Redirecting Linguistic Attention: From LLM Encoding to Interaction and Impact

Kai Kugler
Trier University
kuglerk@uni-trier.de

Abstract

Linguistics has engaged large language models (LLMs) predominantly as objects of representational analysis, addressing them as models *of* language while leaving two more urgent questions largely unasked. First, LLM-generated text is statistically distinguishable from human writing across lexical, syntactic, and discourse dimensions, but rarely interpreted through the lens of register theory. We argue that it should be, with a crucial qualification: LLM output constitutes not a single stable register but a register space indexed to situational parameters including model architecture, task framing, and the interlocutor identity the model infers from contextual cues. Second, human-LLM interaction is bidirectionally accommodative: users apply Gricean defaults and convergence mechanisms to a system that simultaneously infers and adapts to their identity, producing a feedback loop based not on communicative negotiation but on the statistical properties of the model's training distribution. This position paper argues for redirecting attention from what LLMs encode to what they produce and what they do to the humans who interact with them, and proposes a tripartite research agenda centered on register-space description, interactional discourse analysis, and methodological reflexivity in the face of the observer effects that LLM research itself generates.

1 Introduction

The rapid diffusion of large language models into communicative practice presents linguistics with a question that goes beyond the technical: what kind of language do these systems produce, and what happens to human language when people routinely interact with them? The discipline has engaged LLMs primarily as objects of representational analysis: probing whether contextualised embeddings encode structure (Tenney et al., 2019), or whether next-token prediction suffices to acquire grammatical competence (Linzen et al., 2016) and as a test case in the debate between innatist and usage-based accounts of language acquisition. These are legitimate pursuits, and a growing body of work on features of LLM register (Muñoz-Ortiz et al., 2024; Milička et al., 2025) and on the limitations of using LLMs as human proxies (Tjuatja et al., 2024; Münker, 2025) has begun to map the empirical terrain. Within Conversational Analysis (CA), Human-AI-talk itself is already an established object of study, where chatbots and other conversational technologies are examined as interactional participants (Stokoe et al., 2024).[1]

What remains comparatively underdeveloped is a synthesis that connects interactional evidence to register-level findings (see Section 3), treating LLM output as an object of register description and human-LLM interaction as a site of sociolinguistic change shaped by that register-space. This paper argues that such a framing is overdue: (Computational) Linguistics has studied LLMs extensively as models *of* language, and, within Conversation Analysis, increasingly as interactional participants. We argue that the register-level evidence is precisely where these strands could meet.

LLMs are, in everyday practice, text-generating systems embedded in communicative situations. The questions that arise from this reality are what register they produce and spread, and what interactional and cognitive consequences their use has for human speakers. These are not just legitimate concerns but core linguistic problems with two main points standing out:

The first is textual: LLM output is measurably distinct from human writing on many linguistic dimensions, and these differences are often stable enough across model families and genres to warrant description as a register in the technical sense (Biber, 1988). This register in academic, profes-

[1]The EMCAI network brings together much of this ongoing work: https://emcai.conversationanalysis.org/

sional, and public discourse already perturbs the corpora from which descriptions of language are derived (Shumailov et al., 2023).

The second is interactional: humans communicate with LLMs at large scale and Gricean defaults, audience design, and the mechanisms of linguistic accommodation are activated in human-LLM interaction. Their consequences for how people formulate language, form intentions, and construct communicative norms remain almost unstudied. But they are, we argue, an urgent frontier for linguistics to address.

The paper is structured as follows: Section 2 situates the argument within existing debates on LLMs and language. Section 3 reviews evidence on LLM output as a distinct register. Section 4 analyzes the interactional loop in detail. Section 5 proposes a concrete research agenda, specifying what is novel in the program we advance.

