Semina □ Nr 12
Scientiarum 2013
s. 83–96

Maria Wilkowska

A philosophical investigation in machine understanding

The case of implicit meaning

When it comes to machine understanding, the field abounds in various definitions of what it actually means *to understand*. Different approaches to the topic aided by specific techniques [which struggled to implement the insubstantial meaning of words and phrases into the substantial ware of a computer] have been devised by scientists over the years. However, the research has not progressed beyond the stage of mere symbol processing. In other words, contemporarily, computers excel at information processing without understanding what the content of their processing *really* is. The case of human understanding has puzzled Artificial Intelligence (AI) researchers and philosophers alike. What parts or mechanisms of our human brains are responsible for the meaning comprehension? Moreover, what [most probably] neural operations are in charge of the communication and effective interpretation of indirect meanings despite their absence in the read or heard input? The problem of how to make a machine really *understand* the semantic meaning the way humans do and not merely make it process numerical signs remains unanswered as for now.

This paper is a philosophical investigation into machine understanding. It looks into the difference between syntax and pragmatics. Specifically, the point of interest here is the phenomenon of implicature interpretation by humans. Both terms together with the notion of understanding and inference will

be described in a greater detail in the introductory parts of the paper by way of prefatory remarks. The concluding part features a discussion on the subject of the discrepancy between syntax and pragmatics. The inspiration for this paper was prompted by the question what determines and what it may be in the explicit form of a message which makes the implicature understandable and renders the entire message accurate and acceptable despite its incongruity with the syntactical pattern entailed by the interrogative form.

1. Machine understanding

The domain of AI can be divided into two subfields of the so called weak and strong intelligence. The strong AI hypothesis claims that it will be possible for machines to have self-consciousness and experience *qualia* but also that their intelligence and cognitive capabilities will far exceed those of humans[1]. In this context intelligence is equaled with linguistic skills (which are believed to be the base thereof) and so the research focuses on building artificial dialogue systems. Conversational agents or chatbots, as the afore-mentioned systems are also referred to, are exemplary realizations of strong AI in practice. They emulate human linguistic behavior by the use of natural language. A lingubot will thus serve for chat purposes and its aim will be to stay unrecognized as a machine[2,3].

When it comes to information processing chatbots use two methods of analyzing text input: these are called semantic and syntactic parsers. Due to the binary character of computing devices, the very base for both parsers is algorithm. When it comes

[1] J. Searle, *Minds, brains and programs*, "Behavioral and Brain Sciences", 1980, 3, s. 417–457.

[2] H. Henderson, *Encyclopedia of computer science and technology*, New York 2009, s. 83–84.

[3] S. Russell, P. Norvig, *Artificial Intelligence. A modern approach*, New Jersey 2010, s. 25–28.

to data processing the mechanism works in the following way. First, the input data, e.g. a question, is segmented into smaller meaningful bits. This is where the lexical analysis takes place. Subsequently, a token is ascribed to each of the parts what allows the computer to translate the inquiry into its own numerical language. At this stage the syntactic parser replaces the semantic one. The next step is where the machine computes the digitalized question according to a stored grammatical pattern. This is also the phase where a syntactic tree is created according to the pattern yielded by the semantic parser. This ensures the grammatical correctness of the answer. The computer analyzes the generated symbol sequence of the input query and searches for a matching pattern. Lastly, the output answer in a digital form is compiled (translated) into lexical form and so the output is displayed on the computer screen[4,5,6].

Another aspect which requires a more thorough explication with respect to machine understanding is the very notion of what does it mean to understand. The theory offers diverse perspectives on the subject. The ones of interest with regard to this paper will be, first, a human-oriented and, second, machine-focused definition. On the one hand, the cognitive view on the notion of understanding maintains that understanding is embodied, i.e. humans build an internal model of the external world, and thus the meanings, via senses[7]. The human subject is not an isolated entity but someone who can interact with the world. Meaning as acquired in such a way is not numerical and there is neither LAD nor UG as Chomsky maintains. Meaning is understood in terms of mental representations whose precise specifications as for the neurological location of these representations which would allow

---

[4] G. Antoniou, F. van Harmelen, *A semantic web primer*, Cambridge 2008, s. 1–23.

