

Doing Multilingualism with Generative AI: Recentering the Users

Naomi Truan (Leiden) & Daniel Knuchel (Zurich)

Call for Papers

**International Workshop, University of Zurich
10 and 11 March 2027**

Since the public release of ChatGPT in November 2022, generative AI has moved from experimental novelty to everyday communicative infrastructure. Chatbots and other AI systems are now used for many tasks in a wide range of languages, and they appear multilingual: these systems answer in dozens of languages, translate between them on request, and are marketed as dissolving language barriers. Yet this multilingualism is of a particular kind. Online language provision has been organized by selectable categories since well before generative AI (Kelly-Holmes 2019). Ramati and Pinchevski (2018) described this configuration for statistical machine translation as “uniform multilingualism”, where languages are conceptualized as homogeneous and interchangeable units. **These systems hold languages side by side, each separately selectable and treated as internally uniform, which is closer to a drop-down menu than to a repertoire.** Generative AI inherits this menu logic and extends it: the system now answers, evaluates and returns text, so that the menu is no longer only offered but negotiated. This is what we tentatively call “mono-multilingualism” (Truan 2026).

Taking the need for looking at what people do seriously (Erdocia, Migge & Schneider 2024), **our international workshop shifts the focus to multilingual speakers’ practices, experiences, and ideologies.** Moving away from testing LLM outputs, we are interested in what speakers bring to these interactions: fluid repertoires, moving across languages, varieties and registers. **We ask what multilingual speakers do with these systems, what these systems do to multilingual practices, and whether the concepts we use to describe multilingual interactions still hold when one of the ‘participants’ is a machine.**

The ‘multilingualism’ of generative AI cannot be understood independently of the linguistic infrastructures and ideologies on which these systems are built. The consequences of these arrangements have indeed been documented in detail. Coverage is unequal, and the inequality is systematic: over 90% of the world’s languages have little to no support in terms of language technology (Bender et al. 2021: 612; see Joshi et al. 2020), and language technology itself occupies a marginal position in policy, having been “ignored or missed or left out” of almost all national AI strategies across thirty European countries (Rehm & Way 2023: 388). The corpora on which these systems rest are predominantly monolingual, reproducing and amplifying hierarchies that construct English as unmarked and more valuable (Schneider 2022: 371–372). Erdocia, Migge and Schneider (2024) show how ideologies of dataism recast language as an extractable resource; the same authors (2025) trace how language authorities are repositioned as commercial platforms acquiring de facto control over linguistic norms. Model-side evidence converges: outputs systematically disadvantage non-standardized varieties (Fleisig et al. 2024), as does automatic speech recognition (Koenecke et al. 2020). As Schneider (2024) argues, the apparent fluency of these systems is itself an artifact of standardization. Machine-generated texts pass as human because centuries of print literacy and language policing have made human writing probabilistically modelable. What, then, happens

when this particular organization of language encounters speakers whose multilingualism does not operate according to the same logic?

Research on earlier communication technologies established that multilingualism under technological mediation is neither preserved nor destroyed but reorganized: text messaging and social media reshaped code-switching, orthographic norms and the distribution of semiotic resources (Deumert 2014; Androutsopoulos 2015; Lee 2016). Where users of language technology have been studied since, this has largely happened in translation studies and human-computer interaction (e.g. Vieira et al. 2023; Gao et al. 2022). This work frames the interaction as one of transfer and the multilingual speaker accordingly as someone who lacks a language rather than as someone who moves across several. It also reports site by site, so the question of how findings depend on the language constellation in which they arise is not posed.

Generative AI opens a qualitatively different terrain, because linguistic practices here are not only mediated but produced in exchange with a system that answers, evaluates, and returns text. AI output is edited by users according to their audiences and language ideologies, and that edited language feeds back into the systems (Erdocia, Migge & Schneider 2024). **This makes generative AI a particularly productive site for examining multilingualism as practice: not simply as the presence or absence of particular languages in a system, but as something speakers actively do across languages, varieties, registers, and communicative contexts.**

Drawing on the idea that AI technology can be reconceptualized “as human interaction” (Schneider et al. 2026), the workshop therefore brings together work that takes multilingual speakers and their practices as its starting point. **We particularly welcome ethnographic approaches** (Erdocia, Migge & Schneider 2024: 20–21; Lamoureaux, Castelle & Weichselbraun 2025). We invite contributions organized around the practices, experiences, and language ideologies of multilinguals.

