



Mason A. Wirtz\*, Simon Pickl, Konstantin Niehaus, Stephan Elspaß,  
Robert Möller

# Gebrauchsstandard in der deutschen Alltagssprache: Eine integrative Modellierung räumlicher und sozialer Variation

## Standard of usage in everyday German: An integrative modelling of spatial and social variation

<https://doi.org/10.1515/zgl-2025-2003>

**Abstract:** While previous dialectometric analyses of colloquial German have explored the spatial distribution of variants in order to sketch a picture of how German-speaking areas are linguistically structured, there have been no investigations into the extent to which reported everyday language use converges to or diverges from the standard language. What is more, dialectological perspectives have largely neglected the role of social factors such as age, gender, and mobility on patterns of language variation in colloquial German in lieu of strictly examining geography as a predictor for differences in the spatial distribution of variation in everyday language use. In this paper, we attempt to reconcile dialectometric and sociolinguistic perspectives by examining, in an integrative manner, the influence of social and spatial variables in predicting differences in patterns of variation in colloquial German. Specifically, we analyze ten phonetic-phonological variables from the *Atlas zur deutschen Alltagssprache* ('Atlas of Colloquial German') and address the question of where informant reports of colloquial German resemble (non-)standard language. To do so,

---

\***Kontaktperson:** Dr. Mason A. Wirtz: Universität Zürich, Englisches Seminar, Plattenstrasse 47, CH-8032 Zürich, E-Mail: mason.wirtz@es.uzh.ch. <https://orcid.org/0000-0002-9408-1993>

**Assoz. Prof. Dr. Simon Pickl:** Universität Salzburg, Fachbereich Germanistik, Erzabt-Klotz-Straße 1, A-5020 Salzburg, E-Mail: simon.pickl@plus.ac.at. <https://orcid.org/0000-0003-0267-1177>

**Dr. Konstantin Niehaus:** Universität Salzburg, Fachbereich Germanistik, Erzabt-Klotz-Straße 1, A-5020 Salzburg, E-Mail: konstantin.niehaus@plus.ac.at. <https://orcid.org/0000-0002-2489-5764>

**Univ.-Prof. Dr. Stephan Elspaß:** Universität Salzburg, Fachbereich Germanistik, Erzabt-Klotz-Straße 1, A-5020 Salzburg, E-Mail: stephan.elspass@plus.ac.at. <https://orcid.org/0000-0002-3870-0817>

**Prof. Dr. Robert Möller:** Université de Liège, Faculté de Philosophie et lettres, Département de langues modernes: linguistique, littérature et traduction, Place Cockerill 3–5, B-4000 Liège, E-Mail: r.moller@uliege.be. <https://orcid.org/0000-0002-7996-1714>

we employ generalized additive mixed-effects models (GAMMs), which can simultaneously detect underlying (non-)linear aggregate geographical patterns and identify the relative importance of social predictors. Our results indicate that geography is the strongest predictor of differences in colloquial German, and that the interaction between age and gender most notably affects the number of (non-)standard variants informants report for their respective locality. Methodological implications for applying GAMMs to the investigation of social and spatial variation are discussed.

- 1 Hinführung: eine integrative Modellierung räumlicher und sozialer Variation als Forschungsdesiderat
- 2 Methodologie
- 2.1 Atlas zur deutschen Alltagssprache
- 2.2 Variablenauswahl
- 2.3 Informant:innen und Ortsnetz
- 2.4 Statistische Analyse
- 3 Ergebnisse
- 3.1 Zur Modellierung der Rolle von Geographie
- 3.2 Zur Integration von sozialen Prädiktoren
- 4 Diskussion
- 5 Fazit und Ausblick
- Danksagung
- Literatur

## **1 Hinführung: eine integrative Modellierung räumlicher und sozialer Variation als Forschungsdesiderat**

Lange lag der Fokus in geostatistischen bzw. dialektometrischen Verfahren überwiegend auf der Dialektgeographie (vgl. Nerbonne et al. 2010), ungeachtet der möglicherweise ausschlaggebenden Rolle von sozialen Faktoren im Hinblick auf ihre Vorhersagekraft von Unterschieden in der räumlichen Distribution sprachlicher Variation. Die Soziolinguistik hingegen fokussierte den Einfluss soziodemographischer Kategorien auf den inter- und intraindividuell unterschiedlichen Gebrauch von (primär soziolinguistisch-funktionalen) Varianten und Varietäten (vgl. Bülow & Pfenninger 2021; Johnson 2009; Wieling et al. 2014), ungeachtet etwaiger diatopischer Faktoren. Auch wenn weitgehend ein Konsens herrscht, dass die Dialektologie und Soziolinguistik seit Längerem „converge at their deepest point“ und im Wesentlichen denselben Gegenstand haben – Sprachgebrauch und -wandel in der Gesellschaft

schaft – (Chambers & Trudgill 1998, 187), so unterliegen die zwei Bereiche oft weiterhin einer deutlichen Trennung, was die Verwendung von Methoden und Techniken für die Analyse von Sprachvariation und Sprachwandel betrifft (Wieling et al. 2014, 670). Während die Soziolinguistik seit ihrer Begründung (Labov 1966) von überwiegend quantitativ ausgerichteten Ansätzen zur Modellierung verschiedener sozialer Einflussgrößen (wie Alter, Geschlecht, sozioökonomischer Status – neuerdings auch Persönlichkeitsfaktoren [vgl. Steiner, Jeszenszky, Stebler et al. 2023] sowie (sozio-)affektiver Variablen [vgl. Steiner, Jeszenszky & Leemann 2023] und kognitiver Fähigkeiten [vgl. Wirtz i. Dr.]) Gebrauch macht, stützten sich traditionelle dialektologische Verfahren vorwiegend auf qualitative Analysen kategorischer Karten, indem subjektive Kartenbilder einzelner oder weniger sprachlicher Variablen und geographische Verteilungen von Varianten durch bloßen Augenschein interpretiert werden. Erst in den letzten vierzig Jahren wurden quantitative Methoden im Rahmen dialektometrischer, v. a. aggregativer Analysen zur Erforschung geographischer Muster sprachlicher bzw. dialektaler Variation angewandt (vgl. etwa Goebel 1984; Nerbonne 2009; Nerbonne et al. 2021; Pickl 2013; Pröll et al. 2015).

Auch wenn sich die quantitativen Ansätze in Dialektologie und Soziolinguistik ergänzen und voneinander profitieren können, beklagen etwa Wieling et al. (2014, 670), dass „social and spatial analyses of language are still far from being integrated“. Auf der einen Seite kommen in der Soziolinguistik räumliche Faktoren zu kurz, und auf der anderen Seite bleiben häufig soziale Einflussgrößen in dialektometrischen Verfahren zugunsten eines Hauptfokus auf die Dialektgeographie unberücksichtigt (vgl. Britain 2004). In den letzten Jahrzehnten werden über Geographie bzw. Soziodemographie hinausgehende Prädiktoren stärker berücksichtigt (siehe für diachronische Perspektiven etwa Wieling et al. [2007] oder für die Einbeziehung von Alter und/oder Geschlecht als Kovariaten etwa Leinonen [2010], Steiner, Jeszenszky & Leemann [2023], Vergeiner & Elspaß [i. Dr.]; Vergeiner [2023]; Wirtz et al. [i. Dr.]). Ansätze zur gleichzeitigen Berücksichtigung geographischer und mehrerer sozialer Variablen sind jedoch noch immer verhältnismäßig selten.

Als Erste haben Wieling et al. (2011) das Zusammenspiel von sozialen, geographischen und sprachlichen Faktoren integrativ modelliert und untersucht. Anhand von *generalized additive models* und *mixed-effects models* haben sie soziolinguistische und geolinguistische Perspektiven zusammengeführt, um regionale Ausspracheunterschiede vom Standardniederländischen in Abhängigkeit vom geographischen Ort sowie verschiedenen wort- und gemeinschaftsbezogenen Faktoren zu untersuchen. Bemängelt haben allerdings Wieling et al. (2014), dass aufgrund der Stichprobengröße (ein:e Proband:in an jedem Ort) sprecher:innenbezogene Effekte (etwa Alter, Geschlecht, Wohnmobilität) vernachlässigt wurden, weswegen sie weiterführend das Potenzial von *generalized mixed-effects models* (GAMMs)

zur gleichzeitigen Modellierung räumlicher, demographischer und personenbezogener Variation präsentieren. Dabei haben GAMMs den besonderen Vorteil, dass sie viele verschiedene Arten von (Sprach-/Atlas-)Daten modellieren und dabei komplexe Beziehungen zwischen einer Reihe möglicher Einflussfaktoren berücksichtigen können. Spezifisch können sie (a) komplexe, (nicht-)lineare Beziehungen zwischen mehreren Kovariaten und der abhängigen Variable erfassen, (b) eine Reihe an Distributionsfamilien (binomial, Beta etc.) berücksichtigen und (c) gleichzeitig mehrere Zufallseffekte etwa für Informant:innen und/oder Items zur Behandlung von Quellen idiosynkratischer Variation und damit zur Förderung der Generalisierbarkeit integrieren (vgl. Wood 2006; Wieling 2018; Winter & Wieling 2016; Wirtz & Pfenninger 2024).

Die vorliegende Studie nutzt das Potenzial von GAMMs als statistisch strinriger Methode, um geolinguistische und soziolinguistische Perspektiven zu kombinieren. In der herkömmlichen Dialektometrie wurden bzw. werden primär basis-dialektale Formen erfragt und dokumentiert. In der neueren Forschung hingegen geraten realistischere Verwendungsszenarien zunehmend in den Blick, in denen sowohl vertikale Variation als auch soziale Faktoren (potenziell) eine wichtigere Rolle spielen, und die Alltagssprache stellt insbesondere vor diesem Hintergrund einen wichtigen Anwendungsfall für die Kombination geo- und soziolinguistischer Perspektiven dar. Eine übergreifende Frage ist dabei, wie sich die alltagssprachliche Variation im deutschsprachigen Raum strukturiert. Dabei lassen sich für das Deutsche ‚Alltagssprache‘ und ‚Umgangssprache‘ unterscheiden (vgl. Möller & Elspaß 2015, 520): Während in der deutschsprachigen Dialektologie und Soziolinguistik ‚Umgangssprachen‘ in der Regel als mehr oder minder homogene und abgrenzbare Zwischenvarietäten zwischen dem (Basis-)Dialekt- und Standardpol bzw. den jeweiligen Varietäten begriffen werden, wird ‚Alltagssprache‘ im Atlas zur deutschen Alltagssprache (AdA) – welcher die Datenbasis für den vorliegenden Beitrag liefert – eher situativ definiert, d. h. als Sprachgebrauch „im sozialen und funktionalen („Nähe“-)Bereich des Privaten, des spontanen Gesprächs unter Freunden, Verwandten oder Bekannten oder im informellen Austausch unter nicht näher Bekannten aus demselben Ort, etwa im örtlichen Lebensmittelgeschäft“ (Möller & Elspaß 2014, 122). Somit geht es beim Begriff ‚Alltagssprache‘ darum, was man am betroffenen Ort in informellen Situationen im lokalen Rahmen „normalerweise hören würde – egal, ob es mehr Mundart oder Hochdeutsch ist“ (so die Aufforderung im Fragebogen des AdA). Es können also Varianten einbezogen und als Alltagssprache eingestuft werden, „die basisdialektal oder standarddeutsch sind, sofern sie dem ortsüblichen Gebrauch entsprechen“ (Möller & Elspaß 2015, 516).