[5] C. Brewster, Y. Wilks, *Natural language processing as a foundation of the semantic web*, "Foundations and Trends in Web Science", 2006, 1, s. 201–313.

[6] R. Grishman, *Computational linguistics. An introduction*, Cambridge 1994, s. 90–139.

[7] G. Lakoff, *Metaphors we live by*, Chicago 1980.

for their replication on non-biological grounds such as computers have not been detected yet. These representations have the status of "mental spaces" which are very troublesome in themselves[8]. On the other hand, machine-oriented definition views understanding as the processing of information which are numerical, and thus symbolic, in character. The implicit premise here is that human information processing in the brain would also be conducted in a symbolic manner.

Searle explains computer understanding in terms of his self-devised thought experiment called the Chinese Room. It is designed to refute certain arguments for machine thinking and understanding. Elements present in the experiment are a room with a book of codes (Chinese characters) and two people, person A, who is inside the room and who does not speak the Chinese language and person B, outside the room, who does speak Chinese but who does not know whether person A does or does not speak Chinese. Person B gives person A cards with messages written in Chinese, say questions, through an inlet in the door. Person A answers the inquiries correctly using the book of codes. Ultimately, person B is convinced that the person in the room knows Chinese. In fact, this is not the case because despite the successful communication person A merely performed a few manipulations on the signs. Nevertheless, s/he did understand neither what was written nor what s/he has written. The experiment is supposed to weaken (if not completely reject) the premise that human thinking and cognition may be computational in nature. As far as the brain is concerned, the experimentation implicitly supposes that it is a kind of a BlackBox, a computing machine whose precise operations are unknown but generating the correct output suffices to deem it as understanding, thinking and intelligent[9].

---

[8] "Mental spaces" as a way to organize knowledge in the mind as claimed by cognitive linguistics; as found in G. Fauconnier, *Mental spaces: Aspects of meaning construction in natural language*, Cambridge 1994.

[9] J. Searle, *Minds, brains*..., dz. cyt., s. 417–457.

2. Pragmatics

Pragmatics is a field of linguistics which describes the use of language with relation to the context it appears in. As Yule states, it is "concerned with the study of meaning as communicated by speaker (or writer) and interpreted by a listener (or reader)" therefore "it necessarily involves the interpretation of what people mean in a particular context and how it influences what is said"[10]. Another integral facet studied by pragmatics is the interlocutors' inference, namely, the implicit pre- and post-utterance knowledge shared by the speaker and hearer (presuppositions and implications respectively) which is not found in the overt written or spoken form of a message. This is the reason why implicature or inference as such is so challenging to be represented in numeral terms (in the computer programming language). Consequently, this entails the analysis of how meaning which is not visibly present in the explicit form of a message (i.e., the implicature) is successfully deciphered by the human speakers despite the fact that it cannot be extracted from the given oral or written data of a message. The notion of inference is closely related to implicature[11]. The term subsumes different kinds of implicatures (general, particularized or scalar to name just a few), however, for the purposes of this paper only the broad sense will be used.

3. Discussion

The first generation of AI solves the issue of language understanding by means of symbol operations. The system, i.e., a chatbot, manipulates on a prearranged set of signs where each of which is attributed to appropriate lexical entity. The strong AI hypothesis views the brain as a Black Box. To be more exact, it is perceived as a device whose internal operating is unknown. As a result, it is to be regarded in terms of its in- and output.

---

[10] G. Yule, *Pragmatics*, Oxford 1996, s. 3.

[11] P. Grice, *Logic and Conversation*, "Cole and Morgan", 1975, s. 41–58.

