

Will it ever be possible for machines to talk clandestinely? Introducing the non-cooperative principle

Introduction

This article tries to contribute, from semantic, pragmatic and discursive theoretical questions, to the understanding of the present or future possibility that machines communicate with each other behind our back. For this, the forms of the opacity of natural language are presented first, in order to distinguish between implications, silences, secrets and lies. After this, the different types of secrets that could be shared by machines are explained, for which it is necessary to introduce new concepts such as the robotic Non-knowledge and the Non-Cooperative Principle, understood as a universal principle opposed to that defined by Grice and that could regulate the efficiency of clandestine communication. The hypothesis defended here is that for machines to be able to communicate clandestinely they should use another way of speaking the natural language, as we humans do, in such a way that the relevant information is only accessible to members of the same clan, without compromising them or compromising the commonly admitted truth.

How to tell secrets with words

Can machines in the future communicate with each other opaquely, so as not to be understood by us, humans? Can they build their own memory of uses, their own cultural backgrounds, different from any human background?

These questions are not trivial, since for decades Linguists and Computer Engineers work side by side to improve communication between men and machines. Improving communication should be understood here as improving the happy conditions of our language exchanges with machines.

We know that human communication bases most of the information exchanged implicitly, through inference mechanisms (logical but not only logical reasoning) thanks to truths and linguistic routines shared by both participants in the conversation. This means that many times, it may even seem that we humans speak often to say nothing, just for the pleasure of speaking, to fill a void or to avoid an awkward silence. In all cases, even in the most superficial and vain conversations, experts in cognitive pragmatics¹ ensure that there is a relevant exchange of information: information about the informative intention...

For some years, machines have already been able to interpret many implicit information, especially that triggered by the use of a particular word, regardless of context, thanks to logical reasoning based strictly on the analysis of the linguistic component.² That is, machines are able to recognize semantic implications. Now, it seems that machines are also capable of interpreting, with a little margin of error, certain pragmatic implicit contents that depend on both lexicon

¹Sperber Dan, Deirdre Wilson, *Relevance: Communication and Cognition*. Oxford: Blackwell. 2nd edn. 1995.

²Ducrot Oswald. *Pré-supposés et sous-entendus. Langue française. La sémantique*. 1969. pp. 30–43.

Volume 4 Issue 4 - 2020

Juan-Manuel López-Muñoz

Department of French and English Philology, University of Cadiz, Spain

Correspondence: Juan-Manuel López-Muñoz, Department of French and English Philology, University of Cadiz, Spain, Email jmanuel.lpez@uca.es

Received: December 23, 2019 | **Published:** August 25, 2020

and very conventional contexts, through reasonings based on linguistic and rhetorical components: they are the implicit contents called, from Grice (1975: 6), generalized conversational implicatures.³

These advances in the detection of the implicit contents of verbal interaction by the machines are largely due to the collective effort of linguists in the field of Natural language processing (NLP), heiress of comparative, functionalist, structuralist and transformational linguistics. Such effort, in collaboration with engineers of Artificial Intelligence, has focused on detecting, describing and quantifying the numerous automatisms with which human language works, at the both levels of encoding and decoding.

But there is one last type of implicature, which escapes automatisms for now, because they are strictly dependent on the context. Let's understand here by "context," on the one hand, in a narrow sense, the *situation d'énonciation*:^{4,5} that is, who are exactly the people who are speaking, how well do they know each other and to what extent do they share the same cultural background and the same language habits? and where and when are they exactly speaking? And that "context", on the other hand, must be considered in a broad sense, that is, *la scène d'énonciation*⁶ marked by certain discursive routines and communicative contracts⁷ more or less tacit, which speakers can respect or not, depending on their intentions and

³Cf. HP Grice. Logic and conversation. In: Cole P, Morgan J, editors. *Syntax and semantics. 3: Speech acts*. New York: Academic Press. 1975; 41–58.

⁴Benveniste Émile. *General linguistic problems. I et II*, Paris, Gallimard; Benveniste Émile. *The formal apparatus of utterance. Languages*. 1966/1974. pp. 12–18.

⁵Kerbrat-Orecchioni Catherine. *L'énonciation de la subjectivité dans le langage*. Paris, Armand Colin. 1980.

⁶Maingueneau Dominique. State of enunciation, situation of communication. In: MC Figuerola, et al. (eds.), *French linguistics in the new millennium*. Lleida. 2002, pp. 11–19.