Erste dialektometrische Untersuchungen hinsichtlich der räumlichen Strukturierung zur Alltagssprache im Deutschen liegen bereits vor (vgl. Pickl et al. 2019, die ausgehend von einer geostatistischen Faktorenanalyse die Gestalt und Relevanz

bekannter sprachgeographischer Strukturen überprüfen; siehe auch Pröll et al. 2021). Dabei zeichnet sich u. a. eine Nord-Süd-Teilung auf Höhe der Mainlinie ab, welche sich bereits in ersten Untersuchungen auf Basis des Wortatlas der deutschen Umgangssprachen (WDU) als wichtige sprachgeographische Grenze nachweisen ließ (vgl. Durrell 1989; Möller 2003). Pickl et al. (2019) zeigen zudem, dass sich für den Süden des deutschsprachigen Raums einerseits eine südlich-bundesdeutsche Raumstruktur und andererseits eine südöstliche, zum bairisch-österreichischen Dialektraum zugehörige Struktur beobachten lassen, ebenso wie ein gemeinsames eigenständiges hoch- und höchstalemannisches Areal, für dessen Ausdehnung die Schweizer Staatsgrenze maßgeblich ist. Interessanterweise ähneln die Verhältnisse im österreichischen Bundesland Vorarlberg – trotz alemannischer Prägung vor allem im (basis-)dialektalen Bereich – im Bereich der alltagssprachlichen Lexik eher denen des bairisch-österreichischen Raums. Pickl et al. (2019, 56) weisen bei der Zusammenfassung ihrer Ergebnisse darauf hin, dass die Alltagssprache „keine klar definierte diaphasische Varietät“ sei – vielmehr bündele sie „verschiedenartige Realisationsmöglichkeiten entlang des Standard-Dialekt-Kontinuums“ (Pickl et al. 2019, 56). Alltagssprache kann also nicht als eigene Varietät gefasst werden, sondern stellt einen primär außersprachlich definierten Ausschnitt aus der sprachlichen Wirklichkeit dar. Dabei stellt sich aber zum einen die geolinguistische Frage, inwiefern sich die Alltagssprache in unterschiedlichen Regionen in das Dialekt-Standard-Kontinuum einordnen lässt – das heißt, wo genau der alltagssprachliche Gebrauch Varianten des (Gebrauchs-)Standards beinhaltet und wo nicht (zur ausführlicheren Diskussion vom Konzept ‚Gebrauchsstandard‘ siehe Abschnitt 2.2). Zum anderen fragt sich aus soziolinguistischer Perspektive, inwiefern die räumliche Strukturierung der Alltagssprache selbst Variation unterliegt, wenn soziale, vor allem personenbezogene, Einflussgrößen wie Alter, Geschlecht und Wohnmobilität einbezogen werden.

Ziel dieses Beitrags ist es daher zum einen, empirisch festzustellen, wie Standardvarianten gegenüber anderen Varianten innerhalb der Alltagssprache des Deutschen regional verteilt sind – und zwar zunächst mit Blick auf den Bereich Aussprache –, und zum anderen, inwiefern diese Strukturierung durch soziale Prädiktoren wie Alterskohorte, Geschlecht und Wohnmobilität neben der gleichzeitigen Modellierung von Geographie als einer komplexen nicht-linearen Variable prognostiziert werden kann. Konkret gehen wir im vorliegenden Beitrag den folgenden Forschungsfragen nach:

1. Wie ist die jeweilige Wahrscheinlichkeit der Verwendung von Nonstandard- bzw. Standard-Varianten in der Alltagssprache des Deutschen regional strukturiert?
2. Welche sozialen Prädiktoren (Alterskohorte, Geschlecht, Wohnmobilität [operationalisiert als Wohndauer am Ort]) prognostizieren die Angabe von (Non-)Standard-Varianten?

## 2 Methodologie

### 2.1 Atlas zur deutschen Alltagssprache

Der AdA ist ein Online-Sprachatlas-Projekt, welches seit 2003 anhand von Crowd-sourcing-Methoden in bisher 14 Fragerunden die alltagssprachliche Variation im deutschsprachigen Raum erfasst (s. Elspaß & Möller, 2003–; vgl. auch Möller & Elspaß 2014, 2015). Dabei steht die regionale Alltagssprache im Fokus: Informant:innen werden nach lokal gebräuchlichen Varianten in Bezug auf Lexik, Aussprache, Grammatik, Phraseologie und Pragmatik oder auch nach der Üblichkeit eines gewissen Ausdrucks befragt. Die Fragebogen-Items folgen großteils einem onomasiologischen Ansatz, bei dem den Informant:innen meist ein Konzept in Form einer kurzen Beschreibung und/oder einer Abbildung präsentiert wird. Dazu wird ihnen eine Liste an möglichen zum Konzept passenden Varianten angeboten, wobei i. d. R. Mehrfachantworten möglich sind und gegebenenfalls nicht aufgeführte Varianten in ein Freifeld eingetragen werden können.

Anders als bei Erhebungen zur basisdialektalen Variation ist das Ortsnetz des AdA zu einem gewissen Grad dynamisch, d. h. das Ortsnetz ist prinzipiell von Runde zu Runde – je nach Rücklauf – veränderlich. Zudem ist der Proband:innenpool breiter gefasst als in traditionellen dialektologischen Befragungen, sodass nicht primär der Sprachgebrauch von NORMs/NORFs (d. h. *nonmobile, older, rural males/females*) erhoben wird – vielmehr zielt der AdA auf eine hohe Rückmeldequote der eher städtisch geprägten jüngeren und mittleren Generationen ab. Grundsätzlich sind die Fragerunden aber nicht an soziale Ausschlusskriterien gebunden. Wichtig in diesem Zusammenhang ist, dass Informant:innen nicht nach ihrem *eigenen* Sprachgebrauch befragt werden, sondern als Expert:innen zum *ortsüblichen* Sprachgebrauch fungieren, und zwar durch die Fragestellung (angelehnt an jene des WDU): „Bitte geben Sie bei den folgenden Fragen jeweils an, welchen Ausdruck man in Ihrer Stadt normalerweise hören würde – egal, ob es mehr Mundart oder Hochdeutsch ist.“ Zusätzlich werden Sozialdaten wie etwa Alter, Geschlecht, Wohnmobilität und Beruf durch entsprechende Zusatzfragen am Ende des Fragebogens erhoben (vgl. die aktuelle Fragerunde über die Startseite <https://www.atlas-alltagssprache.de>).

Es mag nicht unmittelbar einleuten, wieso im AdA auch Sozialdaten erhoben werden, denn die Informant:innen sollten ja Angaben zu ortsüblichen Varianten machen. Allerdings haben schon vorige Untersuchungen gezeigt, dass die Wahrnehmungen der Informant:innen an einem Ort aufgrund verschiedener sozialer Netzwerke durchaus verschieden sein können. So lassen schon Vergleiche der Antworten älterer und jüngerer Informant:innen signifikante Unterschiede erkennen,

die i. S. der *apparent-time*-Hypothese als Wandel interpretierbar sind (s. z. B. Möller & Elspaß 2015, 529–530).

## 2.2 Variablenauswahl

Für den vorliegenden Beitrag wurden zehn Variablen aus Runde 2 des AdA aus dem Bereich der Aussprache gewählt.<sup>1</sup> Begründet wird unser Fokus auf die Aussprache dadurch, dass sich lautliche – im Vergleich zu etwa lexikalischen – Varianten in vielen Fällen eindeutiger als gebrauchsstandardsprachlich einordnen und davon ausgehend mit non-standardsprachlichen Varianten kontrastieren lassen.

Bei der Kategorisierung der in der Alltagssprache vorkommenden Varianten als Standard- bzw. Nonstandard-Formen haben wir uns am Konzept des Gebrauchsstandards orientiert, um nicht lediglich Vorlesevarianten geschulter, professioneller Sprecher:innen als ‚standardsprachlich‘ zu interpretieren, sondern gerade auch regionale bzw. areale Standardvarianten zu erfassen, die im aktiven Gebrauch weit verbreitet sind, in einer größeren Region in formellen und öffentlichen Situationen als angemessen angesehen und benutzt werden und daher als Bestandteile der deutschen Standardsprache betrachtet werden müssen (vgl. Kleiner 2019, 60–61; Elspaß & Kleiner 2019, 159–160; vgl. auch AADG-Web, Konzept Gebrauchsstandard, siehe <https://prowiki.ids-mannheim.de/bin/view/AADG/KonzeptGebrauchsstandard>). Im Bereich der Aussprache kann also als Gebrauchsstandard gelten, „was in einer größeren Region den Aussprachegewohnheiten der Gebildeten in formal hochstehenden Situationen entspricht oder was in der gleichen Sprechweise in hinreichender Häufigkeit ohne spezifische regionale Verteilung vorkommt“ (König 1997, 266).

Zur Operationalisierung der gebrauchsstandardsprachlichen Geltung einer Variante wurde das *Duden-Aussprachewörterbuch* (8. Auflage) herangezogen, welches auch Ergebnisse aus dem *Atlas zur Aussprache des deutschen Gebrauchsstandards* (AADG) zur Feststellung der Standardsprachlichkeit (verschiedener) Varianten heranzieht. Zur Ermittlung gebrauchsstandardsprachlicher Formen wurde beim AADG ein klar definierter Gebrauchsbereich festgelegt, nämlich überwiegend die Vorleseaussprache von Oberstufenschüler:innen.<sup>2</sup> Einige der im Material vor-

---

1 Ursprünglich wurden zusätzliche Aussprachevariablen aus Runden 10 und 11 gewählt, allerdings konnte aufgrund der dadurch deutlich gestiegenen Anzahl an personenbezogenen Zufallseffekten (> 22.000 statt ca. 2.000) kein Modell generiert werden, auch mit 32 GB RAM und 8 CPU-Kernen. Die vorliegende Analyse ist daher auf die hier präsentierten zehn Aussprachevariablen aus Runde 2 beschränkt.

2 Unser Begriff von ‚Gebrauchsstandard‘ in der Aussprache, der sich an König (1989, 1997) sowie an den Autoren des AADG und des *Duden-Aussprachewörterbuch* orientiert, geht dabei über den

kommenden Varianten werden im AADG als schwache Formen dem Gebrauchsstandard zugerechnet. In diesen Fällen wurde als Entscheidungskriterium herangezogen, ob der im AdA-Fragebogen angegebene Beispielsatz die Verwendung einer schwachen Form erwartet lässt oder nicht. Im ersten Fall wurden die betreffenden Formen als Gebrauchsstandard eingestuft (so *sons*, *sonst* und *sin*, *sind* – Beispielsätze: *Sonst wird er wütend!* und *Sind wieder alle besetzt?*), im letzteren Fall als Non-Standard (so *au*, *auch*, das im angegebenen Satz *Das weiß ich auch nicht!* betont ist). Auch für *das* ist in diesem Satz kaum mit einer schwachen Form zu rechnen, mit der Buchstabenfolge *des* dürfte in den Antworten die regiolektale Form [de:s] (mit Langvokal) statt der schwachen Form [dəs] gemeint sein.

Tabelle 1 präsentiert die im Folgenden analysierten Aussprachevarianten. Dabei werden die als (gebrauchs-)standardsprachlich kategorisierten Varianten in der dritten Spalte und Beispiele für non-standardsprachliche Varianten in der vierten Spalte angeführt. Eine vollständige Visualisierung der Nonstandard-Varianten kann über die entsprechenden Links zu den Online-AdA-Karten eingesehen werden.

Die Teilnehmer:innen der AdA-Runde 2 hatten hinsichtlich Items 1–5 die Aufgabe, den Satz *Das weiß ich auch nicht* Wort für Wort in die alltagssprachlich üblichen Formen an ihren Orten zu übersetzen; dabei standen ihnen freilich nur lateinische, keine phonetischen Schriftzeichen zur Verfügung. Bei den übrigen Items waren verschiedene Aussprachevarianten zum Ankreuzen aufgeführt (bei Item 7 etwa *sons*, *sonst*, *sonz*, *süscht* (*suscht*, ...), *sonsch*, *sonscht*), und zudem hatten die Teilnehmer:innen die Möglichkeit, in ein Freifeld („anders: ...“) weitere Varianten einzufügen – wovon sie auch reichlich Gebrauch machten. Für die vorliegende Untersuchung wurden Varianten, die im Freifeld fünfmal oder mehr aufgetreten sind, berücksichtigt.

