The language of thought as a logically perfect language

Andrea Bianchi*

Abstract

Between the end of the nineteenth century and the first twenty years of the twentieth century, stimulated by the impetuous development of logical studies and taking inspiration from Leibniz's idea of a *characteristica universalis*, the three founding fathers of the analytic tradition in philosophy, i.e., Frege, Russell, and Wittgenstein, started to talk of a *logically perfect language*, as opposed to natural languages, all feeling that the latter were inadequate to their (different) philosophical purposes. In the second half of the twentieth century, however, the very idea of a logically perfect language ceased for various reasons to seem attractive to analytic philosophers. Thus, it might appear that this idea could be classified together with the many other bizarre ideas that from time to time surface in the history of philosophy – an idea that perhaps had a beneficial impact on the development of twentieth century logic, but which can now be put to rest. In this brief note, I contend that this conclusion may be too hasty. Indeed, if a well–known empirical hypothesis advanced in 1975 by Jerry Fodor turns out to be true, then there is a logically perfect language, after all. More precisely, I argue that, if it exists, Fodor's *language of thought* possesses the main characteristics a logically perfect language is required to have.

Keywords: Language of thought; Logically perfect language; Logical form; Analysis; Understanding.

1.

As is well known, the search for a (sometimes, *the*) perfect language, which Umberto Eco (1993) quite convincingly showed to be a recurring theme in Western culture, acquired a specific form in the hands of a group of philosophers working between the end of the nineteenth century and the first twenty years or so of the twentieth century. Indeed, stimulated by the impetuous development of logical studies and taking inspiration from Leibniz's idea of a *characteristica universalis*, Gottlob Frege, Bertrand Russell, and Ludwig Wittgenstein, namely the three founding fathers of the analytic

* University of Parma.

tradition in philosophy, started to talk, each in his own way, of a *logically perfect language*. In fact, they all felt natural languages to be inadequate to certain philosophical purposes they had[1]. Natural languages are affected by vagueness and ambiguity (both lexical and structural), and they contain expressions apparently without reference ("Pegasus")[2]. Because of this, they can easily lead to mistakes in reasoning. In *Über Sinn und Bedeutung*, after presenting a problematic case for his account of subordinate clauses, Frege for example comments:

> This arises from an imperfection of language, from which even the symbolic language of mathematical analysis is not altogether free; even there combinations of symbols can occur that seem to stand for [*bedeuten*] something but (at least so far) are *bedeutungslos*, e.g. divergent infinite series. [...] A logically perfect language [*eine logisch vollkommene Sprache*] (*Begriffsschrift*) should satisfy the conditions, that every expression grammatically well constructed as a proper name out of signs already introduced shall in fact designate an object, and that no new sign shall be introduced as a proper name without being secured a *Bedeutung*. The logic books contain warnings against logical mistakes arising from the ambiguity of expressions. I regard as no less pertinent a warning against apparent proper names that have no *Bedeutung*. The history of mathematics supplies errors which have arisen in this way. This lends itself to demagogic abuse as easily as ambiguity – perhaps more easily. 'The will of the people' can serve as an example; for it is easy to establish that there is at any rate no generally accepted *Bedeutung* for this expression. It is therefore by no means unimportant to eliminate the source of these mistakes, at least in science, once and for all. Then such objections as the one discussed above would become impossible (Frege, 1892, p. 41).

One can find similar observations in Russell's writings, and in Wittgenstein's *Tractatus*. Russell, in particular, insisted on many occasions, starting from *On Denoting* (1905), on the fact that the surface grammar of a sentence may mislead us: it is not immediately evident, for example, that a sentence such as "The queen of England is aged" involves a quantification. For these philosophers, a logically perfect language would be one not having any of these defects of natural language[3]. In fact, «[a] logically perfect language has rules of syntax which prevent nonsense, and has single symbols which

---

1. For more on the uses that have been made of the idea of a logically perfect language in the analytic tradition, see HYLTON (2013).

