, angefangen von einem kompletten Varietätenspektrum des Deutschen plus Italienisch in Südtirol bis hin zu multilingualen Konstellationen mit einem deutschen Dialekt und weiteren Sprachen und Dialekten im ukrainischen Transkarpatien (Franceschini/Riehl 2025).

In einem zweiten Teil werden einige wesentliche Entwicklungen der deutschen Sprache in den Minderheitenregionen dargestellt. Der Fokus liegt dabei auf Sprachkontaktphänomenen, besonders auf semantischem und grammatischem Transfer aus den Umgebungssprachen. Daneben werden sprachinterne Entwicklungen beschrieben, die in allen Diaspora-Varietäten auftreten, wie etwa der Kasusabbau oder der Abbau der Verbendstellung im Nebensatz. Der Umfang der Veränderung ist im Wesentlichen durch den Status des Deutschen in der jeweiligen Region und die Mehrsprachigkeit der Sprecherinnen und Sprecher geprägt (vgl. Riehl 2026).

In einem letzten Punkt wird auf die Bedeutung von Deutsch als Zweit- bzw. Fremdsprache im europäischen Sprachraum eingegangen: Aus historischer Sicht wird u.a. die Verbreitung der deutschen Sprache österreichischer Prägung als Lingua Franca in der Donaumonarchie angesprochen (vgl. Newerkla 2022). Daran anschließend werden aktuelle Tendenzen zum Gebrauch des Deutschen im Rahmen der modernen Migration diskutiert: der Erwerb des Deutschen als Fremdsprache von Nachfahren der deutschen Minderheiten, die Rolle des Pendlertums, die Rückwanderung von deutschsprachigen Nachkriegsaussiedlern, die Remigration von Arbeitsmigrantinnen und -migranten und der Einfluss der sozialen Medien.

Referenzen:

Beyer, R./Plewnia, A. (Hgg.) (2019): Handbuch des Deutschen in West- und Mitteleuropa. Sprachminderheiten und Mehrsprachigkeitskonstellationen. Tübingen: Narr.

Franceschini, R./Riehl, C.M. (2025). Sprachkontakt und Mehrsprachigkeitskonstellationen. In: *Deutsch in Europa. Vielfalt, Sprachnormen und Sprachgebrauch. Vierter Bericht zur Lage der deutschen Sprache.* Tübingen: Narr, 285-322.

Newerkla, S. M. (2022). Reconstructing multilingualism in the Habsburg state: Lessons learnt and implications for historical sociolinguistics. *Sociolinguistica*, 36(1-2), 151-167.

Riehl, C.M. (2025). Zur Entstehung deutschsprachiger Gemeinschaften außerhalb deutschsprachiger Staaten: ein Überblick. In: *Deutsch in Europa. Vielfalt, Sprachnormen und Sprachgebrauch. Vierter Bericht zur Lage der deutschen Sprache.* Tübingen: Narr, 149-169.

Riehl, C.M. (2026). *German as a Heritage Language: Contact, Variation, and Maintenance.* Cambridge: Cambridge University Press.

Riehl, C.M./Beyer, R. (2021). Deutsch als Minderheitensprache. *Lublin Studies in Modern Languages and Literature*, 45(1), 7-20.

DIENSTAG, 10. MÄRZ 2026, 12.00 UHR

Deutsch aus der Perspektive der European Federation of National Institutions for Language (EFNIL)

Sabine Kirchmeier (Europäische Föderation Nationaler Sprachinstitutionen (EFNIL))

EFNIL[1] ist ein Zusammenschluss offizieller Sprachinstitutionen in Europa, die sich mit Sprachpolitik, Sprachforschung, Mehrsprachigkeit und Sprachressourcen in Europa beschäftigen. Die Föderation unterstützt das Studium der offiziellen europäischen Sprachen und das Erlernen von Muttersprache und Fremdsprachen als Mittel zur Förderung der sprachlichen und kulturellen Vielfalt.

EFNIL hat 43 Mitgliedsinstitutionen aus 31 europäischen Staaten (24 Staaten innerhalb der EU und 7 assoziierte Staaten), die sich unter Mitwirkung der Sprachinstitutionen des Europa-Parlaments und der Europäischen Kommission jährlich mehrmals über Mehrsprachigkeit, Sprachpolitik und Sprachforschung in Europa in Webinaren und in einer Jahrestagung austauschen und Forschungsprojekte initiieren.

Der deutschsprachige Raum wird durch das Leibniz-Institut für Deutsche Sprache in Mannheim, die Deutsche Akademie für Sprache und Dichtung in Darmstadt, das Austrian Centre for Digital Humanities der Österreichischen Akademie der Wissenschaften in Wien, das Österreichische Sprachen-Kompetenz-Zentrum in Graz und durch das Institut für Mehrsprachigkeit der Universität Freiburg repräsentiert.

Da die meisten EFNIL-Mitglieder offizielle, staatliche Sprachinstitutionen oder zentrale akademische Institutionen in ihrem Land sind, ist EFNIL in einer ausgezeichneten Position, Informationen über die Sprachsituation in jedem Land zu sammeln und zu analysieren.

Das Projekt *European Language Monitor* (ELM)[2] sammelt in 4-jährigen Abständen Informationen über die sprachpolitische Situation in Europa insbesondere über die Regulierung der Sprachen im Schulbetrieb, an den Universitäten, in der Wirtschaft und im Kulturbereich.

Das Projekt *European Languages and their Intelligibility in the Public Sphere* (ELIPS)[3] untersucht die Kommunikation der öffentlichen Institutionen mit den Bürgerinnen und Bürgern in den einzelnen Ländern. Die Ergebnisse dieser Projekte nutzen Forschende und Entscheidungsträger, insbesondere Mitarbeitende der Institutionen der EU, um einen Überblick über den aktuellen Stand der Sprachen in den Ländern Europas zu bekommen.

---

1 https://efnil.org
2 https://efnil.org/projects/european-language-monitor-elm/
3 https://efnil.org/projects/european-languages-and-their-intelligibility-in-the-public-sphere-elips/

EFNIL prämiert seit 2020 jedes Jahr die drei besten Diplomarbeiten zu den Themen Mehrsprachigkeit und Sprachpolitik mit dem *EFNIL Master's Thesis Award* (MTA)[4], um junge Forschende in diesem Bereich zu fördern.

In meinem Vortrag werde ich die Rolle der deutschen Sprache in Europa aus der Perspektive beleuchten, die sich aus den sprachpolitischen Untersuchungen ergibt, die im Rahmen von EFNIL vorgenommen wurden. Den Abschluss bilden einige Betrachtungen zum aktuellen Einfluss des Deutschen auf die europäischen Sprachen.

DIENSTAG, 10. MÄRZ 2026, 14.15 UHR

Deutsch(e) in Ungarn – aktuell

Koloman Brenner (Eötvös-Loránd-Universität, Budapest (Ungarn))

Die deutsche Minderheit in Ungarn gehört zu den relativ großen deutschen Gemeinschaften in Ost-Mittel-Europa, die eine lange historische Entwicklung vorweisen kann. Im heutigen Ungarn gibt es drei größere Siedlungsgebiete, wo Angehörige der deutschen Minderheit in höherer Anzahl leben: Westungarn entlang der österreichischen Grenze, das Ungarische Mittelgebirge (vom Ofner Bergland bis zum Plattensee-Oberland) mit den Zentren Ofen/Buda und Zirc bzw. Südungarn mit dem Zentrum Fünfkirchen/Pécs. Die Tatsache, dass ab dem 16. Jahrhundert bis 1918 das ungarische Königreich in unterschiedlichen juristischen Konstruktionen mit dem Habsburger-Reich verbunden war, brachte ebenfalls in mannigfaltiger Weise einen sprachlichen und kulturellen deutschen Einfluss mit sich.

Die diversen Identitätsmuster unter Angehörigen der deutschen Minderheit variieren auf der individuellen Ebene stark, falls die Revitalisierung der deutschen Sprache und Kultur gelingt, wird dadurch die deutsche Identitätskomponente modernisiert und verstärkt. Eine neue deutschsprachige Gemeinschaft ist ebenfalls zu erwähnen, die infolge der diversen Firmenniederlassungen aus deutschen Landen in den letzten Jahrzehnten entstanden ist. Eine bessere Sprachpolitik der deutschsprachigen Länder wäre ebenfalls wünschenswert, nicht nur bezüglich der Stellung und Rolle der deutschen Sprache in Ungarn, sondern generell in vielen Ländern von Europa, wo traditionell und geschichtlich eine gute Basis dafür vorhanden wäre.

[4] https://efnil.org/projects/masters-thesis-award/

DIENSTAG, 10. MÄRZ 2026, 15.00 UHR

Sprachbiographien, Sprachpraktiken und Identitäten bei deutschen Minderheiten im östlichen Europa zwischen Tradition und heutiger Mehrsprachigkeit

Katharina Dück und Heiko F. Marten (beide IDS)

Seit einigen Jahren hat sich das IDS zur Aufgabe gemacht, in seinem Programmbereich „Sprache im öffentlichen Raum" soziolinguistische Strukturen in Regionen mit Deutsch als Minderheitensprache zu untersuchen. Dazu werden im Projekt „Sprachminderheiten- und Mehrsprachigkeitskonstellationen unter Beteiligung des Deutschen" Interviews mit Personen geführt, die sich einer deutschen Minderheit zugehörig fühlen, auf deutsche Familiengeschichten zurückblicken und/oder in einem deutschen Minderheitenverein engagiert sind. Diese Interviews werden durch Daten ergänzt, die durch ethnographische Beobachtungen, Auswertungen von Diskursen zu Sprachen, durch historische Berichte, Linguistic-Landscape-Studien und andere Methoden gewonnen wurden.

Der Vortrag fasst einige Ergebnisse dieser hauptsächlich im östlichen Europa und in Zentralasien durchgeführten Projekte zusammen. Dabei handelt es sich um Regionen, in denen Deutsch historisch die Sprache von zahlreichen L1-Sprecher/-innen war, deren Zahl im Laufe des 20. Jahrhunderts aber stark zurückgegangen ist. Exemplarisch werden Daten aus Lettland und Georgien vorgestellt – zwei Ländern, in denen nach Vertreibung, Deportationen und Aussiedlung heute nur noch Reste germanophoner Minderheiten existieren, Deutsch in unterschiedlichen gesellschaftlichen Funktionen aber durchaus eine Rolle spielt.

Nach einer Einführung in den Hintergrund der Forschung stellt der Vortrag die heutigen Sprachpraktiken der Interviewten vor. Diese werden in Bezug zu individuellen Biographien gesetzt. Unsere Daten zeigen, dass Deutsch heute immer Teil einer komplexen gesellschaftlichen und persönlichen Mehrsprachigkeit ist. Dabei sind für ältere Informant/-innen zumeist Traumatisierung und Verheimlichung einer deutschen Identität in der Nachkriegszeit als Folge einer Repressionspolitik zentral, die zu einer Unterbrechung der Weitergabe der deutschen Sprache an die nachfolgenden Generationen geführt hat. Für jüngere Interviewte überwiegen hingegen zumeist pragmatische Erwägungen, weshalb Deutsch als Teil eines größeren Sprachrepertoires wertgeschätzt wird. Diese verschiedenen Biographien und Sprachpraktiken führen zu ausgesprochen heterogenen Identitätswahrnehmungen der Informanten, die zwischen lokaler Verwurzelung, (post-)sowjetischer Prägung, Betonung deutscher Ethnizität und Teilhabe an globalen Sprach- und Kulturtrends oszillieren.

Vor dem Hintergrund dieser diversen Sprachpraktiken und Identitäten möchte der Vortrag schließlich einen Ausblick darauf geben, wie auf der Grundlage der bisherigen Forschung eine Erweiterung von Methoden und Themenschwerpunkten in der Untersuchung von Deutsch in Mehrsprachigkeitskontexten vorgenommen werden kann. Als

Ergänzung der traditionellen Erforschung von Deutsch als Minderheitensprache erhält dabei die Verbindung der Minderheit mit anderen Personengruppen Aufmerksamkeit, für die Deutsch in unterschiedlichen Funktionen von Bedeutung ist. Dazu gehören neue deutschsprachige Minderheiten und Rückkehrer/-innen mit oftmals transnationalen wie transmigrationalen Lebensentwürfen ebenso wie Personen, die Deutsch als Fremdsprache in unterschiedlichen Funktionen gebrauchen. Diese Änderungen in der Betrachtung von Deutsch im Zusammenspiel unterschiedlicher Akteure hat nicht zuletzt Auswirkungen auf die Sprachenpolitik, in der Deutsch als eine Sprache in unterschiedlichsten Mehrsprachigkeitskonstellationen Berücksichtigung erfährt.

DIENSTAG, 10. MÄRZ 2026, 16.15 UHR

Empirisch, gegenstandsbezogen, anwendungsorientiert: Forschungsschwerpunkte und aktuelle Entwicklungen im Fach Deutsch als Fremd- und Zweitsprache

Christian Fandrych und Katrin Wisniewski (beide Universität Leipzig)

Das Fach Deutsch als Fremd- und Zweitsprache (DaF/DaZ) hat sich als akademische Disziplin an vielen Universitäten im deutschsprachigen Raum etabliert. Auch wenn die vielleicht etwas missverständliche Bezeichnung immer wieder verkürzend als reine Praxis des konkreten Unterrichts- und Lerngeschehens verstanden wird, ist DaF/DaZ ein wissenschaftliches Fach, das in den letzten Jahrzehnten – nach verschiedenen Strukturfindungsphasen – eigenständige Forschungsperspektiven entwickelt hat und zu beträchtlicher akademischer Reife gelangt ist. Dies ging einher mit einer stetig zunehmenden Ausdifferenzierung von Methoden, Gegenständen und Fragestellungen.

In unserem Vortrag wollen wir zentrale Forschungsansätze und -methoden des Faches anhand ausgewählter aktueller Beispiele verdeutlichen. Zweitens werden wichtige Herausforderungen und Aufgaben benannt, mit denen sich die Disziplin derzeit konfrontiert sieht. Dazu zählen insbesondere auch Fragen, die die internationale Stellung und Vernetzung des Fachs betreffen.

„Sprache" muss als Gegenstand von DaF/DaZ-Forschung in einem umfassenden, den Sprachgebrauch, verschiedene Sprachhandlungsdomänen und kommunikative Gattungen mit einbeziehenden Sinn verstanden werden: Ziel von Spracherwerb und Sprachvermittlung ist die Handlungsfähigkeit in einem für die Lernenden relevanten Zielsprachenkontext – ob dies Alltagskommunikation auf Reisen, berufliche Handlungskompetenz oder wissenschaftliche Lesekompetenz ist. Für Sprachvermittlungszwecke ist es daher zunächst erforderlich, dass der in bestimmten Zieldomänen relevante Sprachbedarf und die Sprachverwendungsbedürfnisse der Lernenden ermittelt werden – durch die Erhebung von Korpora, der domänenspezifischen Handlungsanforderungen, sprachlichen Mittel, sozialen Rollen und damit verbundenen sozialen Konventionen. Besonders im Mittelpunkt standen hier zuletzt die kommunikativen Domä-

nen Bildungssprache, Wissenschaftskommunikation und berufliche Kommunikation. Im Vortrag wird exemplarisch gezeigt, welche Forschungsinteressen und Fragestellungen hier aus Sicht des Faches DaF/DaZ relevant sind. Im Mittelpunkt stehen einerseits Korpora, die spezifische kommunikative Gattungen (auch im Sprachvergleich) erhoben haben (etwa dem Bereich der Wissenschaftskommunikation), zum anderen Projekte, die es erlauben, für die Sprachvermittlung relevante Sprachgebrauchsphänomene auszuwählen und zu untersuchen.

In der L2-Erwerbsforschung zum Deutschen wiederum kristallisiert sich derzeit das dynamische und komplexe Spannungsverhältnis zwischen Systematizität und Variation als eine Art Metathema heraus. Zwar spielen diese Stellgrößen schon lange eine wichtige Rolle für die erwerbsbezogene Forschung, doch wird in jüngster Zeit die Frage des Ausmaßes und der Rolle von Variation in stark verschärfter Form deutlich. Ermöglicht werden entsprechende Befunde v.a. auf Grundlage einer sich stetig ausdifferenzierenden Lernerkorpuslandschaft. Inwieweit vermeintlich robuste Befunde zu geordneten Erwerbsverläufen angesichts drastischer lernersprachlicher Variation überhaupt haltbar sind, ist im Licht aktueller Befunde, auf die im Vortrag ausschnittweise eingegangen wird, fraglich.

Ein weiterer zentraler Forschungsbereich im Fach DaF/DaZ betrifft die Beurteilung von Deutschkompetenzen. Die Bedeutung und Zahl diagnostischer Verfahren hat in den letzten Jahren stark zugenommen, die mit solchen Tests einhergehenden Konsequenzen haben sich teils erheblich verschärft. Forschungsdesiderata dazu, wie sich Sprachkompetenzen valide diagnostizieren lassen, bestehen unterdessen fort. Dieses an Schnittstellen von Sprachwissenschaft, Testforschung und empirischer Bildungs- bzw. Hochschulforschung situierte Thema, das immer wieder auch politische Fragestellungen berührt, ist geprägt von einer besonders tiefen Kluft zwischen Forschung und Praxis und gekennzeichnet von einer nur punktuellen akademischen Validierungskultur. Der Vortrag illustriert dies und unterstreicht anhand einiger Beispiele die Bedeutung einer unabhängigen akademischen Testkultur in DaF/DaZ.

Trotz der in diesem Rahmen nur im Ansatz darstellbaren Vielfalt der DaF/DaZ-Forschung steht das Fach auch vor einer ganzen Reihe an strukturellen, ressourcenbezogenen und fachlichen Herausforderungen und offenen Fragen. Einige wichtige werden im Vortrag thematisiert, z.B. die Frage aktueller Unwuchten hinsichtlich des ja im Kern international ausgerichteten Charakters des Fachs, der Lückenhaftigkeit und teils mangelhaften Sichtbarkeit von Grundlagenforschung in DaF/DaZ, seiner problematischen infrastrukturellen Anbindung und aktuell beobachtbaren Tendenzen zu Lagerbildungen im Fach.

Das Deutsche in mehrsprachigen europäischen Öffentlichkeiten. Programmatik der Vergleichenden Diskurslinguistik und Anforderungen an offene digitale Forschungsinfrastrukturen

Julia Krasselt und Philipp Dreesen (beide Zürcher Hochschule für Angewandte Wissenschaften, Zürich (Schweiz))

Die Frage nach der Stellung des Deutschen im europäischen Sprachraum berührt grundlegende gesellschaftliche Verständigungsprozesse: Wie wird in Europa über Sprache, Kultur und Identität gesprochen? Und wie formen diese Verständigungsprozesse den Raum, in dem die europäische Mehrsprachigkeit diskursiv hergestellt, stabilisiert oder auch infrage gestellt wird? Unser Vortrag setzt an dieser Schnittstelle von konstruiertem Sprachraum, mediatisierter Öffentlichkeit und konkreter Kommunikationspraktik an. Der Vortrag verfolgt zwei Ziele: Erstens argumentieren wir programmatisch für die Stärkung der Vergleichenden Diskurslinguistik (VDL). Zweitens werden verfügbare digitale Tools, Methoden und Sprachdaten kritisch beleuchtet, um die vorhandenen Potenziale ebenso sichtbar zu machen wie die bestehenden Desiderata.

Ad 1) Ausgangspunkt ist die Beobachtung, dass die Position des Deutschen im europäischen Sprachraum nicht nur durch demografische, politische oder institutionelle Faktoren geprägt ist, sondern maßgeblich durch jene öffentlichen Diskurse, in denen Europa Sprache, Zugehörigkeit und gesellschaftliche Selbstverständnisse verhandelt. Die VDL versteht diesen Raum als relationales Bedeutungsfeld, das durch parallel verlaufende nationale und transnationale Kommunikationsprozesse entsteht. Europäische und internationale Krisen verdeutlichen, dass zentrale Themen wie politische Spannungen, wahrgenommene Polarisierungen, Kriegsgefahren und Digitalisierungseffekte in unterschiedlichen Öffentlichkeiten Europas thematisiert werden. Eine vergleichende Perspektive ist notwendig, um sowohl geteilte Muster wie auch divergierende Diskursordnungen sichtbar zu machen, um so zu einer europäischen Öffentlichkeit beizutragen.