## 2.3 Informant:innen und Ortsnetz

Für die vorliegende Studie wurden die Antworten von insgesamt 2.103 Informant:innen analysiert. Abbildung 1 visualisiert in Form eines Alluvial-Plots die Distribution der Informant:innen im Hinblick auf die sozialen Metadaten. Im AdA wurde Alter als kategoriale Variable mit sechs Ausprägungen erfasst (10–19 Jahre:  $n = 122$ ; 20–29 Jahre:  $n = 1.368$ ; 30–39 Jahre:  $n = 326$ ; 40–49 Jahre:  $n = 147$ ; 50–59 Jahre:  $n = 102$ ; 60+ Jahre:  $n = 38$ ), Geschlecht mit zwei Ausprägungen (weiblich:  $n = 1.130$ ; männ-

---

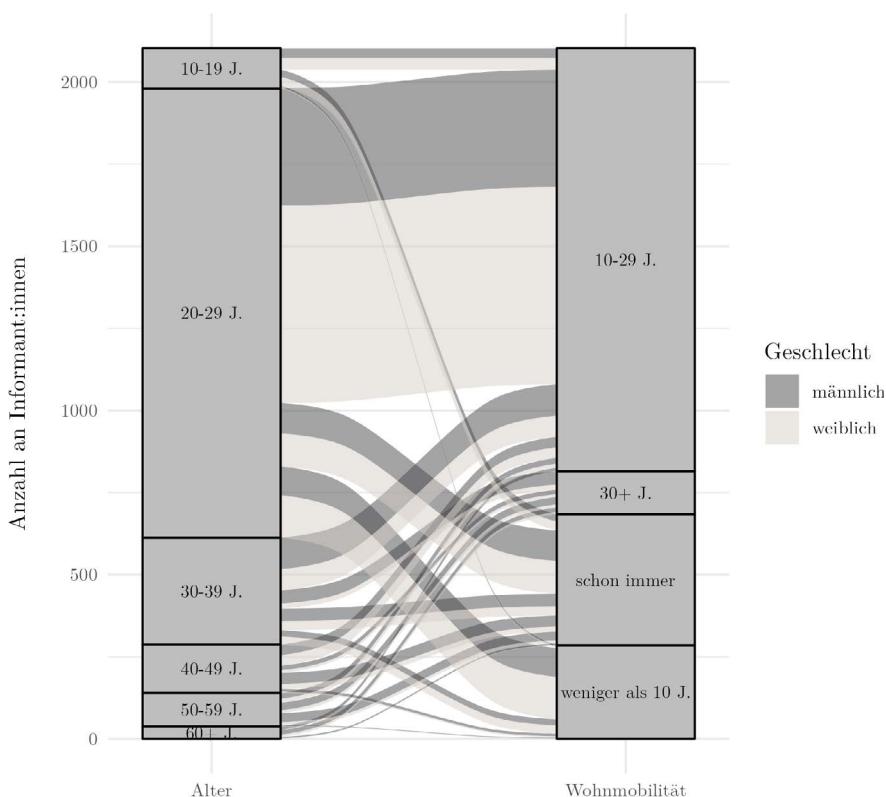
des *Variantenwörterbuchs* (2016, LXIV) hinaus, der sich auf die Aussprache „professioneller oder geübter SprecherInnen“ – wohl auch in formellen Situationen – bezieht.

**Tabelle 1:** Aussprachevarianten aus dem AdA (Runde 2)

Item	Variable (online AdA-Karte)	Gebrauchsstandard-Variante (Duden 6 [2023] Seitenzahl)	Beispiele für Non-standardvarianten
1	Aussprache von <i>das</i> ( <a href="https://www.atlas-alltagssprache.de/runde-2/f25a/">https://www.atlas-alltagssprache.de/runde-2/f25a/</a> )	<i>das</i> (S. 77)	<i>des, dat, det/dit, dis, dos, deis</i>
2	Aussprache von <i>weiß</i> in der 1. Pers. Sing. ( <a href="https://www.atlas-alltagssprache.de/runde-2/f25b/">https://www.atlas-alltagssprache.de/runde-2/f25b/</a> )	<i>weiß</i>	<i>woaß, weeß, wääß, waafß, woijß, wiss</i>
3	Aussprache von <i>ich</i> ( <a href="https://www.atlas-alltagssprache.de/runde-2/f25c/">https://www.atlas-alltagssprache.de/runde-2/f25c/</a> )	<i>ich</i> (S. 72)	<i>i, ech, esch, isch, ik, ek</i>
4	Aussprache von <i>auch</i> ( <a href="https://www.atlas-alltagssprache.de/runde-2/f25d/">https://www.atlas-alltagssprache.de/runde-2/f25d/</a> )	<i>auch</i> (S. 78)	<i>a, ooch, au, aach, oo, o(ch)</i>
5	Aussprache von <i>nicht</i> ( <a href="https://www.atlas-alltagssprache.de/runde-2/f25e/">https://www.atlas-alltagssprache.de/runde-2/f25e/</a> )	<i>nicht, nich</i> , (S. 78)	<i>net, nit, nisch, niat, nöd, ni</i>
6	Aussprache von <i>sonst</i> ( <a href="https://www.atlas-alltagssprache.de/runde-2/f23a-c/">https://www.atlas-alltagssprache.de/runde-2/f23a-c/</a> )	<i>sonst, sons, sonz</i> (S. 78)	<i>suns(t), süscht, sonsch(t), sunsch(t)</i>
7	Aussprache von <i>weg</i> ( <a href="https://www.atlas-alltagssprache.de/runde-2/f22d/">https://www.atlas-alltagssprache.de/runde-2/f22d/</a> )	<i>weg/weck</i>	<i>wech</i>
8	Aussprache von <i>sind</i> ( <a href="https://www.atlas-alltagssprache.de/runde-2/f23a-c/">https://www.atlas-alltagssprache.de/runde-2/f23a-c/</a> )	<i>sind, sin</i> (S. 78)	<i>san, sand, send, sen, si</i>
9	Vokallänge des Stammvokals von <i>Krebs</i> ( <a href="https://www.atlas-alltagssprache.de/runde-2/f22a-c/">https://www.atlas-alltagssprache.de/runde-2/f22a-c/</a> )	Vokalquantität: lang	Vokalquantität: kurz
10	Vokallänge von <i>Obst</i> ( <a href="https://www.atlas-alltagssprache.de/runde-2/f22a-c/">https://www.atlas-alltagssprache.de/runde-2/f22a-c/</a> )	Vokalquantität: lang	Vokalquantität: kurz

lich:  $n = 973$ ) und Wohnmobilität, operationalisiert durch die Wohndauer am Ort, für den die jeweiligen Antwortvarianten gelten, mit vier Ausprägungen (weniger als 10 Jahre:  $n = 285$ ; 10–19 Jahre:  $n = 1.288$ ; 30+ Jahre:  $n = 131$ ; schon immer:  $n = 399$ ).

Im Hinblick auf die Erfassung der geographischen Daten ist es wichtig anzumerken, dass bis zur fünften AdA-Runde lediglich die ersten drei Ziffern der Postleitzahl erhoben wurden. Dies schränkt zwar die Granularität in der Lokalisierung der Informant:innen ein, doch konnten immerhin insgesamt 641 einzelne Orte erfasst werden. Zusammengefasst wurden diese 641 Orte entsprechend den Karten



**Abbildung 1:** Alluvial-Plot der soziodemographischen Informationen der Informant:innen.

des AdA zu einem übergreifenden, an jenes des WDU angelehnten Ortsnetz, welches in dieser Untersuchung aus insgesamt 347 Lokalitäten besteht.<sup>3</sup> Dabei wurden nur Antworten aus Deutschland ( $n = 1.885$ ), Österreich ( $n = 107$ ) und der Schweiz ( $n = 111$ ) berücksichtigt, denn aus Ostbelgien und Südtirol lagen aus der zweiten AdA-Runde auswertbare Daten von lediglich 37 bzw. 7 Informant:innen vor.

<sup>3</sup> Wir differenzieren terminologisch zwischen ‚Ort‘ und ‚Lokalität‘. Der Begriff ‚Ort‘ versteht sich als der konkrete Einzelort (z. B. eine Gemeinde), für den die Antworten eines:r Informant:in vorliegen und wird als Fixeffekt (nämlich zur Erfassung des Einflusses von Geographie, operationalisiert als die zweidimensionale Interaktion zwischen dem Längen- und Breitengrad in Form eines *smooth terms*; siehe Abschnitt 2.4) in die GAMMs integriert. Der Begriff ‚Lokalität‘ hingegen bezieht sich auf einen abstrakten Ortspunkt, d. h. eine kleinräumige Aggregation naheliegender Orte zu einer übergreifenden räumlichen Einheit, welche sich am Ortsnetz des WDU orientiert – diese Lokalitäten werden als Zufallseffekte in die GAMMs mit einbezogen (vgl. Abschnitte 2.4 und 3.1).

## 2.4 Statistische Analyse

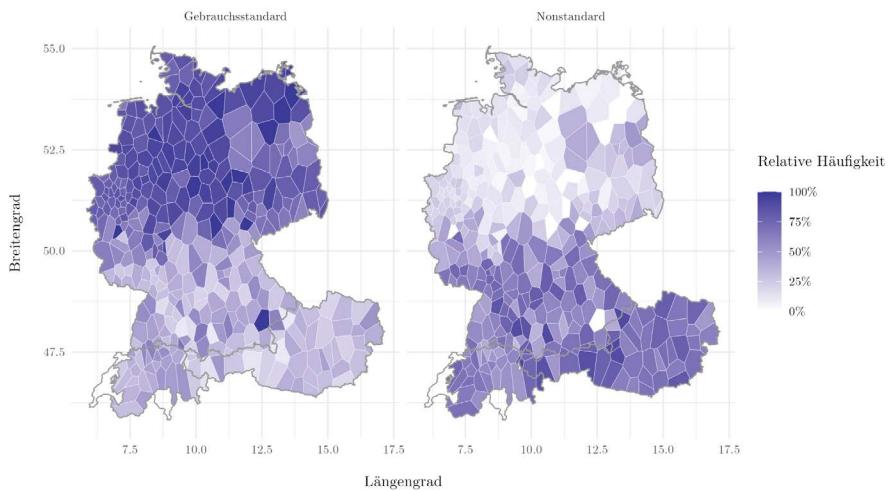
Zur Beantwortung der Forschungsfragen haben wir binär logistische GAMMs herangezogen. Im Folgenden wird in aller Kürze die statistische Prozedur skizziert. An den relevanten Textstellen werden dann die genaueren Modellspezifikationen und deren Begründungen detaillierter ausgeführt. Die Errechnung aller GAMMs erfolgte mithilfe des Pakets *mgcv* (Wood 2006) in R (R Core Team 2020), und die Ergebnisse wurden mit dem Paket *itsadug* (van Rij et al. 2020) visualisiert.

- **Modell 1** (Abschnitt 3.1): Das erste Modell erfasst die Rolle der Geographie als Prädiktor für den alltagssprachlichen Gebrauch von (Non-)Standard-Varianten. Dabei wurde ein *smooth term* vom Längen- und Breitengrad definiert, welcher die komplexe nicht-lineare Interaktion zwischen dem Längen- und Breitengrad berücksichtigt. Zufallseffekte für Informant:innen, Lokalitäten und Items wurden ebenfalls integriert.
- **Modell 2** (Abschnitt 3.2): Zur Beantwortung der zweiten und dritten Forschungsfrage hinsichtlich der Rolle von sozialen Faktoren (Alter, Geschlecht, Wohnmobilität) – bei gleichzeitigem Einbeziehen der Geographie als Einflussgröße – wurden schrittweise GAMMs mit den sozialen Faktoren als Fixeffekte errechnet. Soziale Variablen, welche nicht zur Verbesserung des Modells beitragen, wurden in weiterer Folge vernachlässigt. Modellvergleiche basieren auf dem *Akaike Information Criterion* (AIC), wobei niedrigere AIC-Werte auf ein verbessertes Modell mit einer gesteigerten (relativen) *goodness-of-fit* hinweisen.

