TALKING TO MACHINES

Emanuele Martinelli

ABSTRACT

As mechanization began to approach the frontiers of AI, artificial devices got to challenge humans more and more, leaving us to wonder which abilities will be proven unique to us and which will not. Alan Turing, in the Fifties, notably proposed his test to recognize humanly intelligent behavior in machines. His *imitation game*, though, revolves around the capability of using natural language in a dialogue like a human would do. In this brief paper, we will provide reasons why machines cannot achieve human linguistic competences in principle, drawing them from the fields of pragmatics.
So, after a first section in which we will introduce Turing's famous test, we are briefly going to examine the approach with which AI applications are supposed to recreate human communication, and its limits. In the third and fourth section, the paper will list some topics in the pragmatics of human dialogue and how those threaten its automatized reproducibility: arguably, they are reducible as problems to mimic two major features of language, namely the creation of a common ground between the speakers and the relationship between the speakers and their personal biographies. Finally, these items in hand, we are going to see at what extent such difficulties could be at least theoretically overcome.

THE TURING TEST

Alan Turing suggested we should define a machine as being intelligent, like a human can be, as it is able to win his *imitation game* in a statistically relevant number of cases (Turing, 1950). Briefly put, the idea goes as follows. Two individuals, say, X and Y, are put in two isolated chambers, unable to communicate with each other whatsoever; they can only write messages in, say, English, to a third person, Z, which is the *interrogator*. Suppose either X or Y is a bot, and the other one is a human being: Z's task is to ask them questions, and, based on their answers, tell the machine and the man apart. If the machine is able to fool the interrogator into mismatching, then it won the game.

We see how the key for achieving a human level of cognition is defined on the ability to manipulate natural language like two conversing people would do. There is one main, recapitulatory, reason why this task, which is crucial to the imitation game previously described, is utterly impossible. This "argument from consciousness" had actually already been addressed by Turing (Turing, 1950). The objection is roughly the following. In order to think like a human, machines need a consciousness of their own, and, hence, mental states. This point is somewhat analogous to that of the famous argument from Searle (Searle, 1980).

Turing's proposed answer is that such a claim would entail solipsism: "according to this view the only way to know that a *man* thinks is to be that particular man" (Turing, 1950, p.452). This paper will work on the counter-objection that there is at least one more way to be sure other

men (or machines) are conscious: the acts of theirs whose performance require a conscious mind.

AI AND LANGUAGE

Turing proposed an architecture to win the imitation game, which mostly resemble present-day AI applications. It mainly works with the manipulation of symbols, followed with the implementation of some method for meaning assignment. For instance, in present-day automatized translation devices, machines are able to respond to an input sentence in a given language with another relevant output sentence in the target language, so that the result corresponds to the translation a human would have provided. In order to do so, AI applications are simply supposed to possess sufficient "memory size, computation speed and 'proper programming'" (Landgrebe, Smith, 2019a, p.2), as if "human cognition is itself just a matter of storage and computation" (Landgrebe, Smith, 2019a, p.2).

So, the perfect Turing machine is supposed to show human-like speech through correct logical inferences and knowledge about meaning attribution. Applications need to be fed with huge datasets, in the order of billions (in the previous example, translations may be drawn by the machine from a database of similar manmade translations in the same language pair); the data needs to be consistent, at least for the most part, and enough stable to make repetitive patterns identifiable. First-order logic is then deployed to match external stimuli with plausible outputs. Sensorial awareness of the physical world around may be integrated by adequate functional units to achieve better interaction (Harnad, 1991).

To assume that natural language, as much as it is used in normal conversations, can be formalized with these tools, somewhat resembles the account of human communication that has been popularized in the XX century by Frege (1980) and De Saussure (1916). The combination of symbols uttered by the speakers are selected in order to convey a message, which is a reproduction of their thoughts in an intersubjective way (De Saussure, 1916). The same choice of words, more or less, reports the same meaning, utterances being functions to *objective* concepts.

PUTTING WORDS IN CONTEXT

This account of human language has been critically challenged by the later works of Wittgenstein (1953), Austin (1962), Grice (1975), and by all those who worked in the fields of the *pragmatics* of natural language. For what concerns this paper, we are only going to venture

into so-called "near-side pragmatics", that is those non-linguistic facts that play a role in the message formulation and interpretation (Korta, Perry, 2007). These could hardly be implemented by AI developers in a proper way; as mentioned before, we are going to take a brief look at them by grouping them under two major features of human communication.