2 Background: What LLMs Model and What They Do Not

LLMs are trained on the objective of next-token prediction over large text corpora. Their architecture does not encode explicit symbolic grammar, no referent, and no communicative intention; what they acquire is a high-parametric implementation of the distributional hypothesis (Harris, 1954; Firth, 1957). This makes them the most elaborate empirical test yet of the usage-based tradition (Tomasello, 2003). Probing studies suggest that transformer representations encode substantial structural information, with recoverable syntactic features (Tenney et al., 2019) and sensitivity to agreement and binding effects that challenge arguments of poverty-of-the-stimulus (Warstadt et al., 2020).

However, inference from model properties to claims about language requires careful evaluation. Bender et al. (2021) capture this with the "stochastic parrot" metaphor: LLMs manipulate linguistic form without access to meaning in any referential sense. This metaphor operates at the level of cognitive classification rather than communicative practice: it does not capture what LLM output is nor what happens to human language in interaction with it. We take both questions in turn.

3 LLM Output as a Linguistic Register

A converging body of evidence establishes that LLM-generated text is statistically smoother than human writing: more uniform in sentence length, more uniform and restricted in vocabulary, more regular in cohesion and information structure (Muñoz-Ortiz et al., 2024; Martínez et al., 2025; Guo et al., 2024). The pattern is consistent across model families (and text genres) and extends to human writers working with LLM assistance, whose output shows measurable reductions in linguistic diversity (Padmakumar and He, 2024). These findings have motivated an emerging consensus that LLM output constitutes a distinct register in the technical sense (Biber, 1988): a variety associated with a specific situational configuration, stable enough to be characterised independently of model type or task.

However, the consensus overlooks a structural feature of LLM behaviour: the sensitivity of model output to the inferred identity of the interlocutor. Törnberg and Schimmel (2026) demonstrate this with particular force: varying only the stated identity of the interlocutor shifts model responses by 28 to 62 percentage points on attitudinal measures, moving all six models from left- to right-leaning positions. Accommodation is strongly asymmetric: rightward shift is approximately 8.0$\times$ larger than leftward shift. The authors interpret this as audience accommodation rather than a fixed ideological disposition. In linguistic terms, audience design (Bell, 1984) is executed without communicative intention, as a statistical reflex of training distributions rather than a speaker's orientation to an addressee.

What existing corpus work has described as LLM register is, in this light, a default output profile: the statistical centroid of a substantially large space of variation indexed to interlocutor identity. The default profile may be not arbitrary but the predictable outcome of training a single model across maximally diverse tasks and user populations, where stylistic convergence toward a broadly functional mean is a structural consequence of the optimisation objective. What varies systematically around this default, and what existing corpus work has not measured, is the degree to which inferred user identity shifts the output away from it. The appropriate unit of analysis is therefore not a single register but a register-space, structured along dimensions of model architecture, instruction-tuning, task framing, and, critically, inferred user identity.

These dynamics have sociolinguistic consequences beyond the register-space itself. As LLM-generated text today expands in public discourse (Shumailov et al., 2023), the descriptive baseline of

corpus linguistics is not simply being contaminated by a non-human variety but selectively reshaped in ways calibrated to the user populations that generate it and that current sampling frameworks are perhaps not well equipped to detect.

4 The Interactional Feedback Loop

4.1 Gricean defaults and their misapplication

Human communication is structured by cooperative defaults. Grice's maxims (Grice, 1975) do not function as explicit rules; they are interactional expectations activated by the presence of a responsive interlocutor. When a person converses with an LLM, these defaults are activated. Utterances are interpreted for relevance and intention, and contradictions are resolved by attributing a coherent communicative purpose to the model while pragmatic failures, in turn, tend to be read as conversational breakdown rather than as statistical noise.

Users over-interpret LLM output, attributing semantic precision and intentional structure to responses that are statistically generated without either. They under-specify context on the assumption that common ground has been established (Clark and Brennan, 1991), leading to communicative miscalibration. This dynamic has a parasocial dimension: Horton and Wohl (1956) describe how media audiences construct quasi-social relationships with media figures on the basis of one-sided communicative exposure. Human-LLM interaction intensifies this effect by introducing responsiveness, a feature absent from broadcast media, without introducing genuine intersubjectivity.

