Technolinguistics in practice: Socially situating language in AI systems

University of Siegen, | SFB 1187 | Herrengarten 3 | 57072 Siegen | www.mediacoop.uni-siegen.de

Date: May 24-27, 2023 Where: University of Siegen

Keynote Speakers: Paul Kockelman (Yale University); Ilana Gershon (Rice University); Morana Alač (University of California, San Diego) and Michael Castelle (University of Warwick)

New language technologies give rise to new technolinguistic practices, demanding a reconsideration of earlier questions and disciplinary commitments concerning the study of language and technology. The field of artificial intelligence (AI) has led to new communicative repertoires and ideologies for imagining, designing and interacting with machines as well as with humans. In the spirit of an 'ethnography of "cooperation" (cf. Hymes 1964) which situates communicative cooperation in the context of a wider community of practice, we are interested in: (1) how the fields of artificial intelligence (AI) and natural language processing (NLP) conceptualize and operationalize "language," by reproducing, regressing to, building on, challenging, updating, or otherwise engaging with the intellectual history of the field and its numerous critics, as well as in (2) how this operationalization transforms or is transformed by the socially-situated engagements between humans and machines in the sociocultural, political or economic contexts in which AI and ML models materialize. We aim to assemble scholars from a variety of fields to document and analyze evolving language and semiotic *practices* - the constitutive work that constructs "language" itself as a technology of artificial intelligence both within and surrounding AI and ML technologies by researchers, developers or other users.

Since Alan Turing, language has figured centrally in how the field of AI has understood the "intelligence" of machines even as the field has struggled to define "intelligence"-and with it, "language"-and to turn those definitions into machines that think, act, and speak. Recent critiques from across industry and academia have drawn widespread attention to the problem of "bias" in data sets and to the normative assumptions that shape researchers' engagements with language data (Bender et al 2021; Bolukbasi et al 2016; Sweeney 2013; Speer 2017; Noble, 2018; Nelson 2021). However important these critiques have been, we want to conceptualize language more broadly. By bringing together disparate strands of research from a variety of disciplines, we aim to develop a more complete picture of AI's commitments to particular conceptions of language that may structure research and guide the development and deployment of NLP technologies, and of how and why those commitments have evolved to incorporate or resist critiques from outside the field (see e.g. Bender and Kohler, 2020; Castelle, 2018; Hovy and Yang 2021; Suchman 1993; 2021; Mitchell, 2018; Slater 2018). Outside of AI, there is a long history of debate surrounding the nature of language that includes the generative Chomskian tradition (Chomsky, 1957) and its challenges ranging from philosophy of language (Wittgenstein 1958) and conversation analysis (Garfinkel and Sacks 1970) to functionalism (Givón 1989) and linguistic anthropology (Hymes 1964) to name only a few prominent examples. We ask if pragmatic approaches can help us make sense of the development and practice of language in AI today, especially given the increasingly ubiquitous presence of AI in our lives.

In view of the rapid uptake, diversification and influence of AI/NLP technologies, we call for broader engagements in critical AI research by inviting a range of practice-oriented perspectives in which language and other semiotic practices figure prominently. This can help to map out how materialized and socially-situated AI language technologies manifest or challenge conceptions of language both inside and outside the laboratory. In addition to the early work of interactional analysis in CSCW/HCI (e.g. Suchman 1987; 1993;



Button and Sharrock 1995) and contemporary applications in analyses of chatbots, or intelligent personal assistants (IPAs) (Alač et al. 2020; Pelikan and Broth 2016; Porcheron et al. 2018; Hector and Hrncal 2020), technolinguistic practices are scattered across a range of disciplines, and a host of "new media" practices (Gershon 2017). This includes work on the use of gendered registers in IPAs or robots (Lamoureaux and Hagerty forthcoming; Steele 2021), NLP benchmarks (Donahue, 2021a), and speech synthesis and recognition systems (Bell forthcoming; Li and Mills 2019), to studies of linguistic categories to taxonomize non-linguistic data, as seen in visual object, facial, and emotion recognition algorithms (Barocas and Selbst 2016; Denton et al. 2021, Donahue 2021b). Beyond the communicative capacities of AI systems themselves, we are interested in a broad take on technolinguistics, to include everything from the aggregate content of search engines generated by globe-spanning communities to the "metapragmatic" discourses and functions (Silverstein 1993) by researchers and publics of machine language to language ideologies inscribed in the structures of programming languages. Human languages and registers are inseparable from the interactional relations between humans, society and all of the "artificial" systems above. Discourses about "language" from casual conversation, to textbooks, or content moderation (Rieder and Skop 2021; Castelle 2018) feed into the design of language technologies, and the technologies too shape our modes of interacting with them.