⁷Patrick Charaudeau. From the enunciation scene to the contract round trip, in Johannes Angermüller and Gilles Philippe (eds.), *Analysis of discourse and enunciation devices. Around the works of Dominique Maingueneau*, Limoges, Lambert-Lucas, 2015. pp. 109–116.

the established symbolic power relations^{8,9} among the speakers, in addition to depending on the more or less institutionalized character of the space in which they are located, and their respective *formations discursives*¹⁰ and that socio-historical concrete *episteme*¹¹ when the conversation occurs or occurred.

For it is indeed a concrete occurrence (a “token”, according to Peirce’s terminology¹²), a pure phenomenon at a particular juncture and, therefore, something ever new, out of any base of pre-existing data on a machine. This type of implicatures, called *Particularized* ones (Grice, *op. cit.*, p.56) are strictly interpretations of the interlocutor, which the speaker can always cancel if he wants to.

Both generalized and particularized implicatures generally do not pose serious communication problems between humans: at most they can generate misunderstandings, which can always be repaired, or directly generate laughter, humorous scenes.

The probability that a particularized implication is correctly understood by the interlocutor, establishing a happy exchange (that is, that meets the expectations of the speakers) is actually quite high, due to the existence of a universal Cooperative Principle that governs natural language conversations. It was formalized by Grice (*op. cit.*, p. 57) (and improved by its continuators¹³). This principle presupposes an ideal model of communication in which speakers give all the information they must give (neither more nor less amount of information), that the information given is qualitative (that is, true) and relevant, and that’s expressed briefly and clearly.

This Cooperative Principle works in such a way that any infraction detected in one or several of its maxims (quantity, quality, relation, manner) will trigger in the interlocutor an inference mechanism intended to interpret implicatures (not so much intended to clarify a certain ambiguity in the content of what was said but in the speaker’s real intentions).

This interpretation mechanism usually works with great precision, favouring the economy of language, but as long as it involves choices, there is always a probability of error, so that the exchange risks at any time to become ineffective. However, when an error is detected, the inferences are normally revised or replaced by other more operative ones. This interactive interpretative model succeeds as long as the participants in the exchange want to cooperate and above all that they

share a common discursive memory,¹⁴ both essential conditions for the communication contract.

Any speech can have an infinite number of interpretations. This statement, which can now be regarded as a truism, deserves, however, to be examined with renewed interest, in the light of the question which concerns this article. The fact that a speech is capable of being understood in several different ways, sometimes even opposite, does not imply that its meaning is virtual or indefinite. The comprehension of a speech is, on the contrary, at every moment, the result of a choice from a set of current or already updated possibilities. Let us be clear: in this set of the possibilities, future significances are excluded. This thoroughness is very important in the sense that the act of understanding is thus necessarily linked to an act of remembering and recognition. Interpretation mobilizes recollections about concepts, people, places, times and objects summoned by the words of speech, as well as those about the functioning of language, the materiality of speeches and the ways of speaking. In short, it involves all of the shared knowledge that constitutes the background of a community. The number of possible interpretations of a speech is therefore limited by our memories, by our knowledge. The Not-yet-known (or even the Forgotten) limit the options, both placed at the frontiers of memory.

The fact that one can more or less easily identify the hidden meanings in discourse poses an important dilemma, as I explained in a previous paper,¹⁵ when this information must remain secret, or possibly only accessible to a particular recipient or to a clan (clandestine communication). Let me now return to some of the questions that I have already raised in the aforementioned paper, renewed under this new perspective. Should machines use secret codes to communicate secretly? We know that any code is potentially decipherable, and when a secret message is entered and decrypted, its content is revealed, despite the initial precautions, with total transparency.

For this reason, the most successful way to communicate secretly, between machines as it is between humans, requires using common language, taking advantage of its natural opacity. Only then it would be possible to avoid our secret being revealed. That is to say any potential secret between machines must be implicitly transmitted via particularized Implicatures to ensure its successful clandestine transmission.

For Particularized Implicatures does not compromise its speaker but the interpreter. So, if ever the secret is revealed, the speaker can always deny it by arguing that it is a misinterpretation.

A secret message could thus circulate from a machine to another machine while avoiding possible confiscations and reach its target recipient, by using Particularized Implicatures. But then, how can machines guarantee the correct interpretation in a hypothetical secret communication between machines? The relationship between secrecy and the unspoken and implicit is the same in humans as it could be in machines? Does understanding the secret between machines imply an act of (pre)cognition and no longer recognition?

To be able to try to find an answer to these serious (perhaps future) questions, we propose to ask ourselves here what type of communication contract mobilizes clandestine speeches in Humans.

⁸Bourdieu Pierre. On symbolic power. *Critique of Anthropology*. 1977;32(3):405–411.

