Homework 1: Computers as a New Species of Language-Capable Beings

Edward Hernández

College of William & Mary

Homework 1: Computers as a New Species of Language-Capable Beings

Will computers ever achieve the ability to produce and understand human language? To answer this seemingly simple question, I will break it into three of its (many) component questions. Can a computer produce language? Can a computer understand language? If it can, will the language that it can produce or understand be *human*?

For the purposes of this paper, I will assume that any breakthrough which is conceptually possible will be eventually technologically realized. If a computer can, in principle, produce and understand language, I will assume that someone[1] will eventually build a computer which does. I also assume that the computational theory of mind (termed "strong AI" by Searle, 1980) falls to (at least) embodied cognition arguments.[2] If it is the case that the human brain (or its mind) is a computer, the above question is moot, as human language has then never been produced or understood by anything but computers, and building non-human computers which understand and produce language by the very same computational processes is fully possible and will inevitably be accomplished.

First, can a computer produce language? Yes. It can. This does not need to be justified conceptually, because computers already do produce language. Computers routinely produce outputs in formal languages (like assembly languages). They also produce strings of language which are understandable by humans, like error messages or application feedback. These outputs can be printed to screens or pages, or spoken aloud. They also produce novel strings of language,[3] like haiku (Cope, 2011), or conversations (Baraniuk, 2014; Biever, 2012).

Can a computer understand language? This is a much larger question, and requires some clarification before it can be answered. What does it mean to understand? Does producing language necessitate understanding it?

---

[1] ~~Google.~~ Alphabet.

[2] I leave open the possibility that CTM is not false (specifically that it is trivially true and compatible with embodied cognition). However, if I held it to be true, the answer to the question would be trivial, and the focus of any paper on the topic would be to prove CTM. In order to write a topical essay, I caveat embodied cognition is true and incompatible with CTM.

[3] I shy away from saying "in a language" because of the superdiversity literature I discuss in more detail below.

Searle (1980) explores exactly these questions. To use his thought experiment, imagine a person who does not speak Chinese in a room with books of Chinese. That person is passed notes, written in Chinese, through a door, and writes notes in response based on the characters in the books. She has no idea what the characters on the notes or her responses mean (to the people outside the room), but eventually she learns to use the books so well that everyone outside the room believes her to be a fluent (human) speaker of Chinese. In this case we are inclined to say that the person does not *understand* Chinese or the language which constitutes her inputs and outputs.

This corresponds directly to the question of whether computers understand language. If we say of the person in the Chinese Room that she does not understand the notes, then we must say of a computer that it does not understand language which it produces based on machine-learning, even if it passes other measures, like the Turing test. The human, after all, would also pass as a fluent speaker of Chinese, and this does not move us intuitively to say she understands.

Searle (1980) takes this thought experiment to show us a divide based on intentionality. Brains, he argued, posses special causal capacity, enabling them to be intentional in a way which a computer (or, other computers, on his view) cannot. Another possible explanation of our reticence is that the operations within the room are (arguably) exclusively syntactic(Marconi, 1996). Words are chosen from books, fit together such that they adhere to the rules of Chinese and returned. The semantic content of both incoming and outgoing notes is completely opaque to the person writing the responses. Both of these are robust attacks, I here argue that they are incorrect about it is to understand.

Imagine that you are asked whether a given human understands language. How do you come to an answer? If you are like me, you ask them questions, and when they answer in correct and comprehensible ways, you conclude that they do. This means that computers which pass the Turing test *do* understand language by our everyday definition of the word understand. I hold that this is entirely sufficient to say that they do understand, based on arguments out of Wittgenstein, 1969/1975, 1953/1986. There is no language game for determining understanding which involves assessing the causal effect of that language on the brain or whether that brain engages with the semantic content rather than solely the words' syntactic information. Obviously this cannot be our

criteria for other humans, because that data is never available to us as interlocutors. If a human can talk about something, I judge that she understands both language and the semantic content contained in our utterances. Thus neither their supposed lack of intentionality or alleged failure to engage with semantic content, computers should be taken to understanding language if we would say of them that they understood language were they human. Computers understand language (and anything else) in that they satisfy the criteria in our language-games for understanding.

Additionally, since I am operating under the assumption that all possible technological advances *will* be made, I expect that biological computers (which are possible in at least principle) will be produced such that they have any possible biological property. I see no reason that such a computer would be theoretically incapable of intentionality in the way that Searle requires, and might store and access semantic information in ways (functionally, if not physically) identical to brains, satisfying the other concern.

Lastly, is the language which computers understand or produce human? This may seem like a trivial question to ask, but it does not fall within the scope of either of the above component questions, and is a requirement of the overall question. I here argue that, currently, computers do not produce or understand human language, and that there are significant conceptual and technical hurdles to their doing so.

Currently, computers produce a small subset of the language that is created by humans, and they are very rarely judged to be human speakers. Recently, the first computer arguably passed the Turing test by convincing 33% of judges that it was human ("Turing Test success marks milestone in computing history", 2014). By my above argumentation, that computer produced and understood language. However, it was only able to do so in a specific context: a conversation via keyboard in which the other interlocutor believed it to be a 13 year old Ukranian boy. This ruse was designed to cover up the computer's errors and failures. Those failures were sufficient that we might well say of a human who made them did not understand language (well). Computers also routinely fail to understand language on speech-to-text and text tasks, fail to respond correctly to sarcasm and metaphor, and are horribly bad at handling dialectal variation. Until these hurdles are overcome, none of these capabilities constitute production or understanding human language.