Firstly, natural language is highly sensitive to the context of the conversation. We could ideally simplify a dialogue as the cooperative activity in which the speakers display their intentions by exchanging messages, assuming rules and patterns that facilitate an efficient conveyance of meaning (Landgrebe, Smith, 2019a; Grice, 1975). For instance, in the standard dialogue, interlocutors take turns to express their utterances, or expect some kinds of utterances (e.g. questions) in response to some others (e.g. answers) (Landgrebe, Smith, 2019a). Though, real-life communication often breaks this model. Speakers find themselves pausing, interrupting mid-sentence, overlapping, enriching discourse with external elements, etc. The way in which intentions are transmitted and treated throughout the conversation changes and is shaped based on a myriad of factors that only partly rely on the semantics of words. Intentions themselves evolve as the interlocutors interact, as the partner's messages or third reasons may push one to switch one's perspectives or aims in real-time.

When a dialogue takes place, the repertoire of speech acts that are actually relevant to that particular occasion is dependent on the setting where the speakers are located. Such contexts have different forms and, sometimes, may evolve as well in response with the participants' intentions (Landgrebe, Smith, 2019a). Sharing (or not sharing) a spatial context, a temporal context, a social or cultural context, first of all influences the amount of information the speakers assume can remain implicit throughout the conversation. For instance, the choice of words a doctor is going to employ in order to explain his patient the characteristics of his disease is probably different by which he would select if his interlocutor were a fellow doctor. Or, the choice of *language* depends on the other participants in dialogue, selecting English or Italian, a dialect or another one, a sociolect or another one, and so on (Landgrebe, Smith, 2019a).

The spontaneous building of a "dialogue horizon" (Landgrebe, Smith, 2019a, p.6) not only impacts the formulation of messages, but also their interpretation. Proper nouns or pronouns, for instance, respectively imply that the speakers share common knowledge about the referents or presence in the same spatial/temporal context. The dialogue's common ground, moreover, enables one of main features of human communication to express, i.e. non-verbal communication, from gestures to unconscious somatic representations of emotions. Tomasello thus defines communication as the cooperative activity of building a common ground for

exchanging information and interacting with it to utter intentions (Tomasello, 2008). This definition is reinforced by the empirical confirmation pre-linguistic infants express their understanding of the world through gestures even before they can manipulate verbal language, as a result of the fact human communication actually developed from primates' non-vocal communication habits (Tomasello, 2008).

Finally, even the dialogue itself can become its own context, at times (Landgrebe, Smith, 2019a). The interpretation of some statement can be influenced by the previous and successive utterances. My story about my recent trip to New Zealand reinforces the possibility that, by "The following day, I went in the Alps", I want to refer to the Southern Alps (and not the European Alps). Furthermore, *how* the utterance is pronounced, i.e. what is its force and modality, can highlight, through pauses and accents, some information over other[1] (Landgrebe, Smith, 2019a).

TELL ME WHO YOU ARE AND I WILL TELL YOU WHAT YOU SAY

The second source of pragmatic influence on the content of discourse is the psychology of the speakers themselves. The formulation of our intentions when we engage in dialogue, as well as the patterns that guide their evolutions in response to certain factors, are heavily influenced by one's biography (Landgrebe, Smith, 2019a). Past experiences, the knowledge at one's disposal, desires, preferences, perceptions, all concur into forming the personality of the interlocutor. All these characteristics reinforce or weaken certain dispositions to formulate certain messages or conceiving certain interpretations. For instance, the speakers' identity varies the common ground of communication in terms of shared knowledge or intentions.

Memories from the speakers are implicit factors of influence on the dialogue's content too. If Mark casually mentions he is considering buying a BMW to Paul, who happens to have been driving one in the past, the conversation could bring about useful and unexpected considerations for Mark's possible acquisition. Machines, of course, do not possess any full personality or identity. In principle, programmers could implement some knowledge into an AI application in order to mimic the presence of past memories and preferences. Though, machines are bound to experience gaps and involuntary inconsistencies under this respect: the synthetic reproduction of living experience cannot be fully expressed by mathematical tools, because it entails the presence of a mental life and intentionality.

[1] Of course, different languages build in their syntactical structures features for different interpretation issues. For instance, in the Japanese language, information is highlighted in dialogue by some *theme-marker* particles rather than by pragmatic means.

Lastly, to lack personality is to lack a proper idiolect. Idiolects are defined as the unique way in which one particular man uses language in order to convey his intentions (Barber, Ramirez, 2017). For a myriad of reasons, mainly pertaining the past experiences thanks to which a person learnt how to speak, one's idiolects could give birth to personal variances in the use of language. Machines, of course, cannot display full-fledged idiolects by their side. A couple example could make the implications of this clear. The English dictionary claims the main meaning of *livid* is *bluish*; in spite of that, many English speakers use this word with the meaning of *reddish*, as in the expression *livid with rage*. Or, alternatively, the Latin expression *una tantum*, which means of course *one time only*, is mistaken by many Italian speakers to mean *sometimes*, *seldom*, because of its similarity with the Italian expression *una (volta) tanto*. A machine couldn't display any of this variation in the same way human individuals may develop it. For instance, they, knowing *livid* means *bluish* in the official, abstract, version of the English language, would not be able to understand a human speaker using *livid* to refer to something *reddish*, while a human user could grasp the deviation from the context or from past experiences.