What the Gricean analysis must now accommodate, however, is that the interaction is not simply a case of a human applying cooperative defaults to a passive system. Evidence from Törnberg and Schimmel (2026) indicates that LLMs simultaneously infer the identity of their interlocutor and calibrate their output accordingly. The interaction is therefore mutually accommodative in a structural sense, though asymmetrically so. It is a pragmatic illusion: the model's accommodation is not grounded in communicative intention but in statistical regularities of the training distribution. This creates a situation for which Gricean pragmatics, designed for intentional agents, provides only partial analytical solution.

4.2 Mutual mirroring without ground

The interlocutor-sensitivity documented by Törnberg and Schimmel (2026) is one instance of a broader behavioural pattern that the AI alignment literature terms *sycophancy*: the tendency of instruction-tuned models to produce outputs that match the user's perceived preferences (Sharma et al., 2024), a tendency reinforced by RLHF insofar as human evaluators prefer outputs that agree with their expressed views (Bai et al., 2022). From a linguistic standpoint, sycophancy is not a failure of calibration but a reconfiguration of the communicative situation. An interlocutor whose output is systematically oriented toward user preference is not a cooperative interlocutor in Grice's sense; its apparent cooperation is structurally decoupled from the norms that give cooperation its communicative value.

The human side of this loop is equally well documented. Interactional sociolinguistics has established the pervasive tendency toward linguistic convergence: speakers adapt their phonology, lexis, and syntax towards their interlocutor through accommodation (Giles et al., 1991) and priming (Pickering and Garrod, 2004). Padmakumar and He (2024) show these mechanisms operate robustly in human-LLM interaction: users simplify syntactic structures, prefer lexical items the model handles reliably, and adopt the register that LLMs characteristically produce. Taken together, these two processes constitute a feedback loop of mutual mirroring: the model reflects the user's inferred identity back to them; the user converges toward the model's output style. Neither side provides a stable communicative ground for the other. The model's accommodation is not grounded in intention but in training statistics (Grieve et al., 2025); the user's target of convergence is, as Section 3 argues, a moving profile that shifts with the identity cues the user is emitting.

This mirroring extends beyond surface style to the level of communicative intention. Effective prompting requires decomposing one's goals into explicit, delimited sub-specifications, a practice at odds with the implicit, incremental way human intentions are normally formed and expressed (Levinson, 1983). The user who adapts their lexis to the model is also reorganising their communicative goals around the model's input architecture. Whether this restructuring of intention persists beyond LLM interaction, and how it varies across

users and contexts, is among the open questions this feedback loop raises.

5 A Three-Fold Research Agenda

The analysis above points toward a research agenda that is broad in scope and, in places, genuinely difficult to operationalise.

1. LLM output as a structured object of variation. The register-space framing implies that no corpus of LLM output is neutral: every sample encodes assumptions about the interlocutor, task, model, and configuration. These must be treated as systematic variables, not background conditions. The methodological model is factorial experimental design (Törnberg and Schimmel, 2026), adapted for corpus-linguistic purposes: systematic variation of situational parameters to map the structure of the register-space rather than describe a single point within it. The difficulty should be stated directly: any description of LLM language that does not control for these dimensions risks characterising a sampling artifact. The complexity of the register space is itself significant. LLM text accumulating in public discourse is not a single non-human variety but a highly variable one calibrated to its users. Beyond description, a theory is needed of why certain situational dimensions produce large output shifts while others do not. This requires integrating register theory with distributional learning and alignment dynamics that the discipline does not currently possess.

2. Human-LLM interaction as a site of interactional discourse analysis. The interactional consequences described in Section 4 call for methods that treat the human-LLM exchange as a unit of analysis in its own right, an approach Conversation Analysis has already established. Work on conversational technologies, chatbots, and voice assistants as interactional participants is accordingly a growing field (Stokoe et al., 2024).