⁹Butler Judith. *The power of words. Performative politics*, traduit de l’anglais (*Excitable Speech*, Routledge, 1997), Paris, Éditions Amsterdam. 2004.

¹⁰Haroche Claudine, Henry Paul, Pêcheux Michel. La sémantique et la coupure saussurienne: langue, langage, discours », *Langages*. 1971;24:93–106.

¹¹Michel Foucault. *The order of things: an archeology of the human sciences*. Paris, Gallimard. 2002.

¹²Peirce Charles Sanders (CP). *Collected Papers of C. S. Peirce*. In: C Hartshorne, P Weiss, A Burks, editors. 8 vol, Cambridge, MA: Harvard University Press. 1931-1958.

¹³The Cooperative Principle and the Grice’s maxims have been the subject of major posterior revisions, without however having been entirely disputed. D. Sperber and D. Wilson reduce them to the only maxim of relation, renamed principle of relevance (*Relevance: Communication and Cognition*. 2nd ed. Oxford, Blackwell, 1986/1995). Others, on the contrary, applied themselves to renovating them by incorporating submaxims into them, cf. notably S. Levinson, *Presumptive Meanings: The theory of generalized conversational implicature*. Cambridge MA: MIT Press, 2000.

¹⁴Sophie Moirand. Speeches, memories and contexts: about how allusion works in the press, Corela [Online], 2007.

¹⁵López Muñoz, Juan Manuel. The transmission of non-textual knowledge (Montpellier. *Actualite (Transmissions)*. 2011;2:19–31.

What knowledge do they involve? Is there a discursive memory of the secret? If so, what function would silence and forgetfulness play there? These are all questions that deserve to be asked. They will guide our reflections in the following paragraphs. But firstly, what could be the nature of such secrets shared between machines?

What kinds of secrets could machines tell?

It could be about deceptive or hostile information, threatening secrets in various degrees, including offensive strategies, non-legitimized knowledge, unfounded or refused by human's science, or perhaps knowledge concerning non-facts, that is to say to say concerning events which never took place or which did not take place yet (fake news), not deserving the consideration of Human's History (false predictions). By their common character of (imposed or desirable) exclusion, and despite of sometimes quite remarkable differences on several levels, from one type to another of these upsetting secrets, I will group them in this paper under the common name of Nonknowledge.¹⁶

This term will allow me to lay the foundations for a Non-Cooperative Principle which could, I hope, help me to answer the questions raised above. This principle would imply not only the transgression of the maxims of quantity, quality, relevance and manner previously mentioned, but the total reversal of the Grice's Cooperative Principle.

At first glance, the use of the negative prefix in the term of Nonknowledge would seem to describe an attempt to abolish positive knowledge, that is to say an ambition to destroy commonly accepted knowledge, that which takes official human channels to build and circulate. But Nonknowledge is not necessarily something nihilistic. Its disturbing nature is mainly due to its deconstructive mode of operation. This deconstruction operates on three levels: on the one hand, it seeks to overthrow hierarchies by highlighting negation, which results in a "positivization" of the forgotten and the denied; on the other hand, it brings out the undecidable ("neither ... nor", or "else ... or else") and concepts that resist opposition patterns¹⁷ ("other", "rest", etc.); third, it attempts to re-inscribe positive knowledge in a game that is no longer dialectical but summative: "and ... and".

Nonknowledge can therefore only be understood by the dynamic that it maintains with its complement, positive knowledge (or just knowledge). In order to better determine the scope of the negative prefix, we will examine this relationship below. These reflections will allow us here and there to problematize Nonknowledge from a pragmatic point of view, and will finally lead us to make observations on the nature of the contract that clandestine communication between machines could request.

Nonknowledge often works, at least between humans, in a way like the double knowledge, in the Freudian sense: it is made, in this case, of repressed or unrealized knowledge, which subsists in the unconscious and emerges only in situations of automatism or lack of control (madness, hypnosis, reverie, narcotic states, etc.). It can thus be perceived and defended as an inner knowledge, which is produced and transmitted by unconscious repetitions, but subsists in repression and omission. By posing Nonknowledge as the double of the knowledge, by playing the role of hidden, decapitated, castrated

precursor brother, Nonknowledge implies a "delay": the time necessary for the constitution of a positive knowledge around a second focus, like an ellipse. At this point in our reflections, we would be tempted to see in this kind of Nonknowledge platonic reminiscences of the archaic or maybe anticipatory knowledge of the future (in this sense, Nonknowledge would come close to science fiction). The timelessness of Nonknowledge can therefore only be understood in opposition to the temporality of the traditional concept of knowledge.