Auch bildungspolitisch gewinnt die VDL an Bedeutung. Die sinkende Nachfrage an geisteswissenschaftliche Studiengänge im In- und Ausland steht in Spannung zu einem wachsenden Bedarf an reflexiver sprachlicher und kultureller Kompetenz. Die VDL kann hier auch mittels digitaler Kompetenzvermittlung neue Attraktivität entfalten, weil sie den Wert sprachlicher, diskursiver und hermeneutischer Expertise sichtbar macht und diese mit digitalen Methoden, Tools und Sprachdaten verbinden kann.

Aus angewandter Perspektive wird die Relevanz der VDL in jenen professionellen Kommunikationsräumen sichtbar, in denen Organisationen, Verwaltungen und Unternehmen in mehreren Sprachen und Öffentlichkeiten zugleich agieren. Praktiker/-innen in Sprachberufen profitieren von der Fähigkeit, kommunikative Ereignisse in

ihre diskursiven Kontexte einzuordnen, vergleichend zu interpretieren und strategisch zu nutzen. Dies ist ein Profil, das im europäischen, mehrsprachigen Umfeld nach wie vor gefragt ist.

Ad 2) Zweitens beleuchten wir kritisch verfügbare digitale Infrastrukturen der europäischen OpenScience-Community, um auf Potenziale und Desiderata hinzuweisen. Die europaweite Verfügbarkeit offener Sprachdaten und digitaler Werkzeuge kann ein wesentlicher Motor für die Weiterentwicklung der VDL sein, erfordert aber zugleich Reflexion über Interoperabilität, Datenäquivalenz und rechtliche Rahmenbedingungen.

Der Vortrag greift diese Fragen anhand des mehrsprachigen Korpusprojekts Swiss-AL sowie aktueller Entwicklungen in der europäischen Open-Science-Community auf und zeigt, welche infrastrukturellen Voraussetzungen für eine vergleichende Diskurslinguistik notwendig sind. Der Vortrag baut auf den Ergebnissen des EU-geförderten Projekt MORCDA („Making Open Research Data suitable for Comparative Discourse Analysis"), dem an der ZHAW angesiedelten CLARIN Knowledge Centre for Applied Comparative Discourse Analysis (CLARIN-APPLIED), der Gründung des Forschungsnetzwerks comparative discourse studies sowie dem mehrjährigen Aufbau und Einsatz der Plattform Swiss-AL für die angewandte Diskursanalyse in der mehrsprachigen Schweiz auf.

MITTWOCH, 11. MÄRZ 2026, 9.45 UHR

Wasserlaufmetaphern im Flüchtlingsdiskurs der Wikipedia aus sprachübergreifender und mikrodiachroner Perspektive

Janja Polajnar Lenarčič (Universität Ljubljana (Slowenien))

Die intensivierte Flüchtlingsbewegung nach Europa seit 2015 wird bis heute europaübergreifend an unterschiedlichen Thematisierungsorten (im öffentlich-medialen Diskurs, in wissenschaftlichen Publikationen, in der Wikipedia usw.) diskursiv relevant gesetzt. Hierbei weist die Wasserlauf-Metaphorik neben Naturkatastrophen-, Kriegs-, Krankheits-, Reise- und Wegmetaphern u.a. eine beeindruckende Dynamik auf und wird durch Diskursakteure zu einem vielfältigen Metaphernfeld des Wasserslaufs ausgebaut, das als zentrales Bildfeld den Flüchtlingsdiskurs konstituiert und strukturiert. Die diskursive Dominanz von Wasserlauf- und Naturkatastrophen-Metaphern geht aus zahlreichen diskurslinguistischen Studien hervor, die Wasserlauf-Metaphorik im Flüchtlingsdiskurs meist in der Presseberichterstattung einzelner Länder oder kontrastiv in unterschiedlichen Ländern untersuchen und sprachkritisch beleuchten (vgl. Spieß 2017, Issel-Dombert/Wieders-Lohéac 2018, Gruber 2018, Fijavž/Fišer 2020 usw.).

Der vorliegende Vortrag ergänzt die Erforschung von Wasserlauf-Metaphorik im Flüchtlingsdiskurs mehrfach, indem er im Gegensatz zum bisherigen Fokus auf Presseberichterstattung ein relativ wenig beachtetes, jedoch aus der Perspektive der kontrastiven Diskurslinguistik hoch relevantes digitales Diskursfragment untersucht

(vgl. Flinz/Gredel 2019, Gredel 2020, 2024). Im Fokus stehen Dynamiken von Wasserlauf-Metaphern zum Zielbereich Flüchtlinge und deren Bewegung nach Europa. Sie werden aus multimodaler und sprachübergreifender Perspektive untersucht, wie sie sich mikrodiachron in zwei Zeitscheiben auf deutschen und englischen Artikel- und Artikeldiskussionsseiten der Diskursfragmente *Europäische Flüchtlingskrise 2015* sowie *2015 European migrant crisis* rekonstruieren lassen. Digitale Diskursfragmente in Wikipedia sind aufgrund von Links non-linear und sehr komplex, was für die vorliegende Untersuchung von zweifacher Bedeutung ist: Die Verlinkung von Artikel- und Artikeldiskussionsseiten zu einem Thema legt es nahe, die Artikelseiten nicht isoliert sondern in Relation zu den Diskussionsseiten zu betrachten. Zudem stellen die Artikelseiten das Produkt diskursiver Aushandlungen auf den Artikeldiskussionsseiten dar. Die Verlinkung von Artikelseiten zu einem Thema über die Sprachen hinweg legt es nahe, diese verlinkten Diskursfragmente kontrastiv zu analysieren. Hierbei werden Metaphern als dynamische, kontextuelle, situationsgebundene und kulturell verortete Phänomene mit ihren diskursspezifischen Ausprägungen, Bedeutungen und Funktionen (vgl. Liebert 1992) aufgefasst, die diskursabhängig beschrieben werden sollen. Als sprachliche Muster sind Metaphern Zeichenkomplexe auf verschiedenen sprachstrukturellen Ebenen. Da sie in unterschiedlichen Zeichenmodalitäten realisiert werden können, werden neben sprachlichen auch visuelle und multimodale Metaphernrepräsentationen (vgl. Forceville 2006) berücksichtigt.

Im Vortrag werden im Rahmen einer korpusgestützen kontrastiven Diskursanalyse zunächst auf den Artikeldiskussionsseiten unterschiedliche Reflexionsgrade der Wikipedia-Autorinnen und -Autoren zum sensiblen Metapherngebrauch analysiert. Es wird davon ausgegangen, dass sich die sprachkritische Metadiskussion auf den Diskussionsseiten auf den Metapherngebrauch auf den Artikelseiten auswirkt. Folglich werden in einem nächsten Schritt Metapherninventare zum Quellbereich Wasserlauf textorientiert analysiert, indem auf den Artikelseiten sprachübergreifend rekonstruiert wird, wie multimodale Wasserlauf-Metaphern im englischen und deutschen Diskursfragment zu einem vielfältigen, multimodalen Metaphernfeld ausgebaut werden, das bestimmte Diskurspositionen stärkt. Schließlich werden über zwei Versionsgeschichten (2015 und 2025) mikrodiachrone Dynamiken im Metapherngebrauch intralingual und kontrastiv erarbeitet.

Referenzen:

Fijavž, Zoran/Fišer, Darja (2020): Corpus-assisted analysis of water flow metaphors in Slovene online news migration discourse of 2015. In: Fišer, Darja/Smith, Philippa (Hg.), The dark side of digital platforms: linguistic investigations of socially unacceptable online discourse practices, 1st ed. Ljubljana: University Press, Faculty of Arts, S. 56-84.

Flinz, Carolina/Gredel, Eva (2019): Bildinventare und konkurrierende Termini im Flüchtlingsdiskurs in der Wikipedia. Eine kontrastive Diskursanalyse der deutschen und der italienischen Sprachversion. In: Schiewe, Jürgen/Niehr, Thomas/Moraldo, Sandro M. (Hg.): Sprach(kritik) Kompetenz als Mittel demokratischer Willensbildung. Sprachliche In- und Exklusionsstrategien als gesellschaftliche Herausforderung. Bremen: Hempen Verlag, S. 177-196.

Forceville, Charles (2006): Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. In: G. Kristiansen et al. (Hg.): Cognitive Linguistics: Current Applications and Future Perspectives. Berlin: de Gruyter, S. 379-402.

Gredel, Eva (2020): Digitale Diskursanalysen. In: Marx, Konstanze/Lobin, Henning/Schmidt, Axel (Hg.): Deutsch in Sozialen Medien. Interaktiv – multimodal – vielfältig. Jahrbuch des Leibniz-Instituts für Deutsche Sprache. Berlin/Boston: de Gruyter, S. 247-263.

Gredel, Eva (2024): Chapter 5. Sock puppets, Wikifants, and Honeypots: Metaphorical patterns in digital discourses on Wikipedia talk pages. In: Poudat, Celine/ Lüngen, Harald/Herzberg, Laura (Hg.): Investigating Wikipedia: Linguistic corpus building, exploration and analysis. Amsterdam/Philadelphia: Benjamins (= Series in Corpus Linguistics), S. 134-154.

Gruber, Teresa (2018): Migration, Metaphern und Medien: Metaphorische Konzeptualisierungen der 'Flüchtlingskrise' (2014/2015) in der spanischen, italienischen und deutschen Presseberichterstattung. In: Kohlrausch, Laura/Schoeß, Marie/Zejnelovic, Marko (Hg.): Krise: mediale, sprachliche und literarische Horizonte eines viel zitierten Begriffs. Languagetalks, Würzburg: Königshausen & Neumann, S. 59-85.

Issel-Dombert, Sandra/Wieders-Lohéac, Aline (2018): Auf Flüchtlingswellen surft man nicht vor Lampedusa – Zur Metaphorik des öffentlichen Diskurses zur Flüchtlingskrise Italiens. In: metaphern.de 28/2018, S. 77-97.

Liebert, Wolf-Andreas (1992): Metaphernbereiche der deutschen Alltagssprache. Kognitive Linguistik und die Perspektiven einer Kognitiven Lexikografie. Frankfurt a.M.: Peter Lang.

MITTWOCH, 11. MÄRZ 2026, 11.00 UHR

Geschlechterbilder im deutsch-französischen Kontrast: eine (Meta-)Analyse digitaler Korpustools

Hélène Vinckel-Roisin (Universität von Lothringen, Nancy (Frankreich))

In den letzten Jahren wurden verschiedentlich Umfragen zu Geschlechterrollen und -stereotypen durchgeführt, insbesondere auch in ländervergleichender Perspektive: Aus einer Studie des Deutschen Instituts für Wirtschaftsforschung (DIW – Menkhoff/Wrohlich 2024) ergibt sich, dass im europäischen Vergleich das gesellschaftliche Rollenbild der Frau in Bezug auf Familie, Bildung, Politik und Arbeitsmarkt in Deutschland in den letzten Jahrzehnten deutlich moderner geworden sei, aber unter der EU-Spitze bleibe; dem neuesten Eurobarometer-Bericht (2024), *Gender Stereotypes*, zufolge sind europaweit andererseits manche Geschlechterstereotype noch tief verankert, nicht zuletzt solche wie 'Frauen = Hausfrauen, Männer = Geldverdiener'. Besonders auffällig sind deutliche Unterschiede zwischen Deutschland und Frankreich: Der Behauptung „Kinder zu haben ist für Frauen erfüllender als für Männer" stimmen 42% der deutschen Befragten zu, jedoch nur 24% der französischen – ein Ergebnis, das an bereits früher aufgezeigte Divergenzen in den Auffassungen weiblicher Rollenmuster beidseits des Rheins erinnert (vgl. Badinter 2013 sowie u.a. Wiegel 2016).

Vor diesem Hintergrund verortet sich der Vortrag an der Schnittstelle zwischen Genderlinguistik, Lexik und Korpuslinguistik im deutsch-französischen Vergleich: Es werden Ergebnisse einer überwiegend qualitativen Pilotstudie zu Geschlechterbildern von Frauen und Männern vorgestellt, wie sie über automatisch ermittelte Ko-

okkurrenzen von dt. <Frau> bzw. <Mann> und fr. <femme> bzw. <homme> in digitalen Informationsangeboten bzw. Korpusanalysetools vermittelt werden. Untersucht werden die ersten zwanzig Kookkurrenzen (auf der Basis von LogDice) von <Frau> bzw. <Mann> und <femme> bzw. <homme> in ausgewählten Vergleichskorpora (wie 'news', 'wikipedia' bzw. 'reference/encyclopedia', etc.) der Leipzig Corpora Collection (Software: NoSketch Engine; vgl. u.a. Eckart/Quasthoff 2013) und der TenTen23-Korpora von Sketch Engine (Software: Sketch Engine; vgl. u.a. Kilgarriff et al. 2014). Das Hauptaugenmerk liegt auf Kookkurrenzpartnern innerhalb der Nominalphrase: attributiv verwendete Adjektive und Partizipien sowie attributiv verwendete Nomen rechts des nominalen Kopfes im Französischen (*femme artiste*).

Im Anschluss an überwiegend einzelsprachliche Studien zu Geschlechterdarstellungen, darunter insbesondere zum Einfluss der Korpusgrundlage in Wörterbüchern (vgl. u.a. Krötzsch-Viannay 1979, Ackermann 2016, Lautenschläger 2017, Müller-Spitzer 2024, Müller-Spitzer/Lobin 2022) sowie in Sketch Engine (vgl. Pearce 2008, neuerdings Schmuck/Vinckel-Roisin 2025) verfolgt der Vortrag ein doppeltes Ziel:

– Im Rahmen einer Mehr-Ebenen-Analyse wird das Deutsche unter Berücksichtigung sprachstruktureller, kulturell-geschichtlicher und textsortenspezifischer Kriterien im Hinblick auf Konvergenzen und Divergenzen mit dem Französischen verglichen. Als *tertium comparationis* fungieren ausgewählte Charakterisierungsdimensionen wie beispielsweise Aussehen/physische Eigenschaften, Charakter, Ehe/Familie/Elternschaft.

– Im Sinne einer Meta-Analyse wird genauer auf Auffälligkeiten bei den Kookkurrenzpartnern eingegangen, die korpustoolbedingte Asymmetrien bzw. thematische *biases* im deutsch-französischen Kontrast deutlich machen; dabei werden auch mögliche Ursachen (Korpuszusammensetzung, d.h. Auswahl der Quellen, die computergenerierten digitalen Korpora zugrunde liegen) in den Blick genommen.

Referenzen:

Ackermann, Fabian (2015). Gehören nun die Männer an den Herd? Anmerkungen zum Wandel der Rollenbilder von Mann und Frau. In: *Sprachreport*, Jg. 31, Heft 4, 12-15.

Badinter, Elisabeth (2013). *Der Konflikt. Die Frau und die Mutter.* [Aus dem Französischen von Ursula Held und Stephanie Singh]. München: Beck.

Eckart, Thomas/Quasthoff, Uwe (2013). Statistical Corpus and Language Comparison on Comparable Corpora. In: Sharoff, Serge/Rapp, Reinhard/Zweigenbaum, Pierre/Fung, Pascale (eds): *Building and Using Comparable Corpora*. Berlin, Heidelberg: Springer, 151-168.

Kilgarriff, Adam/Baisa, Vít/Bušta, Jan/Jakubíček, Milo/Kovář, Vojtěch/Michelfeit, Jan/Rychlý, Pavel/Suchomel, Vít (2014). The Sketch Engine: ten years on. In: *Lexicography*, 1: 7-36.

Krötzsch-Viannay, Monique (1979). Sexisme et Lexicographie. Les mots femme et homme dans le dictionnaire. In: *Osnabrücker Beiträge zur Sprachtheorie (OBST)* 9, 109-143.

Lautenschläger, Sina (2017). *Geschlechtsspezifische Körper- und Rollenbilder: Eine korpuslinguistische Untersuchung*, Berlin, Boston: De Gruyter.

Menkhoff, Lukas/Wrohlich, Katharina (2024). Einstellungen zu Geschlechterrollen sind in Deutschland im Laufe der Zeit egalitärer geworden. In: *DIW Wochenbericht* 46 / 2024, 717-724. <https://www.diw.de/documents/publikationen/73/diw_01.c.925667.de/24-46-3.pdf> (abgerufen: 30.11. 2025).

Müller-Spitzer, Carolin (2024). Ein *Mann* ist *reich*, *stark* und *bewaffnet* und eine *Frau schön*, *nackt* und *schwanger*. Stereotype Zuschreibungen zu Mann und Frau in Pressetexten und ihr Einfluss auf korpusbasierte Wörterbücher des Deutschen. In: Ewels, Andrea-Eva/Nübling, Damaris (Hgg.): *Geschlechterbewusste Sprache. Argumente, Positionen, Perspektiven*. Hildesheim, Zürich, New York: Olms, 269-294.

Müller-Spitzer, Carolin/Lobin, Henning (2022). Leben, lieben, leiden: Geschlechterstereotype in Wörterbüchern, Einfluss der Korpusgrundlage und Abbild der sprachlichen 'Wirklichkeit'. In: Diewald, Gabriele/Damaris, Nübling (Hgg.): *Genus – Sexus – Gender*. Berlin, Boston: De Gruyter, 35-63.

Pearce, Michael (2008). Investigating the collocational behaviour of *man* and *woman* in the BNC using Sketch Engine. In: *Corpora* 3(1), 1-29.

Schmuck, Mirjam/Vinckel-Roisin, Hélène (2025). Gender bias in digital tools and language awareness – A cross-linguistic perspective. Vortrag beim Symposium *Multicultural and Interdisciplinary Approaches to Language Awareness: Current Research Perspectives and Challenges*. ATILF, Nancy, 26.05.2025.

Special Eurobarometer 545 (2024). *Gender Stereotypes*. Eurobarometer Report, Jan.-Feb. 2024. European Commission. <https://europa.eu/eurobarometer/surveys/detail/2974> (abgerufen 30.11.2025).

Wiegel, Michaela (2016). Super maman allemande et mauvaise mère française : les rôles de femme et de mère chez la voisine. In: Demesmay, Claire/Pütz, Christine (éds): *Images et Stéréotypes. Perceptions franco-allemandes en temps de crise*. Paris: Heinrich Böll Stiftung, 77-87.

