

LARGE LANGUAGE MODELS AND THE VARIETIES OF MEANING

Mitchell Green¹ 

¹ University of Connecticut, Department of Philosophy

Abstract

This article responds to Steffen Koch's, 'Babbling stochastic parrots? A Kripkean argument for reference in large language models' (this journal). Koch attempts by appealing to the causal theory of reference to show that when LLM's use (or seem to use) proper names, they refer to what those names standardly refer to in the language from which they were drawn; Koch then suggests ways to generalize the strategy beyond proper names to kind terms, both natural and social. We argue here that Koch's argumentative strategy is needlessly controversial and even if successful, would be of limited scope. We offer instead a more direct argument for the conclusion that LLM's use language (rather than merely appearing to do so) because they meet the conditions required to perform rudimentary acts that Austin (1975) termed *phatic acts*. The argument applies to language generally, and implies that when so used, that language carries its usual linguistic meaning.

Keywords

Large language models, Causal Theory of Reference, chatbots, speech act theory, phatic act, locutionary act, speaker meaning, linguistic meaning.

1. Preliminaries: Two kinds of meaning and three ways of using language

Different notions of meaning are intuitively familiar yet apt to confuse us when we try to theorize about language and communication.¹ To safeguard against such confusion, it helps to draw two sets of distinctions: (1) between linguistic meaning and speaker meaning, and (2) among phatic, locutionary, and illocutionary acts. To elucidate the first of these: Linguistic meaning concerns what morphemes and/or the words and phrases built from them are conventionally used to say in a particular community. 'Crepuscular' linguistically means 'active during dawn and dusk'; 'Buffalo buffalo buffalo buffalo' has a linguistic meaning enabling it to be a grammatical, indicative sentence. By contrast, 'speaker meaning' refers to what a speaker aims to get across in a particular communicative act, often by means of their addressees' recognition of that aim.² Linguistic meaning is a useful abstraction from large patterns of communicative behavior, and has a limited degree of autonomy from it.³ The linguistic/speaker meaning distinction enables us to say, for instance, that 'crepuscular' has a definite meaning even if most English speakers don't know how to use that

1 My thanks to Dr. Jan Michel for insightful comments on an earlier draft of this essay.

2 'Speaker meaning' the term more commonly used than its synonymous predecessor, 'non-natural meaning', coined by (Grice, 1989). We also say for instance that the verdigris on this bronze statue means that it has been exposed to the elements for decades. Yet this notion of meaning—often called natural meaning—has little to do *per se* with communication, though it can be recruited for such purposes. See Green (2016) for further discussion.

3 For further discussion on the autonomy of linguistic meaning, see Green (2001).

word. Also, the linguistic/speaker meaning distinction enables us to observe that a single, or even a small number of speakers, cannot change a word's or phrase's linguistic meaning by using it in a non-standard way. Yet, if enough speakers start using a word or phrase in a new way, it will likely come to have that meaning. The phrase 'beg the question' once meant (only), 'assume a premise that is up for debate.' For some time, speakers misused that phrase in using it to mean 'raise the question.' However, as a result of a pervasive pattern of (what were at first) misuses, the phrase now also means 'raise the question,' and thus is polysemous.

A special case of linguistic meaning is semantic reference: this concerns the conventional meaning of a referring expression like a pronoun or proper name. Similarly, a special case of speaker meaning is speaker reference (Kripke, 1977), what a speaker deliberately refers to (with or without the aid of a referring expression). Semantic reference and speaker reference can come apart, as in everyday cases in which a speaker mis-describes what she means to refer to. "That deerhound is beautiful," says my neighbor, intending to draw my attention to what is in fact a wolfhound in a nearby yard. She refers—"speaker refers"—to the wolfhound even though her words semantically refer to dogs of a different breed.

In addition to the linguistic meaning (reference/speaker meaning) distinction, it is also helpful to recall Austin's distinctions among phatic, locutionary, and illocutionary acts (Austin, 1975). In a *phatic act*, a speaker makes an utterance in a way that is sensitive to the part(s) of speech contained in the grammatical structure (if any) of the expression used. Without knowing its meaning, I might utter the Swahili sentence, 'Mbwa anabweka,' in a way sensitive to the fact that its first term is a noun and the second is a verb, and that the entire string is an indicative sentence. (That sensitivity might be shown in my agreement that if 'anakimba' is another verb in Swahili, then 'Mbwa anakimbia' is also a grammatical sentence of that language.) In so doing, I have performed a phatic act.⁴

