LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

GS L&L Graduate School Language & Literature Munich

www.symp.gwi.uni-muenchen.de

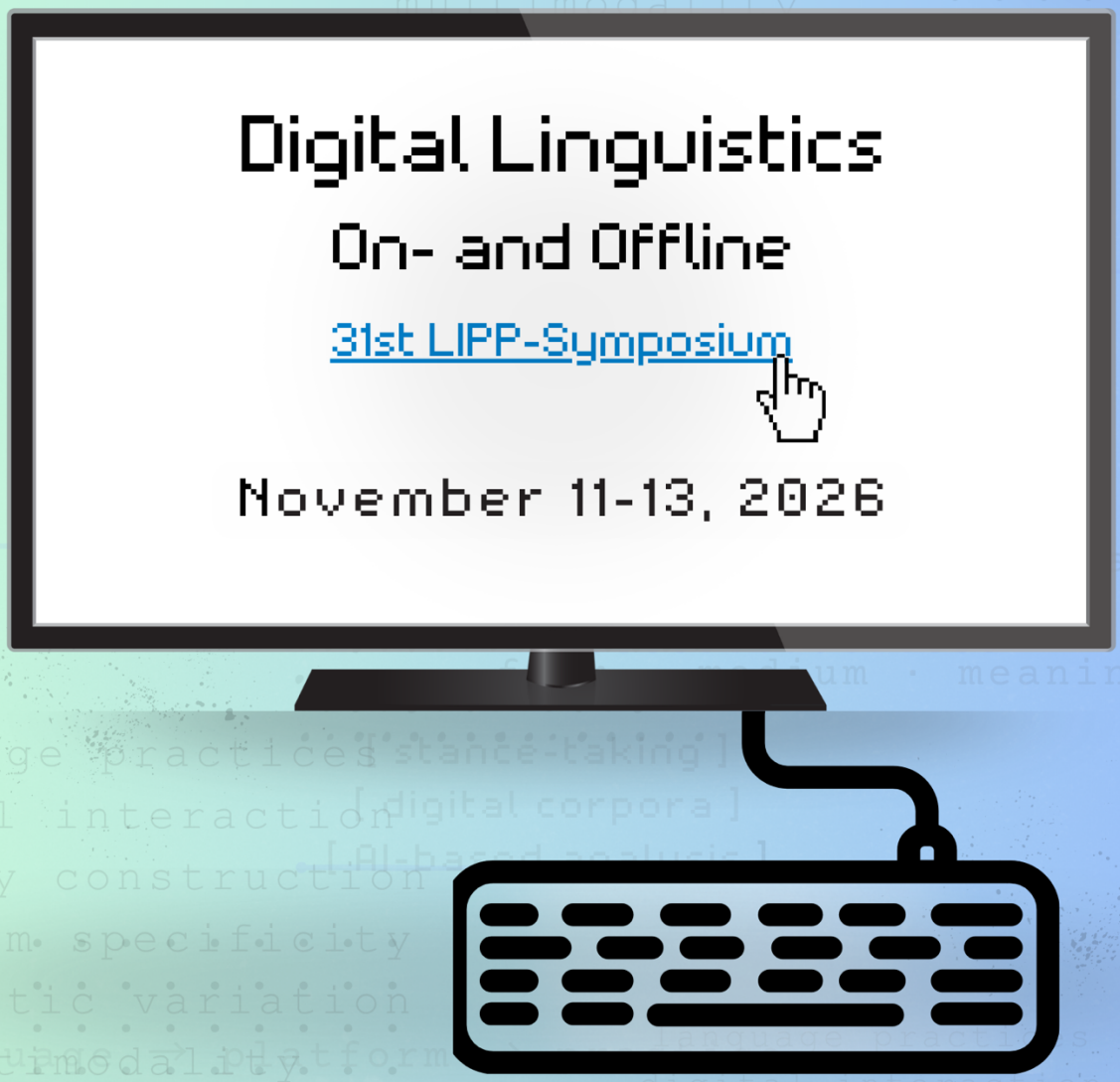
symposium@lipp.uni-muenchen.de

Digital Linguistics

On- and Offline

31st LIPP-Symposium

November 11-13, 2026



Location:

Fachbibliothek Philologicum
Ludwigstraße 25
80539 München

Keynotes:

Jannis Androutsopoulos (Hamburg)
Lars Bülow (Munich)
Janina Wildfeuer (Groningen)

SCAN ME

Organized by (in alphabetical order): Nicole Benker, Susanna Chang, Laura Hahn, Diana Lepka, Nicolas Peyrou, Yuemeng Zhu

Graphics created with Canva

31st LIPP Symposium
"Digital Linguistics On- and Offline"
November 11 – 13, 2026
Ludwig-Maximilians-Universität München

Book of Abstracts

Organizing Committee
Nicole Benker
Susanna Chang
Laura Hahn
Diana Lepka
Nicolas Peyrou
Yuemeng Zhu

Patronage
Prof. Dr. Andreas Dufter
Prof. Dr. Saskia Kersten
Prof. Dr. Claudia Maria Riehl
Prof. Dr. Johanna Wolf

Contents

Keynotes 5

Memes als metapragmatische Kommentare zum Sprachgebrauch des Piefkes in Österreich 6
Lars Bülow 6

Unpacking multimodality: Material, canvases and semiotic modes – and the specific case of short video content 8
Janina Wildfeuer 8

Scaling the nexus: How online and offline blend together, and what's in there for linguistics 9
Jannis Androutsopoulos 9

Talks 11

Eye-dialect usage for othering Spanish-speaking migrants in the USA: A digital corpus 12
Fatima Zahra El Balrhi, Jose Rolania Navarro 12

***nicht nachhaltig (?)*: Zur Funktion syntaktischer Negation beim WIDERSPRECHEN zur Aushandlung des Begriffs der Nachhaltigkeit 13**
Alexander Dübbert 13

Constructing (Alt-)European identity online: A methodological framework for social media discourse analysis 16
Xandra Knappe 16

"Got raw milk?" On dogwhistles and (covert?) social positioning in digital discourse 18
Andreas Osterroth 18

Multimedia and multimodal perspectives on the French pension reform 2023: Newspapers, YouTube, and comment sections in comparison 20
Sara Kristina Roters 20

Early detection of online child sexual grooming 23
Veronika Aures 23

Translingual (post)digital practices among newly arrived students as an interactional resource across online and offline learning settings 25
Dariia Orobchuk 25

Using online data as a resource for diachronic corpus analysis: The COSAE as a digital research tool for Contemporary Appalachian English 27
Maria Wißner 27

Cross-border lexicography for an endangered Tungusic variety: A computational approach to the dialectal classification of Aoluguya Evenki **29**
Jeremy Kwok *29*

Bulgarian web-based system for creating, visualizing and comparing associative dictionaries **32**
Velka Popova, Dimitri Popov, Krasimir Kordov *32*

What would (A)I say in my science podcast? A linguistic evaluation of corpus-based AI-recreated podcast episodes **33**
Christina Sanchez-Stockhammer, Dominic Piazza *33*

***1.000 Seiten in 1 Sekunde sichten:* Sprachgebrauch im Kontext digitaler Fach- und Werbekommunikation** **35**
Margret Mundorf *35*

Evaluating acceptability in LLMs: Evidence from human-model comparisons in German **37**
Nicholas Catasso *37*

Same prompt, different answer? On the temporal stability of large language models **39**
Olga Zagovora, Ljubiša Bojić, Ana Jovančević, Velibor Ilić *39*

Six hundred and sixty-six reasons to (not) save Te Reo Māori: Language activism and online discourse on X/Twitter **42**
Leon Grausam *42*

Digital translanguaging on the r/Morocco subreddit: A quantitative corpus study of submission **44**
Marie Flesch *44*

Using AI for replication studies on social media data in (Romance) linguistics **46**
Robert Hesselbach *46*

Cross-cultural differences in the perception of politeness strategies in conversational artificial intelligence feedback on trust, perceived anthropomorphism and cognitive appraisal **47**
Cynthia Mueller, Julia Hofweber *47*

Multimodal analysis of cold-calling in synchronous online classrooms: Evidence from Chinese Junior High EFL students **49**
Xi Wang *49*

Exploring Informal Digital Learning of English (IDLE) with digital language learning diaries **51**
Ralf Gießler *51*

Keynotes

Memes als metapragmatische Kommentare zum Sprachgebrauch des Piefkes in Österreich

Lars Bülow
(University of Munich)
Lars.buelow@germanistik.uni-muenchen.de

Internet-Memes sind nicht nur Träger sprachlicher Variation, sondern zugleich kommunikative Praktiken, in denen Sprachgebrauch explizit reflektiert, bewertet und zur sozialen Positionierung genutzt wird. Der Vortrag untersucht Formen metapragmatischen Stancetakings (vgl. Spitzmüller 2013) in Internet-Memes aus österreichischen Subreddits. Konzeptionell verbindet die Studie Ansätze der Metapragmatik und Sprachideologieforschung (vgl. z.B. Spitzmüller 2013, 2022; Silverstein 2003) mit dem Konzept der Enregistrierung (Agha 2007) und dem Stance-Modell von Du Bois (2007). Die Datengrundlage bilden 150 als metapragmatisch klassifizierte Internet-Memes aus mehreren Subreddits. Das Kategoriensystem wurde in einem kombinierten induktiv-deduktiven Verfahren entwickelt und erfasst unter anderem die thematisierten sprachlichen Ebenen (z.B. Lexik, Syntax, Orthografie etc.), evaluative Positionierungen (positiv, negativ, neutral) sowie Praktiken sozialer Stilisierung (vgl. Bucholtz & Hall 2005).

Die Ergebnisse zeigen, dass insbesondere lexikalische Formen zu metapragmatischen Stance-Objekten werden. Ausdrücke wie *Krapfen*, *Marillenknödel* oder *oida* fungieren als enregisterte Embleme regionaler und nationaler Zugehörigkeit. Besonders häufig werden Praktiken der Abgrenzung eingesetzt; zugleich spielen Authentifizierung und Illegitimierung eine zentrale Rolle. Negative Bewertungen bundesdeutschen Sprachgebrauchs überwiegen, während kontrastive Positionierungen das dominante Muster der untersuchten Memes bilden.

Die Memes bilden sprachliche Unterschiede somit nicht lediglich ab. Vielmehr wirken sie aktiv an der Herstellung und Stabilisierung sprachideologischer Grenzen zwischen Österreich, Deutschland sowie Bayern mit. Durch wiederkehrende Bewertungen und Kontrastierungen tragen sie zur Enregistrierung sprachlicher Formen als sozial erkennbare Zeichen von Zugehörigkeit, Authentizität und Differenz bei.

References

Agha, Asif. 2007. *Language and Social Relations*. Cambridge University Press. https://doi.org/10.1017/CBO9780511618284.

Bucholtz, Mary & Kira Hall. 2005. Identity and interaction: A sociocultural linguistic approach. *Discourse Studies* 7(4–5). 585–614.

Du Bois, John 2007. The stance triangle. In Robert Englebretson (ed.), *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, 139–182. John Benjamins. https://doi.org/10.1075/pbns.164.07du.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3–4). 193–229. https://doi.org/10.1016/S0271-5309(03)00013-2.

Spitzmüller, Jürgen. 2013. Metapragmatik, Indexikalität, soziale Registrierung: Zur diskursiven Konstruktion sprachideologischer Positionen. *Zeitschrift für Diskursforschung* 1(3). 263–287.

Spitzmüller, Jürgen. 2022. *Soziolinguistik: Eine Einführung*. J. B. Metzler. https://doi.org/10.1007/978-3-476-05861-4.