Practices

Speech recognition performs measurably worse for speakers of non-standardized varieties (Koencke et al. 2020), and generative systems reproduce the same asymmetry in text (Fleisig et al. 2024); users of voice interfaces describe the consequence as a system that “do[es] not allow [for] bilingual practices or [nonstandard] accents and offer[s] languages in a packaged form” (Leblebici 2021: 9, cited in Schneider 2022: 372). **What speakers do with those terms is the open question.** What do multilingual speakers actually do with these systems, and what do these systems do to multilingual practices? We are interested in the level of the interaction: which language(s) or variety(ies) for which task, and what reasons do people give when asked? What happens when a prompt returns something unusable in one language and passable in another: do people switch, translate, give up, shift their communicative standards? We take the position that this cannot be answered from the system side alone, and that the social organization of multilingualism is where the analytical work has to happen.

Language ideologies

Language ideologies are central to this inquiry, and they operate at several levels simultaneously. They are **inscribed in the discourse** through which these systems are designed, marketed, and regulated, where notions of linguistic adequacy, support, and quality are established without being made explicit. They are **enacted in the outputs themselves**, which realize particular registers, standards, and orthographic conventions while presenting them as unmarked. And they are

articulated by users, whose accounts of what a system understands, prefers, or handles well constitute a body of metapragmatic commentary in its own right.

Concepts

Addressee, audience, accommodation, interlocutor, speech community, or repertoires were built for exchanges between human beings with common ground and social accountability. A chatbot has none of that. Users nevertheless orient to it, adjust for it, judge it, and build up expectations about what it ‘understands’. **This is not a new observation: people extend social responses to machines and interaction with artefacts is organized as an accountable, repairable achievement.** What do these and similar findings imply for the sociolinguistic terms specifically? Our provisional position is that the vocabulary may be used beyond its warrant. Accommodation, in its sociolinguistic sense, presupposes a co-present other whose evaluation matters and who may evaluate in return. What users do with a chatbot may have the observable form of accommodation, but none of its social conditions. That mismatch, we argue, is not simply terminological, but an empirical question that becomes visible once we look at situated use.

Methods

These interactions are textual, private, ephemeral and almost never observed. They leave no community of practice in the usual sense and no corpus that occurs naturally anywhere. Yet “it is impossible to know what people do and why without engaging with people” (Erdocia, Migge & Schneider 2024: 22). We therefore **welcome methodological experimentation** and expect contributions to give an account of how their data were produced—e.g. interaction data, screen recordings, logs, digital ethnography, interviews, or combinations of these. One of our aims is to leave behind a usable methodological repertoire for a kind of data that currently has no settled protocol.

Expected scope of the contributions

Contributions are empirical work in sociolinguistics, linguistic anthropology, discourse studies and related disciplines. The collection does not include model evaluation or benchmarking without an account of situated use, nor programmatic commentary unanchored in language data. Data are in any language; we welcome contributions that engage with scholarship published in languages other than English.

Workshop concept

All contributions will be pre-circulated in draft form. **Previously published papers may not be presented at the workshop.** Contributions will be discussed in full during the 1.5-day workshop in Zurich on 10–11 March 2027, with half a day devoted to discussing the coherence of the collection as a whole. Following the workshop, contributors will revise their papers for submission to a journal. We are currently approaching journals regarding the publication of the collection.

Keynotes

Bettina Migge, University College Dublin
Anna Weichselbraun, University of Vienna

Abstract submission

Please submit your abstract (max. 500 words, excluding bibliography) **by 30 October 2026** to both of our email addresses: n.a.l.truan@hum.leidenuniv.nl and daniel.knuchel@ds.uzh.ch.