## 3 Ergebnisse

### 3.1 Zur Modellierung der Rolle von Geographie

In einem ersten explorativen Schritt wurden die relativen Häufigkeiten von (non-)standardsprachlichen Varianten errechnet, die in Abbildung 2 dargestellt sind. Hier zeichnet sich eine klare Nord-Süd-Teilung etwa auf Höhe der Mainlinie ab, welche sich bereits in dialektometrischen Untersuchungen sowohl zu WDU- als auch AdA-Karten, vornehmlich zur Lexik, als wichtige Nord-Süd-Grenze erwiesen hat (vgl. Möller 2003; Pickl et al. 2019). In diesem Zusammenhang lässt sich nun auch für den untersuchten Bereich der Aussprache die Hypothese aufstellen, dass gebrauchs-standardsprachliche Varianten im Norden durchaus geläufiger sind, während im Süden Deutschlands, in Österreich und der Schweiz Nonstandard-Varianten den alltagssprachlichen Gebrauch dominieren.



**Abbildung 2:** Relative Häufigkeiten von (Non-)Standardvarianten.

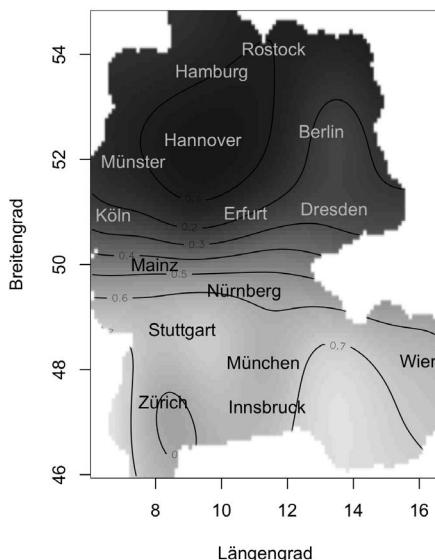
Anmerkung: Das Ortsnetz besteht aus 347 Lokalitäten, welche jeweils mit einer Voronoizelle visualisiert werden.

Während die relativen Frequenzen die räumliche Distribution von (Non-)Standard-Varianten auf den ersten Blick brauchbar skizzieren, gibt es einige Unzulänglichkeiten dieser rein deskriptiven Auswertung. Die in Abbildung 2 vorliegende Darstellung basiert auf aggregierten Frequenzwerten in einem vorgegebenen Ortsnetz, und durch die unterschiedlichen Anzahlen an Informant:innen pro Ort können insbesondere niedrige Antwortzahlen zu einer Über- oder Unterschätzung der tatsächlich in der Lokalität vorzufindenden Variation in der Alltagssprache führen. Daher bedarf es Methoden, welche etwa durch ein datengetriebenes *smoothing*-Verfahren (wie bei GAMMs) die Errechnung der Wahrscheinlichkeit des Gebrauchs einer (Non-)Standard-Variante vom vorgegebenen Ortsnetz entkoppeln (bei GAMMs: die Integration eines *smooth-terms* für die zweidimensionale Interaktion zwischen dem Längen- und Breitengrad, siehe den Text unten).

Daneben liegt aber ein weiteres grundlegendes Problem darin, dass in einer solchen deskriptiven Darstellung die Struktur der Zufallseffekte vernachlässigt wird: Erstens wurden die Daten zur Bildung des Ortsnetzes mit insgesamt 347 Lokalitäten aggregiert. Da es sich allerdings bei diesen Lokalitäten um eine verhältnismäßig kleine Stichprobe aller Gemeinden und Ortschaften im zusammenhängenden deutschsprachigen Gebiet handelt, die in die Stichprobe hätten einzbezogen werden können, gilt Lokalität als eine Quelle idiosynkratischer Variation, die als Zufallseffekt einbezogen und kontrolliert werden muss. Zweitens handelt es sich bei den Informant:innen um ein *convenience sample* und letztlich um eine

Stichprobe aus einer viel größeren Menge an möglichen Informant:innen. Außerdem liegen für jede:n Informanten:in mehrere Datenpunkte vor, da sie jeweils zu mehreren Items im AdA-Fragebogen Angaben gemacht haben; dies führt zu Verstößen gegen die Unabhängigkeitsannahme, was – wenn unberücksichtigt – große Auswirkungen auf die Typ-I-Fehlerrate (falsch positiv) einer Studie haben kann (vgl. etwa Winter 2011). Um die Unabhängigkeitsstruktur aus mehreren vorliegenden Datenpunkten pro Informant:in zu kontrollieren, müssen also auch Zufallseffekte für Informant:innen berücksichtigt werden. Drittens stammen die hier abgefragten und ausgewählten AdA-Items aus einer größeren Population möglicher Aussprachevarianten und stellen daher ebenfalls einen Zufallsfaktor dar. Denkbar ist es also freilich, dass manche Items, wie der Text unten erläutert, idiosynkratischer Variation unterliegen (vgl. auch Pickl & Rumpf 2012) – auch daher müssten mögliche Modelle die einzelnen Items als zufällige Effekte berücksichtigen, um etwaige Idiosynkrasien zu neutralisieren. M. a. W. werden Zufallseffekte als Quellen „zufälligen Rauschens“ betrachtet, welche mit gewissen Beobachtungseinheiten verknüpft sind, in unserem Falle Lokalitäten, Informant:innen und Items. Beispielsweise können verschiedene Informant:innen am selben Ort über unterschiedliche soziale Netzwerke verfügen – ihr Alltag könnte daher von eher (gebrauchs-)standardsprachlichen oder non-standardsprachlichen Varianten geprägt sein, was insgesamt die Einschätzung der Ausgangsfrage aus dem AdA beeinflussen würde, welchen Ausdruck man am jeweiligen Ort „normalerweise hören würde“ (vgl. Möller & Elspaß 2015, 520). Nachdem solche Verzerrungen errechnet und so geschätzt wurden, können verlässlichere Aussagen über die Rolle der Hauptprädiktoren (d. h. Fixeffekte wie Geographie, Alter, Geschlecht, Wohnmobilität etc.) gemacht werden.

In diesem Abschnitt verfolgen wir das Ziel, zunächst die geographische Variation in Bezug auf die Verbreitung von (Non-)Standard-Varianten im deutschsprachigen Gebiet zu skizzieren. Daher soll in einem ersten Schritt ein einfaches Modell mit Geographie als einzigem Prädiktor errechnet werden. Wie aber etwa Wieling et al. (2011, 2014) und Ko et al. (2014) bemängeln, verfügen standardmäßige lineare bzw. logistische Modelle kaum über die Flexibilität zur Modellierung der Ausbreitung von (Non-)Standard-Varianten in Abhängigkeit vom Längen- und Breitengrad – höchstens könnte ein solches Modell Längen- und Breitengrad als zwei Fixeffekte integrieren. Dabei wird aber die Regressionsfläche zwingend planar, und die multiplikative Interaktion zur Bildung einer hyperbolischen Ebene würde den Längen- und Breitengraden eine sehr begrenzte funktionale Form aufzwingen, die aber für Sprachdaten völlig ungeeignet ist. GAMMs auf der anderen Seite können „complex geographical patterns directly via a two-dimensional smooth of longitude and latitude“ modellieren (Nerbonne & Wieling 2018, 409), d. h. sie berücksichtigen das komplexe nicht-lineare Verhältnis zwischen dem Längen- und Breitengrad im Sinne



**Abbildung 3:** Contour plot der Distribution von (Non-)Standard-Varianten im deutschsprachigen Raum.

Anmerkung: Der contour plot zeigt eine Regressionsfläche zur Vorhersage der Unterschiede zwischen gebrauchsstandardsprachlichen bzw. non-standardsprachlichen Varianten als Funktion von Längen- und Breitengrad, ermittelt mit einem GAMM. Die schwarzen Linien stellen Isolinien dar und zeigen die vom GAMM geschätzten Wahrscheinlichkeiten der Angabe einer non-standardsprachlichen Variante. Eine hellere Schattierung repräsentiert eine höhere Wahrscheinlichkeit, dass eine vom Gebrauchsstandard abweichende Variante angegeben wird, und umgekehrt. Die weißen Flächen kennzeichnen Kombinationen von Längen- und Breitengraden, für die es keine (nahe gelegenen) Daten gibt.

einer dynamischen Interaktion (vgl. etwa Wieling et al. 2014; Wieling & Nerbonne 2015). Weiters können GAMMs durch die Integration von mehreren Zufallseffekten Quellen idiodynamischer Variabilität berücksichtigen (etwa Informant:in, Lokalität und Item).

Um die Distribution von (Non-)Standardgebrauch in der Alltagssprache verlässlicher als in Abb. 2 zu skizzieren, wurde ein binär logistisches GAMM mit einer nicht-linearen Interaktion zwischen dem Längen- und Breitengrad spezifiziert, welches zufällige Effekte für Informant:innen (2.104 Ebenen), Lokalitäten (347 Ebenen) und Items (10 Ebenen) integriert:

Nonstandard-Variante ~ s(Längengrad, Breitengrad) + s(Informant:in, bs = "re") + s(Lokalität, bs = "re") + s(Item, bs = "re")

Dabei prognostiziert das Modell die Wahrscheinlichkeit, mit der eine non-standardsprachliche Variante im Vergleich zu einer gebrauchsstandardsprachlichen

**Tabelle 2:** Ergebnisse der Modellierung von Geographie

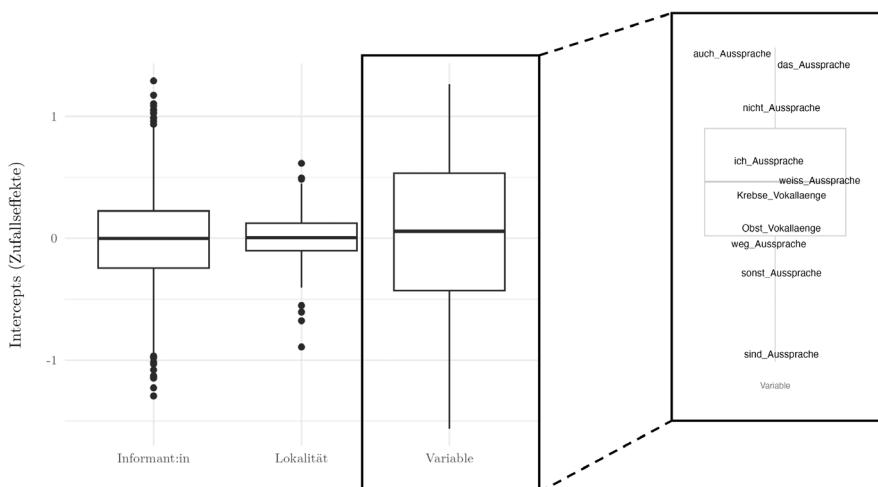
Parametric Coefficients	Estimate	Std. Error	z-Value	Pr(> t )
(Intercept)	-0,67	0,28	-2,42	0,02 *
Smooth Terms	Edf	Ref.df	F	p
s(Längengrad, Breitengrad)	21,86	23,49	1173,60	< 0,001***
s(Informant:in)	659,00	2100,00	1218,30	< 0,001***
s(AdA Lokalität)	99,91	346,00	334,20	< 0,001***
s(Item)	8,96	9,00	1707,00	< 0,001***

Notiz: *Edf* = effective degrees of freedom; *Ref.df* = reference number of degrees of freedom. \*\*\*  $p < 0,001$ .

angegeben wurde, und zwar in Abhängigkeit von der Geographie, d. h. vom Längen- und Breitengrad des Ortes, für den die Antworten eines:r jeweiligen Informant:in gelten.