Syntactical patterns allow the machine to adjust the right output to a given input. Then, roughly speaking, a simple query (input, here: interrogative sentence), as for a computer, will have the following form of Q → V N (*Does it?, Is John?*) where the symbols stand for a question, verb and noun respectively. To this basic interrogative sentence pattern the system will search its database for a matching output sequence which in this case will be of the S → N V (*It does*, *John is*) form (S denotes a sentence, other symbols being the same). Nevertheless, it is easy to mistake sign manipulation with true semantic understanding. The correctness of the output messages generated by the system can mislead someone into thinking that it actually understands the content it processes. The emphasis is put on syntax, or to put it in different words, on grammatical patterns. Such approach neglects the semantic part, however. This can be observed in cases where the system fails to deliver an even acceptable answer despite its grammatical well-formedness and congruency with the input, e.g. *Colorless green ideas sleep furiously*[12]. An alternative situation supposes that the sentence pattern S → N V N yields the following output of *Maggie likes Bart* contrary to *Maggie sleeps Bart*[13]. From the discussion above it emerges that syntax does not suffice when it comes to a successful reconstruction of linguistic skills in a machine.

Although it is frequently referred to as the wastebasket of linguistics, pragmatics may come as aid to issues to which syntax is helpless. The field of pragmatics incorporates all those linguistic aspects which the other fields (semantics, syntax, etc.) do not have sufficient tools to describe. As a result "the subject matter and therefore the data of pragmatics was seen as made up of bits and pieces that could not conveniently be accommodated elsewhere"[14]. With connection to inferred meaning specifically there is no one-to-one relation between the input and output. In other words, because it is not the spoken or written form of the input and output

[12] N. Chomsky, *The logical structure of linguistic theory*, Chicago 1975, s. 15.
[13] R. K. Larson, *Grammar as science*, Massachusetts 2011, s. 87.
[14] Ch. Siobhan, *Pragmatics*, Houndsmills 2011, s. 11.

which is of interest to the interlocutors in transmitting indirect information (but the covertly present inference), in theory, the system should not be able to fit an appropriate pattern to something which it cannot compute (the machine is not capable of processing something invisible as the inference or implication are). Still, there are examples to be found which support the contrary like the one below:

1) User: Some academics are lazy.

2) Cleverbot: Some???[15,16]

Additionally,

1) User: Can you tell me the time?

2) Cleverbot: 11 07 pm.[17]

In both examples the bot's replies are acceptable and grammatical too. Despite the absence of information in the lexical form as well as the lack of syntactical incongruity between the input and output (the answer from the first sample is incomplete as for a sentence of English while the second, provided the input was interpreted directly, should display a *yes* or *no* reply; it does not, however, since this was not the information of interest to the User) a human interlocutor will be eager to admit that the bot managed to grasp the disguised meaning. As far as we can argue that the second instance is a matter of convention, and thus may be more or less formalized there arises a lot of doubt as for the in/correctness of such replies as the one below:

1) User: Can you tell me the time?

2) Cleverbot: Ceeelebrate good times c'mon![18]

The ambiguity which Cleverbot's response gives rise to renders the message difficult to inspect in connection with syntactic correctness. On one hand, it may be that the bot revealed not only

---

[15] M. Wilkowska, *Pragmatic analysis of language understanding and use by Artificial Intelligence systems (the case of chatbot language)*, pre-print, Uniwersytet Jagielloński, Kraków 2012, s. 75.

[16] Cleverbot is a state-of-the-art digital dialogue system said to be one of the best so far.

[17] Tamże.