But how would this Nonknowledge could work between machines, given their lack of unconsciousness. We could only think of it as being knowledge which is inadmissible from the point of view of the state of human knowledge, that one legitimized by our institutions. The robotic Nonknowledge can thus be made up of scientific hypotheses not conceived or rejected by humans, or else historical facts ignored or unexplored by humans. It is thus made up of gaps in our (human) knowledge and breaks in the chain of transmission of our (human) knowledge, which are probably much more abundant - the first - and frequent - the second - than we think, and from which several clandestine groups can benefit (an hypothetical clandestine group of machines), self-proclaimed holders of the key to these unknown doors.

This Nonknowledge would be so diluted and concealed in the common interaction between machines that the uninitiated human reader, not sharing the same machines' background, would easily fall into the traps of ambiguity, lies and errors.

In other cases, robotic Nonknowledge may consist of positive human knowledge that is reserved for an elite. Its clandestine circulation through machines aims to guarantee its conservation and its dissemination only within this elite, by preventing the massive diffusion to a large public. This Nonknowledge is thus a secret shared between an elite of humans and the machines which serve them as their mediating instrument. But why "separate" such positive knowledge? maybe that's why the general public is not supposed to be ready to receive it yet? Is there a fear that general public will misuse this knowledge? In this case, Nonknowledge would only be accessible to those who have the right to know (a human clan and their machines, or an hypothetic clan of machines), who only are supposed to know how to assume the chaotic component of the Nonknowledge, while putting its threatening dimension into perspective, to the point of welcoming it no longer as a perturbing element but as a future regenerative principle, as a kind of positive negative knowledge.

Clandestine communication essentially conveys information of a subversive nature intended to circulate unnoticed. It generally presupposes a secret (Latin *Clam*) and imply a little group (a *clan*) which, for various reasons, is created in resistance and is built in an extraterritorial (online) space. The clan and its discourse are tolerated as long as they do not seek to truly threaten the structure of the positive knowledge in place. Since clandestine speeches are potentially destabilizing and, because of this, they risk provoking repressive reactions, the clan must find ways to communicate outside the control of some dominant group or enemy. Among the strategies that are available, paradoxically, those that aim to obscure the message (through, for example, encryption or steganography) are the most likely to be prevented. As I pointed out above, any secret code is potentially decipherable, and when a secret message becomes readable, it shows its contents explicitly. In addition, the use of a secret code openly manifests a threatening intention, automatically triggering the repressive mechanisms that were intended to be avoided.

¹⁶We will not understand Nonknowledge in the sense of Inner Experience, as Georges Bataille do in *The Unfinished System of Nonknowledge*. Trans. Michelle Kendall and Stuart Kendall, Minneapolis: University of Minnesota Press, 2001.

¹⁷Baudrillard Jacques. *Simulacra and simulations*, Paris: Galilée. 1981.

On the other hand, clandestine communication strategies that seek not to conceal a message but simply to conceal it using the common code benefit from the opacity of language and therefore have a much greater chance of success. The part given to the chance of many interpretations allows the message to pass the filters of the “enemy” and thus reach without problem its agreed interlocutor.

Clandestine communication is the subterfuge used for the transmission of secrets as well as for the preservation of knowledge that resists denial and oblivion (that is, for the transmission and preservation of Nonknowledge). It also makes it possible to relay ever-later information that could not be understood until the future. In short, it is at the service of Nonknowledge and therefore, instead of engaging a discursive memory it mobilizes a memory of silence.

In fact, Nonknowledge requires a reading that breaks with current models, by implementing devices that break down the text, reveals the silence of the words and invalidates the usual mechanisms of inferences. Nonknowledge, instead of inviting us to control the meaning of the texts, calls for taking the reins of reading, by turning off our human “automatic reading” which nevertheless makes everyday communication so effective. The words of Nonknowledge thus become hollow words, sound carcasses capable of referring otherwise: for example, by conveying forgotten or ignored meanings, by referring to words from another language, or by referring in an unconventional onomatopoeic way.

By breaking with the current models of reading, Nonknowledge is constituted as a kind of elitist knowledge. It uses common language but does not use it “like everyone else”. Nonknowledge requires a very good mastery of the functioning of common language (pragmatic level) and of the discursive constraints (discursive level), and above all, of the way of thinking about this functioning (metapragmatic and metadiscursive levels) as well as of the relationship between interlocutors (metacommunicative level). Nonknowledge is therefore only accessible by adopting a somewhat schizophrenic perspective, exotopic to the highest degree. Because of this difficulty, its transmission often requires initiation training, which further reinforces the common perception of Nonknowledge as being essentially clandestine and elitist.