2. According to Andrea Iacona, another alleged defect of natural languages is that some of its expressions are context sensitive. In contrast, he claims, «a logically perfect language is free from context sensitivity» (2018, p. 23). This, however, seems dubious to me. I do not see any reason why a logically perfect language should not count indexicals and demostratives among its expressions. In fact, Russell went so far as to write that «[t]he only words one does use as names in the logical sense are words like 'this' or 'that'» (1918–19, p. 201).

3. For a discussion of Russell's *On Denoting* relating it to the idea of a logically perfect language, see KAPLAN (1970) and MARCONI (1981).

always have a definite and unique meaning» (Russell, 1922, p. X)[4]. Hence, any such language would not be affected by vagueness and lexical or structural ambiguity, it would not contain expressions without reference, and it would be such that the inferential potential of any sentence – the logical relations of any sentence with any other sentence – is not hidden, so to speak. In a word, it would be a language in which there is no mismatch between syntactic structure and logical form. And, of course, it would be a language capable of expressing everything that can be expressed: for any meaningful sentence of a natural language there would be a sentence of the logically perfect language having exactly the same (truth–conditional) meaning. Actually, Frege dreamed of building an *artificial* language with such characteristics: his *Begriffsschrift* was meant to be a first step in this direction. And, some years later, Russell made some further steps. As he wrote in the second lecture of *The Philosophy of Logical Atomism*, «[t]he language which is set forth in *Principia Mathematica* [the book on the foundations of mathematics he wrote together with Alfred Whitehead] [. . .] aims at being that sort of language that, if you add a vocabulary, would be a logically perfect language» (Russell, 1918–19, p. 198). Interestingly for my purpose here, the passage continues in the following way:

> Actual languages are not logically perfect in this sense, and they cannot possibly be, if they are to serve the purposes of daily life. A logically perfect language, if it could be constructed [Note the caution here, as well as in the passage quoted in the preceding footnote], would not only be intolerably prolix, but, as regards its vocabulary, would be very largely private to one speaker (*ibidem*).

Now, there is no doubt that with their work related to these issues Frege, Russell, and Wittgenstein made long–lasting contributions to both philosophy and logic. The fact is, however, that the very idea of a logically perfect language soon ceased for various reasons to seem attractive to analytic philosophers – as is well known, Wittgenstein himself changed his mind on this. To use Richard Rorty's (1967) apt terminology, "Ideal Language Philosophy" was quickly supplanted by "Ordinary Language Philosophy". And even those philosophers who, like W. V. Quine, were not hostile to formal methods and took the logical regimentation of natural language to be helpful for doing science, stopped dreaming about a logically

---

4. It is perhaps worth noting here that this passage immediately precedes the one with Russell's famous, and contested, interpretation of the aim of Wittgenstein's *Tractatus*: «Mr Wittgenstein is concerned with the conditions for a logically perfect language – not that any language is logically perfect, or that we believe ourselves capable, here and now, of constructing a logically perfect language, but that the whole function of language is to have meaning, and it only fulfils this function in proportion as it approaches to the ideal language which we postulate» (*ibidem*).

perfect language: we can perhaps improve a bit on natural languages, but that is all.

Thus, it may seem that the idea of a logically perfect language can be classified together with the many other bizarre ideas that from time to time surface in the history of philosophy – an idea that perhaps had a beneficial impact on the development of twentieth century logic, but which can now be put to rest. We must just take what we have, natural languages with all their logical imperfections, because there is, and there can be, no logically perfect language.

2.

But is it really so? In this brief note, I want to suggest that no, perhaps it is not. Perhaps, there is a logically perfect language, after all. And, if there is, it is not a fancy artificial language that some extremely smart logician has been able to construct. Not at all. On the contrary, it is a language that we all have, and have always had, at our disposal, and that we use constantly, even though it is not a natural language, at least if a natural language is a spoken, or otherwise public, language.