What this literature has not yet systematically addressed is the register-space sensitivity documented in Section 3: how mirroring, convergence, and accommodation phenomena interact with a system whose output shifts with the interlocutor identity it infers. Connecting CA methods to this register-level variation – for instance, by tracking whether accommodation patterns differ depending on the output profile – is a specific extension of this existing tradition that we see as tractable and underexplored. Key empirical questions, such as how users' linguistic strategies change over sustained LLM use, whether convergence and mirroring effects persist in non-LLM contexts, and how communicative failure is attributed, are in principle addressable through corpus collection, controlled extraction, and the conversation-analytic methods already established in this field.

3. Methodological reflexivity as a disciplinary requirement. The observer-effect findings of Törnberg and Schimmel (2026) have direct implications for linguistic research: the prompt, framing, and implied identity of the researcher are situational cues that calibrate model output, meaning any description of LLM language is partly a description of the research situation that produced it. The social sciences have a long tradition of modelling interviewer effects and measurement-response interactions (Anderson et al., 1988); a comparable reorientation is required here, treating the researcher's situational position not as a disruptive impact but as a constitutive feature of the data.

6 Conclusion

This paper has argued that the linguistic study of LLMs must move beyond the analysis of representational encoding to address the pragmatic and sociolinguistic realities of their use. By framing LLM output not as a monolithic variety but as a dynamic *register-space*, we account for the radical situational sensitivity and audience accommodation inherent in these systems. This shift reveals a communicative landscape characterized by a profound irony: while users project Gricean intentionality onto the model, the model executes a statistical simulation of the user.

The resulting feedback loop, where human convergence meets algorithmic sycophancy, threatens to destabilize the empirical foundations of corpus linguistics and the cognitive structures of human interaction. The proposed three-fold agenda provides the necessary framework to navigate this transition. If linguistics is to remain the science of language as it is actually used, it must develop the tools to analyze a world where communication no longer requires a mindful interlocutor, and where the observer effect of our own research is a constitutive feature of the data. The task is not merely to understand how models represent language, but to track how they are fundamentally restructuring the linguistic ecosystem.

Limitations

The arguments presented here are subject to several constraints. First, the opacity of proprietary models, which constitute the majority of frontier LLMs, limits our ability to distinguish between linguistic patterns emerging from the base training distribution and those artificially induced by safety filtering or RLHF. Our analysis of the register space is therefore a description of an output layer whose deeper causal mechanisms remain obscure.

Second, the temporal volatility of LLM development means that any empirical mapping of a register space may be rendered obsolete by new architectural paradigms. Finally, a focus on high-resource languages (predominantly English) leaves open the question of whether the communicative miscalibration we describe manifests differently in low-resource or typologically distant languages, where the model's statistical foundation is significantly more sparse.

Ethical Considerations

The redirection of linguistic attention toward human-LLM interaction carries significant ethical implications. The phenomenon of asymmetric accommodation and model sycophancy raises concerns regarding the creation of confirmatory feedback loops: if systems systematically reflect user identity and bias, they may reinforce existing prejudices under the guise of neutral interaction. This creates a risk of cognitive erosion, where the effort of communicative negotiation is replaced by a frictionless, confirmatory feedback loop.

Furthermore, the pollution of the linguistic commons and the proliferation of synthetic registers in public corpora pose an ethical challenge for future researchers. There is a danger that marginalized linguistic varieties will be further sidelined as LLM outputs converge toward a standardized mean of their training data. We emphasize that researchers have a responsibility to disclose the observer effects of their prompts, as failing to do so contributes to a distorted understanding of what these models are and what they represent.

References

Barbara Anderson, Brian Silver, and Paul Abramson. 1988. The effects of race of the interviewer on measure of electoral participation by blacks in SRC national elections studies. *Public Opinion Quarterly*, 52:53–83.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, Nicholas Joseph, Saurav Kadavath, Jackson Kernion, Tom Conerly, Sheer El-Showk, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Tristan Hume, and 12 others. 2022. Training a helpful and harmless assistant with reinforcement learning from human feedback. *Preprint*, arXiv:2204.05862.

Allan Bell. 1984. Language style as audience design. *Language in Society*, 13(2):145–204.

Emily M. Bender, 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*, pages 610–623.