Unpacking multimodality: Material, canvases and semiotic modes – and the specific case of short video content

Janina Wildfeuer
(University of Groningen)
j.wildfeuer@rug.nl

Let's face it: The multimodal analysis of digital media, especially social media content, can be a daunting task. A bunch of material across complex and varied layers of meaning-making must be untangled, often without knowing which elements are actually relevant to the research question at hand. Digital linguistics might like to focus strongly on language aspects in digital contexts, but can we really ignore elements like music, animations, etc.? Where to start? When to stop? And how not to lose track of things?

This talk addresses these questions by introducing and demonstrating the canvas approach to multimodal analysis (Bateman et al. 2017, Wildfeuer et al. 2020, Bateman et al. 2026). This approach is meant to build a strong basis for any kind of multimodal analysis by carving out the material surfaces on which meaning is made in complex communicative artefacts and situations. The approach does not start with the analysis of language only, but takes into consideration all expressive forms that may contribute to the meaning-making process. It helps apply a systematic and conceptual differentiation of the different units and layers of such meaning-making.

The talk will also show how this approach directly integrates with further research methods. By applying the method to the complex case of short-form video content on Instagram, TikTok or YouTube shorts, we will see how the fine-grained analysis of the material layers, i.e. canvases of TikTok videos, directly informs the development of an annotation scheme for corpus analysis or the qualitative analysis of discourse structures in specific TikTok genres. Ultimately, the goal is to demonstrate how this method offers practical ways to navigate – and make sense of – the rich environment of contemporary digital communication.

References

Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2017. *Multimodality: Foundations, Research and Analysis – A Problem-Oriented Introduction*. de Gruyter.

Bateman, John A., Janina Wildfeuer & Tuomo Hiippala. 2026. *Multimodality: A Hands-On Guide*. de Gruyter.

Wildfeuer, Janina, John A. Bateman & Tuomo Hiippala. 2020. *Multimodalität: Grundlagen, Forschung und Analyse – Eine problemorientierte Einführung*. de Gruyter.

Scaling the nexus: How online and offline blend together, and what's in there for linguistics

Jannis Androutsopoulos
(University of Hamburg)
jannis.androutsopoulos@uni-hamburg.de

The interplay between online and offline dimensions of human action is currently discussed across disciplines, from communication studies and urban sociology to multimodal interaction analysis (e.g. de Souza e Silva et al. 2025; Hsiao et al. 2023; Olbertz-Siitonen et al. 2021). In sociolinguistics, recent work converges towards the term 'online/offline nexus' ('OON'; Blommaert 2019, Androutsopoulos 2027) and includes case studies on the enregisterment of stereotypes (Blitvich 2022), messenger interaction among families or friends (Staehr and Ag 2026), and political protest that oscillates between on- and offline spaces (Silva and Carvalho Lopez 2023). Following up on this work, my main point is the need to go beyond the study of online interaction and discourse and towards understanding how it articulates with on-site, embodied interaction. My talk unfolds in three steps. First, I discuss OON processes at the intersection of embodied action, sociality, and media affordances. Second, I propose to think of this interplay as scaled, i.e. unfolding on different orders of temporality and participation. Here I distinguish between: (a) synchronized online/offline action; (b) a stepwise interlocking on online and offline practices towards reaching a target; and (c) diffuse processes whereby online discourse leads to decentralized, memetic responses. Third, I discuss how different types of OON require different methods of research, effectively calling for new interdisciplinary alliances in linguistics and beyond (Bolander and Locher 2020).

References

Androutsopoulos, Jannis. 2027. Online-offline nexus: Genealogies and trajectories, scales and sites. In Alexandra Georgakopoulou & Tereza Spilioti (eds.), *The Routledge Handbook of Language, Digital and AI Communication.* Routledge.

Blitvich, Pilar G.-C. 2022. Karen: Stigmatized social identity and face-threat in the on/offline nexus. *Journal of Pragmatics* 188. 14-30.

Blommaert, Jan. 2019. From groups to actions and back in online-offline sociolinguistics. *Multilingua* 38(4). 485-493.

Bolander, Brook & Mirimam A. Locher. 2020. Beyond the online offline distinction: Entry points to digital discourse. *Discourse, Context and Media* 35. 100383-.

de Souza e Silva, Adriana, Scott W. Campbell & Rich Ling. 2025. Hybrid Space revisited: From concept toward theory. *Communication Theory* 35. 14-24.

Hsiao, Yuan, John Leverso & Andrew V. Papachristos. 2023. The corner, the crew, and the digital street: Multiplex networks of gang online-offline

conflict dynamics in the digital age. *American Sociological Review* 88(4). 709-741.

Olbertz-Siitonen, Margarethe, Arja Piirainen-Marsh & Marko Siitonen. 2021. Constructing co-presence through shared VR gameplay. *Journal for Media Linguistics* 4(2). 85–122.

Silva, Daniel N. & Adriana Carvalho Lopes. 2023. The seeds of Marielle Franco: Afrodiasporic agency at the online-offline nexus. *Journal of Multilingual and Multicultural Development* 47. 44-57.

Stæhr, Andreas C. & Astrid Ag. 2027. Digitally mediated interaction as part of family life. In Alexandra Georgakopoulou & Tereza Spilioti (eds.), *The Routledge Handbook of Language, Digital and AI Communication*. Routledge.

Talks

Eye-dialect usage for othering Spanish-speaking migrants in the USA: A digital corpus

Fatima Zahra El Balrhi, Jose Rolania Navarro
(University Hassan II of Casablanca, Technical University of Valencia)
fatimazahraelbalrhi@gmail.com

Eye-dialect, the use of non-standard orthography to represent perceived accents or 'broken' speech, serves as a powerful discursive mechanism for the construction of social identity and the enforcement of linguistic hierarchies. Rather than a neutral phonetic transcription, this practice functions as a strategic tool for othering, effectively delegitimizing the voices of migrants by framing them as intellectually or culturally deficient.

Through the lens of Critical Discourse Analysis (CDA), this research focuses on Spanishspeaking migrants within the United States. It argues that phonetic spellings (e.g. "ju" for "you" or "ebree" for "every") shift the reader's focus from the speaker's message to their perceived deviance from Standard American English, thereby reinforcing systemic power imbalances. To investigate these patterns, a large-scale corpus of YouTube comments was collected using the software TexMiLaB. From this large dataset, the analysis focuses on 41 specific comments that thoroughly showcase Mock Spanish and aggressive eye-dialect usage.

The findings reveal that vowel epenthesis and the addition of paragogic vowels (e.g. 'Steppo outo') are exaggerated to create a linguistic caricature that emphasizes the Otherness of the speaker. Furthermore, the data reveals a trend of Anglicized Mock Spanish where expressions like "See Lahn trow" (Cilantro) replace Spanish phonemes with familiar English morphemes, a technique that signals a refusal to engage with the actual rules of the language. These orthographic choices function as a racializing mechanism, as they contribute to the systemic dehumanization of migrant populations in contemporary American sociopolitical discourse.

nicht nachhaltig (?): Zur Funktion syntaktischer Negation beim WIDERSPRECHEN zur Aushandlung des Begriffs der Nachhaltigkeit

Alexander Dübbert
(University of Paderborn)
alexander.duebbert@uni-paderborn.de

Das Konzept der Nachhaltigkeit mit seiner langen Begriffsgeschichte (vgl. Brasch 2024) bildet in seiner politischen Bedeutung (vgl. Rödel 2013: 121–122) ein normatives Leitbild. Dafür liegt seit dem sogenannten Brundtland-Bericht aus dem Jahr 1987 eine internationale Bestimmung als „kleinste[r] gemeinsamer Nenner" (Kropp 2019: 6) vor. Auch infolge von daraus abgeleiteten Entwicklungszielen, die die Vereinten Nationen wiederholt ausgeben, wird gesellschaftlich nach wie vor kontrovers ausgehandelt, was als (nicht) nachhaltig gilt. Das bedeutet, es wird einander widersprochen, was eine nachhaltige Entwicklung ausmacht. Das WIDERSPRECHEN kann als eine kommunikative Praktik eingeordnet werden (vgl. Dübbert 2025: 60–62; Schuster 2022: 101–104), deren Vollzug durch unterschiedliche Grade der Sprachreflexivität (vgl. Busch et al. 2022: 2) gekennzeichnet ist. Eine zentrale Art des WIDERSPRECHENS besteht im Negieren mithilfe der Negationspartikel nicht als dem „mit Abstand wichtigste[n] Negationsausdruck des Deutschen" (Blühdorn 2012: 27).

Dieser Beitrag untersucht, welche Aspekte welcher Dimensionen nachhaltiger Entwicklung (vgl. Pufé 2017: 99–105) in digitaler Kommunikation besonders umkämpft sind. Dazu werden Diskussionen in der Wikipedia und Kommentare auf der Videoplattform YouTube, die jeweils als textformen- und internetbasierte Kommunikation (vgl. Beißwenger 2020: 279–280, 304) bzw. als koorientiertes Schreiben (vgl. Merten 2025: 60–88) erfasst werden können, in den Blick genommen. An diesen Beispielen wird der Frage nachgegangen, wie Artikelautor:innen der Wikipedia sowie Nutzer:innen von YouTube vor dem Hintergrund jeweiliger Affordanzen (vgl. Marx/Weidacher 2020: 105–108) die Praktik des WIDERSPRECHENS vollziehen, um den Begriff der Nachhaltigkeit weiter zu konturieren oder auch zu veruneindeutigen (vgl. Völker/Spieß 2022: 545).

Grundlage dieses Beitrags ist ein mit MAXQDA verwaltetes Korpus aus Daten des wdd24-Korpus des Instituts für Deutsche Sprache, welches alle Wikipedia-Diskussionen zu Artikeln bis 2024 beinhaltet, sowie einem selbsterstellten Subkorpus YT-N aus YouTube-Kommentaren zu Videos mit thematischem Nachhaltigkeitsbezug. Wie ein quantitativer Zugang zeigt (s. Tab. 1), wird nicht (so/nur/sehr/besonders/…) nachhaltig im politischen Sinne im thematischen Spezialkorpus YT-N häufiger gebraucht als in Referenzkorpora. Im wdd24-Korpus ist die relative Häufigkeit zwar nicht in jedem Vergleich häufiger, in absoluten Zahlen zeigt sich aber, dass dieses Ausdrucksmuster in Referenzkorpora kaum verwendet wird.

Tabelle 1: Gebrauchshäufigkeiten nicht (so/nur/sehr/besonders/...) nachhaltig (pMW = pro Millionen Wörter)

Korpus	Token gesamt	absolute Häufigkeit	relative Häufig-keit (pMW)
wdd24	478.136.775	124	0,2593
YT-N-Korpus	6.315.935	147	23,2745
DWDS-Kernkorpus 21 (2000–2010)	15.406.611	1	0,0649
DWDS-Kernkorpus (19000–1999)	121.494.241	1	0,0082

Auf dieser Grundlage wird eine qualitative Auswertung der Belegstellen angeschlossen: Indem der jeweilige Widerspruchsskopus sowie die Ausprägung der Sprachreflexivität manuell codiert werden, werden Cluster gebildet, die Aufschluss darüber geben, wie Autor:innen und Nutzer:innen zur Begriffsaushandlung Sachverhaltsdarstellungen in Bezug auf Nachhaltigkeit anfechten. Die Beispiele (1), wo mit Bezug zur sprachlichen Einheit Wort explizit, jedoch ohne metasprachliche Begründung für den Sprachgebrauch (sprachunbewusst) widersprochen wird, und (2), wo sowohl sprachunbewusst als auch ohne expliziten Bezug auf sprachliche Einheiten widersprochen wird, geben einen Einblick in das Untersuchungskorpus.