More precisely, what I want to contend is that if a well–known empirical hypothesis advanced in 1975 by Jerry Fodor turns out to be true, then there is a logically perfect language. The hypothesis comes in the area of what Fodor (1975, pp. VII–IX) calls "speculative psychology". In the last few decades it has informed thousands of studies in the cognitive sciences, and is still alive – it has fierce opponents, but also staunch defenders[5]. In a nutshell, the hypothesis says that thinking is couched in a symbolic system realized in the brain. This symbolic system is like a language in many respects – in particular, it has a syntax and a (truth–conditional) semantics. No wonder, then, that Fodor called it the "language of thought". My claim is simply that, if it exists, the language of thought is a logically perfect language, since in order to play the roles it is supposed to play in Fodor's theory (the so–called *representational theory of mind*) it has to have the main characteristics that a logically perfect language is required to have: there must be no mismatch between syntactic structure and logical form, and it has to be capable of expressing everything that can be expressed, without any ambiguity.

---

5. Generally, the hypothesis is rejected by those who oppose functionalism in philosophy of mind and classical computationalism in cognitive sciences, such as *eliminative materialists* (e.g., CHURCHLAND, 1981), *connectionists* (e.g., CLARK, 1989; see also the essays collected in MACDONALD and MACDONALD, 1995), and *embodied cognition theorists* (e.g., BARSALOU, 1999, PRINZ, 2002, CLARK, 2008, and CHEMERO, 2009). But functionalism and classical computationalism are still alive and well. See BIANCHI (2006) for some critical considerations on Prinz's book.

As far as I know, Fodor never mentions the idea of a logically perfect language. In fact, he arrives at the idea of a language of thought through entirely different considerations. Thus, in a sense it is a striking coincidence that he found exactly what Frege, Russell and Wittgenstein were searching for. Nevertheless, it is surprising to my eyes that almost no one noticed this coincidence. Both Frege's, Russell's and Wittgenstein's considerations related to the idea of a logically perfect language and Fodor's considerations related to the idea of a language of thought are well known among analytic philosophers, but none of them seem to have connected them. Indeed, it is curious that the only work I am aware of reaching conclusions to a certain extend similar to mine comes from a semiotician rather than an analytic philosopher. In fact, at the time when Eco was working on his book on perfect languages, one of his students, Giovanna Cosenza, wrote her PhD dissertation on Fodor's hypothesis. The title of the dissertation was *Il linguaggio del pensiero come lingua perfetta* ("The language of thought as a perfect language"). Cosenza summarizes the conclusions she drew there in an article with the same title published in *Versus* (1994). Basically, she construes Fodor's language of thought as a perfect language similar to those theorized in the seventeenth century by philosophers such as Francis Bacon, George Dalgarno, and John Wilkins. As should be clear by now, I believe Cosenza was quite right, and I see myself as simply elaborating on her suggestions. The only difference is that I focus on Frege's, Russell's, and Wittgenstein's projects, which Cosenza did not discuss, rather than on Bacon's, Dalgarno's, and Wilkins'. On the one hand, they are projects I am far more familiar with. On the other, it seems to me that Fodor's language of thought, if it exists, is not just a perfect language in a generic sense, but specifically a *logically* perfect language: as I mentioned above, it has almost all the characteristics that make a language perfect from a logical point of view[6].

Unfortunately, I do not have the space here to go into details and to explain why Fodor's language of thought, if it exists at all, must have all these characteristics. Very briefly, what is crucial is that, according to Fodor's representational theory of mind, mental processes (transitions from mental states to mental states) are computations over sentences of the language of thought. These sentences have semantic properties (truth–conditions), but the computations are sensitive only to their syntactic properties. Under these assumptions, the rationality of our mental processes can only be explained by the fact that the syntactic properties of the language of thought

6. I say "almost" because according to Russell (1918–19, p. 197) a logically perfect language does not allow for synonymity («In a logically perfect language, there will be one word and no more for every simple object»), whereas according to Fodor the language of thought does. However, it seems dubious to me that a language allowing for synonymity should *ipso facto* count as *logically* imperfect.

sentences *mirror* their semantic properties, which means, roughly, that their syntactic structure *is* their logical form[7].

3.