You can expect a response by the end of November 2026.

Conference fees and funding

Participation in the workshop is free of charge, and catering will be provided free of charge for all participants. Travel costs and accommodation may be covered, depending on available funding.

References

- Androutsopoulos, Jannis. 2015. Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism* 19(2). 185–205.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623. New York: ACM. <https://doi.org/10.1145/3442188.3445922>.
- Deumert, Ana. 2014. *Sociolinguistics and mobile communication*. Edinburgh: Edinburgh University Press.
- Erdocia, Iker, Bettina Migge & Britta Schneider. 2024. Language is not a data set — Why overcoming ideologies of dataism is more important than ever in the age of AI. *Journal of Sociolinguistics* 28(5). 20–25. <https://doi.org/10.1111/josl.12680>.
- Erdocia, Iker, Britta Schneider & Bettina Migge. 2025. Language in the age of AI technology: From human to non-human authenticity, from public governance to privatised assemblages. *Language in Society*. <https://doi.org/10.1017/S004740452500017X>.
- Fleisig, Eve, Genevieve Smith, Madeline Bossi, Ishita Rustagi, Xavier Yin & Dan Klein. 2024. Linguistic bias in ChatGPT: Language models reinforce dialect discrimination. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 13541–13564. Miami, FL: Association for Computational Linguistics.
- Gao, Ge, Jian Zheng, Eun Kyoung Choe & Naomi Yamashita. 2022. Taking a language detour: How international migrants speaking a minority language seek COVID-related information in their host countries. *Proceedings of the ACM on Human-Computer Interaction* 6(CSCW2). 1–32.
- 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 *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293. Online: Association for Computational Linguistics.
- Kelly-Holmes, Helen. 2019. Multilingualism and technology: A review of developments in digital communication from monolingualism to idiolingualism. *Annual Review of Applied Linguistics* 39. 24–39.
- Koenecke, Allison, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky & Sharad Goel. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences* 117(14). 7684–7689.
- Lamoureaux, Siri, Michael Castelle & Anna Weichselbraun. 2025. Language machines: Toward a linguistic anthropology of large language models. *Journal of Linguistic Anthropology* 36(1). <https://doi.org/10.1111/jola.70033>.
- Leblebici, Didem. 2021. *Language ideologies in human-machine interaction: A qualitative study with voice assistant users*. Frankfurt (Oder): Europa-Universität Viadrina MA thesis.
- Lee, Carmen. 2016. *Multilingualism online*. London: Routledge.
- Ramati, Ido & Amit Pinchevski. 2018. Uniform multilingualism: A media genealogy of Google Translate. *New Media & Society* 20(7). 2550–2565.
- Rehm, Georg & Andy Way. 2023. Strategic Research, Innovation and Implementation Agenda for Digital Language Equality in Europe by 2030. In Georg Rehm & Andy Way (eds.), *European Language Equality: A Strategic Agenda for Digital Language Equality*, 387–412. Cham: Springer International Publishing.
- Schneider, Britta. 2022. Multilingualism and AI: The regimentation of language in the age of digital capitalism. *Signs and Society* 10(3). 362–387. <https://doi.org/10.1086/721757>.
- Schneider, Britta. 2024. A Sociolinguist's Look at the “Language” in Large Language Models. *Critical AI* 2(1). <https://doi.org/10.1215/2834703X-11205168>.
- Schneider, Britta, Bettina Migge, Marie-Theres Fester-Seeger & Christoph Purschke. 2026. AI technology as human interaction: linguistic and cultural framings of AI as communicative infrastructure. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-026-03193-5>.
- Truan, Naomi. 2026. AI is not multilingual. It ‘speaks’ many languages. That’s not the same. <https://languageunbothered.substack.com/p/ai-is-not-multilingual-it-speaks> (14 August, 2026.)
- Vieira, Lucas Nunes, Carol O’Sullivan, Xiaochun Zhang & Minako O’Hagan. 2023. Machine translation in society: Insights from UK users. *Language Resources and Evaluation* 57(2). 893–914.