MITTWOCH, 11. MÄRZ 2026, 11.45 UHR

Sprachübergreifende Forschung an natürlichen Gesprächsdaten

Beatrice Szczepek Reed (Universität Lancaster (Vereinigtes Königreich))

Der Vortrag beschäftigt sich mit den Herausforderungen sprachübergreifender Forschung an Sprache im natürlichen Gespräch. Am Beispiel von Form- und Funktionsvergleichen wird gezeigt, welche neuen Einblicke durch natürliche Gesprächsdaten gewonnen werden können, aber auch welche Problematiken sich beim sprachübergreifenden Vergleich ergeben. Es wird diskutiert werden, inwieweit die existierende Forschung von interaktionsrelevanten und funktionalen 'Universalien' ausgeht, die es möglicherweise noch zu beweisen gilt. Gleichzeitig wird gezeigt werden, dass ein interaktionsbasierter Sprachvergleich weitgehend akzeptierte typologische Annahmen in Frage stellen kann. Eine erste Fallstudie wird der Vergleich von 'newsmarks' im Arabischen, Deutschen, Englischen und Hebräischen sein. Newsmarks markieren vorangehende Redebeiträge als neue Information; so z.B. deutsch 'echt' und englisch 'really'. Wie andere diskursive Kategorien wurden auch newsmarks ursprünglich in konversationsanalytischen Arbeiten am Englischen definiert (Heritage, 1984), ohne dass dort eine Sprachgebundenheit explizit gemacht wurde. Im Vortrag werden neue Studien zu newsmarks in verschiede-

nen Sprachen und deren Rezeption herangezogen werden (Gubina & Betz 2021; Marmorstein & Szczepek Reed 2023; Szczepek Reed & Marmorstein 2025), um Möglichkeiten und Problematiken eines solchen funktionalen Vergleiches zu demonstrieren. Hierbei wird unter anderem auch die Theorie der 'collateral effects' von Enfield und Sidnell (2017) kritisch diskutiert werden, nach der der Gebrauch verschiedener pragmatischer Formate je nach Sprache unterschiedliche interaktionale 'Nebeneffekte' mit sich bringt. Eine zweite Fallstudie wird sich mit phonetischen Formen der Wortanbindung im Deutschen und Englischen befassen. In beiden Sprachen kann die direkte artikulatorische Anbindung einer mit Vokal beginnenden neuen Redebeitragseinheit an die vorige Einheit die gleiche diskursive Funktion erfüllen; ebenso die klare phonetische Trennung der beiden Einheiten durch einen Glottalverschluss. Im ersten Fall gestalten Sprecher/-innen die zweite, angebundene Einheit als Fortsetzung der Ersten; im zweiten Fall initiiert die durch Glottalverschluss abgegrenzte Einheit eine neue Gesprächshandlung (z.B. Szczepek Reed 2014; Szczepek Reed & Cantarutti 2024; Bergmann & Groß 2025). Diese Tendenzen gelten in beiden Sprachen, obwohl für das Deutsche die Glottalverschlussepenthese für wortinitiale Vokale beschrieben wurde (z.B. Kohler 1994), während das Englische zur direkten Wortanbindung neigt (z.B. Cruttenden 2014). Hier zeigt die Analyse von natürlichen Gesprächsdaten, dass ein Verständnis von Artikulation als interaktionale Ressource für den Sprachvergleich vielversprechend ist.

Literatur:

Bergmann, P., & Groß, A. (2025). The prosodic marking of discourse functions: German *genau* 'exactly' between confirming propositions and resuming actions. *Open Linguistics*, *11*(1), 20250064.

Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). Routledge.

Enfield, N. J., & Sidnell, J. (2017). *The concept of action*. Cambridge University Press.

Gubina, A., & Betz, E. (2021). What do newsmark-type responses invite? The response space after German *echt*. *Research on Language and Social Interaction*, *54*(4), 374-396.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 299-345). Cambridge University Press.

Kohler, K. J. (1994). Glottal stops and glottalization in German. *Phonetica*, *51*, 38-51.

Marmorstein, M., & Szczepek Reed, B. (2023). Newsmarks as an interactional resource for indexing remarkability: A qualitative analysis of Arabic *wallāhi* and English *really*. *Contrastive Pragmatics*, 238-273.

Szczepek Reed, B., & Marmorstein, M. (Eds.). (2025). *Crosslinguistic perspectives on newsmarks* (Virtual special issue). *Journal of Pragmatics*. https://www.sciencedirect.com/special-issue/10X37V8155S.

Szczepek Reed, B., & Cantarutti, M. (2024). Turn continuation in *yeah*/*no* responding turns: Glottalization and vowel linking as contrastive sound patterns. In M. Selting & D. Barth-Weingarten (Eds.), *New perspectives in interactional linguistic research* (pp. 73-102). Benjamins.

Szczepek Reed, B. (2014). Phonetic practices for action formation: Glottalization versus linking of TCU-initial vowels in German. *Journal of Pragmatics*, *62*, 13-29.

METHODENMESSE POSTER/EXPONATE/ SYSTEMDEMONSTRATIONEN

Definition virtueller Vergleichskorpora und Nutzung von Annotationen mit EuReCo

Nils Diewald, Eliza Margaretha Illig, Marc Kupietz, Franck Bodmer, Helge Stallkamp, Jennifer Ecker und Beata Trawiński (alle IDS)

Das Europäische Referenzkorpus (EuReCo) ist eine seit 2012 bestehende offene Initiative, die sich zum Ziel gesetzt hat, langfristig den Bedarf an vergleichbaren Korpora in verschiedenen Sprachen für die vergleichende Sprachforschung zu decken (Kupietz et al. 2017). Die Kernidee von EuReCo liegt darin, keine neuen Korpora zu erstellen, sondern bestehende große National- und Referenzkorpora nach zu nutzen und diese über eine einheitliche und interoperable Infrastruktur so zu verbinden, dass Nutzer/-innen in die Lage versetzt werden, domänen- und fragestellungsspezifisch vergleichbare, virtuelle Subkorpora selbst anhand von Metadaten-Restriktionen zu definieren und zu nutzen.

Um eine interoperable Infrastruktur bieten zu können, wurde ein implementationsbasierter (gegenüber einem spezifikationsbasierten) Ansatz verfolgt (Kupietz et al. 2024), was bedeutet, dass dieselbe Korpus-Analyse-Plattform KorAP an allen Standorten der Korpuseigner (gegebenenfalls ergänzend zu bestehender Software) für den einheitlichen Zugang zu den vergleichbaren Korpora zur Verfügung steht.

Ein Hindernis bei diesem Ansatz ist, dass die zugrundeliegenden Korpora oft keine universellen, sondern sprachspezifische Tag-Sets zur Auszeichnung von Wortarten und morphosyntaktischen Features verwenden, was gerade die explorativen Anteile sprachvergleichender Studien erheblich erschweren und verlangsamen kann. Die scheinbar naheliegende Lösung, Korpora einfach mit sprachübergreifenden Tag-Sets zusätzlich nach zu annotieren, ist dabei oft keine Option, da etwa die vollständige Nachannotation von DeReKo, als deutschem Partner der EuReCo-Initiative, mit udpipe2 (Straka 2018) für Universal Dependency Annotationen rund 12 GPU-Jahre in Anspruch nähme.

Eine weitere Herausforderung besteht bereits bei der Definition virtueller Vergleichskorpora darin, dass Ausgangskorpora oft zwar ähnliche Metadaten wie thematische Domäne und Texttyp oder Genre zur Kategorisierung von Texten verwenden, die Werte-Mengen oder Taxonomien aber häufig sehr unterschiedlich sind, was bei der explorativen Justierung von Vergleichbarkeitskriterien zu Problemen führen kann.

In diesem Beitrag stellen wir unseren Ansatz zur Lösung dieser beiden Herausforderungen vor: einen in KorAP integrierten Dienst, genannt *Koral-Mapper*, der sowohl Annotationsanfragen in komplexen Syntaxsuchen als auch Metadatenkriterien als Teil komplexer Definitionen virtueller Korpora nach bestimmten Regeln von einem Kate-

gorieninventar in ein anderes übersetzen kann. Auf diese Weise können beispielsweise Anfragen nach UPOS-Annotationen beantwortet werden, auch wenn das Korpus nur mit dem Stuttgart-Tübingen Tagset (STTS; Schiller et al. 1999) annotiert wurde.

Im Kern basiert dieser Mechanismus auf der Serialisierung aller KorAP-Anfragen in das KoralQuery-Format (Bingel/Diewald 2015). Vor der Beantwortung durch eine Datenbank können diese Anfragen durch so genanntes *Query Rewriting* verändert und ergänzt werden. Im Falle von Koral-Mapper werden die Suchkonstrukte zunächst ausgewertet und in Suchkonstrukte des Transformationsziels abgebildet. Nach der Beantwortung durch die Datenbank können die gemappten Annotationen dann zur Anreicherung der Ergebnisrepräsentation verwendet werden (*Response Rewriting*; s. Abb. 1).

Die Integration des Dienstes in die KorAP-Benutzeroberfläche ist optional und wird durch ein Plugin-System ermöglicht. Dieses bietet für den geschilderten Anwendungsfall die Möglichkeit zur Auswahl und Aktivierung des gewünschten Mappings (s. Abb. 2), erlaubt aber darüber hinaus auch weitere Optionen zur Anpassung und Erweiterung von Korpusanfragen sowie zur Ergänzung um Funktionalitäten wie die Visualisierung und Extraktion von Korpusdaten und die Einbindung externer Ressourcen.

Auf der Methodenmesse stellen wir den Mapping-Mechanismus und seine Integration in KorAP anhand verschiedensprachiger Instanzen des EuReCo-Verbunds vor.

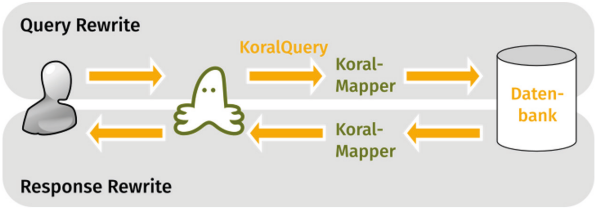

Abb. 1: Ablauf einer Korpusanfrage mit Koral-Mapper

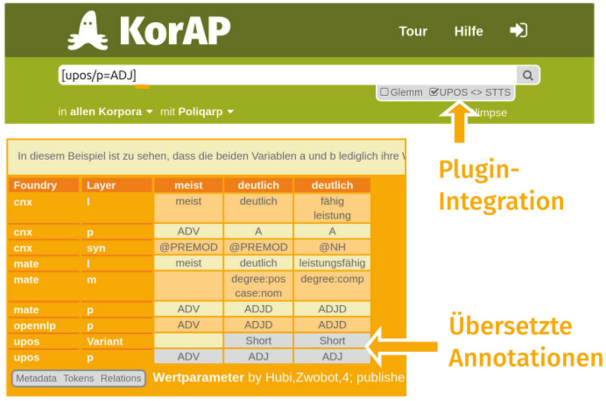

Abb. 2: Integration von Koral-Mapper in die KorAP-Benutzeroberfläche

Referenzen:

Bingel, Joachim/Diewald, Nils (2015): KoralQuery – a General Corpus Query Protocol. In: Proceedings of the Workshop on Innovative Corpus Query and Visualization Tools at NODALIDA 2015. Vilnius, Lithuania, S. 1-5.

Kupietz, Marc/Bański, Piotr/Diewald, Nils/Trawiński, Beata/Witt, Andreas (2024): EuReCo: Not building and yet using federated comparable corpora for cross-linguistic research. In: Zweigenbaum, Pierre/Rapp, Reinhard/Sharoff, Serge (Hgg.): Proceedings of the BUCC 2024: The 17th workshop on building and using comparable corpora. Torino, Italia: ELRA and ICCL, S. 94-103. https://aclanthology.org/2024.bucc-1.10.pdf.

Kupietz, Marc/Witt, Andreas/Bański, Piotr/Tufiş, Dan/Cristea, Dan/Váradi, Tamás (2017): EuReCo – Joining Forces for a European Reference Corpus as a sustainable base for cross-linguistic research. In: Bański, Piotr/Kupietz, Marc/Lüngen, Harald/Rayson, Paul/Biber, Hanno/Breiteneder, Evelyn/Clematide, Simon/Mariani, John/Stevenson, Mark/Sick, Theresa (Hgg.): Proceedings of the Workshop on Challenges in the Management of Large Corpora and Big Data and Natural Language Processing (CMLC-5+BigNLP) 2017 including the papers from the Web-as-Corpus (WAC-XI) guest section. Birmingham, 24 July 2017. Mannheim: Institut für Deutsche Sprache, S. 15-19. https://nbn-resolving.org/urn:nbn:de:bsz:mh39-62580.

Schiller, Anne/Teufel, Simone/Stöckert, Christine/Thielen, Christine (1999): Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset). IMS-CL, University Stuttgart.

Straka, Milan (2018): UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task. In: Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. Brussels, Belgium: Association for Computational Linguistics, S. 197-207. https://doi.org/10.18653/v1/K18-2020.

Universal Dependency Contributors (2024): Tagset de::stts conversion to universal POS tags and features. https://universaldependencies.org/tagset-conversion/de-stts-uposf.html (zuletzt abgerufen: 9. Juli 2025).

LAPIS: ein Portal zu Sprachressourcen in Österreich

Markus Pluschkovits (Austrian Centre for Digital Humanities & Universität Wien (Österreich)), Anja Wittibschlager (Universität Wien (Österreich)), Daniel Schopper (Austrian Centre for Digital Humanities), Jakob Bal (Austrian Centre for Digital Humanities), Kilian Kukelka (Austrian Centre for Digital Humanities) und Alexandra N. Lenz (Austrian Centre for Digital Humanities & Universität Wien (Österreich))

Mittlerweile haben sich in der germanistischen Forschungslandschaft verschiedene Plattformen und Software etabliert, die mehrere Datensätze bzw. Korpora vereinen oder einzelkorpusunabhängige Analysemöglichkeiten bieten. Die Datenbank für Gesprochenes Deutsch (DGD) stellt ein solches Beispiel für ein mehrere Korpora umfassendes Interface dar (vgl. Schmidt 2017), KorAp (Kupietz et al. 2019) und ZuMult (Fandrych et al. 2022) zwei rezente Beispiele für Korpusanalyseinterfaces. Ein verbleibendes

Desiderat ist eine ähnliche Infrastruktur, die verschiedene sprachliche Varietäten und verschiedene Daten für Österreich vereint. Das *Linguae Austriacae: Plattform und Informationssystem zu Sprache in Österreich* (LAPIS) Projekt versucht diesem Desiderat gerecht zu werden.

LAPIS ist ein Projekt am Austrian Centre for Digital Humanities (ACDH) in Kooperation mit dem Spezialforschungsbereich „Deutsch in Österreich (DiÖ). Variation – Kontakt – Perzeption" (FWF F 060). Ziel des Projekts ist eine projektübergreifende Infrastruktur für verschiedene Daten zu Sprache(n) in Österreich bereitzustellen, allen voran der deutschen Sprache (aber auch darüber hinaus). Das übergeordnete Portal LAPIS soll dabei als Such- und Organisationsoberfläche dienen, um verschiedene Ressourcen zu Sprache in Österreich zugänglich zu machen. Gleichzeitig soll ein einheitliches Datenmodell die Interoperabilität der verschiedenen Ressourcen- und Datentypen in LAPIS gewährleisten. LAPIS beinhaltet unterschiedliche Module, die – je nach Datentyp – verschiedene Funktionen bieten. Eines dieser Module wurde im LexAT21-Projekt (*Atlas zur lexikalischen Variation in Österreich im 21. Jahrhundert*, s. Lenz et al. 2025) entwickelt und dient der Darstellung und Kartierung von annotierten Fragebogendaten mit Fokus auf Lexik. Dieses Modul wird momentan im Rahmen des LexVAD20-Projekts (siehe Kunzmann, i.E.) um weitere Kartierungsmöglichkeiten und historische Datensätze erweitert. Ein weiteres LAPIS-Modul ist das sich momentan in Entwicklung befindliche Corpus-Modul, das zur Erschließung von TEI-basierten Korpora genutzt werden kann. Datenbasis für dieses Modul stellt das DiÖ-Konversationskorpus (siehe z.B. Lenz 2023) da. In weiterer Folge soll LAPIS um ein Modul zur Erschließung von Experimentaldaten und Citizen Science Daten erweitert werden.

Möglich wird eine solche Interoperationalisierung von Daten durch ein gemeinsames Datenmodell. Dieses muss flexibel genug sein, um verschiedene Anwendungsfälle zu ermöglichen und zukünftige zu antizipieren. LAPIS stellt zusätzlich zu diesem flexiblen Datenmodell dabei generische Module zur Verfügung, die häufige Anwendungsfälle in der sprachwissenschaftlichen Forschung, wie etwa Datenvisualisierung und -präsentation von linguistischen Daten (wie beispielsweise die Kartierung oder tabellarische Darstellung von Daten im Falle von Fragebogendaten oder die Darstellung von Transkriptionen im Falle des Corpus-Moduls), abdecken.

Fokus des Methodenmessestandes soll einerseits Aufbau, technische Infrastruktur und Datenmodell des LAPIS-Portal sein, andererseits sollen mit LexAT21 und dem sich in Entwicklung befindlichen Corpus-Modul zwei konkrete Module von LAPIS anhand von Anwendungsfällen vorgestellt werden. Die Demonstration von LexAT21 fokussiert dabei lexikalische Variation in verschiedenen Varietäten des Deutschen in Österreich mit Fokus auf inter- und intraindividueller Variation anhand von dynamischer Kartierung der Daten durch das Frontend. Das Corpus-Modul soll anhand der Datenbasis und dem aktuellen Entwicklungsstand sowie durch Einpassung in die LAPIS-Infrastruktur vorgestellt werden. Der Stand umfasst Poster, Hand-Outs und Softwaredemonstrationen.

Referenzen:

Fandrych, Christian; Frick, Elena; Kaiser, Julia; Meißner, Cordula; Portmann, Annette; Schmidt, Thomas; Schwendemann, Matthias; Wallner, Franziska; Wörner, Kai. 2022. „ZuMult: Neue Zu-

gangswege zu Korpora gesprochener Sprache". In: Kämper, Heidrun; Plewnia, Albrecht (Hg.): *Sprache in Politik und Gesellschaft: Perspektive und Zugänge*. Berlin [u.a.]: de Gruyter, 305–312.

Kunzmann, Markus. im Erscheinen. "AI-supported indexing of handwritten dialectal lexis: The pilot study 'DWA Austria' as a case study". *Digital Scholarship in the Humanities*.

Kupietz, Marc; Diewald, Nils; Margaretha, Eliza; Bodmer, Franck; Stallkamp, Helge; Harders, Peter. 2019. „Neues von KorAP". In: Eichinger, Ludwig M.; Plewnia, Albrecht (Hg.): *Neues vom heutigen Deutsch: Empirisch – methodisch – theoretisch*. Berlin [u.a.]: de Gruyter, 345-349.

Lenz, Alexandra N. 2023. Korpora zur deutschen Sprache in Österreich. System- und soziolinguistische Perspektiven. In: Deppermann, Arnulf [et al.] (Hg.): *Korpora in der germanistischen Sprachwissenschaft. Mündlich, schriftlich, multimedial* (IDS-Jahrbuch 2022). Berlin & New York: Walter de Gruyter, 53-70.

Lenz, Alexandra N. (Hg.). 2025ff. *LexAT21 – Atlas zur lexikalischen Variation in Österreich im 21. Jahrhundert*. Konzipiert und entwickelt von Jakob Bal, Kilian Kukelka, Markus Pluschkovits, Daniel Schopper und Anja Wittibschlager. Unter Mitarbeit von Amelie Dorn, Jan Höll, Katharina Korecky-Kröll, Wolfgang Koppensteiner, Claudia Mattes, Markus Pluschkovits, Rita Stiglbauer, Florian David Tavernier, Anja Wittibschlager, Theresa Ziegler, Kerstin Lorenz und Eric Schirl.

Schmidt, Thomas. 2017. „DGD – die Datenbank für Gesprochenes Deutsch: Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim". *Zeitschrift für Germanistische Linguistik* 45(3), 451-463.

DAKODA – Aufbau und Erschließung einer groß angelegten L2-Datenbasis zur Erforschung des Verbstellungserwerbs im Deutschen

Katrin Wisniewski (Universität Leipzig), Lisa Lenort (Universität Leipzig), Annette Portmann (Universität Leipzig), Christine Renker (Universität Leipzig), Josef Ruppenhofer (FernUniversität in Hagen), Matthias Schwendemann (Friedrich-Alexander-Universität Erlangen-Nürnberg) und Torsten Zesch (FernUniversität in Hagen)

Die Erforschung des Erwerbs der deutschen Verbstellung stützt sich bislang überwiegend auf kleinere, heterogene Datensätze, was die Vergleichbarkeit und Generalisierbarkeit empirischer Ergebnisse einschränkt. Zudem sind bestehende Lernerkorpora des Deutschen häufig dezentral organisiert, technisch uneinheitlich erschlossen und nur eingeschränkt interoperabel. Das Projekt DAKODA („Datenkompetenzen in DaF/DaZ: Exploration sprachtechnologischer Ansätze zur Analyse von L2-Erwerbsstufen in Lernerkorpora des Deutschen", 2022–2025, BMFTR-Förderlinie „Datenkompetenzen des wissenschaftlichen Nachwuchses") setzte an dieser Schnittstelle von Infrastruktur und Methodik an (Wisniewski et al., 2023, Schwendemann et al., i.Dr.).