Rather than elaborating on this, I will proceed by briefly considering the relation that, according to Russell, natural language sentences bear to the thoughts we express by them, which should allow us to see from a different perspective the connection between Russell's idea of a logically perfect language and Fodor's idea of a language of thought. Russell's theory of thought is based on his adoption of the so–called Principle of Acquaintance, which he states in various places. Here is the version from *The Problems of Philosophy*: «*Every proposition which we can understand must be composed wholly of constituents with which we are acquainted*» (Russell, 1912, p. 32). In this formulation, «every proposition which we can understand» means *every thought which we can have* or *every content which we can entertain in thinking*. In a nutshell, Russell's idea is that we think by mentally combining, so to speak, certain entities with which we have a special direct and immediate cognitive relation he calls "acquaintance". These entities are the *primitives* of our thoughts, and the mode by which they are combined in a thought is its *logical form*. Due to his empiricist epistemology, however, Russell believed that we cannot be acquainted with material entities such as the notebook on which I am typing now. Rather, we are acquainted with sense data and with certain abstract entities. Hence, the primitives we mentally "combine" when we think are either sense data or abstract entities of a

---

7. On the idea of a syntactic mirroring of semantic properties, see Kaplan (1970, pp. 234–236), who uses it to characterize the notion of a logically perfect language. Fodor writes: «we know from formal logic that certain of the semantic relations among symbols can be, as it were, "mimicked" by their syntactic relations; that, when seen from a very great distance, is what proof–theory is about. So, within certain famous limits, the semantic relation that holds between two symbols when the proposition expressed by the one is implied by the proposition expressed by the other can be mimicked by syntactic relations in virtue of which one of the symbols is derivable from the other. We can therefore build machines which have, again within famous limits, the following property: the operations of such a machine consist entirely of transformations of symbols; in the course of performing these operations, the machine is sensitive solely to syntactic properties of the symbols; and the operations that the machine performs on the symbols are entirely confined to alterations of their shapes. Yet the machine is so devised that it will transform one symbol into another if and only if the symbols so transformed stand in certain *semantic* relations; e.g., the relation that the premises bear to the conclusion in a valid argument. Such machines – computers, of course – just *are* environments in which the causal role of a symbol token is made to parallel the inferential role of the proposition that it expresses. [. . .] Computers are a solution to the problem of mediating between the causal properties of symbols and their semantic properties. So *if* the mind is a sort of computer, we begin to see how you can have a theory of mental processes that succeeds where associationism (to say nothing of behaviorism) abjectly failed; a theory which explains how there could regularly be nonarbitrary content relations among causally related thoughts» (Fodor, 1985, pp. 22–23).

certain sort, which are, then, the only constituents of our thoughts. Note, however, that natural language sentences never contain names for those very peculiar entities which Russell calls "sense data". If these sentences are meaningful, they express a thought, but they express it in a nontransparent way: their syntactic structure does not mirror the structure of the thought, since their syntactic primitives (roughly, the words that occur in them) do not correspond to the constituent of the thought. In fact, as Wittgenstein famously wrote in the *Tractatus*:

> Man possesses the ability to construct languages capable of expressing every sense [...].
>
> Everyday language is a part of the human organism and is no less complicated than it.
>
> It is not humanly possible to gather immediately from it what the logic of language is.
>
> Language disguises thought. So much so, that from the outward form of the clothing it is impossible to infer the form of the thought beneath it, because the outward form of the clothing is not designed to reveal the form of the body, but for entirely different purposes.
>
> The tacit conventions on which the understanding of everyday language depends are enormously complicated (Wittgenstein, 1921, prop. 4.002).