(1) *Wie ich schon andeutete: Es gibt so gut wie nichts mehr, das* ***nicht nachhaltig*** *ist, was auch immer sich jeweils hinter dem Wort verbergen soll.* (wdd24_'nicht nachhaltig', S. 62)
(2) *Was nicht wirtschaftlich ist, ist auch* ***nicht nachhaltig***. (YT-N_"Bald reißen wir ..d | Prof Rieck - Kommentare 2001 - 3000, Spalte: 2 | Zeile: 997)

References

Beißwenger, Michael. 2020. Internetbasierte Kommunikation als *Textformen-basierte Interaktion*: Ein neuer Vorschlaf zu einem alten Problem. In Konstanze Marx, Henning Lobin & Axel Schmidt (Hrsg.), *Deutsch in sozialen Medien: Interaktiv – Multimodal – Vielfältig*, 291-317. de Gruyter.

Blühdorn, Hardarik. 2012. *Negation im Deutschen: Syntax, Informationsstruktur, Semantik*. Narr.

Busch, Florian, Pepe Droste & Elisa Wessels. 2022. Sprachreflexive Praktiken: Empirische Perspektiven auf Metakommunikation. In Florian Busch, Pepe Droste & Elisa Wessels (Hrsg.), *Sprachreflexive Praktiken: Empirische Perspektiven auf Metakommunikation*, 1-19. J. B. Metzler.

Dübbert, Alexander. 2025. WIDERSPRECHEN in der Wikipedia zur Aushandlung des Begriffs der Nachhaltigkeit: Eine sprachreflexive, kommunikative Praktik. *Research on Language* 2(1). 56-72.

Kropp, Ariane. 2019. *Grundlagen der Nachhaltigen Entwicklung: Handlungsmöglichkeiten und Strategien zur Umsetzung*. Springer Fachmedien.

Marx, Constanze & Georg Weidacher. 2020. *Internetlinguistik: Ein Lehr- und Arbeitsbuch*. Narr.

Merten, Marie-Luis. 2025. *Soziale Positionen – Soziale Konstruktionen: Stancetaking im Online-Kommentieren*. de Gruyter.

Pufé, Iris. 2017. *Nachhaltigkeit*. UVK Verlagsgesellschaft.

Rödel, Michael. 2013. Die Invasion der ‚Nachhaltigkeit': Eine linguistische Analyse eines politischen und ökonomischen Modeworts. *Deutsche Sprache* 41(2). 115–141.

Schuster, Britt-Marie. 2022. Widersprechen als diskursive Praktik: Illustriert an der Widerstandskommunikation gegen den Nationalsozialismus (1933-1945). In Mark Dang-Anh, Dorothee Meer & Eva L. Wyss (Hrsg.), *Protest, Protestieren, Protestkommunikation*, 93-116. de Gruyter.

Völker, Hanna & Constanze Spieß. 2022. Veruneindeutigen. *Zeitschrift für Literaturwissenschaft und Linguistik* 52(4). 543–548.

Constructing (Alt-)European identity online: A methodological framework for social media discourse analysis

Xandra Knappe
(University of Duisburg-Essen)
xandra.knappe@uni-due.de

The construction of identity in the digital sphere has emerged as a key area of research within linguistics in recent years (cf. Bös et al. 2018, Liu 2022, Werani 2023, Hnit & Almanna 2025). Social networking services open new possibilities for self-positioning, negotiation of social belonging, and the representation of individual and collective identites. Against the backdrop of advancing digitalisation, globalisation, and increasing political communication online, this project examines how an (Alt-)European identity is discursively constructed and negotiated on the platforms X and Facebook (cf. McMahon 2022). The study quantitatively examines the distribution of language accessed by Populist Radical Right (PRR) Members of European Parliament (MEP) and Englishlanguage posts are subsequently investigated qualitatively.

The fundament of the analysis is a specifically compiled corpus that comprises both multimodal posts by MEP and their party and fraction accounts – including texts, images and videos. The technical compilation of the corpus is carried out using Python scripts, which access the respective platforms via interfaces (APIs) and automatically collect relevant data. The number of accounts examined (245), and the time span the corpus intends to cover (a total of nine months), make this methodological approach essential. This process raises not only practical but also methodological and research-ethical questions, for example regarding selection criteria, data availability, algorithmic prestructuring, and post-processing.

The retrieved data is then analysed by conducting a mixed-methods approach. Corpus linguistics provides the empirical foundation by identifying salient patterns, frequencies, and co-occurrences within the dataset. Building on this, semantic network analysis (Segev 2022) is used to model relationships between key terms. The qualitative analysis of the multimodal corpus is then conducted using the software MAXQDA. In a last step, Critical Discourse Analysis (Fairclough 2003) situates these findings within their broader social and communicative contexts, and analyses how language constructs meaning, frames issues, and reflects power relations.

This project demonstrates how to compile and process a corpus that investigates transnational dynamics and linguistic strategies of identity construction by PRR actors in European Parliament, often tied to national affiliations and language use (cf. Wodak 2007, 2021, Karner & Kopytowska 2017, Stavrakakis 2026). Additionally, while critically reflecting on prerequisites and limitations of internet-based research, impetus for future research in this field is provided.

References

Bös, Birte, Sonja Kleinke, Sandra Mollin and Nuria Hernández (eds.). 2018. *The Discursive Construction of Identities On- and Offline: Personal – Group – Collective*. John Benjamins.

Fairclough, Norman. 2003. *Analysing Discourse*. Routledge.

Hnit, Hussein & Ali Almanna. 2025. Constructing identity through narratives: Personal, social, and digital dimensions. *Social Sciences & Humanities Open* 12. 1-7.

Karner, Christian and Monika Kopytowska. 2017. *National Identity and Europe in Times of Crisis: Doing and Undoing Europe*. Emerald Publishing.

Liu, Kai. 2022. *Identity Construction Online: Identities and Ideologies in Reader Comments in Online News Fora*. Victoria University of Wellington. (Doctoral dissertation).

McMahon, Richard. 2022. Is Alt-Europe possible? Populist radical right counternarratives of European integration. *Journal of Contemporary European Studies* 30(1). 10-25.

Segev, Elad. (ed.). 2022. *Semantic Network Analysis in Social Sciences*. Routledge.

Stavrakakis, Yannis. 2026. Populism, populism research and their discontents: Strategies of interpretation from a psychosocial perspective. *Current Anthropology* 67(28). 110-119.

Werani, Anke. 2023. *Sprache und Identität: Eine Einführung*. Narr.

Wodak, Ruth. 2007. 'Doing Europe': The discursive construction of European identities. In Richard C. M. Mole (ed.), *Discursive Constructions of Identity in European Politics. Language and Globalization*, 70-94. Palgrave Macmillan.

Wodak, Ruth. 2021. Re/nationalising EU-rope: National identities, right-wing populism, and border- and body-politics. In Jürgen Barkhoff & Joep Leerssen (eds.), *National Stereotyping, Identity Politics, European Crises*, 95-121. Brill.

Wodak, Ruth. 2021. *The Politics of Fear: The Shameless Normalization of Far-Right Discourse*. SAGE.

“Got raw milk?” On dogwhistles and (covert?) social positioning in digital discourse

Andreas Osterroth
(University of Kaiserslautern-Landau)
Andreas.Osterroth@rptu.de

In the right context, the seemingly innocuous question of whether someone has “raw milk” can be interpreted as a far-right signal.[1] What is linguistically interesting is that this interpretation is accessible only to certain audiences. Dogwhistles like “raw milk”, “444”, or even emojis like “ 🧃 ” have become highly relevant in contemporary digital discourse, where political positioning is not only achieved through explicit statements but also through coded, deniable, and audience-specific forms of communication. Dogwhistling can be characterized as a form of strategically layered communication in which meanings are distributed unevenly across audiences, typically through recognizable code words (Meibauer 2025: 19). This makes them “a disturbingly important tool of covert political manipulation” (Saul 2018: 361).

This presentation examines how dogwhistles can be described linguistically in a way that captures their meaning, pragmatic function, and role in digital stance-taking. A theoretical account of dogwhistles is developed, with a particular focus on their circulation in digital and hybrid communication, drawing on examples from (social) media platforms such as TikTok, X, YouTube and 9GAG.

Dogwhistles are neither reducible to ordinary implicatures in Grice’s (1975) sense nor fully captured by lexically fixed meanings. Rather, they are best understood as communicative practices that rely on specific, conventionalized expressions enabling at least two simultaneously plausible interpretations: an unmarked reading available to the general public and an additional coded reading available only to a subgroup of recipients based on shared ideological knowledge.

To explain this, the presentation combines insights from several strands of linguistic theory. Frame semantics (Fillmore 1985) helps show how seemingly neutral lexical items can activate ideologically loaded frames for some audiences while remaining publicly innocuous for others. A pragmatic perspective inspired by Grice clarifies why dogwhistles resemble inferential phenomena (Khoo 2017: 34), yet cannot be reduced to ordinary implicatures, since the relevant inference is available only to a partial audience (Henderson & McCready 2025: 35–40). Drawing on Stalnaker (2002), it is argued that dogwhistles operate less by contributing new at-issue content than by modifying shared background assumptions. This is relevant to democratic

[1] https://www.theargus.ca/current-issue/upmwkjxhs0m11ou1i5yydfnmrw67ma; https://felixonline.co.uk/articles/slurs-hatred-and-nazi-ufos-what-really-hides-beneath-imperials-popular-agarthan-society/ (24.04.2026)

societies since “dogwhistles can poison the common ground” (Henderson & McCready 2025: 183) and change discourse significantly.

Finally, the presentation links these observations to work on social positioning and stance-taking. Witten argues that “dogwhistles are about stance-taking” (2023: 15) and that their purpose is “to signal in-group identity or affinity with an audience” (2023: 16). Dogwhistling is therefore treated as a form of covert or overt linguistic positioning in digital or hybrid discourse: it is anchored in communicative choices, depends on conventionalization within specific communities, and gains force through the unequal distribution of contextual knowledge. In doing so, the presentation contributes to current debates in digital linguistics on how meaning, ideology, and identity are negotiated across online and offline environments.

References

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6(2). 222–254.

Grice, Herbert P. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and Semantics, vol. 3: Speech Acts*, 41–58. Academic Press.

Henderson, Robert & Elin McCready. 2025. *Signaling Without Saying: The Semantics and Pragmatics of Dogwhistles*. Oxford University Press.

Khoo, Justin. 2017. Code words in political discourse. *Philosophical Topics* 45(2). 33–64.

Meibauer, Jörg. 2025. Linguistische Sicht auf Hassrede. In Julia Fuchs-Kreiß & Mustafa Temmuz Oğlakcıoğlu (eds.), *Hassrede an der Schnittstelle von Linguistik und Justiz*, 7–24. de Gruyter.

Saul, Jennifer. 2018. Dogwhistles, political manipulation, and philosophy of language. In Daniel Fogal, Daniel W. Harris & Matt Moss (eds.), *New Work on Speech Acts*, 360–383. Oxford University Press.

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25(5–6). 701–721.

Witten, Kimberly. 2023. The definition and typological model of a dogwhistle. *Manuscrito* 46(3). e20230052.