Im Zentrum des Projekts stand der Aufbau einer umfangreichen, korpusübergreifenden L2-Datenbasis, die Produktionen von Deutschlernenden aus rund 40 unterschiedlichen

Lernerkorpora zusammenführt. Durch die Entwicklung eines einheitlichen Metadatenschemas, das an internationale Standardisierungsinitiativen anschließt (vgl. Paquot et al., 2024), wurde eine Grundlage für korpusübergreifende Analysen geschaffen. Parallel dazu wurden Verfahren der automatischen Annotation linguistischer Merkmale – insbesondere der Verbstellung – erprobt und weiterentwickelt, um die Skalierbarkeit und Replizierbarkeit lernerlinguistischer Analysen zu erhöhen.

Der Beitrag auf der Methodenmesse präsentiert den methodischen Workflow von DAKODA, darunter Schritte zur Datenkonvertierung, Vereinheitlichung und rechtlich-technischen Absicherung, die manuelle Annotation von Sprachdaten sowie die Implementierung sprachtechnologischer Tools für die automatische Analyse von Lernersprache. Exemplarisch werden erste Annotations- und Visualisierungsergebnisse vorgestellt (vgl. Ruppenhofer et al., 2024, 2025).

DAKODA leistet damit einen Beitrag zum Aufbau nachhaltiger Forschungsdateninfrastrukturen und zur methodischen Weiterentwicklung der Lernerkorpuslinguistik. Die im Projekt entstandenen Ressourcen (Repositorium, Dashboard und Jupyter Notebooks zur Datenanalyse), die auch während der Methodenmesse präsentiert werden, wurden zudem gemäß den FAIR-Prinzipien (findable, accessible, interoperable, reusable) bereitgestellt und eröffnen neue Möglichkeiten für korpusbasierte Studien zum Deutschen als L2 auch im europäischen und multilingualen Kontext.

Referenzen:

Paquot, M., König, A., Stemle, E. W. & Frey, J.-C. (2024). The Core Metadata Schema for Learner Corpora (LC-meta). *International Journal of Learner Corpus Research*, 10(2), 280-300. https://doi.org/10.1075/ijlcr.24010.paq.

Ruppenhofer, J., Schwendemann, M., Portmann, A., Wisniewski, K., & Zesch, T. (2024). Every verb in its right place? A roadmap for operationalizing developmental stages in the acquisition of L2 German. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (6655-6670). https://aclanthology.org/2024.lrec-main.589/.

Ruppenhofer, J., Portmann, A., Renker, C., Schwendemann, M., Wisniewski, K., & Zesch, T. (2025). Where it's at: Annotating verb placement types in learner language. In *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025)*. https://doi.org/10.18653/v1/2025.law-1.15.

Schwendemann, M., Wisniewski, K., Lenort, L., Portmann, A., Renker, C., Ruppenhofer, J., & Zesch, T. (im Druck). Zur Entstehung und Erschließung der bislang größten Lernerkorpus-Datenbank des Deutschen: Ein Bericht aus dem DAKODA-Projekt. *Zeitschrift für germanistische Linguistik*.

Wisniewski, K., Zesch, T., Schwendemann, M., Ruppenhofer, J., & Portmann, A. (2023). Exploration automatischer Annotationen von Erwerbsstufen in Lernerkorpora: Das Forschungsprojekt DAKODA. *Korpora Deutsch als Fremdsprache*, 3(2), 179–223. https://doi.org/10.48694/kordaf.3845.

ExpoKo – Werkzeug und Ressource für das wissenschaftliche Schreiben

Andrea Lösel (Universität Leipzig/Technische Universität Dortmund), Annalena Messner (Universität Zagreb (Kroatien)), Matthias Schwendemann (Friedrich-Alexander-Universität Erlangen-Nürnberg) und Franziska Wallner (Technische Universität Dortmund)

Exposés als Skizzen für eigene Forschungsvorhaben spielen im studentischen Qualifikationsprozess eine immer größere Rolle – sei es im Kontext von Abschlussarbeiten oder potentiellen Förderantragen. Für viele Studierende, insbesondere für solche mit Deutsch als Fremd- oder Zweitsprache stellt das Verfassen eines Exposés eine große Herausforderung dar (vgl. Stezano Cotelo 2008). Während Handreichungen mit Tipps zum Verfassen von Exposés zahlreich sind, wird die Textsorte in Lehrkontexten nur selten thematisiert. Es mangelt an exemplarischen Materialien anhand derer eine Didaktisierung erfolgen könnte. Dabei weisen Exposés aufgrund ihrer Kürze eine hohe textuelle Dichte auf und enthalten Textmerkmale, die auch in längeren akademischen (studentischen) Textsorten wie Haus- und Abschlussarbeiten zu finden sind. Das macht sie besonders geeignet für eine gezielte didaktische Aufbereitung.

Das Projekt *ExpoKo* zielt darauf ab, ein intuitiv nutzbares Korpus von Exposés aufzubauen, das ohne größere technische Hürden für Forschungszwecke, vor allem aber auch für didaktische Zwecke eingesetzt werden kann – sei es in der schulischen und universitären Wissenschaftssprachvermittlung oder auch zur individuellen Unterstützung beim Verfassen eigener wissenschaftlicher Arbeiten. Der Korpuszugang ist über die ZuMult-Tools geplant (vgl. Fandrych et al. 2023) und wird aktuell auf einem Testserver erprobt und weiterentwickelt. Die Oberfläche soll dabei u.a. das zielgerichtete Ansteuern didaktisch relevanter Phänomene und Strukturelemente ermöglichen.

Das sich noch im Aufbau befindliche Korpus umfasst derzeit 80 Exposés, die überwiegend von Studierenden mit Deutsch als L2 stammen. Die Daten wurden anonymisiert und durchliefen verschiedene automatische korpuslinguistische Aufbereitungsschritte (Tokenisierung, Lemmatisierung, POS-Tagging). Ergänzend werden manuelle Annotationen vorgenommen – darunter die Annotation von Strukturelementen (z. B. Textblöcke zu Forschungsfragen, Methoden) sowie die pragmatische Annotation von Zitaten und Verweisen als zwei zentrale wissenschaftssprachliche Handlungen.

Auf der Methodenmesse soll vorgestellt werden, mit welchen Herausforderungen internationale Studierende bei der sprachlichen Realisierung von Zitaten und Verweisen konfrontiert sind. Zudem wird ein Ausblick auf die geplante Benutzeroberfläche gegeben und gezeigt, welche didaktischen Nutzungsoptionen bereitgestellt werden sollen.

Referenzen:

Fandrych, C./Schmidt, T./Wallner, F./Wörner, K. (Hgg.) (2023): Zugänge zu multimodalen Korpora gesprochener Sprache. Themenheft der Zeitschrift Korpora Deutsch als Fremdsprache (3) 1. Unter: https://kordaf.tujournals.ulb.tu-darmstadt.de/issue/92/info/ [13.08.2025].

Lösel, A./Messner, A./Schwendemann, M./Wallner, F. (eingereicht): ExpoKo – Eine Ressource für das wissenschaftliche Schreiben. Ein Projektbericht. Eingereicht bei: Korpora Deutsch als Fremdsprache.

Stezano Cotelo, K. (2008): Verarbeitung wissenschaftlichen Wissens in Seminararbeiten ausländischer Studierender. Eine empirische Sprachanalyse. München: Iudicium.

GePaDeU – ein Mehrebenen-Korpus politischer Debatten, angereichert mit semantisch-pragmatischen Annotationen

Ines Rehbein (Universität Münster), Julian Schlenker (Universität Mannheim), Lilly Brauner (Universität Heidelberg) und Simone Ponzetto (Universität Mannheim)

Der Fokus unseres Beitrags liegt auf dem Thema „Sprache im politischen Raum" und untersucht die interaktionalen und diskursiven Eigenschaften des Deutschen in argumentativen politischen Texten. Dazu stellen wir eine neue Ressource vor, die politische Sprache in der BRD seit 1949 dokumentiert. Das GePaDeU-Korpus (*Ge*rman *Pa*rliamentary *De*bates, Unified Corpus) beinhaltet Transkripte von Debatten des Deutschen Bundestags, angereichert mit verschiedenen semantischen und pragmatischen Annotationsebenen in einer Mehrebenen-Architektur. Die Annotationen zielen darauf ab, eine Analyse politischer Diskurse zu ermöglichen, die über thematische Keyword-Suchen und grammatikalische Merkmale hinausgeht und pragmatische Aspekte berücksichtigt.

Die Ressource besteht aus einem kleineren, manuell annotierten Korpus mit Debatten aus der 19. Legislaturperiode (2017-2021) und einem großen, automatisch annotierten Korpus, das alle Bundestagsdebatten von September 1949 bis September 2025 umfasst (aktuell im Aufbau). Die manuell annotierten Daten beinhalten 265 Reden, gehalten von 195 Redner*innen aus sechs Parteien, sowie vier Reden von fraktionslosen Parlamentsmitgliedern. Diese Daten wurden mit mehreren Annotationsebenen angereichert:

- Eigennamen (Named Entity Recognition, NER)
- Referenzen auf soziale Gruppen (z.B. PeopleByNation, PeopleByAge, ...)
- Redewiedergabe
- funktionale Sprechakte (z.B. Accusation, Demand, Self-Representation, Promise, ...)
- Situation Entities
- Moralische Frames, narrative Rollen und Moral Foundations.

Die Auszeichnung von Eigennamen wurde auf einer Teilmenge von 40 Reden vorgenommen und folgt dem in Ruppenhofer et al. (2020) beschriebenen Schema, das 31

feinkörnige Klassen umfasst. Die Markierung von Referenzen auf soziale Gruppen und Instanzen der Elite ist relevant für die Untersuchung von Populismus in politischen Texten (Klamm et al., 2023) und schafft die Möglichkeit einer Extraktion von relevanten Akteuren in politischer Kommunikation, die über die Suche nach Eigennamen hinausgeht. Die Annotation von Redewiedergabe ermöglicht es, gesprochene und geschriebene Botschaften sowie Gedanken zu erkennen und sie ihren jeweiligen Quellen zuzuordnen (Rehbein et al., 2024). Unser Schema unterscheidet hier zwischen Botschaft, Topik, Medium, Evidenz, Quelle und Adressat und berücksichtigt auch Mehrwortausdrücke.

Die pragmatische Annotationsebene der funktionalen Sprechakte folgt der Taxonomie von Kondratenko et al. (2020) und erweitert diese um neue Klassen (Reinig et al., 2024). Darüber hinaus beinhaltet das Korpus die Annotation von semantischen Satztypen, sogenannte Situation Entity Types (Smith, 2003). Diese Annotationsebene operiert auf (Teil-)Satzebene und umfasst Eventualitäten (Zustände, Ereignisse, Berichte), abstrakte Entitäten (Fakten, Propositionen) sowie generische und generalisierende Aussagen. Letztere werden häufig mit Stereotypen assoziiert und eignen sich deshalb als Grundlage für die Untersuchung von Bias in politischer Sprache. Zusätzlich wurde in GePaDeU moralisches Framing aus der Perspektive der Redner*innen kodiert, operationalisiert als komplexe Annotationen moralischer Aussagen (Frames), zusammen mit den im Text ausgedrückten narrativen Rollen (*Hero*, *Villain*, *Victim*, *Beneficiary*), inspiriert vom Narrative Policy Framework (Shanahan et al., 2017). Um die im Frame enthaltenen moralischen Werte besser zu erfassen, wurde jeder moralische Frame zusätzlich entsprechend der Moral-Foundations-Theorie (MFT) (Haidt et al., 2009; Graham et al., 2013) mit „moralischen Grundsätzen" (Moral Foundations) klassifiziert.[1]

Wir präsentieren unsere neue Ressource und demonstrieren anhand von kleinen Fallbeispielen, wie das Korpus für die Untersuchung politischer Sprache und die linguistische Analyse von politischen Diskursen genutzt werden kann.

Referenzen:

Christopher Klamm, Ines Rehbein, & Simone Paolo Ponzetto. 2023. Our kind of people? Detecting populist references in political debates. In Findings of the Association for Computational Linguistics: EACL 2023, pages 1227–1243, Dubrovnik, Croatia. Association for Computational Linguistics.

Natalia V. Kondratenko, Anastasiia A. Kiselova, and Liubov V. Zavalska. 2020. Strategies and tactics of communication in parliamentary discourse. Studies about languages, 36:17-29.

Ines Rehbein, Ines Reinig & Simone Paolo Ponzetto. 2025. Moral Framing in Politics (MFiP): A new resource and models for moral framing. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP 2025).

Ines Rehbein, Josef Ruppenhofer, Annelen Brunner, & Simone Paolo Ponzetto. 2024. Out of the Mouths of MPs: Speaker Attribution in Parliamentary Debates. Proceedings of the 2024 Joint

[1] MFT ist eine deskriptive, pluralistische Moraltheorie, die in der Sozialpsychologie verwurzelt ist und die moralische Werte auf Basis von Intuitionen, sogenannten „Moral Foundations" beschreibt, die unser Verhalten beeinflussen.

International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024).

Ines Reinig, Ines Rehbein, & Simone Paolo Ponzetto. 2024. How to Do Politics with Words: Investigating Speech Acts in Parliamentary Debates. Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024).

Josef Ruppenhofer, Ines Rehbein, & Carolina Flinz. 2020. Fine-grained named entity annotations for German biographic interviews. In Proceedings of the Twelfth Language Resources and Evaluation Conference, pages 4605–4614, Marseille, France. European Language Resources Association.

Elizabeth A Shanahan, Michael D Jones, Mark K Mcbeth, & Claudio M Radaelli. 2017. The Narrative Policy Framework. In C.M. Weible and P.A. Sabatier, editors, The Theories of the Policy Process, pages 173--213. Boulder, CO: Westview Press.

Carlota S. Smith. 2003. Modes of Discourse: The Local Structure of Texts. Cambridge Studies in Linguistics. Cambridge University Press.

Das „Parallel European Corpus of Informal Interaction“ in der Datenbank für Gesprochenes Deutsch: Eine Ressource für sprach- und aktivitätsübergreifende Analysen sozialer Interaktion

Siegwalt Lindenfelser und Jörg Zinken (beide IDS)

Das ***Parallel European Corpus of Informal Interaction (PECI)*** ist eine neue mehrsprachige Ressource, die im Rahmen des von der Leibniz-Gemeinschaft geförderten Projekts „Norms, rules and morality across languages“ (2020-2023, Leitung: Jörg Zinken, IDS Mannheim) in länderübergreifender Kooperation erstellt wurde. Ziel war es, eine moderne Datengrundlage für die **vergleichende Forschung zu sozialer Interaktion** zu schaffen (vgl. Kornfeld et al. 2023; Küttner et al. 2024). Die Hauptinnovation gegenüber bestehenden Vergleichskorpora liegt darin, dass PECI Videoaufnahmen natürlicher sozialer Ereignisse enthält, wodurch es möglich wird, sprachliches Handeln vergleichend in seiner primären Umwelt zu untersuchen: informeller, multimodaler Face-to-face-Interaktion.

Das Korpus umfasst Videoaufnahmen von drei sozialen Aktivitäten in mehreren Sprachen, die teils im Projektzeitraum, teils aber auch von internationalen Partnern und Partnerinnen schon in früheren Jahren in paralleler Vorgehensweise und nach denselben Kriterien erhoben wurden:

- **Familienfrühstück am Wochenende** (zu einem kleinen Anteil auch andere Mahlzeiten)
- **Spieleabend** unter Freunden und Verwandten (insbesondere Brettspiele)
- **Autofahrt** unter Freunden (i.d.R. Hin- und Rückfahrt)

Nach Abschluss der Erhebungen und der Transkription eines Großteils der Aufnahmen wurden die Daten an das Archiv für Gesprochenes Deutsch (AGD) am IDS Mannheim übergeben. Das Herzstück des Korpus, das die qualitativ geeignetsten Aufnahmen aus den vier Sprachen **Deutsch, (britisches) Englisch, Italienisch und Polnisch** umfasst, wurde dort weiter aufbereitet und wird mit Release von Version 2.25 der **Datenbank für Gesprochenes Deutsch (DGD)** (vgl. Schmidt 2017) im Frühjahr 2026 für Forschungs- und Lehrzwecke bereitgestellt.

Die über die DGD zugängliche Version des PECI-Korpus umfasst **knapp 80 Stunden** fast vollständig transkribiertes Audiomaterial mit jeweils zwei Kameraperspektiven sowie umfangreiche Metadaten und Zusatzmaterialien (Wort- und Lemmalisten, Aufnahmesettings, Spieleliste). Die in Anlehnung an die cGAT-Konventionen erstellten Transkripte wurden am AGD sprachspezifisch POS-getaggt. Um eine **internationale Nutzung** des Korpus zu ermöglichen, wurden sowohl bei der Datenaufbereitung als auch an der DGD selbst umfangreiche Arbeiten für eine adäquat **mehrsprachige Repräsentation** des Parallelkorpus vorgenommen. Dazu gehört, dass die Metadaten in allen vier Einzelsprachen des Korpus anzeigbar und recherchierbar sind. Die Transkripte sind neben der sprachspezifischen POS-Annotation zusätzlich auch sprachübergreifend mit (gröberen) POS-Tags des Universal-Dependencies-Frameworks (vgl. de Marneffe et al. 2021) durchsuchbar. Die Zusatzmaterialien werden zweisprachig auf Deutsch und Englisch bereitgestellt, Wort- und Lemmalisten jeweils pro Sprache.

Die Bereitstellung des Korpus eröffnet damit in der DGD völlig neue Möglichkeiten der vergleichenden interaktionalen Analyse über mehrere Sprachen hinweg. Dabei ist es u.a. auch möglich, aus Ereignissen in PECI und inhaltlich ähnlich gelagerten Ereignissen im Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK) (vgl. Reineke et al. 2023) korpusübergreifend virtuelle Korpora zu erstellen und die Datenbasis für Fragestellungen unterschiedlichster Art so noch weiter zu vergrößern. Im Rahmen der Methodenmesse stellen wir das Korpus vor und zeigen **mögliche Verwendungen von PECI** auf.

Referenzen:

de Marneffe, Marie-Catherine/Manning, Christopher/Nivre, Joakim/Zeman, Daniel (2021): Universal Dependencies. In: Computational Linguistics 47(2), S. 255-308.

Kornfeld, Laurenz/Küttner, Uwe-A./Zinken, Jörg (2023): Ein Korpus für die vergleichende Interaktionsforschung. Das 'Parallel European Corpus of Informal Interaction' (PECII). In: Deppermann, Arnulf/Fandrych, Christian/Kupietz, Marc/Schmidt, Thomas (Hrsg.): Korpora in der germanistischen Sprachwissenschaft. Mündlich, schriftlich, multimedial. Berlin/Boston: de Gruyter, S. 103-128.

Küttner, Uwe-A./Kornfeld, Laurenz/Mack, Christina/Mondada, Lorenza/Rogowska, Jowita/Rossi, Giovanni/Sorjonen, Marja-Leena/Weidner, Matylda/Zinken, Jörg (2024): Introducing the 'Parallel European Corpus of Informal Interaction' (PECII). A novel resource for exploring cross-situational and cross-linguistic variability in social interaction. In: Selting, Margret/Barth-Weingarten, Dagmar (Hrsg.): New Perspectives in Interactional Linguistic Research. Amsterdam/Philadelphia: John Benjamins (Studies in Language and Social Interaction 36), S. 132-160.

Reineke, Silke/Deppermann, Arnulf/Schmidt, Thomas (2023): Das Forschungs- und Lehrkorpus für Gesprochenes Deutsch (FOLK). Zum Nutzen eines großen annotierten Korpus gesprochener Sprache für interaktionslinguistische Fragestellungen. In: Deppermann, Arnulf/Fandrych, Christian/Kupietz, Marc/Schmidt, Thomas (Hrsg.): Korpora in der germanistischen Sprachwissenschaft. Berlin/Boston: De Gruyter, S. 71-102.

Schmidt, Thomas (2017): DGD – die Datenbank für Gesprochenes Deutsch. Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim. In: Zeitschrift für Germanistische Linguistik 45(3), S. 451-463.