We can nevertheless understand natural language sentences, but this requires a process of *analysis*, and sometimes the process of analysis goes wrong, and we end up misunderstanding. This could not happen with sentences of a logically perfect language, because a sentence of a logically perfect language requires no analysis: its constituents, which Russell calls "logically proper names", would directly stand for the constituents of the thought it expressed, and its syntactic structure would mirror the logical form of the thought. If there were a logically perfect language, we could characterize the analysis, hence the process of understanding, in terms of a *translation*: one has to map a natural language sentence, which expresses a thought "opaquely", onto a sentence of the logically perfect language that expresses that same thought "transparently"[8]. But this is exactly what one does, if Fodor is right and there is a language of thought: one understands a sentence he or she hears or reads by mapping it onto a sentence of the language by which he or she thinks, which, obviously, "expresses" a thought transparently. With regard to this, Fodor is quite explicit, although his terminology is very different. In discussing «the mechanisms whereby human beings exchange information via natural languages» (1975, p. 116), and,

---

8. On the idea of analysis as translation (paraphrase) and its historical vicissitudes, see Marconi (2019).

more precisely, «the psychological mechanisms involved in *understanding* a natural language» (p. 116n, my italics), he claims that the «the internal representation of a [natural language] sentence is simply its *translation* in the language of thought» (*ibidem*, my italics)[9]. Here is an exemplification:

> Suppose that *F* is that formula of the internal code [the language of thought] which corresponds to the English sentence 'There's an ink–blot on this page' (hereafter, '*S*'). Then, presumably, understanding tokens of *S* involves assigning tokens of *F* as their internal representations, and believing that a certain token of *S* is true involves believing that the corresponding token of *F* is true. A natural account of what is involved in believing that a token of *F* is true is simply that *F* is *taken* to be true in those computations in which it is involved; e.g., that it is treated as a nonlogical axiom in those computations (Fodor, 1975, p. 117).

As I said, as far as I know Fodor never mentions the idea of a logically perfect language. There is a passage in his 1975 book, however, where he connects his views about the language of thought and the tradition of which Frege, Russell and the early Wittgenstein are illustrious exponents:

> The notion that a theory of meaning serves, in effect, to pair natural language sentences with some sort of canonical representation of their truth conditions is, of course, not new. It has been in the philosophical literature for as long as philosophers have distinguished between the surface form of sentences and their 'logical' form. Indeed, the precise point of this distinction has always been that the sentences of a natural language do not provide appropriate domains for the application of logical rules, but that some specifiable translations of such sentences would. To represent the logical form of a sentence is to represent the truth condition of the sentence explicitly, in a way that the sentence itself fails to do (Fodor, 1975, p. 114).

Fodor continues by stressing a "difference" between his account and "this tradition". Contrary to the latter, he writes,

> we are taking the notion of a canonical representation seriously as part of a *psychological* theory; the appropriate canonical representation of a sentence is the one that the speaker has in mind when he utters the sentence and the hearer recovers when he understands what the speaker said; i.e., it is that representation which makes explicit what utterances of the sentence are intended to communicate (Fodor, 1975, p. 114).

In my opinion, however, the difference is not so great as Fodor seems to suggest. I cannot argue for this here, but I believe that Wittgenstein, and perhaps, more confusedly, Russell, although almost certainly not Frege, conceived of a logically perfect language as something *actually* involved in the

---

9. Hence, «linguistic and psycholinguistic theories, insofar as they contribute to accounts of communication, must specify the procedures whereby this translation is affected» (p. 117).

activities of understanding and thinking[10] (What they lacked, in contrast to Fodor, was, of course, a *naturalistic*, *computational*, account of these activities).

Fodor's is an empirical hypothesis. We do not know yet whether it is true or false: as I said, it has fierce opponents and staunch defenders (see Fodor, 2008 for a recent revisitation and development of it). What I have tried to argue here is only that until we get a clear verdict on it, it would not be wise to classify the idea of a logically perfect language as another bizarre product of some excessively creative philosophers. Maybe there is a logically perfect language, after all. And if there is, we use it all the time, even though we are not aware of this[11].

Bibliographical references

BARSALOU L.W. (1999), *Perceptual Symbol Systems*, "Behavioral and Brain Sciences", 22: pp. 577–660.

BIANCHI A. (2006), *Critical Notice of* Furnishing the Mind, "Philosophical Books", 47: pp. 52–61.

CHEMERO A. (2009), *Radical Embodied Cognitive Science*, MIT Press, Cambridge, MA.

CHURCHLAND P.M. (1981), *Eliminative Materialism and the Propositional Attitudes*, "The Journal of Philosophy", 78: pp. 67–90.