[18] Tamże.

an extensive knowledge of language but also of the cultural (quote from a song) and temporal context provided it was an occasion to celebrate (since it might have been for the bot, depending on the time zone the bot was in). On the other, it may well be interpreted as unnatural in a context and as for a question which, by convention, triggers giving the present time. Still, for a human, it would rather be a problematic answer to judge in terms of its in/correctness and ir/relevance. What is more, suppose the first explanation was accepted, what is it in the first line of the exchange that makes its antecedent a well-formed, natural and correct answer? Further, how do we decipher the covert information? Consequently, how to represent it in the machine language? What is it in the answer containing an implicature that makes it understandable? Tests for the understanding of meaning in context by dialogue systems of AI revealed that when it comes to indirect or implied information processing two bots (Alice and Cleverbot)[19] upon complex examination performed very unsatisfactorily[20]. A subject strictly related with this phenomenon is how humans successfully arrive at the proper interpretation of covert meaning? Moreover, how to convert and ultimately implement a kind of implicature parser into a machine when syntactic patterns no longer obtain? Ultimately, what is the difference between syntax and pragmatics that renders the transition and compatible use of both almost impossible?

All of these operations are conducted by the software, not the hardware. This implies a kind of dualism[21] which in the philosophy of mind is known as functionalism[22]. This position on mind maintains that humans consist of two elements, i.e., a material and immaterial one. In computer terminology this corresponds to the hardware and the software respectively. This vision much as it is tempting generates a few, if not more, doubts, however.

[19] Alice is another digital dialogue system, however it is an older lingubot than Cleverbot and has been chosen to the research for comparative purposes.

[20] M. Wilkowska, *Pragmatic analysis…*, dz. cyt.

[21] J. Bremer, *Wprowadzenie do filozofii umysłu*, Kraków, 2010, s. 45–51.

[22] Tamże, s. 121–136.

Namely, is it really as Descartes maintained it that the human is "inhabited" by two substances, *res extensa* and *res cogitans*? Not many will claim that not only is it the wrong way of approaching the human being but more importantly that this exactly is *the wrong reading of the philosopher*! Such kind of described dualism implies a very strong separation of what is substantial and what is not despite occupying one place, that is, the human body. Then, it becomes an extremely challenging task to make these two interact seamlessly if they are to work well in a computer environment. In the field of philosophy of mind, there are plenty of various mind-body problem[23] theories. Why should it be the functionalistic one and not some other, e.g., an embodied one?

The question of "how to make a machine understand the way humans do" is central to this paper. Neuropsychology so far knows the rough answers for the mechanisms responsible for the spoken/read input recognition and those involving the production of oral and written output[24]. In the initial stages of the research on the brain tissue, it has been suggested that the left hemisphere which coordinates those processes. Latest neuroimaging research (e.g. Computer Tomography (CT), Positron Emission Tomography (PET) or Functional Magnetic Resonance Imaging(fMRI)) provides the evidence that linguistic competence requires the cooperation of both parts of the brain in order to perform properly. It is known that when it comes to language perception the left half is context-limited and decodes the literal meaning of words while the other half is the opposite – it is context-free, evokes many alternative meanings of words on grounds of their common features and is capable of detecting implicit meaning. Brain damages to the left hemisphere cause in the distortion and severe difficulties when trying to understand metaphors, proverbs, humor and the general line of a story[25]. The

[23] Tamże, s. 19–37.

[24] J. Binder, J. Frost, T. Hammeke, R. Cox, S. Rao, T. Prieto, *Human brain language areas identified by functional magnetic resonance imaging*, "The Journal of Neuroscience", 1997, 17, s. 353–362.

[25] M. Beeman, Ch. Chiarello, *Right hemisphere language comprehension: Perspectives from Cognitive Neuroscience*, Mahwah 1998.

connections between the two brain parts play a vital role as well. It is the corpus callosum, anterior commissure and interthalmic commissure which secure the flow of information. A question then arises, should a computer brain have a twofold construction in itself? What element, the hardware or the software, should it be? More specifically, how to ensure that the communication between the hemispheres will be satisfactory enough for the emergence of *human-like* understanding? Further, one could also ask what does a "satisfactory understanding" mean and how to measure it? There are no answers to these questions as for today.