One symptom of an act's being phatic is that it can be reported in direct discourse: A third party might report, "Mitch said, 'Mbwa anabweka.'" Further, when so quoted, these words retain their linguistic meaning: In the phatic act of uttering 'Mbwa anabweka,' 'Mbwa' refers to dogs, and so on. That is why we can, for instance, evaluate as true or false things we ascribe to others in direct discourse even if they don't know what those words mean: You can say something true even if you don't know what you are saying, whereas if those words did not carry their usual linguistic meaning, this phenomenon would be mysterious. By contrast, if dust settles on the floor in such a way as to spell out the words, 'Mbwa anabweka,' no phatic act has been performed. The reason is that the etiology of that configuration was not sensitive in any way to its grammatical structure. The result is that when a randomly produced configuration of dust on a countertop appears to spell out 'Mbwa anabweka,' there are in fact no words there.⁵

Although I can perform a phatic act without knowing the meaning of the phrase I utter, if I perform such an act in cognizance of that sentence's meaning, I have performed a *locutionary act*. Locutionary acts may be reported by direct discourse attributions, but may also be reported in indirect discourse. In the Swahili case, that would be, "Mitch said that the dog barks." Note that one can say such things as that a dog barks without committing oneself to that proposition: We see this when actors say their lines on stage. Instead, to perform an *illocutionary act*, the speaker commits herself to the content of the sentence uttered. We report such committal acts with indirect discourse ascriptions employing speech act verbs, such as "Mitch claimed (asserted, reported, etc.) that the dog barks."

Many illocutionary acts require complex cognitive and conative conditions to be met on the part of their producers, and for this reason, there has been controversy over whether

4 J. L. Austin, who introduced the terminology explained here, writes, "The phatic act is the uttering of certain vocables or words of certain types, belonging to and as belonging to, a certain vocabulary, conforming to and as conforming to a certain grammar." (Austin, 1975, page 95).

5 This conclusion is compatible with holding that the countertop configuration could over time be artifactualized so as to acquire linguistic meaning. If so, the situation would be similar to one in which constellations of stars come to be used by human beings to represent bears, ladles, and swans.

machines can illocute (Butlin & Viebahn, 2024; M. Green & Michel, 2022; Gubelmann, 2024; Millièrè & Buckner, 2024). For purposes of this essay, I shall remain agnostic on the capacity of LLM's to perform illocutionary acts. To a lesser degree, we might find controversy over the question whether machines can locute: this will of course turn on whether a machine can understand language, a question we shall also sedulously avoid in what follows. By contrast, the performance of a phatic act requires clearing a lower bar. As we noted, performing a phatic act requires using language in a manner constrained by its grammar. But then a machine designed to convey information by means of conventional language, and in a way sensitive to grammatical structure and different parts of speech, will also perform phatic acts. This is pertinent to recent developments in LLM technology, because we now have strong evidence that these machines can acquire sophisticated syntactic competence based on relatively modest exposure to training data (Millièrè, 2024). Accordingly, we have substantial grounds for concluding that when LLM's produce what appears to be language, they are (at the very least) performing phatic acts, and thus appearances are not deceptive: the word- and phrase-like patterns that these machines generate really are words and phrases, and carry their standard linguistic meaning when so generated. Thus, when LLM's produce proper names and other referring expressions, those expressions do refer in a way dictated by their linguistic meaning. That is, these expressions semantically refer when used by LLM's, even if our philosophical scruples lead us to dress those expressions in quotation marks when reporting these machines' output.

2. Avoiding a tempting muddle

In recent years several authors have argued that in spite of their impressive achievements, LLM's such as Chat GPT3 do not mean anything. A prominent example of this style of argument is Bender *et al* (2021, p. 616); another is Mallory (2023). This line of argument should immediately lead us to wonder what is being denied: That these machines are capable of speaker meaning, that the patterns they produce are words or phrases bearing linguistic meaning, that they can perform locutionary acts, that they can perform phatic acts, or something else? Further, once we have identified one of these claims as being denied, we should watch out for fallacious inferences from the denial of that one to the denial of any others.