Die ZuMult-Plattform als Instrument für sprachvergleichende Analysen auf mündlichen Daten

Thomas Schmidt (Universität Duisburg-Essen), Lotfi Abouda (Laboratoire Ligérien de Linguistique, Université d'Orléans (Frankreich)), Flora Badin (Laboratoire Ligérien de Linguistique), Hans C. Boas (University of Texas, Austin (USA)), Margaret Blevins (University of Texas, Austin (USA)), Kristin Bührig (Universität Hamburg), Celine Dugua (Laboratoire Ligérien de Linguistique, Université d'Orléans (Frankreich)), Christian Fandrych (Herder-Institut, Universität Leipzig), Anne Ferger (Universität Duisburg-Essen), Elena Frick (IDS), Christian Gregor (Freie Universität Berlin), Peter Kompiel (Freie Universität Berlin), Johanna Miecznikowski-Fuenfschilling (Università della Svizzera italiana, Lugano (Schweiz)), Cord Pagenstecher (Freie Universität Berlin), Karola Pitsch (Universität Duisburg-Essen), Matthias Schwendemann (Friedrich-Alexander-Universität, Erlangen), Marie Skrovec (Laboratoire Ligérien de Linguistique, Université d'Orléans (Frankreich)), Darinka Verdonik (University of Maribor (Slowenien)), Franziska Wallner (Technische Universität Dortmund) und Kai Wörner (Universität Hamburg)

> "[Tools widely used by corpus linguists] all offer a different user-experience, because each tool is created in isolation and thus offers a different user interface, control flow, and functionality." (Anthony 2009)

Nur wenige größere Korpora sind aus sich heraus auf sprachvergleichende Analysen angelegt. Zu den von Vorneherein als „Comparable Corpus" konzipierten Ausnahmen gehören das International Comparable Corpus (ICC, Čermáková et al. 2021) oder das mehrsprachige GeWiss-Korpus (Fandrych et al. 2017). Ein alternativer oder ergänzender Ansatz sind *virtuelle* vergleichbare Korpora wie EuReCo (Trawíński & Kupietz 2021), die über Föderation verteilter, auf vergleichbarer technischer Basis stehender Korpora und Korpusplattformen ermöglichen, das Deutsche kontrastiv mit anderen europäischen Sprachen in Beziehung zu setzen; EuReCo ist allerdings auf schriftsprachliche Daten beschränkt.

Unser Beitrag stellt die ZuMult-Plattform (Fandrych et al. 2023) und deren Potential, Vergleichbares für mündliche Korpora zu leisten, vor. ZuMult ist eine offene, flexible, auf etablierten Standards und Technologien basierende Architektur für den Zugang

zu audio-visuellen Korpora. Neben Korpusrecherchen mit CQP unterstützt ZuMult die für eine Analyse gesprochener Sprache notwendige erweiterte Kontextualisierung von Suchergebnissen (Frick & Schmidt 2025) sowie interaktive Transkriptanalysen (Schmidt et al. 2023), die insbesondere für qualitativ orientierte Analysen, z.B. in der Gesprächsforschung, und für didaktische Anwendungen, z.B. in der DaF/DaZ-Lehre, genutzt werden. An der Universität Duisburg-Essen erfolgt eine Erweiterung, die darauf zielt, Sprache auch in ihrem multimodalen Zusammenspiel mit weiteren körperlichen Ressourcen – wie Blick, Gestik etc. – für korpuslinguistische Herangehensweisen zu erschließen.

Über eine ZuMult-Instanz am Archiv für Gesprochenes Deutsch sind bereits seit 2021 neben dem GeWiss-Korpus zwei der wichtigsten Referenzkorpora des gesprochenen Deutsch – FOLK (Deppermann & Hartung 2011, Schmidt 2016, Reineke et al. 2023) und Deutsch Heute (Kleiner 2015) – zugänglich. Mit der Veröffentlichung einer ZuMult-Instanz für das französische ESLO-Korpus (Abouda & Baude 2006, Baude & Dugua 2011, Eshkol-Taravella 2012, Schmidt 2025) ergeben sich nun erste Möglichkeiten für sprachvergleichende Analysen zwischen dem Deutschen und dem Französischen. Wie unser Beitrag anhand von Proof-Of-Concept-Implementierungen zeigen wird, sind beispielsweise auch das Griffith Corpus of Spoken Australian English (Haugh & Chang 2013), das TIGR Corpus des gesprochenen Italienisch (Miecznikowski-Fuenfschilling et al. i.V.), das Training Corpus of Spoken Slovenian (Verdonik 2024) und die Kollektionen aus Oral History Digital (Pagenstecher 2024) in ZuMult integrierbar. Gleiches gilt für das zwölfsprachige EXMARaLDA-Demokorpus, das derzeit im Rahmen eines Text+-Kooperationsprojekts für eine Publikation in einer ZuMult-Instanz an der Universität Hamburg aufbereitet wird.

Auf ähnliche Weise eröffnet ZuMult neue Möglichkeiten der „vergleichenden Sprachinselforschung" (Boas 2016), also der (korpusgestützten) Analyse von Sprachkontaktphänomenen und Sprachentwicklung von Deutsch als Minderheitensprache im Kontakt mit dominanten anderen (i.d.R. europäischen) Sprachen. Seit Dezember 2024 macht eine ZuMult-Instanz an der University of Texas in Austin (Boas et al. 2025) die Daten des Texas German Dialect Projects auf ähnliche Weise verfügbar, wie die IDS-Instanz Zugriff etwa auf Daten zum Australiendeutsch (Clyne 1981), Deutsch in Namibia (Zimmer et al. 2020) oder Mennonitendeutsch in Amerika (Kaufmann et al. 2023) bietet.

Unser Beitrag wird diese verschiedenen Anwendungen vorstellen und illustrieren, sowie einige methodische Herausforderungen der Mehrsprachigkeit (z.B. sprachübergreifendes POS-Tagging) und technische Ansätze zur Aggregation von Suchanfragen an mehrere ZuMult-Instanzen (CLARIN Federated Content Search) thematisieren.

Referenzen:

Abouda, L. & Baude, O. (2006). Constituer et exploiter un grand corpus oral : Choix et enjeux théoriques. Le cas des ESLO. *Corpus en Lettres et Sciences sociales, Des documents numériques à l'interprétation*, 2006, Albi, France.

Baude, O. et Dugua, C. (2011). (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste?, *Corpus*, 10/2011, 99-118.

Anthony, L. (2009). Issues in the design and development of software tools for corpus studies: The case for collaboration. In: Baker, P. (Hg.), *Contemporary corpus linguistics*. London: Continuum Press, 87-104.

Boas, H. C. (2016). Variation im Texasdeutschen: Implikationen für eine vergleichende Sprachinselforschung ('Implications for comparable speech island reseach'). In Alexandra Lenz (ed.), *German Abroad. Perspektiven der Variationslinguistik, Sprachkontakt- und Mehrsprachigkeitsforschung*, 11-44. Vienna University Press.

Boas, H. C., Blevins, M. & Schmidt, T. (2025). A new corpus platform for the Texas German Dialect Project. Erscheint in: *Language Resources and Evaluation*. Springer.

Čermáková, A., Jantunen, J., Jauhiainen, T., Kirk, J., Kren, M., Kupietz, M. & Uí Dhonnchadha, E. (2021). International Comparable Corpus: Challenges in building multilingual spoken and written comparable corpora. *Research in Corpus Linguistics* 9 (1), 89-103.

Clyne, M. (1981): *Deutsch als Muttersprache in Australien. Zur Ökologie einer Einwanderersprache. In Zusammenarbeit mit dem Centre for Migrant Studies*. Monash University. Deutsche Sprache in Europa und Übersee. Band 8. Wiesbaden: Franz Steiner Verlag.

Deppermann, A./Hartung, M. (2011). Was gehört in ein nationales Gesprächskorpus? Kriterien, Probleme und Prioritäten der Stratifikation des „Forschungs- und Lehrkorpus Gesprochenes Deutsch" (FOLK) am Institut für Deutsche Sprache. In: Felder, E., Müller, M., Vogel, F. (Hg..): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen*. Berlin/Boston: de Gruyter, 2011. 414-450.

Eshkol-Taravella, I., Baude, O., Maurel, D., Hriba, L., Dugua, C., Tellier, I. (2012). Un grand corpus oral 'disponible' : le corpus d'Orléans 1968-2012. In: *Ressources linguistiques libres, TAL*. 52,3/2011, 17-46.

Fandrych, C., Meißner, C., Wallner, F. (Hrsg.) (2017). *Gesprochene Wissenschaftssprache – digital: Verfahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg.

Fandrych, C., Schmidt, T., Wallner, F., Wörner, K. (Hrsg.) (2023). *Themenschwerpunkt: Zugänge zu mündlichen Korpora für DaF und DaZ: Das ZuMult-Projekt*. Darmstadt: Korpora Deutsch als Fremdsprache 3(1), 2023.

Frick, E. & Schmidt, T. (2025). Querying spoken language data. In: Bański, P./Heid, U./Herzberg, L. (eds.): *Standards for language data and infrastructures. Series: Digital Linguistics*. Boston: de Gruyter.

Haugh, M. & Chang, W. (2013). Collaborative creation of spoken language corpora. In Greer,T. Kite, Y. & Tatsuki, D. (eds.), *Pragmatics and Language Learning. Volume 13* (pp. 133-159), National Foreign Language Resource Center, University of Hawai'i, Honolulu.

Kaufmann, G., Gorisch, J., Schmidt, T. (2023): Das MEND-Korpus im Archiv für Gesprochenes Deutsch: Entstehung, Möglichkeiten, Grenzen. In: Wolf-Farré, Patrick/Löff Machado, Lucas/Prediger, Angélica/Kürschner, Sebastian (Hrsg.): *Deutsche und weitere germanische Sprachminderheiten in Lateinamerika: Methoden, Grundlagen, Fallstudien.* (= MinGLA – Minderheiten Germanischer Sprachen in Lateinamerika 1). Berlin: Lang, 2023. S. 103-147.

Kleiner, S. (2015): „Deutsch heute" und der Atlas zur Aussprache des deutschen Gebrauchsstandards. In: Kehrein, R., Lameli, A., Rabanus, S. (Hrsg): *Regionale Variation des Deutschen. Projekte und Perspektiven*. Berlin u. a., 489-518.

Miecznikowski-Fuenfschilling, J., Battaglia, E., Schmidt, T., Zehr, J. (i.V.): *The TIGR Corpus of Spoken Italian.*

Pagenstecher, C. (2024). Oral-History.Digital: Eine Erschließungs- und Rechercheplattform für audiovisuelle narrative Forschungsdaten. In: *O-Bib. Das Offene Bibliotheksjournal* 11 (1), 2024, 1-8, https://doi.org/10.5282/o-bib/6007.

Reineke, S., Deppermann, A., Schmidt, T. (2023). Das Forschungs- und Lehrkorpus für Gesprochenes Deutsch (FOLK). Zum Nutzen eines großen annotierten Korpus gesprochener Sprache für interaktionslinguistische Fragestellungen. In: Deppermann, A., Fandrych, C.,Kupietz, M. & Schmidt, T. (Hrsg.): *Korpora in der germanistischen Sprachwissenschaft. Mündlich, schriftlich, multimedial. Jahrbuch des Instituts für Deutsche Sprache 2022.* Berlin/Boston: de Gruyter, 2023. S. 71-102.

Schmidt, T. (2016). Good practices in the compilation of FOLK, the Research and Teaching Corpus of Spoken German. In: Kirk, John M./Andersen, Gisle (Hrsg.): *Compilation, transcription, markup and annotation of spoken corpora.* (= International Journal of Corpus Linguistics 21, Issue 3). Amsterdam/Philadelphia: Benjamins, 2016. 396-418. https://doi.org/10.1075/ijcl.21.3.05sch.

Schmidt, T., Schwendemann, M., Wallner, F. (2023): ZuViel: Transkriptvisualisierung und Arbeiten mit Transkripten. In: Fandrych, C./Schmidt, T./Wallner, F./Wörner, K. (Hrsg.): *Korpora Deutsch als Fremdsprache 3(1). Themenschwerpunkt: Zugänge zu mündlichen Korpora für DaF und DaZ: Das ZuMult-Projekt.* Darmstadt: KorDaF, 2023. S. 72-91.

Schmidt, T. (2025): Représenter et accéder à la parole dans les corpus oraux : diversification et adaptation des méthodes et technologies. In: Kanaan-Caillol, L./Dugua, C./Abouda, L./Gerstenberg, A. (Hrsg.): *Représenter la parole.* Berlin/Boston: de Gruyter.

Trawiński, B. & Kupietz, M. (2021). Von monolingualen Korpora über Parallel- und Vergleichskorpora zum Europäischen Referenzkorpus EuReCo. In: Lobin, H., Witt, A., Wöllstein, A. (Hrsg.): *Deutsch in Europa. Sprachpolitisch, grammatisch, methodisch. Jahrbuch des Instituts für Deutsche Sprache 2020.* Berlin/Boston: de Gruyter, 2021. S. 209-234.

Verdonik, D.; et al. (2024). *Training corpus of spoken Slovenian ROG 1.0.* Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1992.

Zimmer, Christian/Wiese, Heike/Simon, Horst J./Zappen-Thomson, Marianne/Bracke, Yannic/Stuhl, Britta/Schmidt, Thomas (2020): Das Korpus Deutsch in Namibia (DNam): Eine Ressource für die Kontakt-, Variations- und Soziolinguistik. In: *Deutsche Sprache* 3/20. Berlin: Schmidt, 2020. 210-232.

Lang*Reg: Erhebung und Verarbeitung sprachvergleichender Daten für unterschiedliche kommunikative Situationen mit Fokus auf das Deutsche

Nico Lehmann, Elisabeth Verhoeven und Henri Schellberg (alle Humboldt-Universität zu Berlin)

Eine der größten Herausforderungen bei der sprachübergreifenden Erhebung natürlicher Sprachdaten ist die gleichzeitige Kontrolle mehrerer Variationsdimensionen. Wir stellen ein Datenerfassungsprotokoll vor, das zur Erstellung des frei zugänglichen Lang*Reg-Korpus (Adli et al. 2023) verwendet wurde. Dieses Korpus umfasst intra-individuelle Variation in verschiedenen kommunikativen Situationen mit deutschen Muttersprachlern aus dem Berliner Sprachraum. Parallel dazu wurden Daten für weitere Sprachen in gleicher Weise erhoben: Persisch, Südkurdisch (indoeuropäisch), Yukatek Maya (Maya) und Javanisch (Austronesisch). Die Sprachproben wurden ausgewählt, um die Rolle sprachübergreifender und kulturübergreifender Aspekte bei der Registervariation zu untersuchen. Die Teilnehmer durchliefen eine Reihe von Aufnahmesituationen, in denen sie Aufgaben zur Sprachproduktion erfüllten: a) freies Gespräch mit verschiedenen Gesprächspartnern (Freund, Fremder, Taxifahrer oder Professor) und b) mündliches und grafisches Erzählen einer Geschichte (Lehmann et al. 2025). Die ausgewählten Situationen zielen auf interkulturelle Validität ab, d.h. sie kommen in einer Vielzahl von Kulturen und Sprachen natürlich vor. Sie variieren nach den Parametern (i) soziale Hierarchie, (ii) soziale Distanz sowie (iii) Modus und (iv) Interaktivität (siehe z.B. Halliday & Hasan 1989; Neumann 2014).

Eine weitere Herausforderung besteht darin, mehrsprachige Daten so zu verarbeiten, dass eine maximale Vergleichbarkeit gewährleistet werden kann. Wir werden die Herausforderungen und Lösungen diskutieren, die sich bei der Datenverarbeitung und Annotation des Lang*Reg-Korpus ergeben haben. In einem ersten Schritt wurden alle Daten in Zusammenarbeit mit Muttersprachlern einer genauen Transkription in lateinischer Schrift unterzogen, wobei eine grundlegende syntaktische Segmentierung nach Hauptsatzgrenzen (inkl. aller Nebensätze) in ELAN (The Language Archive 2022) mit separaten Ebenen für jeden Sprecher verwendet wurde. Eine syntaktische Segmentierung unterstützt spätere syntaktische Dependenzannotationen und Analysen, erfordert jedoch ein gutes Verständnis der Sprachsyntax. Zur Vergleichbarkeit ist darüber hinaus eine rigorose Dokumentation der getroffenen Entscheidungen über Sprachen und Mitarbeiter hinweg notwendig, beispielsweise in Bezug auf den Umgang mit Wiederholungen, Abbrüchen, Ellipsen und Apo-Koinou-Konstruktionen. Zusätzlich zur engen Transkription wurde eine Normalisierungsebene hinzugefügt, um Such- und automatische Annotationsprozesse zu erleichtern, indem die Rechtschreibung an die Standardschreibweise oder konventionelle Praxis angepasst und so beispielweise Kontraktionen reduziert wurden (z.B. Deutsch *kannste* > *kannst du*).

Auf der Grundlage der normalisierten Ebene wurden für die deutschen Daten halbautomatisierte Prozesse angewendet, um Lemmata, POS-tags sowie Dependenzannotationen (unter Verwendung von UDPipe 2016) zu erhalten. So konnten wir die syn-

taktische Komplexität in Abhängigkeit von verschiedenen sprachlichen Registern untersuchen. Darüber hinaus werden mit INCEpTION (Klie et al. 2018) spezifischere Annotationen für konkrete Forschungsfragen hinzugefügt, unter anderem für die Form direkter Objekte (overtes vs. kovertes Pronomen), wobei weitere Herausforderungen bestehen, z.B. hinsichtlich der Annotation von koverten Einheiten und der referenziellen Distanz bei mehreren Sprechern. All dies umfasst auch eine Reihe von rein technischen Prozessen, v.a. die Konvertierung der Dateien zwischen den verfügbaren Formaten der verwendeten Tools, für die zusätzlich zu den jeweils integrierten Ex- und Importfunktionen auch das Konvertierungstool Annatto (Krause & Klotz 2024) zum Einsatz kam. Wir werden unsere Arbeitsabläufe vorstellen und über einige der Probleme berichten, auf die wir in diesem Zusammenhang gestoßen sind.

Referenzen:

Adli, Aria & Verhoeven, Elisabeth & Lehmann, Nico & Mortezapour, Vahid & Vander Klok, Jozina (eds.). 2023. Lang*Reg: A multi-lingual corpus of intra-speaker variation across situations. Version 0.1.0. [Data set]. Zenodo. Berlin, Köln: Humboldt-Universität zu Berlin, Universität zu Köln. https://doi.org/10.5281/zenodo.7646320.

Halliday, Michael A. K. & Hasan, Ruqaiya. 1989. Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective. Oxford: Oxford University Press.

Klie, Jan-Christoph & Bugert, Michael & Boullosa, Beto & de Castilho, Richard Eckart & Gurevych, Iryna. 2018. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations. 5-9. Association for Computational Linguistics.

Krause, Thomas & Klotz, Martin. 2024. Annatto. https://github.com/korpling/annatto/.

Lehmann, Nico & Mortezapour, Vahid & Klok, Jozina Vander & Farokhnejad, Zahra & Müller, David & Verhoeven, Elisabeth & Adli, Aria. 2025. Lang*Reg corpus: Documenting intra-speaker variation across languages and registers. Language Documentation & Conservation 19. 40-66. https://hdl.handle.net/10125/74810.

Neumann, Stella. 2014. Contrastive Register Variation: A Quantitative Approach to the Comparison of English and German. Berlin: De Gruyter Mouton.

Straka, Milan & Hajič, Jan & Straková, Jana. 2016. UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In Calzolari, Nicoletta & Choukri, Khalid & Declerck, Thierry & Goggi, Sara & Grobelnik, Marko & Maegaard, Bente & Mariani, Joseph & Mazo, Helene & Moreno, Asuncion & Odijk, Jan & Piperidis, Stelios (eds.), Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). Paris, France: European Language Resources Association (ELRA).

The Language Archive. 2022. Elan. Nijmegen: Max Planck Institute for Psycholinguistics. https://archive.mpi.nl/tla/elan.