Das Ergebnis zeigt der *contour plot* in Abbildung 3. Er liefert ähnliche Erkenntnisse zur deskriptiven Auswertung wie die Darstellung relativer Häufigkeiten in Abbildung 2, lässt allerdings differenziertere, vor allem aber statistisch valide Einblicke in die (Non-)Standard-Variation in der Alltagssprache zu. Unter Berücksichtigung idiodynamischer personen-, item- und lokalitätsbezogener Variation ist im nordwestlichen Teil Deutschlands die Wahrscheinlichkeit von non-standard-sprachlichen Aussprachevarianten im Alltag nur gering (circa 10%). Im Nordosten hingegen ist die Probabilität einer non-standardsprachlichen Variante leicht höher (circa 20%). Etwa auf der Höhe der Mainlinie schlägt das Bild um und die Wahrscheinlichkeit einer Nonstandard-Variante steigt rapide, sodass im Süden des deutschsprachigen Raums – besonders aber in (Alt-)Bayern, Österreich und Teilen der Schweiz – der Nonstandard mit einer prognostizierten Wahrscheinlichkeit von bis zu 70 % den alltagssprachlichen Gebrauch dominiert. Es lassen sich zudem interessante Tendenzen im hoch- und höchstalemannischen Raum beobachten, insbesondere dass die Wahrscheinlichkeiten einer non-standardsprachlichen Variante im Verhältnis zu Österreich und Bayern leicht niedriger sind, wobei die Werte im Westen wieder jenen in Österreich und Bayern ähneln.

Tabelle 2 gibt die Modellschätzungen des in Abbildung 3 präsentierten *contour plots* wieder. Wie die obige Visualisierung nahelegt, stellt sich Geographie als höchst signifikanter Prädiktor heraus, und zwar auch auf nicht-lineare Weise, wie der *edf*-Wert zeigt (*edf* = 1 entspricht einem linearen Muster, *edf* > 1 entspricht einem nicht-linearen Muster, wobei höhere Werte höhere Nichtlinearität repräsentieren). Ebenfalls erweisen sich alle Zufallseffekte als hoch signifikant. Abbildung 4 visualisiert diese und liefert Hinweise darauf, dass die mit den Zufallseffekten verbundene

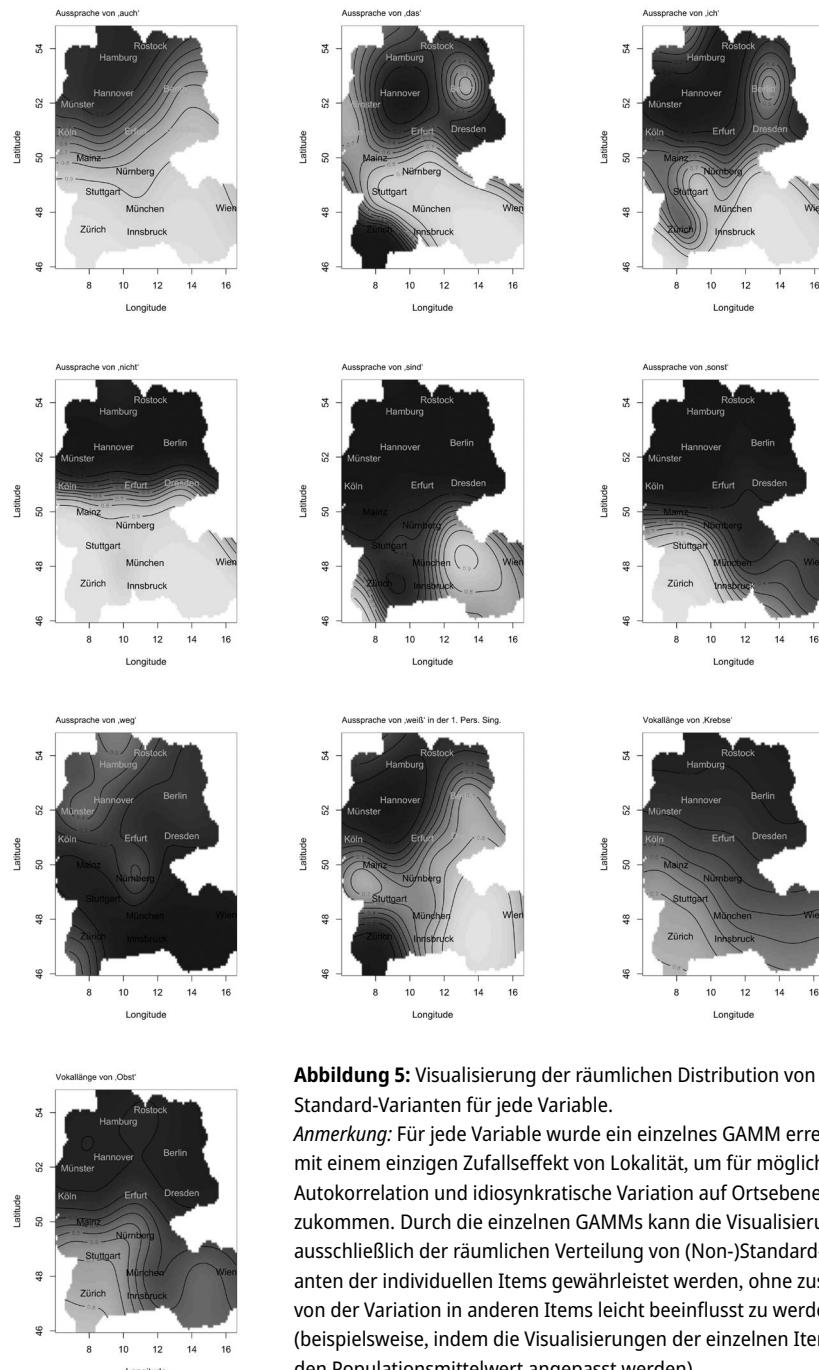


**Abbildung 4:** Boxplot zur Visualisierung der mit jedem der drei im Modell verwendeten Zufallseffekten verbundenen Variation.

*Anmerkung:* Auf der y-Achse wird der Grad der Divergenz von dem vom Modell errechneten Populationsdurchschnitt angezeigt.

strukturelle Variabilität auf Ebene der einzelnen Items – d. h. grob gesagt: Variablen – am ausgeprägtesten zu sein scheint.

Zusätzlich zur Gesamtbetrachtung der Wahrscheinlichkeit von (non-)standardsprachlichen Varianten in der Alltagssprache lohnt aber auch der Blick auf die einzelnen Items bzw. Variablen, zumal diese eine große Quelle für Varianz im vorhin präsentierten GAMM sind. Abbildung 5 visualisiert die räumliche Verteilung der individuellen Items strikt in Abhängigkeit von geographischen Faktoren (d. h. von Längen- und Breitengrad). Dabei hebt sich ein Item, die Aussprache von *sind*, deutlich vom Populationswert im Hinblick auf die Wahrscheinlichkeit einer non-standardsprachlichen Variante ab, mit einem Divergenzwert von -1,56. Das heißt, Informant:innen haben bei dieser Variable deutlich weniger wahrscheinlich eine Nonstandard-Variante von *sind* angegeben. Dieser Effekt scheint allerdings stark regional bedingt zu sein: Vor allem in Bayern und Österreich, aber auch in westlichen Teilen der Schweiz, sind non-standardsprachliche Varianten wie *san(d)* oder *si* im alltagssprachlichen Gebrauch zu weiten Teilen vorhanden (<https://www.atlas-alltagssprache.de/runde-2/f23a-c/>). Auf der anderen Seite des Spektrums zeichnen sich einzelne Items bzw. Variablen wie die Aussprache von *auch* (Divergenzwert: 1,26) und *das* (Divergenzwert: 1,04) durch hohe positive Divergenzwerte aus, welche darauf hinweisen, dass bei diesen Variablen überdurchschnittlich oft non-standardsprachliche Varianten angegeben werden. Im Falle von *auch* ist die standardsprach-



**Abbildung 5:** Visualisierung der räumlichen Distribution von (Non-)Standard-Varianten für jede Variable.

*Anmerkung:* Für jede Variable wurde ein einzelnes GAMM errechnet mit einem einzigen Zufallseffekt von Lokalität, um für mögliche Autokorrelation und idiosynkratische Variation auf Ortsebene aufzukommen. Durch die einzelnen GAMMs kann die Visualisierung ausschließlich der räumlichen Verteilung von (Non-)Standard-Varianten der individuellen Items gewährleistet werden, ohne zusätzlich von der Variation in anderen Items leicht beeinflusst zu werden (beispielsweise, indem die Visualisierungen der einzelnen Items an den Populationsmittelwert angepasst werden).

liche Variante am stärksten im Nordwesten Deutschlands vertreten, während der Großteil des deutschsprachigen Raums von Nonstandard-Varianten geprägt ist (vgl. <https://www.atlas-alltagssprache.de/runde-2/f25d/>). Ebenfalls dominieren im Westen (*dat*) bzw. auch in der südwestdeutschen und in den bairisch-österreichischen Gebieten (*des*) non-standardsprachliche Varianten den Alltagsgebrauch (vgl. <https://www.atlas-alltagssprache.de/runde-2/f25a/>).

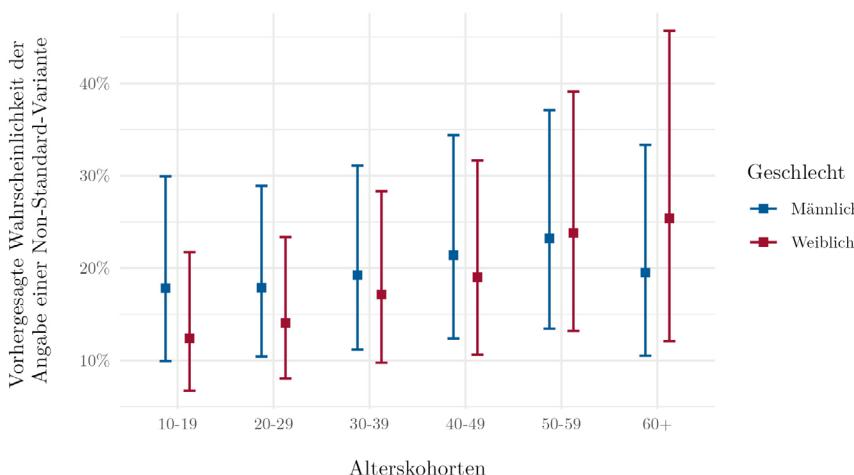
### 3.2 Zur Integration von sozialen Prädiktoren

Bislang wurden lediglich geographische Faktoren im Sinne einer dialektometrischen Analyse berücksichtigt. Aus soziolinguistischer Perspektive lässt sich allerdings die Frage stellen, ob die im vorigen Abschnitt dargestellten räumlichen Tendenzen womöglich auch von sozialen Faktoren wie Alter, Geschlecht und Wohnmobilität abhängen. Hierfür wurden zusätzliche GAMMs errechnet, und Prädiktorvariablen wurden ins Modell nur dann integriert, wenn sie zur Verbesserung des Modells auf Basis eines niedrigeren AIC-Wertes beitragen. Die Modelle mit den sozialen Prädiktorvariablen wurden dem GAMM mit Geographie als einziger Einflussgröße gegenübergestellt. Getestet wurde der Effekt von Alter (6 Ebenen), Geschlecht (2 Ebenen) und Wohnmobilität (4 Ebenen) (siehe Abschnitt 2.3) auf die Wahrscheinlichkeit einer Angabe von non-standardsprachlichen Varianten bei gleichzeitiger Modellierung der Geographie. GAMMs mit Interaktionseffekten von Geschlecht \* Alter und Geschlecht \* Wohnmobilität wurden ebenfalls errechnet und deren Erklärungspotenzial jenem des GAMMs mit Geographie als Prädiktorvariable gegenübergestellt.<sup>4</sup>

Die Interaktion zwischen Geschlecht und Alter zeichnete sich als signifikant für die Wahrscheinlichkeit der Angabe einer Nonstandard-Variante aus und trug zur Verbesserung des Gesamtmodells bei. Im Vergleich zum GAMM mit Geographie als Prädiktor (AIC = 18524,72) schneidet das Modell mit dem Interaktionseffekt zwischen Geschlecht und Alter leicht besser ab (AIC = 18492,04). Auch gegenüber einem GAMM mit allen sozialen Prädiktorvariablen inklusive Interaktionseffekten (AIC = 18497,59) weist das Modell mit dem Interaktionseffekt von Geschlecht \* Alter als soziodemographischer Einflussgröße den niedrigeren AIC-Wert auf. In einem letzten Schritt haben wir Geschlecht und Alter jeweils nach Geographie variieren lassen, um zu überprüfen, ob bzw. inwiefern sich die Effekte der jeweiligen sozialen

---

<sup>4</sup> Eine Interaktion zwischen Alter und Wohnmobilität, operationalisiert als Wohndauer am jeweiligen Ort, wurde hingegen nicht errechnet, da es zu konzeptuellen Konfundierungen in den aus der Regression resultierenden Interaktionseffekten kommen würde (z. B., eine Person im Alter zwischen 10–19 könnte nicht über 30 Jahren an einem Ort wohnhaft sein).