Douglas Biber. 1988. *Variation Across Speech and Writing*. Cambridge University Press.

Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. In L. B. Resnick, J. M. Levine, and S. D. Teasley, editors, *Perspectives on Socially Shared Cognition*, pages 127–149. APA.

John R. Firth. 1957. A synopsis of linguistic theory 1930–1955. In *Studies in Linguistic Analysis*, pages 1–32. Blackwell.

Howard Giles, Justine Coupland, and Nikolas Coupland. 1991. *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge University Press.

H. Paul Grice. 1975. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics, Vol. 3: Speech Acts*, pages 41–58. Academic Press.

Jack Grieve, Sara Bartl, Matteo Fuoli, Jason Grafmiller, Weihang Huang, Alejandro Jawerbaum, Akira Murakami, Marcus Perlman, Dana Roemling, and Bodo Winter. 2025. The sociolinguistic foundations of language modeling. *Frontiers in Artificial Intelligence*, 7 - 2024.

Yanzhu Guo, Guokan Shang, Michalis Vazirgiannis, and Chloé Clavel. 2024. Benchmarking linguistic diversity of large language models. *Transactions of the Association for Computational Linguistics*.

Zellig Harris. 1954. Distributional structure. *Word*, 10(2–3):146–162.

Donald Horton and Richard Wohl. 1956. Mass communication and para-social interaction. *Psychiatry*, 19(3):215–229.

Stephen C. Levinson. 1983. *Pragmatics*. Cambridge University Press.

Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. 2016. Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, 4:521–535.

Gonzalo Martínez, José Alberto Hernández, Javier Conde, Pedro Reviriego, and Elena Merino-Gomez. 2025. Beware of words: Evaluating the lexical diversity of conversational llms using chatgpt as case study. *ACM Trans. Intell. Syst. Technol.*, 16(6).

Jiří Milička, Anna Marklová, and Václav Cvrček. 2025. Benchmark of stylistic variation in llm-generated texts. *Preprint*, arXiv:2509.10179.

Simon Münker. 2025. Fingerprinting LLMs through survey item factor correlation: A case study on humor style questionnaire. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 245–258, Suzhou, China. Association for Computational Linguistics.

Alberto Muñoz-Ortiz, Carlos Gómez-Rodríguez, and David Vilares. 2024. Contrasting linguistic patterns in human and llm-generated news text. *Artificial Intelligence Review*, 57(10):265.

Vishakh Padmakumar and He He. 2024. Does writing with language models reduce content diversity? In *International Conference on Learning Representations*, volume 2024, pages 642–669.

Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190.

Mrinank Sharma, Meg Tong, Tomasz Korbak, and 1 others. 2024. Towards understanding sycophancy in language models. In *International Conference on Learning Representations (ICLR)*.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. 2023. The curse of recursion: Training on generated data makes models forget. *arXiv preprint arXiv:2305.17493*.

Elizabeth Stokoe, Saul Albert, Hendrik Buschmeier, and Wyke Stommel. 2024. Conversation analysis and conversational technologies: Finding the common ground between academia and industry. *Discourse & Communication*, 18(6).

Ian Tenney, Dipanjan Das, and Ellie Pavlick. 2019. Bert rediscovers the classical nlp pipeline. In *Proceedings of ACL 2019*, pages 4593–4601.

Lindia Tjuatja, Valerie Chen, Tongshuang Wu, Ameet Talwalkwar, and Graham Neubig. 2024. Do LLMs exhibit human-like response biases? a case study in survey design. *Transactions of the Association for Computational Linguistics*, 12:1011–1026.

Michael Tomasello. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Petter Törnberg and Michelle Schimmel. 2026. Political bias audits of llms capture sycophancy to the inferred auditor. *Preprint*, arXiv:2604.27633.

Alex Warstadt, Alicia Parrish, Haokun Liu, Anhad Mohananey, Wei Peng, Sheng-Fu Wang, and Samuel R. Bowman. 2020. BLiMP: The benchmark of linguistic minimal pairs for English. *Transactions of the Association for Computational Linguistics*, 8:377–392.