Multimedia and multimodal perspectives on the French pension reform 2023: Newspapers, YouTube, and comment sections in comparison

Sara Kristina Roters
(University of Münster)
s_rote03@uni-muenster.de

In 2023, the French government passed the pension reform by using article 49.3 of the French Constitution, allowing the Prime Minister to bypass key legislative procedures. The already contested reform consequently provoked widespread public backlash, play-ing out not only in the streets of France but also across media platforms, ranging from traditional newspapers to *YouTube*, where the topic was covered in an extensive and polarizing way.

In Romance discourse linguistics, newspaper articles have dominated corpus-based research, while social media and audience reactions remain underexplored (Eibensteiner & Hesselbach 2025). This project addresses both gaps by arguing that a multiplatform, multimodal corpus approach reveals discursive dynamics that single-platform or monomodal analyses cannot capture. The corpus includes about 4.500 newspaper articles from *Le Monde*, *Le Figaro* and *Les Echos* and 3.600 *YouTube* videos, as well as their respective comment sections (~900.000 comments). This design enables a comparison between hegemonic media discourse and vernacular public responses, foregrounding how platform affordances shape political meaning making and resistance.

Building on established and recent discourse-linguistic frameworks from German and Romance philology (Warnke & Spitzmüller 2008; Spitzmüller & Warnke 2011; Weiland 2020; Weiland 2022; Harjus 2023), the project aims to enrich the multimodal facet of linguistic dis-course analysis by integrating multimodal methods (e.g. Wildfeuer et al. 2020) to account for the interplay of linguistic, visual and auditory resources in the data. This dimension has frequently been neglected by Romance discourse linguistics studies. While the quantitative verbal analysis is conducted using Sketch Engine (Kilgarriff et al. 2014), MAXQDA (VERBI GmbH 2025) will be used for the multimodal qualitative analysis.

Preliminary findings suggest that discursive construction differs not only between platforms but also across newspapers, with platform- and outlet-specific affordances emerging as key factors shaping both the multimodal resources used and the nature of audience responses. Pilot studies revealed the politicization of urban space, where lex-emes like Paris or rue are re-semanticized as arenas of political debate and resistance through the interplay of verbal and visual resources as well as discursive creation of spaces as markers of in- and outgroup dynamics (Roters forthcoming; Van Dijk 2006). There is also a general tendency among newspapers for reverse remediation, showing that platform-specific affordances and moderation practices of

comment sections (constraints) systematically shape the linguistic complexity and emotionality of the dis-course, ranging from vernacular outbursts on *YouTube* to highly regulated argumentative exchanges in newspaper comment sections (Roters forthcoming). This contribution focuses on the multimodal construction of political discourse across media types and its role in shaping these discursive differences.

References

Eibensteiner, Lukas & Robert Hesselbach (eds.). 2025. *Social Media: Current Issues in Romance Linguistics and Foreign Language Education*. AVM.

Harjus, Jannis. 2023. *Kontrastive romanistische Diskurslinguistik: Multimediale Rivakitätskonstruktionen in portugiesischen, spanischen, katalanischen und französischen Sportzeitungen*. AVM.

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

Roters, Sara K. forthcoming. "Paris la poubelle ville du monde!": *Urban Linguistics* und die Repräsentation des Urbanen im Diskurs um die französische Rentenreform. In Lidia Becker, Julia Kuhn, Christina Ossenkop & Claudia Polzin-Haumann (eds.), *(Mediale) Sprachliche Räume und Raumkonstruktion: Konzepte, Methoden, Perspektiven in der romanistischen Linguistik*. Narr.

Roters, Sara K. forthcoming. "Oui c'est la nouvelle caisse de retraite spéciale macron 💩: le cercueil…": Ein Vergleich der Kommentarspalten zur französischen Rentenreform 2023 auf YouTube und in Online-Zeitungen. In Felix Fahnroth, Philipp Förster & Alexandra Funk (eds.), *ImPuls der Romania: Beiträge zum 37. Forum Junge Romanistik in Jena (26.-28.03.2025)*. AVM.

Spitzmüller, Jürgen & Ingo H. Warnke. 2011. *Diskurslinguistik: Eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse*. de Gruyter.

Van Dijk, Teun A. 2006. Ideology and discourse analysis. *Journal of Political Ideologies* 11(2). 115-140.

VERBI GmbH. 2025. MAXQDA. https://www.maxqda.com/products/maxqda.

Warnke, Ingo H. & Jürgen Spitzmüller (eds.). 2008. *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*. de Gruyter.

Weiland, Verena. 2020. *Sprachwissenschaftliche Zugriffe auf Diskurse: Ein korpuslinguistischer Ansatz am Beispiel des Themas "Sicherheit und Überwachung" in Frankreich*. Universitätsverlag Winter.

Weiland, Verena. 2022. Diskurslinguistik in der deutschsprachigen Romanistik: Wege und Umwege. *tekst i dyskurs – text und diskurs* 16. 163-170.

Wildfeuer, Janina, John Bateman & Tuomo Hiippala. 2020. *Multimodalität: Grundlagen, Forschung und Analyse – Eine problemorientierte Einführung*. de Gruyter.

Early detection of online child sexual grooming

Veronika Aures
(University of Munich)
veronika.aures@campus.lmu.de

Online grooming in gaming environments (e.g. Faraz et al. 2022) unfolds through subtle, evolving linguistic patterns that conventional detection systems often struggle to recognize. The multi-turn nature of these interactions (see An et al. 2025) challenges models that lack contextual awareness, leading to delayed or missed risk identification (for machine-learning approaches see Leiva-Bianchi et al. 2025; An et al. 2026). Drawing on broader early-detection research (see Chehbouni et al. 2025; Kim et al. 2025; originally Vogt et al. 2021), this work advances *context-aware* AI methods for identifying grooming behaviors in online gaming systems (e.g. Faraz et al. 2024).

Using multilingual gaming interactions spanning multiplayer and dyadic settings, we construct message-level and context-level representations to support supervised grooming detection within a human-in-the-loop safety framework. We compare three modeling approaches: (i) a single-turn, hierarchical MLP that predicts binary grooming status and fine-grained stages; (ii) a context-conditioned MLP that incorporates conversational context; and (iii) a sequence-based MiniLM model fine-tuned for next-stage classification using encoded stage histories together with the current message. Each model's stage predictions provide structured signals that feed into downstream risk scoring and trajectory forecasting.

The stage-aware, sequence-based MiniLM achieves the strongest performance (multi-class macro F1 = 0.91, binary F1 = 0.96, Brier ≈ 0.014), demonstrating strong calibration and the benefit of transformer-based contextualization. The single-turn MLP, by contrast, performs markedly worse due to the absence of contextual cues (macro F1 = 0.42, binary F1 = 0.50, Brier ≈ 0.104). The context-conditioned MLP substantially improves upon the single-turn baseline (macro F1 = 0.87, binary F1 = 0.93, Brier ≈ 0.022) but still falls short of the MiniLM model's calibration, stability, and stage-level discrimination. Ultimately, both contextual variants provide a reliable foundation for downstream risk scoring and trajectory forecasting by offering temporally coherent, intervention-relevant stage signals.

These findings support earlier and more reliable identification of text-based harms, enabling timely human-supervised intervention before grooming escalates. By improving contextual detection of high-risk interactions, this work contributes to the prevention of online child sexual abuse and advances proactive safety measures in digital environments for children and teens.

References

An, JinMyeong, Sangwon Ryu, Heejin Do, Yunsu Kim, Jungseul Ok & Gary Lee. 2025. Revisiting early detection of sexual predators via turn-level optimization. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 4713-4724. Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.naacl-long.241.

An, Heajun, Marcos Silva, Qi Zhang, Arav Singh, Minqian Liu, Xinyi Zhang, Sarvech Qadir, Sang Won Lee, Lifu Huang, Pamela J. Wisniewski & Jin-Hee Cho. 2026. Toward integrated solutions: A systematic interdisciplinary review of cybergrooming research. arXiv. https://doi.org/10.48550/arXiv.2503.05727.

Chehbouni, Khaoula, Martine De Cock, Gilles Caporossi, Afaf Taik, Reihaneh Rabbany & Golnoosh Farnadi. 2025. Enhancing privacy in the early detection of sexual predators through federated learning and differential privacy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(27): 27887–27895. https://doi.org/10.1609/aaai.v39i27.35005.

Faraz, Anum, Fardin Ahsan, Jinane Mounsef, Ioannis Karamitsos & Andreas Kanavos. 2024. Enhancing child safety in online gaming: The development and application of Protectbot, an AI-powered chatbot framework. *Information* 15(4): 233. https://doi.org/10.3390/info15040233.

Faraz, Anum, Jinane Mounsef, Ali Raza & Sandra Willis. 2022. Child safety and protection in the online gaming ecosystem. *IEEE Access* 10: 115895–115913. https://doi.org/10.1109/ACCESS.2022.3218415.

Kim, Dohyeon, Taehoon Kim & Jihoon Yang. 2025. Early detection of online grooming with language models. In *Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing*, 963–970. ACM. https://doi.org/10.1145/3672608.3707796.

Leiva-Bianchi, Marcelo, Nicolas Castillo, César A. Astudillo & Francisco Ahumada-Méndez. 2025. Effectiveness of machine learning methods in detecting grooming: A systematic meta-analytic review. *Scientific Reports* 15(1): 9008. https://doi.org/10.1038/s41598-024-83003-4.

Vogt, Matthias, Ulf Leser & Alan Akbik. 2021. Early detection of sexual predators in chats. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 4985–4999. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.386.

Translingual (post)digital practices among newly arrived students as an interactional resource across online and offline learning settings

Dariia Orobchuk
(University of Hildesheim)
orobchuk@uni-hildesheim.de

The paper examines the (post)digital linguistic practices of newly arrived students within the context of German language acquisition, exploring how multilingualism, multimodality, and the blurring of online and offline worlds manifest in their everyday communication and learning practices. The starting point is the observation that, while the digitally shaped lifeworlds of this group have only been examined selectively and predominantly from an institutional-didactic perspective within German as a Second Language (DaZ) research, their lifeworld linguistic practices remain under-researched as a subject of digital linguistics. Theoretically, the paper builds on work regarding digital langauge practices (Androutsopoulos & Stæhr 2018), postdigital multilingualism (Höfler 2025), and translanguaging (García & Wei 2014), conceptualizing digital spaces as multimodal, multilingual, and cross-contextual spaces of action where verbal, visual, auditory, and interactive layers interconnect.

The data basis consists of semi-structured interviews with six newly arrived students from a secondary school language support class (Sekundarstufe I), their parents, and their teacher. The interview data was evaluated using qualitative content analysis (Mayring 2022). Additionally, ten classroom videos were analyzed using interaction analysis (Markee & Kunitz 2015). In doing so, the article combines reflexive reconstructions of digital practices with the analysis of situated classroom communication involving digital media.

The interview analyses demonstrate that the students do not merely "consume" digital media in their daily lives, but actively use it for linguistic orientation, participation, and self-positioning: via YouTube, TikTok, messengers, streaming, language learning apps, translation tools, and digital gaming worlds. These practices give rise to translanguaging spaces (Wei 2018), where German is combined with Ukrainian, Russian, Arabic, and English, as well as with subtitles, interfaces, comments, and in-game actions. This highlights both the potential for informal learning and self-efficacy, as well as limitations such as linguistic overload, platform-specific restrictions, and unequal access.

The analysis of the classroom videography clarifies that the use of digital technologies in language support classes enables new forms of interaction; however, these must be competently prepared and framed linguistically by the teacher. The students participate actively and support one another, fostering positive group dynamics. At the same time, the findings indicate that the role of the teacher in such sequences partially shifts from a primarily instructive role toward a more supportive and moderating function.