**Abbildung 6:** Visualisierung der konditionalen Effekte der Interaktion zwischen Geschlecht und Alter.

Variablen regional unterscheiden. Diese Modelle unterlagen jedoch höheren AIC-Werten und wurden deswegen nicht weiter berücksichtigt.<sup>5</sup>

Abbildung 6 visualisiert die konditionalen Effekte des Interaktionseffekts zwischen Alter und Geschlecht vom bestgefittenen GAMM. Daraus geht hervor, dass vor allem die unter 50-jährigen Frauen im Vergleich zu Männern dazu tendieren, gebrauchsstandardsprachliche Varianten anzugeben. Dieser Unterschied scheint sich bei der Alterskohorte 50–59 auszugleichen und darüber hinaus ins Gegenteil zu verkehren. Entgegen dem generellen Trend scheint es bei den über 60-jährigen Frauen wahrscheinlicher zu sein, dass sie non-standardsprachliche Varianten angeben, wobei hier insbesondere auf die große Varianz und die daraus resultierende Unsicherheit in den Modellschätzungen hingewiesen werden muss. Ebenfalls ist zu bedenken, dass der Interaktionseffekt nur zu einer mäßigen Verbesserung des Modells beigetragen hat. Die stärkste Einflussgröße auf die Probabilität einer non-standardsprachlichen Variante ist nämlich trotzdem die Geographie (wobei natürlich bei der Integration von zwei Dimensionen zur Modellierung der Geographie als *smooth term* und demgegenüber von zwei Ausprägungen von Geschlecht ein ganz anderes Potenzial des Effekts besteht).

<sup>5</sup> Alternativ wurden mithilfe der *compareML*-Funktion, welche Unterschiede in den fREML-Werten prüft, im Paket *itsadug* die jeweiligen Modelle miteinander verglichen. Dieses Verfahren liefert ebenfalls Hinweise dafür, dass das Modell mit dem Interaktionseffekt zwischen Geschlecht und Alter am besten abschneidet.

## 4 Diskussion

Die vorliegende Studie liefert erste Erkenntnisse zu regionalen Mustern im alltags-sprachlichen Gebrauch von (gebrauchs-)standardsprachlichen gegenüber non-standardsprachlichen Aussprachevarianten sowie zu der Frage, welche sozialen Einflussgrößen sich auf die Wahrscheinlichkeit einer Angabe von (Non-)Standard-Varianten auf Informant:innenebene auswirken. Dabei war es das übergreifende Ziel, in Anlehnung an bisherige Versuche (etwa Ko et al. 2014; Steiner, Jeszenszky & Leemann 2023; Steiner, Jeszenszky, Stebler et al. 2023; Wieling et al. 2011; Wieling et al. 2014; Wirtz et al. i. Dr.) geolinguistische und soziolinguistische Perspektiven in Einklang zu bringen, um einen umfassenderen Blick auf die räumliche und soziale Variation in der deutschen Alltagssprache zu liefern. Hierfür haben wir auf das Potenzial von binär logistischen GAMMs (*binary logistic generalized additive mixed-effects models*) zurückgegriffen. Diese können sowohl das komplexe nicht-lineare Verhältnis von Geographie durch die Modellierung einer zweidimensionalen Interaktion zwischen dem Längen- und Breitengrad als auch das Zusammenspiel von (mehreren) sozialen Variablen in Zusammenhang mit räumlichen Faktoren berücksichtigen. Unsere Studie reiht sich somit in die aktuell sich formierende Forschung zur gleichzeitigen Modellierung von geographischen und sozialen Faktoren ein (vgl. etwa Wieling et al. 2014; Nerbonne & Wieling 2018), bietet aber zudem weiterführende Erkenntnisse zur Variation der deutschen Alltagssprache zwischen Dialekt und Standard sowie ihrer räumlichen Struktur (vgl. etwa Durrell 1989; Möller 2003; Möller & Elspaß, 2008, 2014, 2015; Pickl et al. 2019).

Die erste Analyse in Abschnitt 3.1 ging der Frage nach, in welchen Gebieten im deutschsprachigen Raum (Non-)Standard-Varianten am stärksten ausgeprägt sind. Das Modell wies Geographie als signifikante Einflussgröße auf die Wahrscheinlichkeit der Angabe einer Nonstandard-Variante aus. Insbesondere bestätigte dies die alltags-sprachliche Nord-Süd-Trennung (etwa auf Höhe der Mainlinie) aus früheren Analysen (vgl. etwa Durrell 1989; Möller 2003; Pickl et al. 2019). Dabei liegt im Norden Deutschlands die Wahrscheinlichkeit von non-standardsprachlichen Aussprache-varianten bei circa 10 % bis höchstens 30 %. Überraschend waren die verhältnismäßig niedrigen Wahrscheinlichkeiten einer Nonstandard-Variante in der eigentlich dialektfesten Schweiz. Denkbar wäre in diesem Zusammenhang, dass dieses Ergebnis methodologisch bedingt sein könnte. Möller und Elspaß (2015, 519) unterstreichen, dass gerade für Schweizer Teilnehmer:innen durch die „eher am Standard orientierte phonologische Form der angegebenen Varianten [...] möglicherweise nicht immer klar“ ist, ob „die Fragen sich auf das Schweizer Standarddeutsch („Schriftdeutsch“) oder auf den Dialekt beziehen“, auch wenn die Grundfrage nach dem Gebrauch im informellen Alltag diese methodologische Hürde ausschließen müsste. Weiters ist wohl auch die diglossische Situation in der Deutschschweiz als

mögliche Erklärung für die niedrigen Wahrscheinlichkeiten der Angabe einer Non-standard-Variante etwa im Verhältnis zu den eher diaglossischen Gebieten Bayern und Österreich zu erwägen, sodass der (informelle) Alltag personenabhängig von unterschiedlichen Varietäten geprägt sein und sich unterschiedlich auf das Antwortverhalten bzw. Variantengebrauch auswirken kann (vgl. hierzu etwa Steiner, Jeszenszky & Leemann 2023). Und schließlich ist auf den Sachverhalt zu verweisen, dass in der Deutschschweiz – wie auch in den anderen Gebieten – Standard- und Dialektvarianten dieselben sein können. Die Informant:innen aus der Schweiz haben für die Variablen 1, 2, 7 und 8 fast ausschließlich die Varianten *das*, *weiss*, *we(c)k* und *sind* angegeben – mithin Formen, die im Dialekt und im Gebrauchsstandard übereinstimmen.

Eine eingehende Analyse der Zufallseffekte (Informant:in, Lokalität und Item) im ersten GAMM liefert Hinweise darauf, dass der zufällige Effekt eines Items (d. h. einer sprachlichen Variable) einer großen Spannbreite und damit großer struktureller Variabilität unterliegt. Zusätzlich zur Gesamtbetrachtung der Wahrscheinlichkeit von (Non-)Standard-Varianten in der Alltagssprache wurden die zehn individuellen Variablen einer visuellen Analyse unterzogen. Dabei zeigte sich eine große Idiosynkrasie in der Verteilung regionalsprachlicher (Non-)Standard-Varianten im zusammenhängenden deutschsprachigen Gebiet. Dies unterstreicht, dass – zunächst hier nur in Bezug auf Aussprachevarianten – diese nicht denselben räumlichen Mustern folgen. Dass auf Item-Ebene die größte idiosynkratische Varianz herrscht, ist an und für sich keine neue Erkenntnis und deckt sich beispielsweise mit der Beobachtung in Ko et al. (2014), dass die Variation auf Wortebene bei der Analyse niederländischer Dialekte am deutlichsten ausgeprägt ist (siehe dazu auch Pickl 2013, 16–20). Eine nuancierte Analyse dieser Varianz im Sinne eines ‚variant-based‘- oder ‚bottom-up‘-Ansatzes fördert allerdings einen tiefergehenden Blick auf die „variation among the spatialities of linguistic features“ (Pickl & Rumpf 2012, 212). Zudem ergibt sich aus dieser Erkenntnis ein Plädoyer dafür, dass dialektometrische Auswertungen, etwa mithilfe von Regressionsverfahren mit gemischten Effekten, verstärkt für die idiodynamische Natur einzelner Items in der Analyse aufkommen müssen (vgl. auch Winter & Grice 2021 zu Item-Varianz), statt rein über diese zu aggregieren (vgl. auch Vergeiner [2023] zu den Vor- und Nachteilen von aggregierenden vs. nicht-aggregierenden Verfahren).

In einem nächsten Schritt wurden die sozialen Daten auf Informant:innen-ebene (Alter, Geschlecht und Wohnmobilität) in die Analysen als zusätzliche Prädiktorvariablen miteinbezogen. Zur Verbesserung des Gesamtmodells trug schlussendlich der Interaktionseffekt zwischen Alter und Geschlecht bei. Dabei wurde für Männer eine höhere Wahrscheinlichkeit der Angabe einer Nonstandard-Variante prognostiziert, jedoch konnte dieser Effekt nur für die unter 50-Jährigen festgestellt werden. Grundsätzlich stimmt dieser Befund bei den unter 50-Jährigen mit

Geschlechtereffekten aus der Soziolinguistik, insbesondere aus dem angloamerikanischen Raum, hinsichtlich des variablen Gebrauchs von Nonstandard-Varianten überein (vgl. etwa Labov 1990; Cheshire 2004; Tagliamonte 2012). Im deutschsprachigen Gebiet zeichnet sich ein leicht differenzierteres Bild bezüglich geschlechtspräferierter Sprachvariation ab. Zum einen weisen Männer sich aus subjektiver Sicht eine höhere Dialektkompetenz bzw. -kenntnis als Frauen zu, wobei dieser Effekt von Region zu Region unterschiedlich ist (vgl. Mattheier 2009; Elmentaler 2012). Während die vorliegende Untersuchung Dialektkompetenz nicht erfasst hat, wäre in diesem Zusammenhang durchaus denkbar, dass sie eine kritisch beeinflussende Moderatorvariable darstellt, die sich auf die Stärke des Zusammenhangs zwischen Geschlecht und variablen Antwortmustern auswirkt. Zum anderen deuten auch objektive Daten darauf hin, dass Sprecherinnen tendenziell standardsprachliche Varianten präferieren (vgl. Twilfer 2014). Insgesamt besteht allerdings in der Soziolinguistik und Dialektologie der Konsens, dass Unterschiede im Sprachgebrauch nicht alleine durch das Geschlecht bestimmt werden – andere soziolinguistische Faktoren (etwa Alter, Bildungsgrad, Formalitätsgrad, Beruf) beeinflussen stets zusätzlich die Verteilung der Varianten (vgl. etwa Bülow et al. 2018).

Dass soziale Variablen komplex miteinander verwoben sind und sich gegenseitig beeinflussen können, zeigt sich ebenfalls in den hier vorliegenden Ergebnissen, insbesondere was den Geschlechtereffekt bei Proband:innen über 50 anbelangt. Für die Alterskohorte 50–59 zeigen die konditionalen Effekte keinen Geschlechterunterschied mehr, und bei Informant:innen über 60 wird sogar ein gegenteiliger Trend prognostiziert, nämlich dass Frauen wahrscheinlicher als Männer eine Non-standard-Variante angeben. Im Vergleich zu den jüngeren Alterskohorten sagt das Modell vorher, dass die über 60-jährigen Männer sogar mit einer leicht höheren Wahrscheinlichkeit eine gebrauchsstandardsprachliche Variante angeben. Diese Trends im Vergleich zu jenen der anderen Alterskohorten sind jedoch mit besonderer Vorsicht zu interpretieren, vor allem aufgrund der verhältnismäßig deutlich kleineren Stichprobengröße der über 60-Jährigen ( $n = 38$ ). Davon identifizieren sich lediglich zehn als weiblich (was zur großen Varianz und der damit verbundenen Unsicherheit in den Modellschätzungen beiträgt), und die Verteilung der Berufe und des damit assoziierten sozialen Status unter den Frauen (z. B. Kauffrau, Hausfrau, Psychologin) scheint etwas breiter zu sein als bei den Männern in dieser Stichprobe (z. B. Zahnarzt, Rektor, Architekt). Dieser Überhang an mutmaßlich hoch gebildeten männlichen Informanten in der höchsten der sechs Altersgruppen in der vorliegenden Stichprobe kann möglicherweise auch den hier festgestellten Effekt einer sehr hohen Wahrscheinlichkeit der Angabe einer Gebrauchsstandard-Variante bei den über 60-jährigen Männern erklären.