We hope to expand dialogue across a range of disciplines including but not limited to the social sciences, computational sociolinguistics, information and media sciences, Human-Computer Interaction (HCI), Computer Supported Cooperative Work (CSCW). By connecting disparate case studies we'll develop a coherent understanding of how language is being transformed by, for, and with machines, and to map possible future engagements between language and technology that incorporate the rich histories of thought on both topics in a variety of disciplines. We are particularly interested in two fields not often in dialogue with each other: linguistic anthropology takes up a critical perspective on "language" and sees it as the outcome of "language ideological" practices (Silverstein 1979). STS accomplishes the same for technologies as objectified or inscribed materialities which come to have a material-semiotic life and power of their own in scientific, historical, political and economic constellations. Yet to date neither field attends to language form and use in artificial intelligence (important exceptions include: Suchman 1987; 1993; Kockelman 2014; 2017; 2020; Collins 2018). Linguistic anthropology neglected computer science in favor of "natural" languages (Kockelman 2014), while posthumanist trends in STS have rhetorically favored 'material-semiotics' set in opposition to language-as-representation (Lynch 1994; Pinch 2010; Alač 2011; Lamoureaux forthcoming). And yet, these fields could decisively advance this project. In their respective vocabularies, both ground their methodological approach in practice, attending to the (technologically) mediated semiotic work of meaning-making in situated, embodied and context-bound action, and the recursive processes of objectification and recontextualization. Such approaches invite analyses of discursive, textual or other (human or material) technolinguistic practices that go into designing, testing and implementing "language" in AI technologies and their circulation in society. Further, a feminist STS approach would seek to dissolve the binaries of mind and body, the seeming stability of categories, make visible the labors (or practical work) behind linguistic technologies, and identify sites where they become fetishized and taken for granted as universal, isolable forms of knowledge (Haraway 1991; Suchman 2007).

Contributions can address the following themes (or propose a related one):

- Ethnographic descriptions of practical engagements with NLP/AI systems.
- How sociocultural, political-economic or historical contexts reveal ideologies inscribed in AI/NLP.
- How notions of "language" become taken-for-granted, stabilized, commodified and circulated.
- Reflections on the disciplinary and/or conceptual trends addressing NLP/AI
- The social harms (e.g. race, class, gender) that such language notions help to enact and forms of resistance to them.
- A consideration of "new" vs. "old" communicative practices.
- How actors negotiate autonomy and agency with post-human conceptions of the subject.



• Alternative views on language and communication offering inspiration for new socio-technical imaginings.

Please submit **abstracts of 300 to 500 words** along with a **title** and a **short biography** (max. 150 words) by **December 1, 2022, to** <u>technolinguistics@sfb1187.uni-siegen.de</u>

Organization: Siri Lamoureaux, Evan Donahue, Sarah Bell, David Waldecker, Susanne Förster, Marcus Burkhardt, Yarden Skop

References

Alač, Morana. 2011. Handling Digital Brains: A Laboratory Study of Multimodal Semiotic Interaction in the Age of Computers. Inside Technology. Cambridge, Mass: MIT Press.

Alač, Morana, Yelena Gluzman, Tiffany Aflatoun, Adil Bari, Buhang Jing, and German Mozqueda. 2020. "Talking to a Toaster: How Everyday Interactions with Digital Voice Assistants Resist a Return to the Individual." *Evental Aesthetics* 9 (1), 3-53.

Barocas, Solon, and Andrew D. Selbst. 2016. "Big Data's Disparate Impact." California Law Review 104 (3), 671-732. <u>https://doi.org/10.15779/Z38BG31</u>.