References

Androutsopoulos, Jannis & Andreas Stæhr. 2018. Moving methods online: Researching digital language practices. In Angela Creese & Adrian Blackledge (eds.), *The Routledge Handbook of Language and Superdiversity*, 118–132. Routledge.

García, Ofelia & Li Wei. 2014. *Translanguaging: Language, Bilingualism and Education*. Palgrave Macmillan. https://doi.org/10.1057/9781137385765.

Höfler, Elke. 2025. Postdigitale Mehrsprachigkeit: Versuch einer Legitimation und Definition. In Günther Pallaver, Klaus Rummler, Petra Missomelius, Valentin Dander & Oliver Leistert (eds.), *Streifzüge an den Nahtstellen von Medien, Bildung und Philosophie*, 201–221. MedienWissenBildung. https://doi.org/10.21240/978-3-03978-154-6_10.

Markee, Numa & Silvia Kunitz. 2015. CA-for-SLA studies of classroom interaction: Quo vadis? In Numa Markee (ed.), *The Handbook of Classroom Discourse and Interaction*, 425–440. Wiley-Blackwell.

Mayring, Philipp. 2022. *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz Verlagsgruppe.

Wei, Li. 2018. Translanguaging as a practical theory of language. *Applied Linguistics* 39(1). 9–30. https://doi.org/10.1093/applin/amx039.

Using online data as a resource for diachronic corpus analysis: The COSAE as a digital research tool for Contemporary Appalachian English

Maria Wißner

(University of Munich)

Maria.Wissner@campus.lmu.de

The Corpus of Spoken Appalachian English (COSAE) (Version est. 2023) is a 100,000word digital corpus designed to document contemporary patterns in Appalachian English through online, user-generated media. Building on the historical and sociolinguistic foundations of Appalachian English research (e.g., Wolfram & Christian 1976; Montgomery 2009), COSAE offers a systematic and replicable method for studying dialect speech beyond traditional fieldwork and oral-history archives. Its methodological innovation lies in the systematic incorporation of TikTok and YouTube videos as primary linguistic data, aligning with recent work on digital discourse and platform-mediated communication (cf. Androutsopoulos 2013; cf. Bucher & Helmond 2018). Compiled from transcripts of videos published between 2020 and 2023, the corpus documents naturally occurring speech from 27 creators across Tennessee, Kentucky, North Carolina, and Virginia. This provides access to younger, mobile, and digitally embedded speakers who remain underrepresented in legacy corpora such as the AAPCAppE and whose linguistic practices reflect both local identity and global media participation. COSAE's design—manual verification of auto-captions, detailed speaker metadata, and annotation of phonological and semantic constraints—supports finegrained quantitative analysis of features such as the a-prefix, whose distribution is shaped by well-established syntactic and prosodic conditions (cf. Wolfram 1976; Feagin 1979). Crucially, COSAE captures speech embedded in multimodal performance: creators speak while editing, gesturing, storytelling, and responding to algorithmic incentives, situating linguistic behavior at the interface of dialect, identity, and platform affordances. The corpus thus reveals how Appalachian English is negotiated in digital spaces where authenticity, visibility, and regional belonging are co-constructed. By foregrounding online creators as legitimate data sources, COSAE expands the empirical foundations of variationist research and demonstrates how digital-era corpora can illuminate processes of dialect change, levelling, and persistence. It offers a replicable model for studying regional varieties within contemporary, platform-shaped communicative spaces. An updated and extended version of COSAE is currently being developed.

References

Androutsopoulos, Jannis. 2013. Code-switching in computer-mediated communication. In Susan Herring, Dieter Stein & Tuija Virtanen (eds.),

Pragmatics of Computer-Mediated Communication, 667-694. de Gruyter Mouton.

Bucher, Taina & Anne Helmond. 2018. The affordances of social media platforms. In Jean Burgess, Alice Marwick & Thomas Poell (eds.), *The SAGE Handbook of Social Media*, 233-53. SAGE.

Feagin, Crawford. 1979. *Variation and Change in Alabama English: A Sociolinguistic Study of the White Community*. Georgetown University Press.

Montgomery, Michael. 2009. *The Dialect of the Smoky Mountains: A Linguistic Analysis*. University of Tennessee Press.

Wißner, Maria E. 2023. *Corpus of Spoken Appalachian English (COSAE), 2020-2023*. Unpublished.

Wißner, Maria E. 2023. *The Verb Phrase in Appalachian English. A Study Based on Spoken Data*. Justus-Liebig-Universität Gießen.

Wißner, Maria E. 2025. *The Development, Diffusion or Death of Appalachian English in the 21st Century: A Diachronic Investigation of the A-Prefix and its Meaning for Dialect Change*. Justus-Liebig-Universität Gießen.

Wolfram, Walt. 1976. The linguistic variables in Appalachian English. *American Speech* 51(1-2). 45–56.

Wolfram, Walt & Donna Christian. 1976. *Appalachian Speech*. Center for Applied Linguistics.

Cross-border lexicography for an endangered Tungusic variety: A computational approach to the dialectal classification of Aoluguya Evenki

Jeremy Kwok
(University of Greenwich)
C.Kwok@greenwich.ac.uk

The Evenki (Ewenki, Tungus) belong to the Ewenic group of the Northern Tungusic branch of the Tungusic language family (Janhunen 2023). In China, the state recognises a single "Ewenke" nationality that administratively subsumes three linguistically distinct subgroups—Solon, Khamnigan, and the Aoluguya or "Yakut" Evenki—even though speakers report substantial difficulty in mutual comprehension across these varieties (Baek 2025). The Aoluguya are a small community of reindeer herders whose ancestors migrated from the Lake Baikal region to the Greater Khingan Mountains in the mid-seventeenth century (Kolås & Xie 2015); their variety is conventionally grouped with Siberian Evenki rather than with the numerically dominant Solon, yet this affiliation rests on impressionistic observation rather than systematic lexical comparison. Because linguistic diversity within officially recognised *minzu* categories is routinely under-represented in Chinese ethnic classification (Mullaney 2011), and because Aoluguya Evenki is critically endangered—with an estimated 30 fluent speakers remaining (Bradley 2007)—clarifying its dialectal position is at once a theoretical and a documentary priority.

This paper presents a digital-linguistic workflow for the comparative analysis of Aoluguya Reindeer Evenki. The immediate empirical goal is to clarify the dialectal position of Aoluguya Evenki relative to Siberian Evenki varieties documented in Russia; the broader methodological goal is to show how heterogeneous language resources collected on- and offline can be transformed into a unified comparative dataset for endangered-language research.

The study builds a cross-border lexical database from five currently integrated sources: a large online phonetic dictionary with speaker-linked Siberian Evenki entries, an annotated university speech corpus, a Russian–Evenki reference dictionary, a manually transcribed Aoluguya fieldwork monograph from the Records of China's Endangered Languages series, and a thematic lexicon with voiced materials. These sources differ substantially in transcription system, metadata structure, elicitation format, and language of glossing. To make them comparable, the paper develops a reproducible processing pipeline that combines automated extraction, manual verification, repetition segmentation for elicited tokens, rule-based transcription harmonisation, and semantic alignment via English/Russian translation pivots.

Methodologically, the paper speaks directly to digital linguistics in two senses. First, it addresses digital methods for working with offline documentation: scanned monographs, QR-linked audio, and legacy elicitation

materials are converted into machine-readable lexical data. Second, it addresses online linguistic infrastructure by integrating born-digital dictionaries, corpora, and audio resources across national and script boundaries. The resulting dataset enables systematic string-level comparison of IPA-like forms and the calculation of preliminary dialect-distance measures across named Siberian sub-dialects.

Preliminary results suggest that Aoluguya Evenki aligns more closely with the eastern Siberian Evenki group than with the administratively associated Solon variety in China. This finding is consistent with historical migration accounts (Kolås & Xie 2015), but the main contribution of the paper is methodological: it demonstrates that digital corpus integration can generate new comparative evidence even where documentation is fragmented, multilingual, and unevenly structured. The paper therefore contributes a reusable workflow for digital fieldwork and corpus construction in small-language settings, while also offering an empirically grounded case study in computational lexicography, dialectometry, and endangered-language documentation. The approach is designed to scale as additional Evenki corpora are incorporated in later project stages.

References

Baek, Sangyub. 2025. The language ecology and endangerment of Solon, a Tungusic language spoken in China. *Sibirica* 24(1). 1–28.

Bradley, David. 2007. East and Southeast Asia. In Christopher Moseley (ed.), *Encyclopedia of the World's Endangered Languages*, 349–424. Routledge.

Bulatova, Nadezhda J. 1999. *Jazyk saxalinskix evenkov* [The language of the Sakhalin Evenki]. Nauka.

Däbritz, Chris L., Valentin Gusev & Natalia Stoynova. 2024. *INEL Evenki Corpus*, version 2.0. Universität Hamburg.

Gu, X. & Sirenbatu (eds.). 2011. *Àolǔgǔyǎ Èwēnkè yǔ dúběn* [A reading manual of Aoluguya Reindeer Evenki]. Inner Mongolia Culture Press.

Janhunen, Juha. 1996. *Manchuria: An Ethnic History*. Finno-Ugrian Society.

Janhunen, Juha. 2005. Tungusic: An endangered language family in Northeast Asia. *International Journal of the Sociology of Language* 175/176. 37–54.

Janhunen, Juha. 2012. The expansion of Tungusic as an ethnic and linguistic process. In Martine Robbeets & Hubert Cuyckens (eds.), *Shared Grammaticalization*, 5–16. John Benjamins.

Janhunen, Juha. 2023. Siberian Ewenki. In Alexander Vovin, José Andrés Alonso de la Fuente & Juha Janhunen (eds.), *The Tungusic Languages*, 139–183. Routledge.

Khabtagaeva, Bayarma. 2009. *The Ewenki Dialects of Buryatia and Their Relationship to Khamnigan Mongol*. Harrassowitz.

Kolås, Åshild & Yuanyuan Xie. 2015. *Reclaiming the Forest: The Ewenki Reindeer Herders of Aoluguya*. Berghahn Books.

Mullaney, Thomas. 2011. *Coming to Terms with the Nation: Ethnic Classification in Modern China*. University of California Press.

Nedjalkov, Igor. 1997. *Evenki*. Routledge.

Vovin, Alexander, José Andrés Alonso de la Fuente & Juha Janhunen (eds.). 2023. *The Tungusic Languages*. Routledge.

Whaley, Lindsay J., Lenore A. Grenoble & Fengxiang Li. 1999. Revisiting Tungusic classification from the bottom up: A comparison of Evenki and Oroqen. *Language* 75(2). 286–321.

Xie, Yuanyuan. 2015. *Ecological Migrants: The Relocation of China's Ewenki Reindeer Herders*. Berghahn Books.

Bulgarian web-based system for creating, visualizing and comparing associative dictionaries

Velka Popova, Dimitri Popov, Krasimir Kordov
(Konstantin Preslavsky University)
labling@shu.bg

The focus of the paper is on associative lexicography, which provides empirical data for studying the linguistic picture of the world of a native speaker. In this sense, it could be attributed to the field of online linguistics.

Studying word associations is a key moment in explaining a number of processes and resolving multiple questions in the field of contemporary anthropological linguistics. They are a unique material for the study of the mental lexicon. The poster presents LABLASS (https://lablass.fhn-shu.com/), the first Bulgarian web-based system for studying word associations, for the construction of which some of the most common technologies were used - MySQL database, PHP, JavaScript, HTML, CSS.