In künftigen Untersuchungen erscheint es daher angeraten, soziodemografische Faktoren über die hier untersuchten drei Variablen (Alter, Geschlecht, Wohn-

mobilität) hinaus zu untersuchen. Ebenfalls scheint es – zumal solche statischen sozialen Variablen nur ein gewisses Ausmaß an Varianz erklären können – vielversprechend, die Palette an zu berücksichtigenden Faktoren in soziolinguistischen und dialektologischen Ansätzen zu erweitern: Beispielsweise könnte die prädiktive Stabilität von komplexen psychosozialen Einflussgrößen wie (Sozio-)Affekt, Persönlichkeitsfaktoren (vgl. Steiner, Jeszenszky & Leemann 2023; Steiner, Jeszenszky, Stebler et al. 2023) oder auch kognitiven Ressourcen (vgl. Tamminga, MacKenzie & Embick 2016; Wirtz i. Dr.) im Vergleich zur Stärke des Einflusses von Geschlecht, Alter und Wohnmobilität auf variable alltagssprachliche Muster und deren regionale Verbreitung geprüft werden. Zudem muss – da es uns ja nur um Aussprache ging – in künftigen (etwa meta-analytischen) Ansätzen die Frage in den Blick genommen werden, ob bei verschiedenen sprachlichen Phänomenebenen (man denke etwa an Lexik oder Syntax) die Rolle von Alter und/oder Geschlecht (in Kombination mit weiteren Einfluss- bzw. Moderatorvariablen) unterschiedlich relevant sind.

Schließlich hat diese erste Untersuchung zur Verteilung von (Non-)Standard-Varianten in der alltagssprachlichen Variation auch in quantitativer Hinsicht ihre Grenzen. Zum einen basiert die quantitative Auswertung auf lediglich zehn lautlichen Varianten, welche freilich kaum aussagekräftig für den Bereich der Aussprache im Gesamten sind und auch eine eher opportunistische Auswahl darstellen als ein zufälliges Sample. Die präsentierten Ergebnisse dürfen also allenfalls als erste Beobachtung einer Tendenz interpretiert werden. Inwiefern dieselben Tendenzen bei weiteren Aussprachevariablen und vor allem bei einer größeren Anzahl an Varianten bestehen, bleibt eine offene empirische Frage. Weiters ist in ähnlicher Weise die Einschränkung gegeben, dass in Umfragen wie solchen des AdA nur jene Phänomene erfragt werden können, welche auch für Laien unmissverständlich in literarischer Umschrift wiedergegeben werden können, wodurch „Variation in der Phonologie [...] nur sehr begrenzt einbezogen werden [kann]“ (Möller und Elspaß 2015, 520). Denkbar im Rahmen künftiger Forschung wäre eine feingliedrigere Analyse und sensitivere Messung der Aussprachedistanz von gebrauchsstandard-sprachlichen Varianten, statt rein binär die Varianten zu kategorisieren. Auch ist in Betracht zu ziehen, dass diese Daten aus der zweiten AdA-Runde im Jahr 2004 erhoben wurden – die hier beobachteten Tendenzen wären also mit rezenteren Erkenntnissen aus anders gewonnenen Daten zu vergleichen (in Bezug auf Variable 1, *das*, etwa mit Leemann et al. 2019).

Es ist ebenfalls die Subjektivität in der Auswahl des Kategorisierungsverfahrens anzumerken, auch wenn die Kategorisierungsprozedur an sich möglichst objektiviert war: Wir haben uns strikt am Gebrauchsstandard orientiert, der im *Duden-Aussprachewörterbuch* angegeben ist, welches ab seiner 7. Auflage Ergebnisse aus dem AADG berücksichtigt. Aus unserer Sicht ist ein solches Vorgehen zwar sowohl methodologisch wie auch konzeptuell abgesichert, denn dadurch haben alle Variab-

len standardsprachliche Geltung, die im aktiven Gebrauch weit verbreitet sind und in einer größeren Region in formellen und öffentlichen Situationen als angemessen angesehen und benutzt werden (vgl. Elspaß & Kleiner 2019). Allerdings würden wohl andere Kategorisierungsverfahren entsprechend andere Ergebnisse liefern. Insofern sind unsere Erkenntnisse vor dem Hintergrund des von uns gewählten Kategorisierungsverfahrens zu betrachten und können daher künftig lediglich mit jenen Erkenntnissen verlässlich verglichen werden, die aus einer ähnlichen Kategorisierungsprozedur entstehen.

## 5 Fazit und Ausblick

Ziel des Beitrags war es, die regional bedingte Wahrscheinlichkeit von Gebrauchs-standard- gegenüber Nonstandard-Varianten in der Alltagssprache festzustellen und zudem den Fragen nachzugehen, welche sozialen Faktoren sich auf die Wahrscheinlichkeit der Angabe von (Non-)Standard-Varianten auswirken. Hierfür haben wir auf das Potenzial von GAMMs als einer statistisch stringenten Methode zurückgegriffen, welche die gleichzeitige Erfassung von sozialen und geographischen Einflussgrößen ermöglicht. Im vorliegenden Beitrag wurden GAMMs erstmalig zur Erfassung von sozialen und geographischen Einflüssen auf die alltagssprachliche Variation im Deutschen angewandt. Als die am stärksten ausgeprägte Einflussgröße auf die Verteilung von (Non-)Standard-Varianten erwies sich Geographie, wobei die bereits festgestellte Nord-Süd-Trennung auch hier bestätigt werden konnte. Auch zeichnen sich soziale Variablen als signifikante Korrelate von alltagssprachlicher Variation ab: Dabei geben Männer, insbesondere jene unter 50 Jahren, mit einer höheren Wahrscheinlichkeit Nonstandard-Varianten an.

Die konzeptuellen und statistisch-technischen Vorteile von GAMMs wurden bereits an mehreren Stellen erläutert (vgl. etwa Ko et al. 2014; Nerbonne & Wieling 2018; Wieling et al. 2011, 2014; Wieling & Nerbonne 2015), GAMMs finden zudem bereits in der Phonetik wie auch in der Sozio- und Psycholinguistik aufgrund ihrer Stärken zur Berücksichtigung komplexer dynamischer Beziehungen zwischen Kovariaten weitreichende Anwendung (vgl. etwa Kliesch & Pfenninger 2021; Tamminga, Ahren & Ecay 2016; Verspoor et al. 2021; Wieling 2018; Wirtz & Pfenninger 2024). Im Bereich der Dialektologie kommen GAMMs einigen kritischen Rufen nach einer stärkeren Berücksichtigung von Autokorrelation entgegen, wie etwa Möller und Elspaß (2015, 518) fordern: „Um ein Bild vom Gebrauch in einem Gebiet zu gewinnen, müssen vielmehr immer die Daten für die benachbarten Punkte mit betrachtet werden“. Durch die zweidimensionale und auch granulare Modellierung von Längen- und Breitengrad, und auch die Mitmodellierung von sowohl

komplexen Zufallseffekten als auch Fixeffekten, stellen GAMMs eine geeignete geostatistische Methode zur Generierung statistisch verlässlicher Aussagen über die (in unserem Fall: alltagssprachliche) Variation an einem Ort unter Berücksichtigung benachbarter Lokalitäten dar.

Für die Zukunft erscheint es angeraten, entsprechende Verfahren auf Basis einer umfangreicheren Datenbasis zu Aussprachevarianten durchzuführen. Die hier präsentierte statistische Prozedur lässt sich allerdings leicht auf weiteres Material aus sonstigen Dialektatlanten bzw. geolinguistischen Datensammlungen anwenden und verspricht dabei großes Potenzial für robuste, integrative Auswertungen von geographisch und sozial variationsreichen Datensätzen. Insofern verstehen wir unsere Studie vor allem als methodischen Beitrag zur integrativen Modellierung räumlicher und sozialer Variation.

**Danksagung:** Dieses Projekt wird dankenswerterweise vom Referat Kultur und Wissenschaft des *Landes Salzburg* (Referenzzahl: 20204-WIIS/262/9-2021) als Digital Humanities-Projekt gefördert.

## Literatur

- AADG = Kleiner, Stefan (2011 ff.), unter Mitarbeit von Ralf Knöbl: *Atlas zur Aussprache des deutschen Gebrauchsstandards (AADG)*. <http://prowiki.ids-mannheim.de/bin/view/AADG/>.
- Britain, David, K. Chambers, Peter Trudgill & Natalie Schilling-Estes (2004): Space and Spatial Diffusion. *The Handbook of Language Variation and Change*, 603–637. John Wiley & Sons, Ltd. doi:10.1002/9780470756591.ch24.
- Bülow, Lars, Kees de Bot & Nanna Hilton (2018): Zum Nutzen der Complex Dynamic Systems Theory (CDST) für die Erforschung von Sprachvariation und Sprachwandel. In Helen Christen, Peter Gilles & Christoph Purschke (Hrsg.), *Räume, Grenzen, Übergänge. Akten des 5. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD)*, 45–69. Stuttgart: Steiner.
- Bülow, Lars & Simone E. Pfenninger (2021): Introduction: Reconciling approaches to intra-individual variation in psycholinguistics and variationist sociolinguistics. *Linguistics Vanguard* 7(2s). 1–10. <https://doi.org/10.1515/lingvan-2020-0027>.
- Chambers, Jack K. & Peter Trudgill (1998): *Dialectology*. Cambridge: Cambridge University Press.
- Cheshire, Jenny (2004): Sex and Gender in Variationist Research. In Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (Hrsg.), *The Handbook of Language Variation and Change*, Malden, MA, Oxford, Victoria: Blackwell, 423–443.
- Duden-Aussprachewörterbuch = Kleiner, Stefan & Ralf Knöbl (2023): *Duden – Das Aussprachewörterbuch*. 8., kompl. überarb. u. erw. Auflage. Berlin: Dudenverlag.
- Durrell, Martin (1989): Die „Mainlinie“ als sprachliche Grenze. *Dialektgeographie und Dialektologie. Günter Bellmann zum 60. Geburtstag von seinen Schülern und Freunden*, 89–109. Marburg: Elwert.
- Elmentaler, Michael (2012): Dialectal concepts of space and linguistic variation. In Sandra Hansen, Christian Schwarz, Philip Stoeckle & Tobias Streck (Hrsg.), *Dialectal concepts of space and linguistic variation*, 31–47. Berlin & Boston: De Gruyter.