There are, of course, more advanced computers targeted at approximating human language.

Generally, linguists or other language scientists attempt to run machine-learning programs on these computers. However they feed computers "language" input which does not closely resemble language as it is used socially. It is often downcased, sometimes with punctuation removed. Spelling is standardized. Only forms deemed grammatical in the target language are selected for the input. This is all based on the practices of Chomskyan linguists, who hold that the ideal speaker in a homogeneous speech community is a reasonable model for the language learner (and therefore a language-capable computer). Because this is their model for language, they feed computers homogeneous strings judged to be grammatical in a single language.

However, no speech community is homogeneous. Women speak differently than men, and both speak differently at different ages. There are generational and class differences among speakers. There are also, of course, multilingual speakers, whose communities are obviously not homogeneous. In many places, there are multilingual communities, in which nearly all communication may be heterogeneous and heteroglossic, as the text in Figure 1, which is written ungrammatically and simultaneously in multiple dialects of Chinese, with a mix of two incompatible styles of Chinese characters. These realities have put under threat the very idea of "a language," moving some, like Bommaert and Rampton (2011) to produce and adopt new models of language as a whole. On these views, often broadly labeled *superdiversity*, language is much less predictable than Chomskyan linguists assume (Bommaert & Rampton, 2011, p. 1), varying on every level, based on not only the interlocutors and their languages, but also based on every sort of context available.

Computers are very far from being able to understand (much less respond appropriately to and themselves produce) such variation. Additionally, according to the embodied cognition literature, much of human language processing is handled by the interaction of the sensorimotor capacities of the brain with the body and the environment. This view also holds that understanding semantic content may involve sensorimotor processing (e.g. the concept of kicking may involve the sensorimotor activity related to performing kicking motions; Rueschemeyer, oliver Lindemann, van Rooij, van Dam, & Bekkering, 2010). This conception of understanding is not contradictory with my Wittgensteinian definition above. To assess whether someone understands some semantic content, I might ask them to demonstrate it. I might ask someone to kick, etc., and

I might judge that a person who fails to do so does not understand kicking. If a computer has no sensorimotor functions related to kicking, then, it may be judged to not understand it.[4]

Quite a lot of language as it is spoken by humans has to do with human activities in which a computer cannot (yet) participate. By the above argumentation, computers cannot (yet) understand this language, or any language which deviates from the input on which they were trained. I anticipate that the proportion of human language which can be understood and produced by computers will grow continually, eventually approaching full human competence, but I hold that to fully accomplish this task, they will need bodies capable of performing all the same sensorimotor activities as humans.[5]

Thus, I argue that computers are only capable of producing and understanding human language insofar as they are capable of becoming indistinguishable from humans, but that this is fully possible in principle. I despair at computers learning language as they are currently taught, however. It is only by being embodied and living socially that any being, synthetic or otherwise, can hope to be a fully literate being. I, for one, welcome our new android companions.

[4]But, a robot which could both talk and kick, and could talk about kicking, I would judge to understand both kicking and language.

[5]Eventually, they would become full philosophical zombies, as discussed by **?**.

References

Baraniuk, C. (2014, June 9). How online 'chatbots' are already tricking you. *BBC future*. Retrieved from http://www.bbc.com/future/story/20140609-how-online-bots-are-tricking-you

Biever, C. (2012, June 25). Bot with boyish personality wins biggest Turing test. *New Scientist*. Retrieved from https://www.newscientist.com/blogs/onepercent/2012/06/bot-with-boyish-personality-wi.html

Bommaert, J., & Rampton, B. (2011). Language and superdiversity. *Diversities*, *13*(2).

Cope, D. (2011). *Comes the fiery night*. Epoc Books.

Marconi, D. (1996). On the referential competence of some machines. In P. McKevitt (Ed.), *ntegration of natural language and vision processing: Theory and grounding representations* (Vol. 3). Springer.

Rueschemeyer, S.-A., oliver Lindemann, van Rooij, D., van Dam, W., & Bekkering, H. (2010). Effects of intentional motor actions on embodied language processing. *Experimental Psychology*, *57*, 260-266. doi: 10.1027/1618-3169/a000031

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417-424. doi: 10.1017/S0140525X00005756

Turing test success marks milestone in computing history. (2014, 8 June). *University of Reading News*.

Wittgenstein, L. (1975). *Über gewißheit* [On certainty] (G. E. M. Anscombe & G. H. von Wright, Eds. & D. Paul & G. E. M. Anscombe, Trans.). Basil Blackwell. (Original work published 1969)

Wittgenstein, L. (1986). *Philosophische untersuchungen* [Philosophical investigations] (G. E. M. Anscombe, Trans.). Oxford: Basil Blackwell. (Original work published 1953)

Figure 1. A note in a shop window in Antwerp. © Jan Blommaert