Verortung des Deutschen und seiner Ortsdialekte im EU-Sprachraum sowie weitere Anwendungsmöglichkeiten phonetischer Vektoren

Patrick Brookshire (Akademie der Wissenschaften und der Literatur, Mainz)

Maschinelles Lernen hat in den vergangenen Jahrzehnten für die im digitalen Raum am weitesten verbreiteten EU-Amtssprachen neue Analyse- und Vergleichsmöglichkeiten hervorgebracht. Diese basieren auf einer statistischen Auswertung von Häufigkeit und Abfolge geschriebener Wortformen kodiert als Zahlenvektoren. Sprachmodelle bilden dabei aus dieser Representationsform einen „semantischen Raum" zur Gruppierung „ähnlicher" Konzepte. Obwohl derartige Verfahren bereits beeindruckende Ergebnissen liefern, haben sie den Nachteil, dass sie überwiegend Schriftformen auswerten. So führen etwa bereits Abweichungen von der Standardorthographie zu einer anderen Position im „semantischen Raum", weshalb Sprachmodelle meist nicht von Ähnlichkeiten zwischen verwandten Nachbarsprachen profitieren. Solche Probleme könnten durch Einbindung phonetischer Vektoren, die eine ähnliche Aussprache mitabbilden, reduziert werden. Darüber hinaus dürften Sprachmodelle so noch anschlussfähiger zu Anwendungen der Automatischen Spracherkennung und -synthese werden. Eine Schnittstelle zwischen diesen unterschiedlichen Verfahren ist dabei das Internationale Phonetische Alphabet (IPA) und ein Test zur Umsetzbarkeit wird in diesem Beitrag anhand der Verortung deutscher Varietäten in der Sprachenlandschaft der EU vorgestellt.

In der Pilotstudie wurde die phonetische Ähnlichkeit zwischen 36 deutschen Ortsdialekten sowie den 20 indo-europäischen und drei uralischen EU-Amtssprachen (d.h. alle außer Maltesisch) in Bezug auf Übersetzungsäquivalente aus dem Kernwortschatz hin untersucht. Als Datengundlage wurden die Wortlisten mit IPA-Transkriptionen aus den Projekten PhonD2 (Lameli/Werth 2021) und NorthEuraLex (Dellert et al. 2020) zusammengeführt. Anschließend wurden alle diese IPA-Transkriptionen in einen „phonetischen Raum" vektorisiert, der Merkmale wie Artikulationsort oder Vokalquantität abdeckt. Die Ergebnisse zeigen, dass bereits wenige Wortformen ausreichen, um sowohl die bekannten verwandtschaftlichen Beziehungen zwischen den deutschen Ortsdialekten als auch zwischen den 23 EU-Amtssprachen abzubilden. Dabei unterstreicht eine durchschnittliche Kosinusähnlichkeit von 0,81 bei den Ortsdialekten gegenüber 0,51 bei den EU-Amtssprachen, dass sich alignierte Wortformen bei erstgenannter Gruppe erwartungsgemäß „ähnlicher klingen". Zudem sind die Effektstärken in Bezug auf verwandschaftliche Beziehungen gemessen mit Cohen's D mit 2,96 bzw. 3,10 in beiden Fällen nicht nur sehr hoch, sondern weit über dem üblichen Schwellenwert von 0,80 für einen „hohen Effekt". Interessanterweise ändern sich die Ähnlichkeitsverhältnisse zwischen den EU-Amtssprachen nur minimal, wenn statt der kuratierten Wortlisten, die deutsche automatisch mit dem Google-Übersetzungsdienst in die anderen 22 übersetzt und mit Phonemizer (Bernard/Titeux 2021) transkribiert wird, sodass sich das phonetische Vektorisierungsverfahren auch als robust gegenüber einer solchen (weniger hochwertigen) Hochskalierung herausgestellt hat. Um die deutschen Varietäten auch visuell in der europäischen Sprachenlandschaft zu verorten, lassen sich

geographische Metadaten der Datengrundlage in das System integrieren, um aus den erhobenen Ähnlichkeitswerten neben Cluster- auch Kartendarstellungen automatisch zu erzeugen. Darüber hinaus passen die Ähnlichkeitswerte zu der Beobachtung von Gooskens et. al. (2011), dass Muttersprachler der niederländischen Sprache die deutsche Standardsprache besser verstehen als geographisch näherliegende niederdeutsche Dialekte.

Insgesamt deuten diese Beispielerkenntnisse an, dass sich mit phonetischen Vektoren auch ohne Einbindung in Sprachmodelle bereits vielfältige Fragestellungen beantworten lassen. Weitere Anwendungsszenarien sind etwa Analysen der phonetischen Ähnlichkeit zwischen verschiedenen Wortfeldern oder Einflüsse unterschiedlicher Übersetzungssysteme. Diese könnten anschließend mit Ähnlichkeitswerten bezogen auf weitere linguistische Untersuchungsebenen wie Grammatik oder Semantik verglichen werden.

Referenzen:

Bernard, Mathieu/Titeux, Hadrien (2021): Phonemizer: Text to Phones Transcription for Multiple Languages in Python. In: Journal of Open Source Software 6 (68), S. 3958.

Dellert, Johannes/Daneyko, Thora/Münch, Alla/Ladygina, Alina/Buch, Armin/Clarius, Natalie et al. (2020): NorthEuraLex: A Wide-Coverage Lexical Database of Northern Eurasia. In: Language Resources and Evaluation 54 (1), S. 273-301.

Gooskens, Charlotte/Kürschner, Sebastian/van Bezooijen, Renée (2011): Intelligibility of Standard German and Low German to Speakers of Dutch. In: Dialectologia Special issue II, S. 35-63.

Lameli, Alfred/Werth, Alexander (2021): PhonD2. Marburg: Research Center Deutscher Sprachatlas. http://www.regionalsprache.de/phonD2 (Stand: 01.10.2025).

CorpSum – Visualisierungstool für Korpusabfragen

Wolfgang Koppensteiner (Austrian Centre for Digital Humanities/Österreichische Akademie der Wissenschaften), Claudia Mattes (Universität Wien (Österreich) & Austrian Centre for Digital Humanities/Österreichische Akademie der Wissenschaften) und Christoph Hoffmann (Austrian Centre for Digital Humanities/Österreichische Akademie der Wissenschaften)

Im Fokus des Beitrags steht die Präsentation von CorpSum, einer Webapplikation, die benutzerfreundliche, dynamisch generierte Abfragen in schriftsprachlich basierten Textkorpora entlang verschiedener extralinguistischer Variationsdimensionen (z.B. <Zeit> und <Raum>) anbietet. CorpSum wurde mit der Idee entwickelt, direkt auf Daten aus der Korpussuchmaschine NoSketch-Engine zugreifen zu können und diese ohne weitere Umwege (über Datendownloads o.Ä.) auf vielfältige Art zu visualisieren. Ein wesentliches Augenmerk wurde dabei auch auf einen vergleichsweise niederschwelligen und schnellen Zugang gelegt, der Korpusanalysen auch ohne tiefergehendes linguistisches Detailwissen ermöglicht. Die Analyseoptionen skalieren dabei mit den user*innen-seiti-

gen Anforderungen, so können auf Wunsch beispielsweise auch komplexe Suchabfragen mittels Corpus Query Language (CQL) durchgeführt werden. Die Benutzerfreundlichkeit von CorpSum wird überdies dadurch erhöht, dass die notwendige Prozessierung der Daten, beispielsweise Normalisierung, die bei der Arbeit mit großen Korpora und der Visualisierung nach unterschiedlichen Metadaten notwendig ist, automatisiert im Hintergrund stattfindet. Durch Caching wiederholter Suchabfragen werden überdies schnellere Ladezeiten möglich, was die Anwender*innenfreundlichkeit nochmals erhöhen soll. Ungeachtet dessen bietet CorpSum selbstverständlich weiterhin die Flexibilität der NoSketch-Engine-Eingabemaske.

Ursprünglich wurde CorpSum zur Visualisierung von Daten des Austrian Media Corpus (amc) entwickelt, eines der größten deutschsprachigen (Presse-)Textkorpora, welches am Austrian Centre for Digital Humanities (ACDH) an der Österreichischen Akademie der Wissenschaften (ÖAW) gehostet wird (vgl. Hoffmann/Koppensteiner/Mattes 2025). Das amc bezieht seine Daten aus einer Kooperation des ACDH mit der Austrian Press Agency (APA), bildet die österreichische (Print-)Medienlandschaft seit 1986 ab und macht diese für sprachwissenschaftliche Forschung zugänglich (vgl. Dorn et al. 2023). Mit rund 12,9 Mrd. Token und 51 Mio. Artikeln in seiner aktuellen Version stellt das Korpus eine der größten Ressourcen zur Standardsprache in Österreich dar (vgl. amc_b).

Die Funktionalität von CorpSum ist auf mehreren Ebenen angesiedelt. Das integrierte Such-Interface ermöglicht Abfragen unterschiedlicher Granularität, von einfacher wortbasierter Suchedass eben nicht nur eine einzelne Abfrage, sondern mehrere gleichzeitig möglich sind, was die Abbildung von direkten Vergleichen ermöglicht. Umfangreiche Filter-Optionen unterstützen die Anwender*innen überdies, um möglichst zielgenaue Resultate zu erhalten. Das inkludiert, etwa im Fall des amc-Korpus, eine Fülle an außersprachlichen Kategorien, wie Bezug auf die Regionalität des Mediums, unterschiedliche Medientypen oder Zeitintervalle. Die verschiedenen Visualisierungen, die im Frontend angeboten werden, fußen dementsprechend auf unterschiedlichsten Analysedimensionen, speziell in den Dimensionen <Zeit> und <Raum>, die bei der Betrachtung von variationslinguistischen Phänomenen von großer Relevanz sind. Zur weiteren Verwendung können sämtliche Visualisierungen und Suchergebnisse in typischen Standardformaten heruntergeladen werden.

Aufgrund seines technischen Aufbaus kann das Visualisierungstool mittlerweile auch für andere NoSketch-Engine-Instanzen verwendet und angepasst werden, was CorpSum auch für weitere Korpora interessant macht. Die Grundvoraussetzung hierfür ist eine öffentliche Rest API. CorpSum kommt bereits in der universitären Lehre zum Einsatz, um Studierenden korpuslinguistische Analysen zu vermitteln. Weitere Anwendungsbereiche sind u.a. die Untersuchung beispielsweise von lexikalischen Varianten (siehe Lenz et al. (eingereicht)).

Bei der Methodenmesse werden wir sowohl die technischen Gegebenheiten als auch die konkrete Anwendung des Tools mit dem amc präsentieren. Bei einer Softwaredemonstration können Konferenzteilnehmer*innen überdies selbst ihre persönlichen Abfragen anhand des Korpus testen.

Referenzen:

Austrian Center for Digital Humanities. *CorpSum – Visualisierungstool für Korpusabfragen.* URL: https://www.oeaw.ac.at/de/acdh/forschung/dh-forschung-infrastruktur/ressourcen/applikationen/corpsum-visualisierungstool-fuer-korpusabfragen [zuletzt zugegriffen: 15.9.2025].

amc_a = Austrian Media Corpus (amc), Version 4.24qLTS, zugänglich über http://hdl.handle.net/21.11115/0000-000C-4F08-4, abgerufen am 15.9.2025.

amc_b = Austrian Media Corpus. Das amc! URL: https://amc.acdh.oeaw.ac.at/about-amc/ [zuletzt zugegriffen: 15.9.2025].

Dorn, Amelie, Jan Höll, Theresa Ziegler, Wolfgang Koppensteiner & Hannes Pirker. 2023. Die österreichische Presselandschaft digital: Das Austrian Media Corpus (amc) und sein Potenzial für die Linguistik. In Marc Kupietz & Thomas Schmidt (eds.), *Neue Entwicklungen in der Korpuslandschaft der Germanistik: Beiträge zur IDS-Methodenmesse 2022, Korpuslinguistik und Interdisziplinäre Perspektiven auf Sprache*, 43–55. Tübingen: Narr Francke.

Hoffmann, Christoph, Wolfgang Koppensteiner & Claudia Mattes. 2025. CorpSum – a corpus visualization tool. In Cristina Grisot & Thalassia Kontino (eds.), *CLARIN Annual Conference Proceedings 2025, Vienna, Austria*, 105-107. ISSN 2773-2177 (online).

Lenz, Alexandra N., Wolfgang Koppensteiner, Claudia Mattes & Katharina Korecky-Kröll. (in Print). Variationssensitive Ansätze zur korpuslinguistischen Erforschung lexikalischer Dynamik im 21. Jahrhundert. In Annelen Brunner, Gabriele Diewald, Sandra Hansen, Kristin Kopf & Angelika Wöllstein (eds.), *Deutsch im Wandel. Jahrbuch des Instituts für Deutsche Sprache 2025.* Berlin/Boston: de Gruyter.

Ransmayr, Jutta, Karlheinz Mörth & Matej Ďurčo. 2017. AMC (Austrian Media Corpus) – Korpusbasierte Forschungen zum österreichischen Deutsch. In Claudia Resch & Wolfgang U. Dressler (eds.), *Digitale Methoden der Korpusforschung in Österreich* (Sitzungsberichte / Österreichische Akademie der Wissenschaften, Philosophisch-Historische Klasse 879. Band), 27-38. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Von der Korpusanalyse zur Lern-App: Das Projekt COPLUS und die Förderung der deutschen Wissenschaftssprache im europäischen Kontext

Carolina Flinz (Universität Pavia (Italien)), Marcella Costa (Universität Turin (Italien)) und Sabine E. Koesters Gensini (Sapienza Università di Roma (Italien))

Das vom italienischen Ministerium für Forschung und Universität geförderte Projekt COPLUS (https://sites.google.com/uniroma1.it/coplus/home) widmet sich der Förderung der deutschen allgemeinen/alltäglichen Wissenschaftssprache (AWS) (vgl. u.a. Ehlich 1999; Meißner 2014; Fügert/Richter 2016). Ausgangspunkt ist die Beobachtung, dass Studierende an italienischen Universitäten oftmals über unzureichende Kenntnisse dieser für den akademischen Austausch zentralen Varietät verfügen und daher in ihrer Mobilität und fachlichen Anschlussfähigkeit eingeschränkt sind.

Ziel des Projekts ist es, das Erlernen der deutschen AWS gezielt zu unterstützen, indem die bereits vorhandenen (meta-)sprachlichen Kompetenzen der Lernenden im Italienischen (L1) und Englischen (L2) systematisch einbezogen werden. Methodisch kombiniert COPLUS empirische Korpuslinguistik, kontrastive Sprachanalyse und digitale Didaktik, um Ressourcen, Methoden und Werkzeuge zu entwickeln, die das Deutsche im europäischen Mehrsprachigkeitskontext verorten (vgl. Arnone/Flinz 2025, Koesters Gensini/Costa/Flinz i. Vorb.).

Das Projekt ist in drei aufeinanderfolgende Phasen gegliedert: (a) den Aufbau eines vergleichbaren mehrsprachigen Korpus akademischer Fachtexte (Dissertationen aus unterschiedlichen Disziplinen in Deutsch, Italienisch und Englisch), (b) die Erstellung eines Lernerkorpus deutscher Texte italienischer Studierender und die Analyse typischer sprachlicher Schwierigkeiten, sowie (c) die Entwicklung didaktischer Materialien, die in Form einer interaktiven App umgesetzt werden. Diese App, die im Zentrum der Präsentation steht, überführt die Ergebnisse der kontrastiven Korpusanalysen in ein digitales Lernwerkzeug. Sie wurde im Co-Design-Verfahren von IT-Expert:innen, Linguist:innen und Sprachlehrenden entwickelt und konzipiert den Lernprozess metaphorisch als Weg von der L1- zur L3-Kompetenz. Sprachliche „Hindernisse" (lexikalische, morphosyntaktische und diskursive Besonderheiten) werden durch „Brücken" zu bekannten Strukturen in Italienisch oder Englisch überwindbar gemacht.

Die App integriert drei methodische Komponenten: Beschreibung zentraler Merkmale der deutschen AWS auf unterschiedlichen sprachlichen Ebenen, kontrastive Analyse und Visualisierung ausgewählter Phänomene anhand authentischer Korpusbelege aus den drei Sprachen, interaktive Übungsformate, die metasprachliche Reflexion und Mehrsprachigkeitsbewusstsein fördern. Die im Projekt entwickelten mehrsprachigen Korpora bilden zugleich eine wertvolle Ressource für die kontrastive Erforschung des Deutschen und eröffnen neue Perspektiven für quantitative, datengestützte Analysen in mehrsprachigen Kontexten. Durchdie Integration semantischer Netzwerke, multimedialer Materialien und visueller Navigationsstrukturen wird das Deutsche als Wissenschaftssprache im europäischen Sprachraum erfahrbar und kontextualisiert.

Die Posterpräsentation und die Demo der App zeigen den aktuellen Entwicklungsstand des Projekts und demonstrieren das Potenzial der COPLUS-Korpora und der App als innovatives Werkzeug für den DaF-Unterricht. COPLUS leistet damit einen Beitrag zur Stärkung der Rolle des Deutschen in der europäischen Sprachenlandschaft und zur Erweiterung des methodischen Repertoires empirisch und digital gestützter Sprachlernforschung.

Referenzen:

Arnone C./Flinz C. (2025), Risorse lessicografiche e corpora per la ricerca e la didattica del tedesco come lingua straniera: il caso di *einen Hinweis auf etwas geben*. In: *Italiano LinguaDue* 2. 2025. i.E.

Ehlich K. (1999), Alltägliche Wissenschaftssprache. In: *Informationen Deutsch als Fremdsprache*, 26, 1, pp. 3-24: https://doi.org/10.1515/infodaf-1999-0102.

Fügert N./Richter U. A. (2016), *Wissenschaftssprache verstehen. Wortschatz – Grammatik – Stil – Lesestrategien*, 1., Deutsch für das Studium, Ernst Klett Sprachen, Stuttgart.

Koesters Gensini, S.E./Costa M./Flinz C. (i. Vorb.): Von Korpora zur App: die Entwicklung eines digitalen Lehrgangs zum Erwerb der alltäglichen Wissenschaftssprache des Deutschen als Drittsprache: das Coplus Projekt. In: *Akten der IVG-Tagung*, Grazer Universitätsverlag, Graz.

Meißner C. (2014), *Figurative Verben in der allgemeinen Wissenschaftssprache des Deutschen: eine Korpusstudie*, Stauffenburg, Leipzig.

Digilex^{PLUS} – ein zweisprachiger Blick auf neue Wörter

Inja Skender Libhard und Kristina Hrastov (beide Abteilung für Germanistik, Philosophische Fakultät der Universität Zagreb (Kroatien))

In einer globalisierten Welt sind die kontinuierliche Anpassung und Erweiterung der Lexik jeder Sprache unerlässlich. Fast tagtäglich erscheinen neue Wörter, die erklärt bzw. übersetzt werden müssen. Die lexikographische Plattform Digilex^{PLUS} (https://digilex.ffzg.unizg.hr/) ist das erste zweisprachige Projekt, das sich mit der Übersetzung neuer Wörter aus dem Deutschen ins Kroatische beschäftigt. Dieses Projekt wurde vom Neologismenwörterbuch des Leibniz-Instituts für Deutsche Sprache inspiriert. Zurzeit umfasst die Plattform drei Wörterbücher, und zwar *Deutsch-kroatisches Wörterbuch der lexikalischen Innovationen* (Anić et al., 2015), *Deutsch-kroatisches Wörterbuch der lexikalischen Innovationen 2015-2021* (Čaušević et al., 2022) und *Coronawörterbuch*. Diese Wörterbücher enthalten mehr als 6.500 Stichwörter, und zwar neu gebildete Wörter, Wortverbindungen, Phraseologismen und neue Bedeutungen.