- Elspaß, Stephan & Stefan Kleiner (2019): Forschungsergebnisse zur arealen Variation im Standard-deutschen. In Joachim Herrgen & Jürgen Erich Schmidt (Hrsg.), *Sprache und Raum. Ein internationales Handbuch der Sprachvariation*. (Handbücher zur Sprach- und Kommunikationswissenschaft 30.4), Bd. 4: Deutsch. Unter Mitarbeit von Hanna Fischer und Brigitte Ganswindt, 159–184. Berlin & Boston: De Gruyter.
- Elspaß, Stephan & Robert Möller (2003–): Atlas zur deutschen Alltagssprache (AdA). <https://www.atlas-alltagssprache.de>.
- Goebel, Hans (1984): *Dialektometrische Studien anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. Tübingen: Max Niemeyer.
- Johnson, Daniel Ezra (2009): Getting off the GoldVerb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass* 3(1). 359–383. <https://doi.org/10.1111/j.1749-818X.2008.00108.x>
- Kleiner, Stefan (2019): Die Neuauflage des Duden-Aussprachewörterbuchs – Allgemeine Neuerungen und Umsetzung empirischer Erkenntnisse aus dem Korpus „Deutsch heute“. In: Eichinger, Ludwig M. & Albrecht Plewnia (Hrsg.): *Neues vom heutigen Deutsch. Empirisch – methodisch – theoretisch* (Jahrbuch des Instituts für Deutsche Sprache 2018). Berlin/Boston: De Gruyter, 59–80. doi:10.1515/9783110622591-004
- Kliesch, Maria & Simone E. Pfenninger (2021): Cognitive and Socioaffective Predictors of L2 Microdevelopment in Late Adulthood: A Longitudinal Intervention Study. *The Modern Language Journal* 105(1). 237–266. doi:10.1111/modl.12696.
- Ko, Vinnie, Martijn Wieling, Ernst Wit, John Nerbonne & Wim Krijnen (2014): Social, geographical, and lexical influences on Dutch dialect pronunciations. *Computational Linguistics in the Netherlands Journal* 4. 29–38. <https://clijournal.org/clijnj/article/view/38>.
- König, Werner (1989): *Atlas zur Aussprache des Schriftdeutschen in der Bundesrepublik Deutschland*. 2 Bände. Ismaning: Hueber.
- König, Werner (1997): Phonetisch-phonologische Regionalismen in der deutschen Standardsprache: Konsequenzen für den Unterricht „Deutsch als Fremdsprache“? In Gerhard Stickel (Hrsg.), *Varietäten des Deutschen: Regional- und Umgangssprachen*, 246–270. Berlin & New York: De Gruyter.
- Labov, William (1966): *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, William (1990): The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2(1). 205–254. doi:10.1017/S0954394500000338
- Leemann, Adrian, Curdin Derungs & Stephan Elspaß (2019): Analyzing linguistic variation and change using gamification web apps: the case of German-speaking Europe. *PlosOne*. 14(12): e0225399. doi:10.1371/journal.pone.0225399
- Leinonen, Therese (2010): *An acoustic analysis of vowel pronunciation in Swedish dialects* (Groningen Dissertations in Linguistics 83). Groningen Rijksuniversiteit.
- Möller, Robert (2003): Zur diatopischen Gliederung des alltagssprachlichen Wortgebrauchs. Eine dialektometrische Auswertung von Jürgen Eichhoff: Wortatlas der deutschen Umgangssprachen. *Zeitschrift für Dialektologie und Linguistik* 50(3). 259–277. <https://www.jstor.org/stable/40504953>.
- Möller, Robert & Stephan Elspaß (2008): Erhebung dialektgeographischer Daten per Internet: Ein Atlasprojekt zur deutschen Alltagssprache. In Stephan Elspaß & Werner König (Hrsg.), *Sprachgeographie digital. Die neue Generation der Sprachatlanten (mit 80 Karten)* (Germanistische Linguistik 190–191), 115–132. Hildesheim, Zürich & New York: Georg Olms.
- Möller, Robert & Stephan Elspaß (2014): Zur Erhebung und kartographischen Darstellung von Daten zur deutschen Alltagssprache online: Möglichkeiten und Grenzen. In Fabio Tosques (Hrsg.), *20 Jahre digitale Sprachgeographie*, 121–131. Berlin: Humboldt-Universität, Institut für Romanistik.

- Möller, Robert & Stephan Elspaß (2015): Atlas zur deutschen Alltagssprache (AdA). In Roland Kehrein, Alfred Lameli & Stefan Rabanus (Hrsg.), *Regionale Variation des Deutschen. Projekte und Perspektiven*, 519–540. De Gruyter.
- Nerbonne, John (2009): Data-driven dialectology. *Language and Linguistics Compass* 3(1). 175–198. <https://doi.org/10.1111/j.1749-818X.2008.00114.x>.
- Nerbonne, John, Wilbert Heeringa, Jelena Prokić & Martijn Wieling (2021): Dialectology for Computational Linguists. In Marcos Zampieri & Preslav Nakov (Hrsg.), *Similar Languages, Varieties, and Dialects. A Computational Perspective*, 96–118. Cambridge: Cambridge University Press.
- Nerbonne, John, Jelena Prokić, Martijn Wieling, Charlotte Gooskens, Gotzon Aurrekoetxea & Jose Luis Ormaetxea (2010): Some further dialectometrical steps. *Tools for Linguistic Variation, Supplements of the Anuario de Filología Vasca "Julio Urquijo"*, XIII, 41–56. Bilbao: University of the Basque Country.
- Nerbonne, John & Martijn Wieling (2018): Statistics for Aggregate Variationist Analyses. In Charles Boberg, John Nerbonne & Dominic Watt (Hrsg.), *The Handbook of Dialectology*, 400–414. Wiley-Blackwell.
- Pickl, Simon (2013): *Probabilistische Geolinguistik. Geostatistische Analysen lexikalischer Variation in Bayerisch-Schwaben* (Zeitschrift für Dialektologie und Linguistik, Beiheft 154). Stuttgart: Steiner.
- Pickl, Simon, Simon Pröll & Stephan Elspaß (2021): Zum variationslinguistischen Verhältnis von Stadt und Land. Ein Beitrag zu einer sprachlichen Raumtopologie am Beispiel Österreichs. *Linguistik online* 5(21). 293–329. <https://doi.org/10.13092/lo.110.8147>.
- Pickl, Simon, Simon Pröll, Stephan Elspaß & Robert Möller (2019): Räumliche Strukturen alltags-sprachlicher Variation in Österreich anhand von Daten des „Atlas zur deutschen Alltagssprache (AdA)“. In Lars Bülow, Ann Fischer & Kristina Herbert (Hrsg.), *Dimensions of Linguistic Space: Variation – Multilingualism Conceptualisations. Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung* (Schriften zur deutschen Sprache in Österreich 45), 39–60. Berlin et al.: Peter Lang.
- Pickl, Simon & Jonas Rumpf (2012): Dialectometric concepts of space: Towards a variant-based dialectometry. In Sandra Hansen, Christian Schwarz, Philip Stoeckle & Tobias Streck (Hrsg.), *Dialectometric concepts of space: Towards a variant-based dialectometry*, 199–214. De Gruyter. doi:10.1515/9783110229127.199.
- Pröll, Simon, Stephan Elspaß & Simon Pickl (2021): Areal microvariation in German-speaking urban areas (Ruhr Area, Berlin, and Vienna). In Arne Ziegler, Stefanie Edler & Georg Oberdorfer (Hrsg.), *Urban Matters. Current approaches in variationist sociolinguistics* (Studies in Language Variation 27), 227–252. Amsterdam & Philadelphia: John Benjamins.
- Pröll, Simon, Simon Pickl, Aaron Spettl, Volker Schmidt, Evgeny Spodarev, Stephan Elspaß & Werner König (2015): Neue Dialektometrie mit Methoden der stochastischen Bildanalyse. *Regionale Variation des Deutschen – Projekte und Perspektiven*, 173–194. Berlin: De Gruyter.
- R Core Team (2020): R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
- Rij, Jacolien van, Martijn Wieling, Harald Baayen & Hedderik van Rijn (2020): itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.
- Steiner, Carina, Péter Jeszenszky & Adrian Leemann (2023): Variation and change in Swiss German agreement morphology: Spatial, social, and attitudinal effects. *Journal of Linguistic Geography* 11(1). 8–24. doi:10.1017/jlg.2022.12.
- Steiner, Carina, Péter Jeszenszky, Viviane Stebler & Adrian Leemann (2023): Extraverted innovators and conscientious laggards? Investigating effects of personality traits on language change. *Language Variation and Change* 1–28. doi:10.1017/S0954394523000091.

- Tagliamonte, Sali (2012): *Variationist Sociolinguistics: Change, Observation, Interpretation*. Malden & Oxford: Wiley-Blackwell.
- Tamminga, Meredith, Christopher Ahern & Aaron Ecay (2016): Generalized Additive Mixed Models for intraspeaker variation. *Linguistics Vanguard* 2(s1). doi:10.1515/lingvan-2016-0030.
- Tamminga, Meredith, Laurel MacKenzie & David Embick (2016): The dynamics of variation in individuals. *Linguistic Variation* 16(2). 300–336. doi:10.1075/lv.16.2.06tam.
- Twilfer, Daniela (2014): *Sprachvariation bei Frauen und Männern. Empirische Untersuchungen zur geschlechts-präferierten Lautlichkeit in Norddeutschland*. Westfälischen Wilhelms-Universität Münster.
- Variantenwörterbuch (2016) = Ammon, Ulrich et al. (2016): *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz, Deutschland, Liechtenstein, Luxemburg, Ostbelgien und Südtirol sowie Rumänien, Namibia und Mennonitensiedlungen*. 2., völlig neu bearbeitete und erweiterte Auflage. Berlin & Boston: De Gruyter.
- Vergeiner, Philip C. (2023): Quantitative Perspectives on the Geolinguistic Structures of Dialect Morphology in Austria. *Dialectologia et Geolinguistica* 31(1). 87–115. doi:10.1515/dialect-2023-0005.
- Vergeiner, Philip C. & Stephan Elspaß (i. Dr.): Infinitival constructions in the German dialects of Austria. On variation, constraints and change of a prominent syntactic feature. *Journal of Germanic Linguistics*. doi:10.1017/S1470542724000096.
- Verspoor, Marjolijn, Wander Lowie & Martijn Wieling (2021): L2 Developmental Measures from a Dynamic Perspective. In Bert Le Bruyn & Magali Paquot (Hrsg.), *Learner Corpus Research Meets Second Language Acquisition* (Cambridge Applied Linguistics), 172–190. Cambridge: Cambridge University Press.
- WDU = Eichhoff, Jürgen (1977–2000): *Wortatlas der deutschen Umgangssprachen*. Bd. I/II: Bern: Francke [1977/78]; Bd. III: München u. a.: Saur [1993]; Bd. IV: Bern/München: Saur [2000].
- Wieling, Martijn (2018): Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70(1). 86–116. doi:10.1016/j.wocn.2018.03.002.
- Wieling, Martijn, Wilbert Heeringa & John Nerbonne (2007): An aggregate analysis of pronunciation in the Goeman-Taeldeman-Van Reenen-Project data. *Taal en Tongval* 59(1). 84–116.
- Wieling, Martijn, Simonetta Montemagni, John Nerbonne & Harald Baayen (2014): Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language* 90(3). 669–692.
- Wieling, Martijn & John Nerbonne (2015): Advances in Dialectometry. *Annual Review of Linguistics* 1(1). 243–264. doi:10.1146/annurev-linguist-030514-124930.
- Wieling, Martijn, John Nerbonne & Harald Baayen (2011): Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLOS ONE* 6(9), 1–14. <https://doi.org/10.1371/journal.pone.0023613>.
- Winter, Bodo (2011): Pseudoreplication in phonetic research. *Proceedings of the International Congress of Phonetic Science* 2137–2140.
- Winter, Bodo & Martine Grice (2021): Independence and generalizability in linguistics. *Linguistics* 59(5). 1251–1277. <https://doi.org/10.1515/ling-2019-0049>.
- Winter, Bodo & Martijn Wieling (2016): How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution* 1(1). 7–18.
- Wirtz, Mason A. (i. Dr.): *Dynamics of L2 Sociolinguistic Development in Adulthood*. Bristol: Multilingual Matters.
- Wirtz, Mason A. & Simone E. Pfenninger. (2024): Signature Dynamics of Development in Second Language Sociolinguistic Competence: Evidence From an Intensive Microlongitudinal Study. *Language Learning* 74(3). 707–743. <https://doi.org/10.1111/lang.12634>.

Wirtz, Mason A., Simon Pickl, Konstantin Niehaus, Stephan Elspaß & Robert Möller (i. Dr.): Reconciling the social and spatial: An apparent-time analysis of variation intensity in colloquial German.  
*Journal of Linguistic Geography*.

Wood, Simon (2006): *Generalized additive models: An introduction with R*. Boca Raton: CRC Press.