There are other philosophical consequences, e. g., those concerning syntax insufficiency. These posit that the human mind as something intangible (*res cogitans* in the Cartesian terms) cannot be reduced to some physical, causally working substance. The difficult part of the electronic brain task is that reductionist approaches fail. Reductionism itself poses that it will suffice to minimize all intelligent activity to the workings of the material brain: the change of voltage difference between synapses and the change of neurotransmitters[26]. Then, it is not only the material elements which would be sufficient for the reconstruction of mind. Nevertheless, it seems that the mind is still something of a superstructure, meaning that it is built over on the material base. Such view would imply emergence[27] to be the key to understand the brain with all of its complexities.

A somewhat different issue concerns the fact that the fact that the human apparatus is biological contrary to computers which are electronic. The task will be to convert the latter – the hardware – into the former – the wetware.

It remains to be seen whether strong AI will be possible to build on such foundation as a computing machine. By creating an artificial mind within the realm of a computer we accept what the machine has to offer but also we are then confined to its obvious limitations. The point is that perhaps it is possible to create strong AI but the clue is in the kind of the machine. As was often the case in history,

[26] J. Bremer, *Wprowadzenie*..., dz. cyt., s. 111–121.
[27] Tamże, s. 19, 20, 147.

progress in various other, frequently unrelated to IT or CS fields and scientific disciplines may prompt the construction of such a machine in question (whether the invention will be called a machine or by other neologism (which is most probable) is a task for the future). It may be also that we already have the right theories but not yet the right machinery or devices to implement them into.

Excessive machine precision is another aspect which may hinder the creation of conversational AI. That is to say, human communication can be described in terms of the rule of minimal effort. The approach assumes that not all utterances be completely grammatical. This is contrary to what the strong AI supposes – it is constructed in an idealistic and overly realistic way which does not allow for any mistakes or errors. Nevertheless, this may be a point which requires modification since too exact and hyper-grammatical utterances are not natural for an everyday conversation let alone for the context in which the human-to-bot conversation takes place – the internet chat – which facilitates and promotes the violation of the written discourse rules.

Besides, there is a huge difference with connection to reality and its virtual counterpart. The world with its theories and scientific models is fragmentary and not holistic. It is in a constant flux; so there is little possibility for once and for all (forever) solved issues since everything can be subjected to re-definition, re-arrangement or simply approached from a novel perspective (see: ethical or physical theories). The artificial reality requires something quite the contrary – a complete system of principles and rules operating the world. Otherwise than that it is bound to malfunction. Trying to re-create the world in the artificial substitute is very much like trying to create (or rather simulate) the finite state of the world, that is to say, the state of the world as of its end. Only then will it be possible to definitely state that phenomenon X is this and that or that the definition of Y is this. Nevertheless, the creation of a strong AI system seems to be an attempt at striving to make the impossible to happen.

Of course, further interest should be also devoted to the method thanks to which it will be possible to measure whether

the machines really understand. This meta-task requires efforts not only on the IT part but mostly from the philosophy of science. In view of the presented context a simple Turing Test will not be a reliable manner to perform it.

Alongside all of the above remarks, there is the question, what would we need a strong AI for? Why should we spend the expenses on building a conversational agent which would not be distinguishable from a human and which would handle the so called general conversational skills (i.e., the inference or implicature)? These questions are acute especially when the expert systems (weak AI) are enough for contemporary needs.

Conclusion

As for now, these phenomena remain unanswered and so is the realization of strong AI. Much as the syntactic theories provide a sound foundation to the AI project, they nonetheless require the support of semantic parsers used at present but most importantly the aid in the form of computable pragmatics to complete the picture and the realization of strong AI. Contemporarily it is possible to screen and examine the brain by way of various neuroimaging devices. What they allow to investigate is "only" the material, *res extensa*, part. Still, relatively very little is known about *res cogitans*. And so, as for now and as far as language competence is concerned my hypothesis is the following: provided, it is possible neither to detect nor to track the immaterial part of our cognition – the mind – another premise being that machines are not fully capable of communicating and understanding on the human level it may well be the immaterial part exactly which is responsible for those aspects of language that cause so many complications. An emulation of an electronic brain could thus have the hardware as the material base. The software, however, should be something of a dual-software or di-software. This does *not* mean that a piece of hardware should have two operating systems or so. By di-software I mean that there is one item of software installed on the computer