Bell, Sarah A. forthcoming. Vox ex Machina: A History of Voice Synthesis Cambridge, Mass.: MIT Press.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and 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*, 610–23. Virtual Event Canada: ACM. https://doi.org/10.1145/3442188.3445922.

Bender, Emily M., and Alexander Koller. 2020. "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–98. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.463.

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. 2016. "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings." *ArXiv:1607.06520 [Cs, Stat]*, July. <u>http://arxiv.org/abs/1607.06520</u>.

Button, Graham, and Wes Sharrock. 1995. "On simulacrums of conversation: Toward a clarification of the relevance of conversation analysis for human-computer interaction." In *The social and interactional dimensions of human-computer interfaces*, 107-25.

Castelle, Michael. 2018. "The Linguistic Ideologies of Deep Abusive Language Classification." In *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, 160–70. Brussels: Association for Computational Linguistics. <u>https://doi.org/10.18653/v1/W18-5120</u>.

Chomsky, Noam. 1957. Syntactic Structures. The Hague: Mouton & Co.



Collins, H. M. 2018. Artifictional Intelligence: Against Humanity's Surrender to Computers. Cambridge, UK: Polity Press.

Denton, Emily, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole. 2021. "On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet." *Big Data & Society* 8 (2): 205395172110359. <u>https://doi.org/10.1177/20539517211035955</u>.

Donahue, Evan. 2021a. Towards an Android Linguistics: Pragmatics, Reflection, and Creativity in Machine Language. Proceedings.

Donahue, Evan. 2021b. "Theseus in the Epistemic Labyrinth: A Critical History of the Semantic Differential Method and the Apparent Weight of Color" In Machine Intelligences in Context: Beyond the Technological Sublime, special issue of Culture Machine.

Garfinkel, Harold, and Harvey Sacks. 1970. "On Formal Structures of Practical Actions." In *Theoretical Sociology*, edited by J.D. McKinney, and E.A. Tiryakian, 337–66. New York: Appleton-Century Crofts.

Gershon, Ilana. 2017. "Language and the Newness of Media." *Annual Review of Anthropology*46 (1), 15–31. https://doi.org/10.1146/annurev-anthro-102116-041300.

Givón, Talmy. 1989. *Mind, Code and Context: Essays in Pragmatics*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Hector, Tim M. and Hrncal, Christine. 2020. Intelligente Persönliche Assistenten im häuslichen Umfeld. Universität Siegen: Collaborative Research Center 1187 Media of Cooperation.

Hymes, Dell. 1964. "Introduction: Toward Ethnographies of Communication." *American Anthropologist* 66 (6.2), 1–34. <u>https://doi.org/10.1525/aa.1964.66.suppl_3.02a00010</u>.

Haraway, Donna Jeanne. 1991. Simians, Cyborgs, and Women: The Reinvention of Nature New York: Routledge.

Hovy, Dirk, and Divi Yang. 2021. "The Importance of Modeling Social Factors of Language: Theory and Practice." In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics*, 588–602. Association for Computational Linguistics.

Kockelman, Paul. 2020. "The Epistemic and Performative Dynamics of Machine Learning Praxis." Signs and Society 8 (2), 319–55. <u>https://doi.org/10.1086/708249</u>.

Kockelman, Paul. 2017. The Art of Interpretation in the Age of Computation New York: Oxford University Press.

Kockelman, Paul. 2014. "Linguistic Anthropology in the Age of Language Automata." In *The Cambridge Handbook of Linguistic Anthropology* edited by N.J. Enfield, Paul Kockelman, and Jack Sidnell, 708–33. Cambridge, UK: Cambridge University Press. <u>https://doi.org/10.1017/CBO9781139342872.034</u>.

Lamoureaux, Siri, and Alexa Hagerty. forthcoming. "Women, Machines, and Dangerous Things: Animating Intelligent Personal Assistants as Semantico-Pragmatic Violence." In *The Gender of Things: How Technologies and Epistemic Objects Become Gendered*, edited by Maria Rentetzi and Aida Bosch. New York; London: Routledge.