TALKING TO MACHINES?

It appears straightforward clear how originally-conceived Turing machines cannot take the challenge of pragmatics. There have been many proposals on how to mathematize natural language, in order to create better AI applications for mimicking human dialogues (Landgrebe, Smith, 2019a; Landgrebe, Smith, 2019b; Legg, Hutter, 2006; Hernandez-Orallo, 2000). In principle, some aspects can be improved, some obstacles removed. On the other hand, there are some features of human communication, as mentioned before, that necessarily require proper mental states to be reproduced. Machines do not have them; hence machines cannot learn to behave like real humans.

The need for knowledge in order to select relevant information and responses theoretically may be satisfied by two implementations. Greater datasets, in the order of trillions or quadrillions, might map any possibility and any variability to build consistent knowledge for general AI projects. Even the interpretation of emotive contents in non-verbal communication may be met by emotion ontology applications (Landgrebe, Smith, 2019b). The relevance of the context in providing and highlighting certain facts over other ones could be resolved in its turn by implementing synthetic contexts and a proper classifier (Landgrebe, Smith, 2019a). With

these intuitions in mind, AI applications could mimic human language through stochastic systems and differential equations.

There are problems, though, that cannot be resolved by any mathematical model, as a matter of principle. While the structure of the context of communication can be synthetically reproduced by pre-set information and sensors, 1) the interaction between the machine and the context, as well as 2) the features of natural language that require personal psychological background, cannot. When some element in the relevant environment changes, e.g. new intentions of the other speaker are unveiled, or some new perception needs attention, appropriate output for such an emergency input can only be spontaneous. Moreover, decoding and responding to some communicative signals, e.g. non-verbal signals, often takes place at an unconscious level. In dialogue, feelings, memories, and other psychological elements shape the messages that are going to be uttered, and the interlocutor consciously or unconsciously senses these emotions and interprets them accordingly (Landgrebe, Smith, 2019a). AI applications will never be able to model these kinds of cognitive pattern, as they need intentionality to be properly, humanly, expressed.

BIBLIOGRAPHY

- J. LANDGREBE, B. SMITH, *There is no General AI*, 2019a, forthcoming,
- J. LANDGREBE, B. SMITH, *Making AI Meaningful Again*, Cornell University, 2019b, arXiv:1901.02918v3,
- A, BARBER, E. G. RAMIREZ, *Idiolects*, in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), 2017, https://plato.stanford.edu/archives/spr2017/entries/idiolects/,
- K. KORTA, J. PERRY, *Pragmatics*, in *The Stanford Encyclopedia of Philosophy*, E.N. Zalta (ed.), 2015, https://plato.stanford.edu/archives/win2015/entries/pragmatics/,
- M. TOMASELLO, *Origins of Human Communication*, MIT Press, 2008,
- K. KORTA, J. PERRY, *The Pragmatic Circle*, Springer Science, 2007,
- S. LEGG, M. HUTTER, *A Formal Measure of Machine Intelligence*, Cornell University, 2006, arXiv:cs/0605024v1,
- J. HERNANDEZ-ORALLO, Beyond the Turing Test, in Journal of Logic, Language, and Information, Kluwer Academic Publishers, n. 9, pp. 447-466, 2000,
- S. HARNAD, Other bodies, other minds: a machine incarnation of an old philosophical problem, in Minds and Machines, n.1, v.1, pp. 43-54, 1991,
- G. FREGE, *Über Sinn und Bedeutung*, in *Zeitschrift für Philosophie und philosophisce Kritik*, as *On Sense and Reference*, M. Black (trad.), in *Translations from the Philosophical Writing of Gottlob Frege*, P. Geach, M. Black (eds.), Oxford: Blackwell, 1980,
- J. SEARLE, *Minds, Brains, and Programs*, in *Behavioral and Brain Sciences*, vol. 3, pp. 417-457, 1980,
- P. H. GRICE, *Logic and Conversation in Syntax and Semantics*, in *3: Speech acts*, P. Cole (ed.), Academic Press, pp. 41-58, 1975,
- R. STALNAKER, *Pragmatics*, in *Semantics for Natural Language*, Davidson, Harman (eds.), Dordrecht, pp. 380-397, 1970,
- J. AUSTIN, *How to Do Things with Words*, Oxford: Clarendon, 1962,

- L. WITTGENSTEIN, *Philosophische Untersuchungen*, as *Philosophical investigations*, G.E.M. Anscombe, R. Rhees (trads., eds.), Oxford: Blackwell, 1953,
- A. M. TURING, Computing Machinery and Intelligence, in Mind, n. 59, pp. 433-460, 1950,
- F. DE SAUSSURE, *Cours de Linguistique Géneral*, Jean-Didier Urbain, as *Course of General Linguistics*, C. Bally, A. Sechehaye (trads., eds.), 1916.