The paper includes a schematic presentation of the model of the system, as well as a description of the main modules and the users' options for working with it. The graphic interface of LABLASS is also illustrated. A brief preview of the collections of associative data entered in LABLASS is provided. Its technical capabilities are not limited to the inclusion of ready-made lexicographical sources, but also allow for the creation of new dictionaries, and visualization and comparison of data from different dictionaries.

What would (A)I say in my science podcast? A linguistic evaluation of corpus-based AI-recreated podcast episodes

Christina Sanchez-Stockhammer, Dominic Piazza
(Chemnitz University of Technology)
christina.sanchez@phil.tu-chemnitz.de

Science podcasts are an increasingly popular form of digital communication (MacKenzie 2019). Many widely known linguistics podcasts are hosted by researchers (e.g. *Lingthusiasm* by Gretchen McCulloch and Lauren Gawne), but new AI tools like NotebookLM make it possible to generate podcasts directly from research articles. This raises the question what human-made podcasts can offer by comparison. In this presentation, we will address this issue from an unusual angle and based on a specialised multimodal corpus.

Both presenters of this paper jointly produce the podcast *Linguistics Behind the Scenes*, which targets general audiences. In each of the more than 20 episodes so far, the two hosts have informal dialogic conversations about language, culture and their own linguistic research. The recordings vary in length between 20 and 79 minutes, consist of unscripted spontaneous language with slight post-editing in the form of cuts and can be considered a "mediated staged performance" of linguistic discourse (Bell & Gibson 2011: 558). The audio files of the podcast, which can be downloaded from a repository (https://videocampus.sachsen.de/channel/Linguistics-Behind-the-Scenes/45), and the full orthographic transcripts, which comprise over 120,000 words (https://www.tu-chemnitz.de/phil/english/sections/edling/sciencecommunication/podcast-transcripts.php), jointly constitute the LinguisticsBTS Corpus, a multi-modal corpus of the digital linguistic register of science podcasts. Important amounts of human post-editing in the transcripts ensure e.g. the marking of metalinguistic items with italics and quotation marks, as well as precision and proper spelling (e.g. in the frequent instances of code-switching between English, German, French, Spanish and Japanese), which transcription software cannot easily capture.

What makes the LinguisticsBTS Corpus unusual is that it contains the speech of the corpus compilers themselves, who simultaneously function as podcast hosts of past and future episodes. The post-editing of the automated orthographic transcripts thus puts linguists in the unusual position of reflecting their own spoken output in detail, e.g. concerning sentence boundaries or idiosyncratic language use. The resulting raised level of linguistic self-awareness occasionally influences the hosts' own linguistic performance in subsequent episodes, e.g. with respect to the over- or underuse of specific constructions.

For better readability (and to serve their main function as support for listeners of the podcast), the transcripts do not usually conserve all the

hesitation phenomena and repetitions from the audio. However, they still contain numerous features of spoken English (like broken-off constructions), reflecting the spontaneous language of the science podcast accurately and permitting the corpus-based linguistic study of spontaneous performed speech (e.g. with regard to sentence structures). This also makes the transcripts an ideal source for the AI-based recreation of science podcast episodes.

In the final part of the paper, we present a study that provides a qualitative and partly also quantitative assessment of AI-generated podcast episodes. We first used the transcripts of published episodes from the LinguisticsBTS Corpus as a reference text to produce an AI-generated transcript for a new episode on a yet unpublished linguistic article that was also provided. During this process, we used various prompts to induce ChatGPT to imitate our own style and content as much as possible. We then compared the AI-recreated material to a podcast episode on the same topic that we had previously recorded, but which was not yet available online. For the assessment, we used several commonly used objective measures (like sentence length), but our combined experiences as hosts of the real podcast and author of the linguistic research article furthermore allowed us to evaluate the correctness and authenticity of the AI-generated podcast "transcripts" regarding content and style in a unique way.

References

Bell, Allan & Andy Gibson. 2011. Staging language: An introduction to the sociolinguistics of performance. *Journal of Sociolinguistics* 15(5). 555-572.

MacKenzie, Lewis E. 2019. Science podcasts: Analysis of global production and output from 2004 to 2018. *Royal Society of Open Science* 6(1). 1-18. http://dx.doi.org/10.1098/rsos.180932.

1.000 Seiten in 1 Sekunde sichten: Sprachgebrauch im Kontext digitaler Fach- und Werbekommunikation

Margret Mundorf
(University of Heidelberg)
margret.mundorf@memoscript.de

Mein Beitragsvorschlag widmet sich dem Untersuchungsgegenstand des (fachlichen) Sprachgebrauchs mithilfe von Technologien des Maschinellen Lernens und der Künstlichen Intelligenz im Kontext des Rechts. Entsprechende digitale Produkte, die z. B. bei der Recherche und Verarbeitung von sprachlichen juristischen Daten oder bei Textentwürfen unterstützen können, werden massiv von einer Vielzahl von Produktanbietern auf Unternehmens- und Produktwebsites beworben. Die Websites können als transmediale Storytelling-Plattformen gestaltet sein, die neben Text visuell und multimodal aufbereitete Inhalte, Infografiken, FAQs, Videos und andere Interaktionsmöglichkeiten enthalten. Für meinen ausgewählten Korpus von Websites, die Legal-Tech-Produkte, digitale Rechtsdienstleistungen und juristische KI-Tools, bewerben, lauten meine Untersuchungsfragen:

1. Wie und mit welchen Zeichenressourcen wird über digitale Sprachverwendung gesprochen, wie wird der (fachliche) Sprachgebrauch mit KI-Technologien auf multimodalen Websites konstituiert? Welche Affordanzen prägen und vermitteln multimodal den Einsatz von digitalen Technologien im Spannungsfeld von Information und Persuasion vermittelt?
2. Wie lassen sich dabei klassische korpuslinguistische Analysetools und KI-bassierte Ansätze methodisch sinnvoll kombinieren und welche Beobachtungen sind bei unterschiedlichen kombinativen Verfahren festzustellen? Welche Chancen und Herausforderungen ergeben sich bei der Analyse multimodaler und zugleich fachsprachlich geprägter Werbekommunikation?

Mit den beiden Forschungsfragen verknüpfe ich einen doppelten, empirischen und methodischen, Zugang zu dem Ausschnitt multimodaler Werbekommunikate und des digitalen Sprachgebrauchs im Kontext von Recht und KI.

Mit meinen Ergebnissen möchte ich innerhalb des Rahmenthemas der digitalen Linguistik online zeigen, dass eine Analyse des multimodalen digitalen fachlichen Sprachgebrauchs in spezifischen Domänen (wie im Kontext des Rechts) entsprechender Methodenkombinationen bedarf: Für die Fragestellung meines Beitrags sind nicht nur die multimodalen Besonderheiten der digitalen Werbekommunikation relevant, sondern auch die der Fachkommunikation in Recht und Rechtsinformatik. Denn auch die Praktiken der Fachkommunikation sind durch eine Vielzahl semiotischer Modi und deren

Zusammenwirken beeinflusst und statische Artefakte werden um dynamische Dokumente, Visualisierungen, bewegte Bilder und Virtual-Reality-Umgebungen ergänzt (Wildfeuer 2025: 237). Multimodalität ist somit nicht nur ein Kennzeichen der digitalen Werbekommunikation, sie charakterisiert auch fachkommunikative Praktiken. Gleichzeitig wird das ‚Sprechen (Schreiben, Lesen, Daten verarbeiten) mit KI' über das ‚Analysieren mit KI' erfasst, beides zueinander in Beziehung gesetzt und forschungsmethodisch kritisch reflektiert. Dies wiederum eröffnet Potenzial für sprach- und fachdidaktische Einsatzszenarien im Zeitalter von Künstlicher Intelligenz, die abschließend diskutiert werden.

References

Wildfeuer, Janina. 2025. The multimodal complexity of specialized communication: Examples and approaches. In Thorsten Roelcke, Ruth Breete & Jan Engberg (eds.), *Specialized Communication*, 237-256. de Gruyter.

Evaluating acceptability in LLMs: Evidence from human-model comparisons in German

Nicholas Catasso
(University of Wuppertal)
catasso@uni-wuppertal.de

This paper examines how large language models (LLMs) handle grammaticality and acceptability across sociolinguistic variation in German and what this reveals about their limitations as linguistic tools. Against the backdrop of recent work on the linguistic "fingerprints" and biases of LLMs (Liu et al. 2023; Qiu et al. 2024; Yu et al. 2024; Smith et al. 2024), gradient judgments from five state-of-the-art models (ChatGPT-4, Gemini 2.0 Flash, Claude 3.5 Sonnet, Perplexity AI, DeepSeek) are directly compared with those of 200 native speakers rating a curated set of 150 contextualized sentences spanning standard, regional, and register-marked constructions.

While LLMs closely approximate human judgments in syntactically clear-cut cases, they tend to diverge in contexts shaped by diatopic and idiolectal variation. In line with previous findings (Qiu et al. 2024), models reliably distinguish canonical grammatical from ungrammatical structures. However, the comparison with human data reveals systematic differences in gradient domains: LLMs exhibit more categorical rating behavior and show reduced sensitivity to intermediate acceptability levels, particularly in cases involving regional and register-based variation. In contrast, human judgments display the expected gradience and variability across speakers (Featherston 2007; Schütze 2016; Sprouse 2018).

Models further tend to normalize toward prescriptive norms, underrepresenting variation attested in the human data. This over-regularization suggests that, despite their fluency, current LLMs encode skewed distributions of acceptability, privileging more uniform patterns over the structured variability characteristic of natural language use. These findings align with recent observations on biases and limitations in LLM-based linguistic tasks (Yu et al. 2024; Torrent et al. 2023).

It is argued that such biases must be critically examined when integrating LLMs into linguistic research—whether as annotation tools, modeling aids, or objects of theoretical inquiry. Far from being neutral "informants," LLMs produce judgments that diverge systematically from human data in variation-sensitive domains and may therefore obscure important dimensions of linguistic structure.

This paper contributes to broader discussions about how to responsibly use LLMs in empirical methodology and what their performance reveals (and conceals) about the nature of human linguistic competence. We conclude by proposing guidelines for interpreting LLM-generated linguistic data in light of their reduced sensitivity to gradience and variation, drawing on insights from

both theoretical and experimental work (Featherston 2007; Schütze 2016; Sprouse 2018; Yu et al. 2024; Torrent et al. 2023).

References

Featherston, Sam. 2007. Experimentell erhobene Grammatikalitätsurteile und ihre Bedeutung für die Syntaxtheorie. In Werner Kallmeyer & Gisela Zifonun (eds.), *Sprachkorpora: Datenmengen und Erkenntnisfortschritt*, 49-69. de Gruyter.

Liu, Hanmeng, Ruoxi Ning, Zhiyang Teng, Jian Liu, Qiji Zhou & Yue Zhang. 2023. Evaluating the logical reasoning ability of ChatGPT and GPT-4. arXiv. https://arxiv.org/abs/2304.03439.

Qiu, Zhuang, Xufeng Duan, Zhenguang G. Cai. 2024. Grammaticality representation in ChatGPT as compared to linguists and laypeople. arXiv. https://arxiv.org/abs/2406.11116.

Schütze, Carson T. 2016. *The Empirical Base of Linguistics: Grammaticality Judgements and Linguistic Methodology*. Language Science Press.

Smith, Genevieve, Eve Fleisig, Ishita Rustagi & Xavier Yin. 2024. *Standard language ideology in AI-generated language*. arXiv. https://arxiv.org/abs/2406.08726.