Die Stichwörter entstammen unterschiedlichen Quellen. Als Primärquellen dienten das Neologismenwörterbuch des Leibniz-Instituts für Deutsche Sprache sowie die lexikographischen Onlineportale Wortwarte und Wortwarte Reloaded. Ergänzend wurden neue Wörter durch eigene Recherchen in deutschsprachigen Onlinemedien, Magazinen und sozialen Netzwerken ermittelt. Mit den Stichwörtern aus verschiedenen Fach- und Sachbereichen – darunter Politik, Wirtschaft, Sport, Technologie, Lifestyle, Soziales, soziale Netzwerke u.Ä. – umfassen die Wörterbücher ein breit gefächertes Korpus, das die sprachliche Kreativität und Dynamik der gegenwärtigen deutschen Lexik widerspiegelt. Darüber hinaus liegt ein besonderer Wert von Digilex^{PLUS} im Coronawörterbuch, in dem der deutsche Wortschatz rund um die Coronapandemie dokumentiert und ins Kroatische übersetzt ist.

Digilex^{PLUS} wird regelmäßig aktualisiert, indem neue Stichwörter aufgenommen werden, die nach ihrer Häufigkeit, also nach der Anzahl der Bestätigungen in der Google-Suchmaschine, aufgrund ihrer interessanten Form und/oder Bedeutung ausgewählt werden. Eine solche Herangehensweise stellt sicher, dass das Wörterbuch stets zeitgemäß bleibt.

Alle substantivischen Stichwörter sind mit grammatischen Angaben versehen, die sich auf ihre morphologischen Aspekte beziehen. So ist jedes Substantiv einem oder mehreren Genera zugeordnet. Dazu sind auch die Formen des Genitivs Singular und des Nominativs Plural angegeben, wobei jeweils mehrere Formen aufgeführt sind. Bei einigen kroatischen Äquivalenten sind stilistische Angaben vermerkt, die ihre fachli-

che Zuordnung präzisieren oder den Gebrauchskontext und Wertungsaspekt verdeutlichen. Das Wörterbuch kann sowohl nach den Stichwörtern der Ausgangssprache als auch nach deren Übersetzungsäquivalenten in der Zielsprache durchsucht werden. Die Benutzer können sich aktiv an der Erweiterung des Wörterbuchs beteiligen, indem sie ihre Vorschläge einreichen können.

Dieses innovative und umfassende Sprachressourcentool, durch das ein großer Mangel in der kroatischen Lexikografie behoben wird, richtet sich an ein breites Publikum: Übersetzer:innen, Germanistikstudierende, Sprachwissenschaftler:innen, DaF/DaZ-Lehrkräfte und eigentlich alle Sprachinteressierten. Es bietet ihnen die Möglichkeit, sowohl aktuelle lexikalische Entwicklungen im Deutschen und Kroatischen als auch bisher nicht übersetzte Ausdrücke gezielt nachzuschlagen und die Zusammenhänge zwischen Wörtern, deren Bedeutung und Gebrauch nachzuvollziehen. Durch diese Ausrichtung wird das Wörterbuch zu einem wertvollen Werkzeug für Forschung, Lehre und Sprachpraxis.

Internationales **Promovierenden-Netzwerk** des IDS

POSTERSESSION

Dribbeln oder *dribblen*, *fakt* oder *faked*, *downgeshifted* oder *gedownshiftet* - Triggern phonographische Faktoren die Verwendung spezifischer Flexionsmorpheme?

Merle Benter (IDS)

Im Fokus der korpusgestützten Untersuchung stehen die sprachlichen Realisierungsmöglichkeiten morphologisch eingegliederter englischer Verbstämme, die Chronologie ihrer Einbettung in das deutsche Flexionsparadigma sowie das Verhältnis von Merkmalskonservierung und Integrationsprozessen. Dabei wird untersucht, ob die Art und Anzahl der exogenen Marker im Stamm Auswirkungen auf die Ausbildung von Konjugationsformen haben.

Technologiebezogene Erklärungs- und Instruktionspraktiken in der Interaktion

Helena Konstanze Budde (IDS)

Das Dissertationsprojekt untersucht, wie Erwachsene bei Problemen mit Geräten wie Smartphones im Alltag Unterstützung erhalten – und wie diese Hilfe interaktional umgesetzt wird. Grundlage sind Videoaufnahmen privater Interaktionen, die konversationsanalytisch ausgewertet werden. Im Fokus stehen alltägliche Instruktions- und Erklärungshandlungen sowie räumliche und bildschirmbezogene Praktiken.

Grammatikalisierungstand der Verben *haben* und *sein*

Barbora Brožovičová (Universität Ostrava (Tschechien))

Das Promotionsprojekt untersucht den aktuellen Grammatikalisierungsstand der Verben *haben* und *sein* im heutigen Deutsch. Auf der Basis korpuslinguistischer Analysen werden semantische Ausbleichung, syntaktische Bindung und funktionale Vielfalt erfasst. Ziel ist es, Übergänge zwischen lexikalischer und grammatischer Verwendung empirisch zu bestimmen.

Die Integration englischer Adjektive ins Deutsche

Luisa Cimander (Universität Würzburg/IDS)

In diesem Projekt wird die Integration englischer Adjektive ins Deutsche seit dem 17. Jahrhundert untersucht. Es sollen Adjektivgruppen identifiziert werden, die Integrations- und Gebrauchsmuster abbilden. Der Fokus liegt dabei auf morphologischen Prozessen bei der Entlehnung (z. B. über ein Suffix) und im Gebrauch (Flexion, Steigerung) sowie auf der syntaktischen Funktion der Adjektive.

Profis oder Amateure? Disambiguierung von Personenbezeichnungen aus dem Bereich Kunst und Kultur in deutschsprachigen Pressetexten

Sara Corain (IDS)

Das Ziel dieses Teilprojekts besteht darin, anhand sprachlicher Daten aus regionalen und überregionalen Zeitungen einen genaueren Einblick in die Semantik von Personenbezeichnungen aus den Bereichen Kunst/Kultur und Sport zu gewinnen. Dabei werden die Lexeme entweder als berufliche oder als allgemeine Rollenbezeichnungen manuell annotiert. Untersucht wird auch, ob die Zeitungsart einen Einfluss auf die Interpretation dieser Bezeichnungen hat.

Die Pause als grundlegendes prosodisches Phänomen in der Sprache von polnischen, deutschen, englischen und spanischen Fußballkommentatoren – Versuch einer Typologie

Jakub Dzidek (Universität Wrocław (Polen))

Das Thema dieser Doktorarbeit ist im Bereich der Prosodie verortet. Analysiert werden gesprochene Texte aus Fernsehfußballberichterstattungen, die von Kommentatoren live während des Spiels produziert werden. Im Fokus steht die Pause als prosodisches Phänomen, dessen Funktionen, Distribution und Dauer sowie der Zusammenhang mit anderen prosodischen Erscheinungen in einer vergleichenden Analyse von vier Sprachen untersucht werden.

Sprache unter Jugendlichen in Wien – ein neuer Multiethnolekt?

Zara Fahim (Universität Nottingham (Vereinigtes Königreich))

Diese Promotionsarbeit untersucht den Sprachgebrauch und Spracheinstellungen multiethnischer Jugendlichen in Wien. Obwohl der jugendliche Sprachgebrauch in einigen deutschen Städten dokumentiert ist (z. B. der kommunikative Stil der „türkischen Powergirls" in Mannheim, Keim 2008; „Kiezdeutsch" in Berlin, Wiese 2009), bleibt Wien wenig erforscht. Die geplante Forschungsmethodik zielt darauf ab, diese Lücke zu schließen.

Fugenelemente in N+N-Komposita: Ein integriertes Modell zu Distribution und Funktion

Hiroyuki Ikeda (Tokyo University of Foreign Studies (Japan))

Dieses Forschungsprojekt befasst sich mit Fugenelementen in N+N-Komposita im Gegenwartsdeutschen, wie *-s* in *Bildungsroman* oder *-en* in *Linguistentreffen*. Es untersucht, unter welchen Bedingungen sie auftreten und welche Funktion(en) sie erfüllen. Konkret wird dabei ein integriertes Modell entwickelt, in dem morphologische, prosodische, semantische sowie diachron-erklärende Perspektiven vereint sind, und beschrieben, wie diese miteinander interagieren.

Schweizer Diglossie als Integrationsherausforderung: Schwiizertüütsch aus der Perspektive des Sprachmanagements

Vojtěch Kocourek (Karlsuniversität Prag (Tschechien))

Welche Rolle spielen deutsche Sprachvarietäten im Integrationsprozess von Migrant*innen in der Schweiz? In dieser Dissertation wird mit Hilfe qualitativer und ethnografischer Methoden untersucht, wie staatliche Behörden und einzelne Akteure den Erwerb von schweizerdeutschen Dialekten managen. Der Schwerpunkt der Forschung liegt dabei auf Sprachkursen für erwachsene Migrant*innen.

Sprachliche Repräsentationen älterer Menschen in amtlichen und öffentlichen Diskursen in Deutschland und Frankreich (1960-2025)

Manon Pautard (Universität Straßburg (Frankreich))

In dieser Doktorarbeit geht es in erster Linie um die Art und Weise, wie ältere Menschen in parlamentarischen Texten und Zeitungsartikeln sprachlich dargestellt werden. Im Vordergrund steht eine formale, semantische und diskursive Analyse von Bezeichnungen für diese Altersgruppe aus einer kontrastiven Perspektive; dabei wird auch der Frage nachgegangen, inwiefern diese Bildungen als Träger von Stereotypen und potenzieller Diskriminierung fungieren.

Sprachliche Modellierung der Synästhesie in den Novellen von E. T. A. Hoffmann

Mariana Pelikan (Nationale Iwan-Franko-Universität Lwiw (Ukraine))

Die Arbeit untersucht synästhetische Ausdrucksformen im Werk E. T. A. Hoffmanns aus sprachwissenschaftlicher Perspektive. Im Fokus stehen sprachliche Einheiten, die unterschiedliche Sinnesmodalitäten verbinden und dadurch die Wahrnehmungsoriginalität des deutschen Romantikers ausdrücken. Beschrieben werden deren Typen sowie Funktionen an der Schnittstelle von Literatur, Linguistik und Wahrnehmungspsychologie.

Meinungsartikel als Forum für Ideologien und Diskurse zur Plattformarbeit

Marja Rautajoki (Universität Turku (Finnland))

Die Studie analysiert Diskurse zur Plattformarbeit in Meinungsartikeln deutscher und finnischer Zeitungen. Auf Basis von 22 Beiträgen aus der Süddeutschen Zeitung und Helsingin Sanomat (2020–2024) werden Stimmen, Ideologien und dialogische Mittel vergleichend untersucht. Sie ist Teil der Dissertation *Die Vielstimmigkeit der Plattformarbeit – Eine sprachwissenschaftliche Perspektive auf den Wandel der Arbeitswelt in Europa*.

Un-deutsch schreiben: Sprache als Ort der Aushandlung statt Zugehörigkeit

Krishnapriya Shaju (English and Foreign Language University Hyderabad (Indien))

Das Poster untersucht Yoko Tawadas exophones Schreiben und die Verfremdung des Deutschen aus japanischer Perspektive. Markierte Textstellen zeigen, wie Mehrstimmigkeit, Differenz und Hybridität durch Abweichungen in Bedeutung und Wortbildung Normerwartungen unterlaufen und Leser:innen irritieren. Englische und Malayalam-Übersetzungen zeigen, wie das Fremde verschoben, geschwächt oder neu erzeugt wird und hinterfragen sprachliche Zugehörigkeit.

Verbklassen und metaphorische Merkmalsübertragung

Laura Vicente Antunes (IDS)

In der Dissertation wird untersucht, in welchem Ausmaß innerhalb einer Verbklasse immer dieselben Merkmale metaphorisch übertragen werden, ob es sich dabei um ähnliche metaphorische Prozesse handelt und welche Begründung es für die Antworten dieser beiden Fragen gibt. Hierfür wird eine qualitative Korpusstudie für verschiedene Verbklassen anhand von Gedichten verschiedener Epochen durchgeführt.

Wissenschaftliche Identität im mehrsprachigen Kontext: Verfasserreferenz in Texten vietnamesischer DaF-Studierender?

Thien-Sa Vo (Universität Leipzig)

Wer bin ich in der Wissenschaft? Das Dissertationsprojekt untersucht, wie vietnamesische Studierende in deutschsprachigen Abschlussarbeiten ihre wissenschaftliche Identität sprachlich konstruieren. Im Fokus steht die Verfasserreferenz als zentrales Mittel der Selbstpositionierung, anhand derer analysiert wird, wie Schreibende ihre Rolle als Forschende zwischen Objektivitätsanspruch und Autorschaft ausgestalten.

„aber das wusstet ihr ja längst“ – Zum Gebrauch von Modalpartikeln in Science Slams

Johanna Vogel (IDS)

Das Poster untersucht den Gebrauch von Modalpartikeln in Science Slams im Vergleich zu FOLK- und GeWiss-Daten und betrachtet am Beispiel von *ja* unterschiedliche Verwendungsweisen. Grundlage der Untersuchung bildet ein Korpus wissenschaftskommunikativer Daten, wobei das erste Subkorpus Transkripte der Finalrunden der Deutschen Science-Slam-Meisterschaften der Jahre 2021–2024 umfasst.

Propositionale Argumente in europäischen Sprachen: Form, Funktion, Variation

Beata Trawiński und Kerstin Schwabe (beide IDS)

In deutschen wie auch anderen europäischen Sprachen können propositionale Argumente (PAs) in unterschiedlichen Strukturtypen realisiert werden. Im Sprachvergleich zeigt sich dabei eine systematische Korrelation zwischen der Stärke der semantischen Bindung und dem Grad der syntaktischen Integration (vgl. Givón 2001; Wurmbrand & Lohninger 2020). Zugleich lassen sich sprachübergreifende Parallelen in der (lexikalischen) Lizenzierung verschiedener PA-Typen beobachten (vgl. Noonan 1985; Givón 2001; Cristofaro 2003). Eine detailliertere Analyse offenbart jedoch beträchtliche Variation sowohl in den syntaktisch-semantischen Eigenschaften als auch in der Lizenzierung und im Gebrauch von PAs.

Der Beitrag verfolgt eine multikontrastive Perspektive auf verschiedene Typen von PAs mit dem Fokus auf Subjekt- und nicht-obliquer Objektfunktion und kontrastiert das Deutsche mit den vier europäischen Sprachen Englisch, Italienisch, Polnisch und Ungarisch. Ziel ist es, die zentralen Variationsparameter im Bereich der Nebensatzsyntax zu bestimmen und das Deutsche im europäischen Kontext zu verorten.

Im Bereich der Subjektsätze werden insbesondere folgende Parameter analysiert: (i) der Grad der Nominalität propositionaler Subjekte (finite vs. infinite, mit oder ohne Komplementierer bzw. Determinierer), (ii) ihre syntaktische Position und (iii) der mögliche oder obligatorische Einsatz von Proformen. Weitere Aspekte betreffen Kongruenz, Koordinierbarkeit und Referenzidentität. Auch die lexikalische Lizenzierung wird systematisch einbezogen. Die Ergebnisse zeigen, dass propositionale Subjekte sprachübergreifend entlang einer Hierarchie von evaluativen, wahrheitswertbezogenen und deontischen über emotive und kognitive bis hin zu konnektiven Prädikaten lizenziert werden, was auf Unterschiede in den zugrunde liegenden Argumentstrukturen zurückzuführen ist (vgl. Haiman 1980; Kaltenböck 2004; Diessel 2008). Insgesamt erweisen sich propositionale Subjekte als eine formal variable, aber grammatisch fest etablierte Kategorie in allen untersuchten Sprachen.

Die Analyse nicht-obliquer Argumentsätze zeigt, dass alle finiten PAs als CPs realisiert werden und ihre Abhängigkeit durch Komplementierer markieren. Unterschiede bestehen in der Möglichkeit des Komplementiererwegfalls (nur im Englischen, Italienischen und Ungarischen), in der Verbstellung (Deutsch: OV; andere Sprachen: VO) und in der Modusmarkierung (Indikativ vs. nicht-indikativische Modi). Hauptsatzindikatoren können in allen Sprachen auch in eingebetteten Strukturen auftreten, und kataphorische satzwertige Proformen sind überall belegt. Bei infiniten Argumentsätzen treten sowohl bisententiale (Kontroll-)Konstruktionen mit PRO als auch monosententiale Strukturen ohne CP-Grenze auf.

Insgesamt bestätigen die Ergebnisse die Annahme einer grundsätzlichen Korrelation zwischen syntaktischen und semantischen Dimensionen propositionaler Strukturen, während sprachspezifische Unterschiede deren komplexe, systematisch variierende Ausprägung im europäischen Vergleich sichtbar machen.

Literatur:

Cristofaro, S. (2003): Subordination. Oxford and New York.

Diessel, H. (2008): Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. In: Cognitive Linguistics, 19(3), pp. 465-490. https://doi.org/10.1515/COGL.2008.018.

Givón, T. (2001): Syntax. An introduction. Vol. I/II. Amsterdam and Philadelphia.

Haiman, J. (1980): The iconicity of grammar. Isomorphism and motivation. In: Language, 56(3), pp. 515-540.

Kaltenböck, G. (2004): It-extraposition and non-extraposition in English: A study of syntax in spoken and written texts. Wien.

Noonan, M. (1985): Complementation. In: Shopen, T. (ed.): Language typology and syntactic description. Cambridge, pp. 41-140.

Wurmbrand, S./Lohninger, M. (2022): An implicational universal in complementation – Theoretical insights and empirical progress. In: Hartmann, J. M./Wöllstein, A. (eds.): Propositionale Argumente im Sprachvergleich / Propositional arguments in cross-linguistic research. Tübingen, pp. 183-229.

DONNERSTAG, 12. MÄRZ 2026, 9.45 UHR

Deiktische Verben in europäischen Sprachen: Eine theoretische Neubewertung

Volker Gast (Friedrich-Schiller-Universität, Jena)

Es ist bekannt, dass sich die europäischen Sprachen hinsichtlich der Bedingungen unterscheiden, unter denen COME- und GO-Verben verwendet werden (z. B. Fillmore 1997; Ricca 1993; Lewandowski 2014; Martinková & Janebová 2024). Einige Sprachen, etwa Spanisch, verlangen in der Regel, dass ein COME-Verb (*venir*) eine Bewegung in Richtung des Sprechers kodiert – entweder zur „Kodierzeit" oder zur „Referenzzeit". Andere Sprachen, z.B. die westlslawischen Standardsprachen, sind hinsichtlich des Bewegungsziels weniger restriktiv; in diesen Sprachen unterscheiden sich die betreffenden Verben primär in der Perspektive, aus der das Ereignis betrachtet wird (Aufbruch vs. Ankunft; siehe z.B. Lewandowski 2014 zum Polnischen und Martinková & Janebová 2024 zum Tschechischen). Das deutsche Verb *kommen* ist flexibler als das Spanische *venir*, aber weniger flexibel als die entsprechenden Verben der westslawischen Standardsprachen; allerdings ist es immer deiktisch, wenn es mit den Präverben *hin-* oder *her-* auftritt.

Während der Großteil der Variation innerhalb Europas relativ gut beschrieben ist, fehlt bislang eine Theorie, die sowohl sprachspezifische Fakten als auch sprachübergreifende Unterschiede in diesem Bereich explizit modelliert. In diesem Vortrag schlage ich vor, die Ausdrucksweise von Bewegung analog zur von Klein (1994) entwickelten Tempus- und Aspekttheorie zu modellieren. Klein unterscheidet drei Komponenten der temporalen Referenz:

- **Time of Utterance (TU)**
- **Topic Time (TT)**, „the time span to which the speaker's claim on this occasion is confined" (Klein 1994: 4)
- **Time of Situation (TSit)**

Tempus kodiert die Relation zwischen TT und TU, und Aspekt kodiert die Relation zwischen TT und TSit. Klein schlägt zudem eine Typologie von Situationstypen (*0-state*, *1-state*, *2-state-predicates*) vor, welche die Art von Veränderung erfasst, die mit einer Situation einhergeht.