but it should be dualistic in nature, similarly to the two substances as assumed by dualistic theories of mind. To be more precise, each "part" of the di-software would correspond to one part of the brain. One part could be syntax oriented. All of the present achievements in the field of computational linguistics could be inserted in there, that is, those pertaining to syntax. This would be the digital analogue of the left hemisphere. Its counterpart, the digital right hemisphere, could deal with all the implicit and indirect pragmatic aspects of language. The semantic part could be placed somewhere in the cyber-commissures. This is roughly consistent with the place occupied by our mental lexicons – the language center. Of course, this hypothesis is (crypto-) dualistic. Still, it remains to be seen whether there is another, yet undiscovered element which is imperceptible to both, the neuroimaging devices and the human eye.

References:

1. Antoniou G., Harmelen F. van, *A semantic web primer*, Cambridge 2008.
2. Beeman M., Chiarello Ch., *Right hemisphere language comprehension: Perspectives from Cognitive Neuroscience*, Mahwah 1998.
3. Binder J., Frost J., Hammeke J., Cox R., Rao S., Prieto T., *Human brain language areas identified by functional magnetic resonance imaging*, "The Journal of Neuroscience", 1997, 17, s. 353–362.
4. Bremer J., 2010. *Wprowadzenie do filozofii umysłu*, Kraków 2010.
5. Brewster C., Wilks Y., *Natural language processing as a foundation of the semantic web*, "Foundations and Trends in Web Science", 2006, vol. 1., no. 3–4, s. 199–327.
6. Chomsky N., *The Logical Structure of Linguistic Theory*, Chicago 1975.
7. Fauconnier G., *Mental spaces: Aspects of meaning construction in natural language*, Cambridge 1994.
8. Grice P., *Logic and Conversation*, "Cole & Morgan", 1975, s. 41–58.
9. Grishman R., *Computational linguistics. An introduction*, Cambridge 1994.
10. Henderson H., *Encyclopedia of computer science and technology*, New York 2009, Facts on file.

11. Lakoff G., *Metaphors we live by*, Chicago 1980.

12. Larson R. K.,*Grammar as science*, Massachusetts 2011.

13. Russell S., Norvig P., *Artificial Intelligence. A modern approach*, New Jersey 2010.

14. Searle J., *Minds, brains and programs*, *behavioral and brain sciences*, 1980, vol. 3, no. 3, s. 417–457.

15. Siobhan Ch., *Pragmatics*, Houndsmills 2011.

16. Wilkowska M., *Pragmatic analysis of language understanding and use by Artificial Intelligence systems (the case of chatbot language)*, Cracov 2012.

17. Yule G., *Pragmatics*, Oxford 1996.

A philosophical investigation in machine understanding. The case of implicit meaning

The question of machine thinking and understanding, once initiated by Alan Turing, has puzzled scholars from various disciplines. This paper aims at investigating some of the facets involved in the topic of machine language understanding with particular interest devoted to indirect meaning comprehension (more specifically, the implicature). So far as the subject under examination – the chatbot – manages to understand directly conveyed information, still much is to be done with respect to implicit data in which everyday messages (formulated in ordinary language) abound. This situation generates a number of not only hard-science questions but also, more importantly, given the viewpoint of this paper, it gives rise to a considerable amount of philosophically-oriented and frequently neglected issues and dilemmas too. This study is a brief investigation of exactly those phenomena.

Keywords

sztuczna inteligencja, SI, AI, chatbot, maszyny liczące, językoznawstwo komputacyjne, znaczenie, rozumienie, semantyka, pragmatyka, język naturalny, znaczenie niedosłowne