Sprouse, Jon. 2018. Acceptability judgments and grammaticality, prospects and challenges. In Norbert Hornstein, Howard Lasnik, Pritty Patel-Grosz & Charles Yang (eds.), *Syntactic Structures after 60 Years*, 195-223. de Gruyter.

Torrent, Tiago T., Thomas Hoffmann, Arthur L. Almeida & Mark Turner. 2023. *Copilots for Linguists: AI, Constructions and Frames*. Cambridge University Press.

Yu, Danni, Luyang Li, Hang Su & Matteo Fuoli. 2024. Assessing the potential of LLM-assisted annotation for corpus-based pragmatics and discourse analysis: The case of apologies. arXiv. https://arxiv.org/abs/2305.08339v5.

Same prompt, different answer? On the temporal stability of large language models

Olga Zagovora, Ljubiša Bojić, Ana Jovančević, Velibor Ilić
(University of Kaiserslautern-Landau, University of Belgrade, University of Limerick, University of Limerick)
olgazagovora@rptu.de, ljubisa.bojic@ivi.ac.rs, ana.jovancevic@ul.ie, velibor.ilic@ivi.ac.rs

Large language models (LLMs) are increasingly embedded in digital language practices, functioning both as tools for analyzing and as systems that model the interpretation of meaning in online discourse. In communication science, they are widely used to approximate human judgments of stance, sentiment, and socially embedded meaning (Peng & Yang, 2025; Ziems et al., 2024). However, LLM outputs are stochastic, and there is still little systematic evidence on whether repeated interpretations of identical input yield temporally stable results. This creates a methodological blind spot: if model outputs vary substantially across runs, LLM-based analyses may not only become irreproducible but also reveal deeper instability in how meaning is operationalized in digital contexts.

We address this gap by systematically evaluating the temporal stability of six LLMs (DeepSeek, Gemini, Gemma, Grok, Mistral, GPT-5) across two task families central to the analysis of digital discourse. The first focuses on (a) socially grounded meaning and evaluative judgments, operationalized through stereotype-related assessments based on the Stereotype Content Model (Fiske et al., 2002) and (b) capturing how models reproduce culturally embedded associations (Weidinger et al., 2022; Blodgett et al., 2020). The second examines interpretive text classification tasks that involve varying degrees of semantic and pragmatic ambiguity, including sentiment analysis, political leaning, emotional intensity, and pragmatic inference (sarcasm) (Bojić et al., 2025; Hu et al., 2025). Prompts were administered in three languages (English, German, Serbian), reflecting varying levels of linguistic resource availability (Joshi et al., 2020). Each model completed identical prompts across all conditions, yielding approximately 141,600 responses per model. For each model–task–language combination, temporal stability was assessed by sampling responses (sequentially or randomly) and computing the coefficient across increasing numbers of runs, using Krippendorff's α (Krippendorff, 2019).

The results reveal task-dependent differences in stability. Tasks involving relatively constrained semantic evaluation, such as sentiment, emotional intensity, and political leaning, show uniformly high stability (α = 0.93-1.0) across models and languages. In contrast, tasks requiring context-sensitive interpretation, such as stereotype evaluations (e.g., α = 0.86-1.0 for men and 0.03-0.97 for women) and sarcasm detection (α = 0.05-1.0), exhibit substantially lower and more heterogeneous stability. Cross-model and cross-

language variation reflects the impact of training data, resource density, and architecture on meaning inference (Bodroža et al., 2024; Bojić et al., 2025).

Across tasks, stability typically plateaued after approximately ten repetitions, with only marginal gains beyond thirty runs. Persona prompting did not consistently increase stability compared to neutral prompting, and sequential versus repeated prompting produced divergent stability estimates for some models. These findings suggest that interpretive variability in LLM outputs is not merely noise but reflects the interaction between model design, linguistic context, and the inherent ambiguity of meaning in digital communication.

We propose temporal stability as a key methodological dimension for digital linguistics. We derive practical guidelines for LLM-based content analysis, including minimum repetition thresholds and increased caution for lower-resource languages, and argue that empirical stability testing should become a standard component of research using LLMs to study language and communication in digital environments.

References

Blodgett, Su L., Solon Barocas, Hal Daumé III & Hanna Wallach. 2020. Language (technology) is power: A critical survey of "bias" in NLP. In Dan Jurafsky, Joyce Chai, Natalie Schluter & Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5454–5476. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.485.

Bodroža, Bojana, Bojana M. Dinić & Ljubiša Bojić. 2024. Personality testing of large language models: Limited temporal stability, but highlighted prosociality. *Royal Society Open Science* 11(10). 240180. https://doi.org/10.1098/rsos.240180.

Bojić, Ljubiša, Olga Zagovora, Asta Zelenkauskaite, Vuk Vuković, Milan Čabarkapa, Selma Veseljević Jerković & Ana Jovančević. 2025. Comparing large language models and human annotators in latent content analysis of sentiment, political leaning, emotional intensity and sarcasm. *Scientific Reports* 15(1). 11477. https://doi.org/10.1038/s41598-025-96508-3.

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick & Jun Xu. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82(6). 878–902. https://doi.org/10.1037/0022-3514.82.6.878.

Hu, Yuqi, Longguang Wang, Xian Liu, Ling-Hao Chen, Yuwei Guo, Yukai Shi, Ce Liu, Anyi Rao, Zeyu Wang & Hui Xiong. 2025. Simulating the real world: A unified survey of multimodal generative models. arXiv. https://doi.org/10.48550/arXiv.2503.04641.

Joshi, Pratik, Sebastin Santy, Amar Budhiraja, Kalika Bali & Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In Dan Jurafsky, Joyce Chai, Natalie Schluter & Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.560.

Krippendorff, Klaus. 2019. *Content Analysis: An Introduction to its Methodology*. SAGE. https://doi.org/10.4135/9781071878781.

Peng, Tai-Quan & Xuzhen Yang. 2025. Recalibrating the compass: Integrating large language models into classical research methods. arXiv. https://doi.org/10.48550/arXiv.2505.19402.

Weidinger, Laura, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa A. Hendricks, Laura Rimell, William Isaac, Jilia Haas, Sean Legassick, Geoffrey Irving, Iason Gabriel. 2022. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT 2022*), 214–229. Association for Computing Machinery. https://doi.org/10.1145/3531146.3533088.

Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang & Diyi Yang. 2024. Can large language models transform computational social science? *Computational Linguistics* 50(1). 237–291. https://doi.org/10.1162/coli_a_00502.

Six hundred and sixty-six reasons to (not) save Te Reo Māori: Language activism and online discourse on X/Twitter

Leon Grausam
(University of Bremen)
lgrausam@uni-bremen.de

With imminent language death on the rise and more and more languages vanishing from the face of the earth year after year, governments and politicians in charge of many political colors have started to address this issue. With her protest haka, MP Hana-Rawhiti Maipi-Clarke has brought significant attention on the issues of Te Reo Māori in modern day Aotearoa New Zealand. This paper investigates how language activism takes place in and surrounding social media debates on Te Reo Māori through a mixed-methods and corpus linguistic approach.

X/Twitter has been manually scraped for all mentions of Te Reo Māori in the month of November 2024 (n=666) and the parliament haka happening on 14 November 2024, in the middle of the month. Through the qualitative data analysis tool MAXQDA, the tweets have been annotated with so-called codes representing various categories like political actors, cultural concepts, and historical events, as well as linguistic, educational, social, and political aspects of many kinds. Enhanced with metadata on polarization, opinion on Māori rights, and the dates of posting the data, this paper demonstrates how the issues of Te Reo have (re-)entered social media and are being discussed online. As a further step, these tweets were analyzed for linguistic patterns and interrelated to the aforementioned categories. This provides a comprehensive picture of how those issues and topics are discussed and negotiated online in fighting for or against Māori language rights. Initial abstractions show that the haka instigated by MP Maipi-Clarke has potentially brought the topic back on the X/Twitter agenda with almost 2.5 times more tweets on Te Reo after the viral protest than before.

This analysis is a contribution to a new concept in sociolinguistic and language policy research usually titled "language activism." As language activism currently does not have a well-established framework, smaller case studies like this one play an important role in not only shedding some light on the precarious situations of minoritized language communities but also in narrowing down a possible paradigm of language activism studies.

References

Combs, Mary C. & Susan D. Penfield. 2012. Language activism and language policy. In Bernard Spolsky (ed.), *The Cambridge Handbook of Language Policy*, 461–474. Cambridge University Press.

De Korne, Haley. 2021. *Language Activism: Imaginaries and Strategies of Minority Language Equality*. de Gruyter Mouton.

Florey, Margaret, Susan D. Penfield & Benjamin Tucker. 2009. *Towards a Framework for Language Activism.* https://scholarspace.manoa.hawaii.edu/bitstream/handle/10125/5014/5014.pdf. (27 August 2024).

Lillehaugen, Brook D. 2019. Tweeting in Zapotec: Social media as a tool for language activists. In Jennifer Gómez Menjívar & Gloria E. Chacón (eds.), *Indigenous Interfaces: Spaces, Technology, and Social Networks in Mexico and Central America*, 201–226. University of Arizona Press.

Peters, Murray H. 2014. *Reclaiming the Māori Language for Future Generations: Flax Root Perspectives*. University of Waikato Doctoral Dissertation.

Putra, Kristina A. & Luisa M. Nurani. 2022. Indigenous youth digital language activism. In Jessica Z. Pandya, Raúl A. Mora, Jennifer H. Alford, Noah A. Golden & Roberto S. de Roock (eds.), *The Handbook of Critical Literacies*, 391–401. Routledge.

Digital translanguaging on the r/Morocco subreddit: A quantitative corpus study of submission

Marie Flesch
(University of Lorraine)
marie.flesch@univ-lorraine.fr

In sociolinguistics, digital code-switching has mostly been studied through qualitative approaches such as online ethnography, conversation and discourse analysis, which have focused on the functions of code-switching rather than on its structure (Androutsopoulos 2013). Recently, it has also attracted the attention of computational linguists, who have tackled tasks such as language identification and part of speech tagging (Doğruöz et al. 2021).

This contribution positions itself between these traditions, adopting a computationally assisted sociolinguistic approach that combines manual annotation, quantitative analyses and insights from digital sociolinguistics and translanguaging research. The study is based on r/Morocco, a Reddit forum where users discuss topics related to the country of Morocco mostly in English, but also in Darija, the variety of Arabic spoken in Morocco, and in French, a language brought to Morocco by colonization and still widely taught in the country.

Based on a corpus of 600 comments containing code-switching and posted in 2025-2026, the study aims to provide a fine-grained descriptive account of English/Darija/French code-switching practices. Each comment was manually coded using a tagging system that allows to determine, for each comment, the number, length, and languages of switches.

Results suggest that traditional sociolinguistic code-switching frameworks cannot account for the diversity and fluidity of linguistic practices of r/Morocco users. On the forum, code-switching is generally not a two, or a three-way process, with clearly delimited matrix languages and identifiable insertions. This is because of (1) the high frequency of one-word insertions, which are known to be difficult to classify as either borrowing, nonce-borrowing, or code-switching; (2) the fact that written Darija is still an emergent variety, which can written with two different scripts (Arabic and Latin or Arabizi, Allehaiby 2013) (3) the absence of a standard spelling system for either version of written Darija and the large amount of variation in spelling practices; (4) the fact that many one-word insertions can be classified either as French or English, since the spelling of many words are identical (*representation*, *nation*, for ex.); (5) the frequent use of other semiotic ressources such as emoji, interjection and internet abbreviations (*lol*, *hhhhhhh*).