We have noted that an expression may be used with its linguistic meaning even when not used in an act of speaker meaning. Hence, the fact that an expression is not used in a particular case of speaker meaning cannot on its own show that it lacks linguistic meaning when so used. Unfortunately, this inference seems to animate recent discussions of whether LLM's can mean anything. For instance, in their influential article Bender *et al*, point out that text generated by LLM's, "...is not grounded in communicative intent." (2021, p. 616). They take this to imply that an LLM, "...does not have meaning," which claim is most naturally read as denying that this text when used by LLM's possesses linguistic meaning: that is evidently the force of their image of the babbling parrot. However, by now it should be clear that the absence of an illocutionary or even locutionary act is compatible with the presence of a phatic act. Accordingly, we cannot validly infer from the lack of intent on the part of a machine, that the patterns it produces lack linguistic meaning.

A fallacious line of reasoning leading to a conclusion does not show that that conclusion is false. Bender and colleagues might concede that their appeal to communicative intent was a misstep; then go on to bite the bullet and insist that the patterns produced by LLM's are not even words, and thus lack linguistic meaning. However, we have already explained why this position should be challenged: Contemporary LLM's meet the very minimal standard of performing phatic acts, and such acts by definition entail the use of words with their standard linguistic meaning even if we are not prepared to claim that the machines grasp that meaning. Accordingly, even if the linguistic strings that LLM's produce are "not grounded in communicative intent," they're still the results of phatic acts and thus bearers of linguistic meaning.

3. Koch's argument for LLM reference

Koch notes that he will not attempt to answer the question whether LLM's can engage in speaker reference (Koch, 2025). Instead he will argue that patterns produced by LLM's that seem to be referring expressions are in fact referring expressions, and thus carry semantic reference when so occurring. As such Koch is following a path similar to the one pursued here. However, while the latter part of his essay addresses kind terms, both natural and social, Koch does not consider another language whose meaning resists characterization in referential terms. This includes quantifiers, logical constants, modals, and other expressions whose meanings are arguably best elucidated in terms of their contribution to the truth conditions of the sentences in which they occur. By contrast, our account of how LLM's can produce language bearing its linguistic meaning applies across the board to all words and phrases that they are able to use in phatic acts.

Koch's strategy is to appeal to the Causal Theory of Reference (hereafter CTR) to explain how the patterns that LLMs produce are phrases with linguistic meaning. According to the CTR, a pattern produced by a speaker *S* is a proper name *N* referring to object *O* just in case *S* is part of a causal chain of speakers whose use of *N* can be traced back to an initial baptism using *N* to name *O*; to be part of such a causal chain, each speaker must intend to use *O* with the same referent as the speaker(s) from whom he acquired the name. Koch's appeal to the CTR highlights the manner in which LLM's bear causal relations to objects in the world. In so doing, Koch aims to neutralize the "octopus" objection due Bender and Koller (Bender & Koller, 2020).

Koch also considers two possible interpretations of CTR. On the first interpretation, which he calls '(a)', "the requirement demands that whenever someone uses a given name, say "Obama," then in order to refer to Obama that language user must have the explicit intention of using "Obama" in exactly the same way as the person from whom she heard it." (Koch, 2025, page 25). Koch does not tell us what it is for an intention to be explicit, but it is natural to surmise that it is one that is consciously entertained, perhaps in the form of an utterance in inner speech or a visualization. An intention so understood would in turn, contrast with what are likely the most common types of intentions, namely the implicit. These govern quotidian cases of walking, eating, speaking, and the like, particularly for those cases in which a suite of actions already in motion is continued. However, Koch does not formulate an alternative to (a) making use of implicit intentions. Instead, he offers as an alternative (b), formulated thus: "On the second option (b), the requirement merely demands that users of a name do not have the intention of using the name differently than those from whom they picked it up" (Koch, 2025, page 25).

Koch also differentiates (a) into two readings, (a-1) and (a-2), which correspond to de-re and de-dicto construals, respectively. He argues convincingly that the (a-1) reading is less plausible than the (a-2), but in so doing, neglects to consider whether the (a) and (b) readings are exhaustive. And surely they are not, since there is a neglected third alternative, (c): the requirement that whenever someone uses a given name, in order to refer to its bearer that speaker must intend to use that name in the same way as the person from whom she heard it. (C) leaves it open whether the intention is implicit or explicit. Furthermore, (c) is more plausible than either (a) or (b). Koch gives compelling arguments for rejecting (a), and I will not rehearse them. Concerning the superiority of (c) over (b), observe that if we were to ask a speaker who has used 'Barak Obama' several times in her recent conversations whether she intended to use it to refer to a particular person, presumably she would answer in the affirmative. On the other hand, if she were to answer, "I had no intentions whatsoever to refer to anyone in my use of that name," that would undermine our confidence that in so speaking she is referring to the former U.S. President. Nevertheless, she meets the

conditions of Koch's (b).⁶

Fortunately, Koch does endorse a more demanding requirement for a speaker's using a name in such a way that it bears semantic reference: "What is needed", he writes, "is some kind of mechanism that ensures continuity between how users have used and continue to use the name and one's own future use of this name." This is a stronger requirement than that imposed by his condition (b), and Koch's next aim is to argue that LLM's meet it writing:

Whereas human agents typically secure continuity by forming the intention of using the name in the same way as those from whom they have picked it up, it is the design architecture of the LLM that serves this function for them. LLMs are built to pick up and henceforth apply words *in just the same way* as these words are used in the texts it learned from. This feature of their design architecture secures continuity between what names refer to in the training data and what they refer to when the LLM uses them later on. So, even though an LLM cannot form any intentions, the way it is built ensures that the reference that is undoubtedly present in the texts from which it is trained is transmitted to it. If this is right, then no intentions are needed for the LLM to generate texts with the same reference patterns as the texts from which it is trained. (Koch, 2025, page 27) [italics added]

Let us focus on one remark in particular, which I shall anoint "LLM Design":

LLM Design: "LLMs are built to pick up and henceforth apply words in just the same way as these words are used in the texts it learned from."

LLM Design is a claim about how these machines are designed, which is not quite the same as a claim about what they do. However, their capacities are presumably going to constrain their design because the engineers who build them are aware of those capacities. Either way, *LLM Design* is contentious at best. First of all, at the level of the mechanics of how LLM's are trained, we find them building representations of linguistic units in a continuous vector space, and doing so in a manner that is based on those units' statistical distribution in a large corpus (Millière & Buckner, 2024). That is not what appears to occur when human speakers learn or use language. Presumably, some psycholinguistic theories of how words are learned and used by children make appeal to vectors in this way. But then such a theory would have to be shown to be correct, and its competitors refuted, for the "just the same way" claim in *LLM Design* to be acceptable.

Second, at the level of the performance of these machines, what they do with words, and what human users might do with the very same words, can diverge drastically. One of those texts might have been the record of a human speaker referring to a perceptually salient object, like a beach ball, and describing it as red.⁷ LLM lack perceptual capacities and as a result are unable to describe a perceptually salient object as red, or at least to grasp such a description in anything like the way that a normally sighted human user can do. So too, some of those texts might be records of apologies or warnings, and Koch has chosen to refrain from imputing cognitive sophistication to LLM's of a kind that would enable them to perform such acts as these (Koch, 2025). So here again we have an apparent dissimilarity between how LLM's use language and how ordinary human speakers do, and here again we see the implausibility of the "just the same way contention."

Thus *LLM Design* is dubious. That in turn undermines Koch's inference from that claim to the next step in the quoted argument, namely, "This feature of their design archi-

⁶ I am prescinding here from discussion of a further complication that Koch considers, which concerns the time at which the putative name is used.

⁷ These 'perceptual demonstratives' may be intelligible only to agents with perceptual capacities, and so beyond the grasp of LLM's. See Dickie & Matthen (2014) for further discussion of perceptual demonstratives.

texture secures continuity between what names refer to in the training data and what they refer to when the LLM uses them later on.” “This feature” refers to a phenomenon whose existence we have found reason to doubt, namely the “just the same way” claim contained in *LLM Design*. Might there be a way of weakening that claim to something like, “largely the same way,” so that the argument still goes through? Perhaps, but I doubt that we need to dwell on the question.

The reason is that we have already expounded a method for establishing that patterns produced by LLM’s can bear semantic properties: they do so because and to the extent that they are used in phatic acts in the sense that we have defined. This approach is not limited to proper names, noun phrases, or kind terms, but to meaningful expressions generally. That includes expressions for which it is not plausible that they would have been introduced into the language with any act of dubbing à la CTR. Our approach also does not rest on any controversial claims about how similar LLM’s are to human users. For all we have said, LLM’s might be radically different from human users both in how they acquire language and in how they use it once it has been acquired.