Lamoureaux, Siri. forthcoming. "The Quiet Posthuman: How Social Science Lobotomized Language and Materialized a Monster." ms.



Li, Xiaochang, and Mara Mills. 2019. "Vocal Features: From Voice Identification to Speech Recognition by Machine." *Technology and Culture* 60 (2S), 129–60. <u>https://doi.org/10.1353/tech.2019.0066</u>.

Lynch, Michael. 1994. "Representation Is Overrated: Some Critical Remarks about the Use of the Concept of Representation in Science Studies." *Configurations* 2 (1), 137–49. <u>https://doi.org/10.1353/con.1994.0015</u>.

Mitchell, Christine. 2018. "Whether Something Works: Finding the Human in the Margins of Machine Translation." In: *Amodern*, 8.

Nelson, Laura K. 2021. "Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South." *Poetics* 88 (October): 101539. <u>https://doi.org/10.1016/j.poetic.2021.101539</u>.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression. How Search Engines Reinforce Racism.* New York: New York University Press. <u>https://www.degruyter.com/isbn/9781479833641</u>.

Pelikan, Hannah R.M., and Mathias Broth. 2016. "Why That Nao?: How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4921–32. San Jose, California: ACM. https://doi.org/10.1145/2858036.2858478.

Pinch, Trevor. 2010. "On Making Infrastructure Visible: Putting the Non-Humans to Rights." *Cambridge Journal of Economics* 34 (1), 77–89. <u>https://doi.org/10.1093/cje/bep044</u>.

Porcheron, Martin, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. "Voice Interfaces in Everyday Life." In *Proceedings of the 2018 CHI Conference onHuman Factors in Computing Systems*, 1–12. Montreal,: ACM. https://doi.org/10.1145/3173574.3174214.

Rieder, Bernhard, and Yarden Skop. 2021. "The Fabrics of Machine Moderation: Studying the Technical, Normative, and Organizational Structure of Perspective API." *Big Data & Society* 8 (2): 205395172110461. https://doi.org/10.1177/20539517211046181.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. "A Simplest Systematics for the Organization of Turn-Taking for Conversation." *Language* 50 (4), 696–735. <u>https://doi.org/10.1353/lan.1974.0010</u>.

Silverstein, Michael. 1979. "Language Structure and Linguistic Ideology." In *The Elements: A Parasession on Linguistic Units and Levels*, edited by Paul R. Clyne, William F. Hanks, and Carol Hofbauer, 193-247. Chicago: Chicago Linguistic Society.

Silverstein, Michael. 1993. "Metapragmatic Discourse and Metapragmatic Function." In Reflexive Language: Reported Speech and Metapragmatics, edited by John A. Lucy, 33–58. Cambridge, UK: Cambridge University Press.

Slater, Avery. 2018. "Crypto-Monolingualism. Machine Translation and the Poetics of Automation" In *Amodern*, 8.

Speer, Robyn. 2017. "How to Make a Racist AI without Really Trying." *ConceptNet Blog* (blog), July 13, 2017. http://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/.

Steele, Jamie. 2021. "Playful Disasters: Mimetic Language, Gender and the Fantasy of Agential Machines." Talk presented at the *2021 4S Conference*: Toronto.



Suchman, Lucy. 2007. "Feminist STS and the Sciences of the Artificial." In *The Handbook of Science and Technology Studies*, edited by L. Suchman, E. Hackett, O. Amsterdamska, Mike Lynch, and Judy Wajcman, 3rd ed., 139–64. Cambridge, Mass.: MIT Press.

Suchman, Lucy. 1993. "Do categories have politics? The language/action perspective reconsidered." In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy CSCW 93.* Dordrecht: Springer.

Suchman, Lucy. 1987. Plans and Situated Actions: The Problem of Human-Machine Communication Cambridge, UK; New York: Cambridge University Press.

Sweeney, Latanya. 2013. "Discrimination in Online Ad Delivery." In *SSRN Electronic Journal*. <u>https://doi.org/10.2139/ssrn.2208240</u>.

Wittgenstein, Ludwig. 1958. Philosophical Investigations. Oxford: Basil Blackwell.