The findings suggest that translanguaging, which was introduced as pedagogical tool and has been developing into a comprehensive theoretical framework (Azad 2026), may be a better approach to this corpus. Thus, we propose a quantitative operationalization of digital translanguaging, with

several metrics enabling the systematic measurement of multilingual practices that are typically analyzed qualitatively in digital sociolinguistics.

References

Allehaiby, Wid H. 2013. Arabizi: An analysis of the romanization of the Arabic *script* from a sociolinguistic perspective. *Arab World English Journal* 4(3). 52-62.

Androutsopoulos, Jannis. 2013. Code-switching in computer-mediated communication. In Susan Herring, Dieter Stein & Tuija Virtanen (eds.), *Pragmatics of Computer-Mediated Communication*, 667–694. de Gruyter.

Azad, Mohammed T. 2026. Mapping the literature on translanguaging research: A bibliometric analysis. *International Journal of Multilingualism* 23(1). 1–27. https://doi.org/10.1080/14790718.2025.2545468.

Doğruöz, A. Seza., Sunayana Sitaram, Barbara E. Bullock & Almeida J. Toribio. 2021. A survey of code-switching: Linguistic and social perspectives for language technologies. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 1654–1666. https://doi.org/10.18653/v1/2021.acl-long.131.

Otheguy, Ricardo, Ofelia García & Wallis Reid. 2015. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review* 6(3). 281–307. https://doi.org/10.1515/applirev-2015-0014.

Using AI for replication studies on social media data in (Romance) linguistics

Robert Hesselbach
(University of Erlangen)
robert.hesselbach@fau.de

The increased use of digital tools in linguistic research can be observed in both synchronic and diachronic studies. They help to acquire, structure, analyze, and prepare complex data sets in such a way that the data can be stored in research data repositories, such as *Zenodo*, in accordance with the FAIR principles (Wilkinson et al. 2016) and made accessible to the research community. In addition, digital methods can also help to carry out replication studies ("repetitive research" according to Schöch 2023) with little technical effort, as recently presented by Schöch in the field of Romance literary studies, for example, for a study by Spitzer on Racine's style (2024).

The aim of my presentation is to discuss the extent to which AI tools can be used in the sense of reusable research data for replication studies. In this context, each step of the empirical research process, i.e., data collection, evaluation, and preparation for research data repositories, is simulated using artificial intelligence based on social media data (*Twitter/X*) on the political conflict in Venezuela in 2019 (cf. Hesselbach 2025). By using a comparative analysis, my goal is to assess the potential of AI tools for replication studies based on social media data for Spanish and to highlight the importance of published and reusable research data.

References

Hesselbach, Robert. 2025. Analyzing linguistic patterns in the social media discourse of Juan Guaido and Nicolas Maduro during the 2019 political conflict in Venezuela. In Lukas Eibensteiner & Robert Hesselbach (eds.), *Social Media: Current Issues in Romance Linguistics and Foreign Language Learning*, 59-86. AVM.

Schöch, Christof. 2023. Repetitive research: A conceptual space and terminology of replication, reproduction, revision, reanalysis, reinvestigation and reuse in digital humanities. *International Journal of Digital Humanities* 5. 373-403.

Schöch, Christof. 2024. Repetitive research: Spitzer and Racine. In Robert Hesselbach, José Calvo Tello, Ulrike Henny-Krahmer, Christof Schöch & Daniel Schlör (eds.), *Digital Stylistics in Romance Studies and Beyond*, 119-148. Heidelberg University Publishing.

Wilkinson, Mark D. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3. 160018

Cross-cultural differences in the perception of politeness strategies in conversational artificial intelligence feedback on trust, perceived anthropomorphism and cognitive appraisal

Cynthia Mueller, Julia Hofweber
(Northeastern University London)
cm31041@nulondon.ac.uk, julia.hofweber@nulondon.ac.uk

Despite the growing ubiquity of conversational artificial intelligence (AI) in global higher education, it remains unclear how AI-generated feedback is interpreted across cultures and to what extent its effectiveness depends on culturally specific politeness norms (Luo, 2025). Grounded in Brown and Levinson's (1987) *Politeness Theory*, this study examined how four politeness strategies differing in directness (direct *on-record, positive, negative*, and indirect *off-record*) influenced the perception of AI feedback amongst university students from universities in the United Kingdom (UK) and South Korea (SK), i.e. two cultures differing in politeness strategies (Moon, 2018).

Participants (UK=20; SK=20) completed a mathematics task and received corrective feedback framed in one of the four politeness conditions. Students' perception of the feedback was captured using standardised questionnaires assessing the following dependent variables: trust in the AI, perceived anthropomorphism, and cognitive appraisal. To control for individual differences, we assessed participants' demographic background and personality traits. To analyse our results, we conducted ANOVAs using SPSS statistical software.

Results revealed both differences and similarities in cross-cultural perception of AI feedback. Firstly, the pattern of perception described below was identical across all four dependent variable measures. Secondly, across both groups, *positive* politeness produced the highest ratings. However, there were subtle differences across groups. The UK participants exhibited a clear preference for *positive* politeness with strong dispreference for *off-record* indirect feedback. SK participants responded favourably to both *negative* and *positive* politeness. Crucially, they rated *on-record* feedback lowest and rated *off-record* feedback more highly. Personality traits had limited predictive value, indicating greater importance of cultural or extraneous factors.

Our findings indicate that the social framing of AI feedback significantly shapes its evaluation. When politeness strategies match users' cultural expectations, the AI is perceived as more human-like, which in turn may increase trust and leads to more positive cognitive appraisals of the feedback. These findings highlight the importance of culturally attuned politeness strategies for enabling students from all cultural backgrounds to fully benefit equitably from conversational AI.

References

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press.

Luo, Xiaofei. 2025. Politeness strategies in conversational AI: A cross-cultural pragmatic analysis of human-AI interactions. *Pinnacle Academic Press Proceedings Series* 3. 1-14.

Moon, Chanki, Ayse K. Uskul & Mario Weick. 2018. Cultural differences in politeness as a function of status relations: Comparing South Korean and British communicators. *Journal of Theoretical Social Psychology* 3(3). 137-145.

Multimodal analysis of cold-calling in synchronous online classrooms: Evidence from Chinese Junior High EFL students

Xi Wang
(Shandong University)
202300073112@mail.sdu.edu.cn

Classroom silence remains a persistent challenge in junior high English as a Foreign Language (EFL) classrooms, with many students avoiding voluntary oral participation due to foreign language anxiety, face concerns, fear of mistakes, and cultural norms emphasizing modesty. Cold-calling—directly selecting students who do not volunteer—has been shown to increase engagement in face-to-face classrooms, but its implementation in synchronous online settings has been largely unexplored. Online learning environments introduce unique challenges, including the absence of physical proximity, reduced visibility of body language, and the reliance on platform-mediated cues, making it essential to investigate how traditional pedagogical strategies can be adapted digitally. This study focuses on Chinese junior high EFL students to examine how teachers implement cold-calling in online classrooms, and how different designs affect students' responses in terms of length, grammatical accuracy, and content relevance.

Using multimodal conversation analysis, we analyzed 30 synchronous online lessons conducted by 10 EFL teachers. The analysis captured verbal, prosodic, nonverbal, and platform-mediated resources. Three categories of online cold-calling resources were identified: (a) verbal and prosodic cues, such as slowed speech, lowered pitch, and pauses before naming students; (b) camera-facing nonverbal cues, including sustained gaze at the camera, upper-body posture changes, and pointing gestures within the video frame; and (c) platform-mediated cues, such as chat prompts, reaction icons, hand-raise signals, and deliberate use of mute/unmute to manage turn-taking. Teachers also monitored students' facial expressions and provided affective support through smiles, eye contact, and emoji feedback, prompting students to unmute, respond, and engage collaboratively in discussions.

Silent students were defined as those who did not voluntarily speak in two consecutive online lessons and produced minimal responses when called, while maintaining at least median-level written test scores. Preliminary findings revealed two contrasting online cold-calling patterns: a low-participation mode, in which teachers relied solely on verbal naming with minimal follow-up, resulting in brief or single-word responses; and a high-participation mode, combining camera gaze, verbal naming, chat prompts, and affective cues, which elicited longer, more complete, and topically relevant responses. The study highlights how embodied resources traditionally available in physical classrooms, such as proximity and spatial movement, can be functionally

replaced online through coordinated multimodal and affective strategies, reducing participation barriers and mitigating foreign language anxiety.

These findings have significant implications for digital pedagogy in EFL contexts. They demonstrate that synchronous online classrooms can successfully employ adapted cold-calling techniques, providing guidance for teachers seeking to enhance student engagement and learning outcomes. Furthermore, the study contributes empirically grounded insights into how traditional classroom practices can be translated into online teaching, offering a framework for future research on multimodal interaction, student motivation, and teacher-student coordination in online EFL learning. By focusing on Chinese junior high students, this research also illuminates how cultural and linguistic factors influence online participation patterns and pedagogical strategies, offering a model for integrating local classroom practices into digital teaching contexts.

References

Morek, Miriam, Vivien Heller & Noelle Kinalzik. 2023. Engaging 'silent' students in classroom discussions: A micro-analytic view on teachers ' embodied enactments of cold-calling practices. *Language and Education* 37(6). 731– 749.

Dallimore, Elise, Julie H. Hertenstein & Marjorie B. Platt. 2013. Impact of cold-calling on student voluntary participation. *Journal of Management Education* 37(3). 305 –341.

Goodwin, Charles. 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics* 32(10). 1489–1522.

Satar, Müge. 2013. Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL* 25(1). 122–142.

Exploring Informal Digital Learning of English (IDLE) with digital language learning diaries

Ralf Gießler
(University of Wuppertal)
giessler@uni-wuppertal.de

Lee (2019: 116) defines IDLE as "self-directed, naturalistic digital learning of English in unstructured out-of-class environments." In recent years, there has been considerable research and pedagogical experimentation relating to language learning beyond the classroom (LBC) (cf. Reinders, Lai, & Sundqvist 2022). However, processes and learning outcomes of IDLE are still underexplored (Amengual-Pizzarro & Alonso Alonso (2024). One aim of this contribution is to reflect on suitable research methodologies for exploring learner practices in IDLE contexts as ethnographies of practice (cf. Lee 2022). Which digital tools are suitable for documenting learners' IDLE activities and for elaborating and reflecting on the types of discourse and multi-word units learners encounter? The tools for keeping a digital language learning diary can be conceptualized as an interface between formal and informal language learning which is a reality for most language learners.

In a pilot study at Wuppertal University, students were asked to record their informal encounters with English over a period of three weeks. Students were free to choose from three different tools (Notebook LM, Daylio and LangJournal). In addition, students were asked to reflect in writing on the experience, taking three different perspectives (technological, language acquisition and media related). Results suggest that, with the help of generative AI, input processing and language reflection can be enhanced.

References

Amengual-Pizarro, Marian & Rosa Alonso Alonso. 2024. *Informal Language Learning: Expanding the English Learning Horizons Beyond the Classroom*. Peter Lang Verlag.

Lee, Ju S. 2019. Quantity and diversity of informal digital learning of English. *Language, Learning & Technology* 23(1). 114-126.

Lee, Ju S. 2022. Evaluation of instruments for researching learners' LBC. In Hayo Reinders, Chun Lai & Pia Sundqvist (eds.), *The Routledge Handbook of Language Learning and Teaching Beyond the Classroom*, 312–326. Routledge.

Reinders, Hayo, Chun Lai & Pia Sundqvist. (eds.). 2022. *The Routledge Handbook of Language Learning and Teaching Beyond the Classroom*. Routledge.