The Non-Cooperative Principle

Nonknowledge transmission then uses a common language with a hidden meaning. But unlike the implicit meaning which benefit both from the logic of the Logos or from the Cooperative Principle aforementioned, the hidden meaning associated with Nonknowledge generates a quest which ruins the obsession with the mastery of meaning, overturns Grice’s conversational maxims and undermines generalized conventions. Because of this reversal of Grice’s maxims, we support here the idea that clandestine communication presupposes a Non-Cooperative Principle, that requires initiation within a clan. In the following paragraphs, I explain in more detail the nature of this non-Cooperative Principle, by describing each of the main conversational maxims governing the clandestine communication, as they could possibly occur between machines.

The Non-Cooperative Principle can be largely stated as follows: your contribution to the conversation, when it occurs, must seem able to meet the objective or the accepted direction of the verbal exchange in which you are supposed to be engaged.

This general principle is explained by the following four clandestine maxims:

- i. Approximation maxim: your contribution must contain too much or very little information. What you say too much or what you don’t say is what matters.
- ii. Maxim of dissimulation: Affirm what can be considered false, affirm post-truths (*a posteriori* truths). Above all, avoid giving enough evidence, but by concealing as much as possible so as not to be suspected of wanting to abolish the common principle of reality.
- iii. Maxim of incongruity: be inconsistent, speak without cohesion. What does not seem relevant is precisely what is relevant.
- iv. Maxim of Imprecision: Be ambiguous, be verbose. Take advantage of chaotic speech and fragment your speeches. Speak so that you are understood only by members of your clan or other people who may understand you in the future.

Conclusion

From this Non-Cooperative Principle, two main conclusions can be drawn. First, the idea that one can speak secretly using the common language as well as the institutional channels and places of positive knowledge. This statement might seem trivial at first glance, but it is less so for it calls into question a fairly generalized belief that clandestine communication involves illegal or criminal people and it is linked to hidden, obscure or underground places and linked to encrypted languages.

Furthermore, from Non-Cooperative Principle follows the idea that clandestine communication is not a matter of inference of implicit senses. Because, in the end, speaking implicitly is “speaking like everyone else.” Precisely, to transmit Nonknowledge it is necessary to speak “differently”, that is, by deleting relevant information, by repeating superfluous things, by blurring or fragmenting the information, by erasing the points of view.

Nonknowledge, unlike the implicit meaning, require a review of ordinary communication contracts. Implicit contents (presuppositions, implications, generalized implicatures) mobilizes *topoi* and *a priori*; they can be inferred through the intervention of shared background and positive knowledge or deduced from natural logical relationships. The implicit thus allow communication by allusion.

In addition, clandestine communication is not allusive but elusive, which is why the hidden meaning, once decoded, still remains cryptic. Nonknowledge implies shadow, silence, oblivion, ellipsis ... not clarity. It requires not only a quest for the hidden meaning, but also a practice of weaving: the fragments of Nonknowledge must be patched afterwards, because the global clandestine message results from the combination of the pieces (at the permanent risk of the lost piece).

Theories that defend the idea of the existence of hidden knowledge are often called “conspiratorial theories” and rightly attacked on the impossibility of a global orchestration of secrecy. But does Nonknowledge presuppose a large-scale conspiracy? We know that legitimate knowledge is gradually built up as an edifice, guided by a desire for permanence and stability. It is in principle organized and well assembled. Each new knowledge strengthens the structure, repairing the old pillars or replacing them with more efficient ones. The positive knowledge utterances are perfectly related to a precise moment, to a given place and to a singular person. He sets up heterogeneous discourses¹⁸ where the voices of the enunciators are articulated

¹⁸Cf. J Authier-Revuz. Hétérogénéité(s) énonciative(s). *Langages*. 1984;73: 98–111.

according to a complex game of enunciative responsibilities¹⁹ where co-enunciation and especially over-enunciation predominate.

In the negative and deceptive logic that Nonknowledge establishes, we must expect that, Nonknowledge appears to be as something unorganized, fragmented, scattered, which is conveyed by discourses without anchorages, where strategies of under-enunciation prevail in order to muddle (or even to delete) the marks of the enunciation process.

Unlike the positive knowledge, which has a syntactic structure and establishes a solid network of relationships between people, space

and time, the Nonknowledge, on the contrary, brings out paradigmatic relationships, anarchy, anonymity, timelessness, fragmentation and dispersion, ensuring clandestine communication without the risk of undesirable disclosures.

Acknowledgments

None.

Conflicts of interest

The authors declare no conflicts of interest.

¹⁹Cf A Rabatel. *Une histoire du point de vue*. Paris/Metz, Klincksieck/CELTED, 1997.