We should also take care to ensure that our approach to securing linguistic meaning for LLM’s won’t leave any more indeterminacy of reference than we already encounter in everyday language use.⁸ What happens when an LLM is trained on data that involves text containing the name ‘Aristotle’, some occurrence of which refer to the ancient philosopher (call this Aristotle₁), while others refer to the Greek shipping magnate (Aristotle₂)? My suggestion is that if the training data contains significant numbers of occurrences of both ‘Aristotle₁’ and ‘Aristotle₂’, the name will simply be ambiguous. Acknowledging this fact can thus be no objection to our approach to securing linguistic meaning in the patterns that LLM’s produce. In some cases the context of the LLM’s output will provide a likely disambiguation, such as when it occurs in a sentence referring to the doctrine of the four causes as opposed to one discussing Jackie Onassis. In other cases, there may be no fact of the matter as to how the sign is to be interpreted. But that is no more objection to our theory of how LLM’s mean, than is the observation that competent speakers are sometimes ambiguous in ways of which they may be unaware.⁹

Finally, it is now well established that pragmatic factors help to determine what is said, and not just the further implications of what has been said by means of such processes as implicature. “The metal is not warm enough” would seem only to express a determinate content once it is clear whether the speaker and her interlocutors are interested in keeping a plate of food warm, or instead in ironworking.¹⁰ Speakers generally manage this pragmatic determination of what is said with little trouble. We should not expect LLM’s to exercise this capacity, since doing so will require a controversial imputation of goals and projects to them. However, for many purposes, it may be sufficient for an LLM to produce language that is subject to such pragmatic determination by its human interlocutors.

In this essay, we have seen that it should not be controversial to construe LLM’s as performing phatic acts, and as a result, using language with the linguistic meaning that is conventionally attributed to it. Koch’s argument for a like conclusion, appealing to CTR lacks the generality of the approach espoused here while also undertaking controversial commitments. For this reason, our “Austinian” strategy offers a superior answer to the question that Koch’s essay aimed to answer.

References

Austin, J. L. (1975). *How to do things with words* (J. O. Urmson & M. Sbisà, Eds.; 2nd ed). Clarendon.

8 Green (1999) argues that a certain amount of indeterminacy of reference, and thereby of attitude ascription, should be tolerated. This implies that in certain cases, there is no fact of the matter as to which of two or more ascriptions of attitude is correct, so long as all these systems of ascription are adequate to account for an agent’s behavior.

9 Or indexical, per Pelczar and Rainsbury (1998).

10 See Green (2021) for further discussion as well as references.

- <https://doi.org/10.1093/acprof:oso/9780198245537.001.0001>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198.
- Butlin, P., & Viebahn, E. (2024). AI assertion. *Ergo an Open Access Journal of Philosophy*.
- Dickie, I., & Matthen, M. (2014). *Perception and Demonstratives*. Oxford University Press.
- <https://doi.org/10.1093/oxfordhb/9780199600472.013.041>
- Green, M. (2016). Expressing, Showing and Representing. In *The Expression of Emotion: Philosophical, Psychological and Legal Perspectives* (pp. 25–45). Cambridge University Press.
- Green, M., & Michel, J. (2022). What Might Machines Mean? *Minds and Machines*, 32(2), 323–338.
- Green, M. S. (1999). Attitude Ascription's Affinity to Measurement. *International Journal of Philosophical Studies*, 7(3), 323–348. <https://doi.org/10.1080/096725599341794>
- Green, M. S. (2000). Illocutionary Force And Semantic Content. *Linguistics and Philosophy*, 23 (5), 435–473.
- Green, M. S. (2021). *The philosophy of language*. Oxford University press.
- Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.
- Gubelmann, R. (2024). Large Language Models, Agency, and Why Speech Acts are Beyond Them (For Now) – A Kantian-Cum-Pragmatist Case. *Philosophy & Technology*, 37(1).
- Koch, S. (2025). Babbling stochastic parrots? A Kripkean argument for reference in large language models. *Philosophy of AI*, 1, 19–33. <https://doi.org/10.18716/OJS/PHAI/2025.2325>
- Kripke, S. (1977). Speaker's Reference and Semantic Reference. *Midwest Studies in Philosophy*, 2, 255–276. <https://doi.org/10.1111/j.1475-4975.1977.tb00045.x>
- Mallory, F. (2023). Fictionalism about Chatbots. *Ergo an Open Access Journal of Philosophy*, 10(0). <https://doi.org/10.3998/ergo.4668>
- Millière, R. (2024). *Language Models as Models of Language* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2408.07144>
- Millière, R., & Buckner, C. (2024). *A Philosophical Introduction to Language Models – Part I: Continuity With Classic Debates* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2401.03910>
- Pelczar, M., & Rainsbury, J. (1998). The Indexical Character of Names. *Synthese* 114(2), 293–317.