In Analogie dazu schlage ich vor, dass Bewegungsausdrücke drei Komponenten umfassen:

- **Origo** (deiktisches Zentrum, vgl. TU)
- **Vantage Point** (Ort, über den eine Aussage getroffen wird, vgl. TT)
- **Path** (der gesamte Vektor von Source zu Goal, vgl. TSit)

Darüber hinaus besitzen Bewegungsverben inhärente lexikalische Bedeutungsmerkmale, die spezifizieren, ob jeweils Source, Goal oder der gesamte Path in den Blick genommen wird (inchoativ, terminativ, holistisch).

Jede Beschreibung eines Bewegungsereignisses kann somit anhand folgender Kategorien charakterisiert werden:

- **Deixis:** Relation zwischen Vantage Point und Origo (analog zu Tempus)
  Werte: proximal vs. distal
- **Orientation:** Relation zwischen Vantage Point und Path (analog zu Aspekt)
  Werte: accessive vs. abscessive)
- **Event Type:** Fokus auf einem spezifischen Segment des Paths (analog zu Aktionsart)
  Werte: inchoativ vs. terminativ

Diese Kategorien sind orthogonal, so dass sich acht logisch mögliche Beschreibungen von Bewegungsereignissen ergeben. Wie bei Tempus und Aspekt kodieren jedoch einzelne Sprachen nur bestimmte Kombinationen von Deixis, Orientation und Event Type. Beispielsweise scheinen die europäischen Sprachen keine Verben zu besitzen, die zugleich inchoativ und accessiv sind, obwohl solche Verben in anderen Sprachen durchaus belegt sind (z.B. in einigen Sprachen Papua-Neuguineas).

Die Anwendung von Kleins Tempustheorie auf den räumlichen Bereich ermöglicht es, verschiedene sprachübergreifende Unterschiede einheitlich zu erfassen und Variation sowohl innerhalb Europas als auch darüber hinaus theoretisch strukturiert zu model-

lieren. Ich werde, ausgehend vom Deutschen, unter Verwendung eines Parallelkorpus (InterCorp, Čermák & Rosen 2012) zunächst die europäischen Sprachen vergleichen und anschließend weitere Sprachen außerhalb Europas in den Blick nehmen.

Literatur:

Čermák, František and Alexander Rosen (2012). The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics* 17(3): 411-427.

Fillmore, Charles (1997). *Lectures on Deixis*. Stanford: CSLI.

Klein, Wolfgang (1994). *Time in Language*. London: Routledge.

Lewandowski, Wojciech (2014). Deictic verbs: Typology, thinking for speaking and SLA. *SKY Journal of Linguistics* 27, 43-65.

Martinková, Michaela & Markéta Janebová (2024). English or Czech Venitive Verbs in Contrast: Deictic or not. *Nordic Journal of English Studies* 23(2), 34-60.

Matsumoto, Yo, Kimi Akita & Kiyoko Takahashi (2017). The functional nature of deictic verbs and the coding patterns of deixis. In I. Ibarretxe-Antuñano (ed.), *Motion and Space across Languages: Theory and Applications*, 95-122. Amsterdam: John Benjamins.

Ricca, Davide (1993). *I verbi deittici di movimento in Europa: una ricerca interlinguistica*. Firenze: La Nuova Italia Editrice.

DONNERSTAG, 12. MÄRZ 2026, 10.30 UHR

Zwischen Komplexität und Vereinfachung: Deutsch im europäischen Sprachraum am Beispiel der alpinen Sprachinseln

Livio Gaeta (Universität Turin (Italien))

In letzter Zeit hat das Thema der Komplexität eine sehr lebhafte Debatte ausgelöst (vgl. Miestamo et al. 2008, Sampson et al. 2009, Baerman et al. 2015, Baechler et al. 2016, Meakins et al. 2019, Arkadiev et al. 2020). Insbesondere wurde die Hypothese diskutiert, dass Sprachkontakt Vereinfachungsprozesse fördert und somit der Komplexität entgegenwirkt. Andererseits wurde beobachtet, dass sprachliche Veränderungen (jeglicher Art, unabhängig davon, ob sie eine Vereinfachung begünstigen oder die Komplexität beibehalten) eng mit dem soziolinguistischen Profil der Sprechergemeinschaft verbunden sind, insbesondere hinsichtlich des Status der Mehrsprachigkeit innerhalb dieser Gemeinschaft (Trudgill 2011, 2020). Dieser Beitrag konzentriert sich auf die deutschen Sprachinseln der italienischen Alpen, die auf eine lange Tradition des Kontakts mit den romanischen Sprachen zurückblicken können, mit denen sie seit Jahrhunderten in Verbindung stehen (Angster & Gaeta 2021). Es wird gezeigt, wie die Phänomene des sprachlichen Verfalls – und damit der Vereinfachung –, die nach der Festlegung der Grenzen der Nationalstaaten zu beobachten sind, von der internen Entwicklung der einzelnen Varietäten in Kontaktsituationen unterschieden werden

müssen. Dieser Vergleich wird dank der Überlieferung dieser Sprachvarietäten im 19. Jahrhundert und allgemein vor den in den letzten Jahrzehnten zu beobachtenden Verfallsprozessen ermöglicht (Gaeta im Druck). Insbesondere wird das Thema der Komplexität anhand verschiedener Phänomene aus dem Bereich der nominalen bzw. Verbmorphologie, der Wortbildung sowie der Satzstruktur vertieft (Gaeta 2018, 2020, 2025, Gaeta et al. im Druck). Es ist gerade die soziolinguistische Typologie der jeweiligen Sprechergemeinschaften, die Aufschluss über die Entwicklung von Komplexifizierungs- oder Vereinfachungsphänomenen gibt, die mit anderen Fällen des Kontakts zwischen germanischen und romanischen Sprachen außerhalb des Alpenraums verglichen werden können (Gaeta 2024).

Literatur:

Angster, M. & L. Gaeta, Contact phenomena in the verbal complex: the Walser connection in the Alpine area. *STUF – Language Typology and Universals* 74.1: 73-107.

Arkadiev, P. et al. (eds.) 2020. *The Complexities of Morphology*. Oxford.

Baechler, R. et al. (eds.) 2016. *Complexity, Isolation, and Variation*. Berlin etc.

Baerman, M. et al. (eds.) 2015. *Understanding and Measuring Morphological Complexity*. Oxford.

Gaeta, L. 2018. Im Passiv sprechen in den Alpen. *Sprachwissenschaft* 43.2: 221-250.

Gaeta, L. 2020. Remotivating inflectional classes: an unexpected effect of grammaticalization. In B. Drinka (ed.), *Historical Linguistics 2017*. Amsterdam etc., 205-227.

Gaeta, L. 2024. Intense language contact and collapse of lexical strata: verbs ending with *-urun* in Issime. *Journal of Language Contact* 17: 642-663.

Gaeta, L. 2025. Remodeling inflectional classes: Strong and weak verbs in Walser German. In N. Levkovych et al. (eds.), *Exploring Structures in Languages and Language Contact*. Berlin etc., 63-91.

Gaeta, L. im Druck. Diachronic perspectives on a rich linguistic repertoire: Translations of the Parable of the Prodigal Son in Walser German varieties. In F. Cognola et al. (eds.), *Le traduzioni della Parabola del Figliol Prodigo nelle varietà parlate nelle isole linguistiche germanofone italiane*. Venedig.

Gaeta, L. et al. im Druck. Korpuslinguistik am Beispiel der walserdeutschen Sprachinseln in Italien. Sprachkontakt, Spracherhaltung, Sprachwandel. In A. Prediger et al. (eds.), *Deutsche Sprachminderheiten in der Welt: Empirische Studien zur Sprachvariation und Sprachideologie*. Berlin etc.

Meakins, F. et al. 2019. Birth of a contact language did not favor simplification. *Language* 95.2: 294-332.

Miestamo, M. et al. (eds.) 2008. *Language complexity: typology, contact, change*. Amsterdam etc.

Sampson, G. et al. (eds.) 2009. *Language complexity as an evolving variable*. Oxford.

Trudgill, P. 2011. *Sociolinguistic Typology*. Oxford.

Trudgill, P. 2020. *Millennia of Language Change*. Cambridge.

Mehrsprachige Korpora für die DGD: Potenziale und Herausforderungen

Mark-Christoph Müller und Siegwalt Lindenfelser (beide IDS)

Die „Datenbank für Gesprochenes Deutsch“ (DGD) (vgl. Schmidt 2017) ist die zentrale Distributionsplattform für mündliche Korpora des „Archiv für Gesprochenes Deutsch“ (AGD) am IDS in Mannheim. Mit dem Release 2.25 im Frühjahr 2026 stellt die DGD jetzt 45 inhaltlich erschlossene und mit Metadaten angereicherte mündliche Korpora für die Online-Nutzung (Recherche von Transkripten und Metadaten, Zugriff auf Audios und ggfs. Videos) zur Verfügung. Während manche dieser Korpora – z.B. Variantionskorpora zu von Sprachkontakt geprägten extraterritorialen Varietäten des Deutschen – schon immer (kleinere) fremd- und mehrsprachige Anteile enthielten, bewegten sich diese nicht-deutschsprachigen Anteile jedoch bisher in einem Rahmen, in dem es nicht nötig erschien, spezielle Funktionalitäten für den Umgang mit mehrsprachigen Daten in der DGD zu entwickeln.

Durch die Übernahme und Aufbereitung eines ersten multilingualen Vergleichskorpus zur Nachnutzung in der DGD hat sich dies nun geändert. Es handelt sich um das Korpus „Parallel European Corpus of Informal Interaction“ (PECI) (vgl. Kornfeld et al 2023; Küttner et al. 2024), welches seit dem letzten Release mit seinen vier Aufnahmesprachen Deutsch, Englisch, Italienisch und Polnisch in der DGD zur Verfügung steht. Das Korpus enthält knapp 80 Stunden Aufnahmen informeller Interaktion bei Familienfrühstücken, bei Spieleabenden unter Freunden sowie bei privaten Autofahrten mit jeweils zwei Kameraperspektiven. Außerdem ist für das Release 2.27 die Bereitstellung der multilingualen Version des Korpus „Gesprochene Wissenschaftssprache kontrastiv“ (GWSS) (vgl. Fandrych et al. 2012) in Vorbereitung. Eine rein deutschsprachige Version dieses Korpus ist in der DGD bereits seit 2017 vorhanden, eine basale multilinguale Version ist derzeit über die Plattform ZuMult recherchierbar.

Im Vortrag zeigen wir, welche Funktionalitäten für die Arbeit mit genuin mehrsprachigen Korpora wie PECI und GWSS für die DGD entwickelt wurden, wie sich die vorangehende Datenaufbereitung mit Blick darauf gestaltete und welche Nutzungspotenziale sich hieraus ergeben. Dabei werden auch verschiedene Herausforderungen skizziert, die sich für die Aufbereitung und Bereitstellung solch komplexer multilingualer Daten ergeben. Zuletzt erfolgt ein Ausblick zu weiteren Perspektiven und auch Grenzen hinsichtlich des Funktionalitäts- und Datenangebots für mehrsprachige Daten am AGD und in der DGD.

Referenzen:

Fandrych, Christian/Meißner, Cordula/Slavcheva, Adriana (2012): The GeWiss corpus: Comparing spoken academic German, English and Polish. In: Schmidt, Thomas/Wörner, Kai (eds.): Multilingual Corpora and Multilingual Corpus Analysis. Hamburg Studies in Multilingualism 14. Amsterdam: John Benjamins, S. 319-338.

Kornfeld, Laurenz/Küttner, Uwe-A./Zinken, Jörg (2023): Ein Korpus für die vergleichende Interaktionsforschung. Das 'Parallel European Corpus of Informal Interaction' (PECII). In: Deppermann, Arnulf/Fandrych, Christian/Kupietz, Marc/Schmidt, Thomas (Hrsg.): Korpora in der germanistischen Sprachwissenschaft. Mündlich, schriftlich, multimedial. Berlin/Boston: de Gruyter, S. 103-128.

Küttner, Uwe-A./Kornfeld, Laurenz/Mack, Christina/Mondada, Lorenza/Rogowska, Jowita/Rossi, Giovanni/Sorjonen, Marja-Leena/Weidner, Matylda/Zinken, Jörg (2024): Introducing the 'Parallel European Corpus of Informal Interaction' (PECII). A novel resource for exploring cross-situational and cross-linguistic variability in social interaction. In: Selting, Margret/Barth-Weingarten, Dagmar (Hrsg.): New Perspectives in Interactional Linguistic Research. Amsterdam/Philadelphia: John Benjamins (Studies in Language and Social Interaction 36), S. 132-160.

Schmidt, Thomas (2017): DGD – die Datenbank für Gesprochenes Deutsch. Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim. In: Zeitschrift für Germanistische Linguistik 45(3), S. 451-463.

DONNERSTAG, 12. MÄRZ 2026, 12.30 UHR

Das Deutsche im Repertoire türkeistämmiger Rückkehrer- und TransmigrantInnen

Ibrahim Cindark (IDS) und Serap Devran (Marmara-Universität, Istanbul (Türkei))

Untersuchungen zu kontaktsprachlichen Varietäten des Deutschen außerhalb der deutschsprachigen Kerngebiete in Europa fokussieren bislang Sprachkontaktsituationen, die entweder aus kolonialen Kontexten hervorgegangen sind (u.a. Wiese/Bracke (2021) für Namibiadeutsch, Lindenfelser (2021) für Unserdeutsch auf Papua-Neuguinea) oder auf historisch weit zurückliegende Auswanderungen von Deutschen zurückgehen (u.a. Berend (2011) für Russlanddeutsch, Riehl (2015) für Australiendeutsch, Rosenberg (2016) für Brasiliendeutsch, Boas (2018) für Texasdeutsch, Stolberg (2015) für Pennsylvaniadeutsch und Dück (2024) für Kaukasiendeutsch). Dagegen wurde bisher das Deutsche von AuswanderInnen, die in den letzten Jahrzehnten zum Beispiel nach Spanien oder in die Türkei migriert sind, bislang kaum erforscht.

Mit unserem Beitrag wollen wir diese Forschungslücke ein Stück weit schließen. In der Türkei lassen sich zwei Gruppen von SprecherInnen des Deutschen unterscheiden: zum einen deutschstämmige AuswanderInnen, die die Türkei zu ihrem neuen Lebensmittelpunkt gewählt haben, und andererseits die Gruppe der türkeistämmigen Rückkehrer- bzw. TransmigrantInnen, die entweder als Erwachsene freiwillig migrierten oder als Kinder im Zuge der Migrationsentscheidung ihrer Eltern von Deutschland in die Türkei kamen. Insgesamt wird geschätzt, dass gegenwärtig um die 140.000 deutsche StaatsbürgerInnen in der Türkei leben (Ermagan/Ermagan 2020).

In unserem Beitrag untersuchen wir ein Korpus aus 37 sprachbiografischen Interviews und drei Gruppengesprächen mit der Beteiligung von 14 türkeistämmigen Rückkehrer- bzw. TransmigrantInnen. Die Aufnahmen wurden in den 2010er Jahren gemacht und bislang nur biografie- und interaktionsanalytisch untersucht (Devran 2017, 2019a, 2019b, 2019c). In unserem Beitrag erweitern wir diese Perspektive um sprachkontakt- und variationslinguistische Analysen der erhobenen Daten. Zunächst diskutieren wir die theoretische und analytische Abgrenzung der Kategorien „RückkehrerInnen" und „TransmigrantInnen". Im Anschluss untersuchen wir, inwiefern die Beteiligten in ihren sprachbiografischen Narrativen auf das Konzept der „Muttersprache" zurückgreifen und welche Funktionen diesem Begriff in der Selbstpositionierung und Identitätskonstruktion zukommen. Aus der Sprachkontaktperspektive beleuchten wir in der Folge den Gebrauch der deutschen Präpositionen im gesprochenen Deutsch der Untersuchten. Abschließend diskutieren wir, welche Aspekte der Sprachvariation zwischen Deutsch und Türkisch in den Daten zu beobachten sind. Zum Schluss und als ein Ausblick gehen wir darauf ein, welche neuen Aufnahmen wir seit 2022 gemacht haben und welche noch demnächst ausstehen.

Literatur:

Berend, Nina (2011): Russlanddeutsches Dialektbuch. Halle (Saale): Projekte Verlag.

Boas, Hans (2018): Texas. In: Plewnia, Albrecht/Riehl, Claudia (Hg.): Handbuch der deutschen Sprachminderheiten in Übersee. Tübingen: Narr, S. 171-192.

Devran, Serap (2017): Deutsch-türkische Migration: Die Darstellung narrativer Identitäten von Studentinnen in Istanbul. Eine biografie- und interaktionsanalytische Pilotstudie. (= *amades*. Arbeiten und Materialien zur deutschen Sprache 51). Mannheim: Institut für Deutsche Sprache.

Devran, Serap (2019a): Narrative Bewältigung von einschneidenden Erlebnissen eines Rückkehrers in der deutschen und türkischen Lebenswelt. In: Deutsche Sprache, Jg. 47, Nr. 3, S. 258-282.

Devran, Serap (2019b): Narrativer Entwurf einer positiven Selbstkategorie in unterschiedlichen Sozial- und Sprachwelten. In: Alman Dili ve Edebiyatı Dergisi, 2019/41, S. 55-88.

Devran, Serap (2019c): Biographische Rekonstruktion von Identität im deutschen und türkischen Bildungskontext. In: Diyalog – Interkulturelle Zeitschrift für Germanistik, 2019/2, S. 269-295.

Dück, Katharina (2024): "Vot die drei Sprache kann i gut…nu russische und grusinische sehr […]" – Zu russischen Gesprächspartikeln als Eröffnungs- und Schlusssignal im Kaukasiendeutschen. In: Deutsche Sprache, Jg. 52, Nr. 3, S. 228-245.

Ermagan, Ismail/Ermagan, Elif (2020): Bogaz`dan Alanya`ya - Türkiye`deki Almanlar. In: Cihan, Ahmet/Genc, Hamdi/Ermagan, Ismail (Hg.): Dünyada Göç Yönetimi ve Türkiye'nin Göçmenleri Göçü Nasıl Yönetmeli? Istanbul: Akademik Kitaplar, S. 341-385.

Riehl, Claudia (2015): Language attrition, language contact and the case of a relic variety: the case of Barossa German. In: International Journal of the Sociology of Language 236, S. 261–293.

Stolberg, Doris (2015): Changes Between the Lines. Diachronic contact phenomena in written Pennsylvania German. Berlin/Boston: de Gruyter.

Wiese, Heike/Bracke, Yannic (2021): Registerdifferenzierung im Namdeutschen: Informeller und formeller Sprachgebrauch in einer vitalen Sprechergemeinschaft. In: Csaba Földes (Hg.): Kontaktvarietäten des Deutschen im Ausland. Tübingen: Narr, S. 273-292.

Wir freuen uns sehr, auch bei der 62. IDS-Jahrestagung in Mannheim wieder zahlreiche Verlage begrüßen zu dürfen, die vor Ort mit Verlagstischen und ihren aktuellen Programmen vertreten sind.

De Gruyter Brill

DUDEN

ESV ERICH SCHMIDT VERLAG

Franz Steiner Verlag

GEORG OLMS VERLAG

BUSKE

iudicium
(mit virtuellem Büchertisch vertreten)

IDS IDS-Verlag

1682 J.B. METZLER

narr\f ranck e\atte mpto

PETER LANG
INTERNATIONAL ACADEMIC PUBLISHERS

wbv Publikation

Universitätsverlag WINTER Heidelberg

Frank & Timme

Königshausen & Neumann

Das Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim ist die gemeinsam vom Bund und allen Bundesländern getragene zentrale wissenschaftliche Einrichtung zur Dokumentation und Erforschung der deutschen Sprache in Gegenwart und neuerer Geschichte.

Es gehört zu den über 90 außeruniversitären Forschungs- und Serviceeinrichtungen der Leibniz-Gemeinschaft.

IDS | LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

R 5, 6-13 • 68161 Mannheim
Tel: +49 621 1581-0
Fax: +49 621 1581-200
info@ids-mannheim.de
www.ids-mannheim.de
ids.mannheim
@idsmannheim.bsky.social
ids_mannheim
https://wisskomm.social/@ids_mannheim

Leibniz Gemeinschaft