CLARK A. (1989), *Microcognition. Philosophy, Cognitive Science, and Parallel Distributed Processing*, MIT Press, Cambridge, MA.

——— (2008), *Supersizing the Mind. Embodiment, Action, and Cognitive Extension*, Oxford University Press, Oxford.

COSENZA G. (1994), *Il linguaggio del pensiero come lingua perfetta*, "Versus", 67: pp. 77–120.

ECO U. (1993), *La ricerca della lingua perfetta nella cultura europea*, Laterza, Roma–Bari.

FODOR J.A. (1975), *The Language of Thought*, Harvard University Press, Cambridge, MA.

——— (1985), *Fodor's Guide to Mental Representation. The Intelligent Auntie's Vade–Mecum*, "Mind", 94: pp. 76–100 (Reprinted in J.A. FODOR, *A Theory of Content*

---

10. Diego MARCONI (2019, p. 350) seems to agree, at least as far as Wittgenstein is concerned.

and Other Essays, MIT Press, Cambridge, MA 1990, pp. 3–29 [page numbers given relate to this volume]).

——— (2008), *LOT 2. The Language of Thought Revisited*, Clarendon Press, Oxford.

FREGE G. (1892), *Über Sinn und Bedeutung*, "Zeitschrift für Philosophie und philosophische Kritik", 100: pp. 25–50 (Engl. trans. On Sinn and Bedeutung, in M. Beaney (ed.), *The Frege Reader*, Blackwell, Oxford 1997, pp. 151–171).

HYLTON P. (2013), *Ideas of a Logically Perfect Language In Analytic Philosophy*, in M. Beaney (ed.), *The Oxford Handbook of the History of Analytic Philosophy*, Oxford University Press, Oxford, pp. 906–925.

IACONA A. (2018), *Logical Form. Between Logic and Natural Language*, Springer, Cham.

KAPLAN D. (1970), *What Is Russell's Theory of Descriptions?*, in W. Yourgrau, A.D. Breck (eds), *Physics, Logic, and History*, Springer, Boston, MA, pp. 277–295 (Reprinted in D. Pears (ed.), *Bertrand Russell. A Collection of Critical Essays*, Anchor Books, Garden City, NY 1972, pp. 227–244).

MACDONALD C., MACDONALD G. (eds; 1995), *Connectionism. Debates on Psychological Explanation*, Blackwell, Oxford.

MARCONI D. (1981), *Le ambigue virtù della forma logica*, "Studi di grammatica italiana", 9: pp. 265–284 (Reprinted with a postscript in "Rivista di Estetica", 32: pp. 7–20, 2006).

——— (2019), *Analysis as Translation*, "Topoi", 38: pp. 347–360.

PRINZ J.J. (2002), *Furnishing the Mind. Concepts and Their Perceptual Basis*, MIT Press, Cambridge, MA.

RORTY R.M. (1967), *Introduction. Metaphilosophical Difficulties of Linguistic Philosophy*, in R.M. Rorty (ed.), *The Linguistic Turn. Essays in Philosophical Method*, The University of Chicago Press, Chicago, pp. 1–39.

RUSSELL B. (1905), *On Denoting*, "Mind", 14: pp. 479–493 (Reprinted in B. Russell, *Logic and Knowledge*, edited by R.C. Marsh, Routledge, London 1988, pp. 41–56).

——— (1912), *The Problems of Philosophy*, Williams and Norgate, London (Reprinted by Oxford University Press, Oxford 1998 [page numbers given relate to this volume]).

——— (1918–19), *The Philosophy of Logical Atomism*, "The Monist", 28: pp. 495–527 and 29: pp. 32–63, 190–222, and 345–380 (Reprinted in B. RUSSELL, *Logic and Knowledge*, edited by R.C. Marsh, Routledge, London 1988, pp. 177–281 [page numbers given relate to this volume]).

——— (1922), *Introduction*, in L. Wittgenstein, *Tractatus Logico–Philosophicus*, Routledge and Kegan Paul, London 1922 (Reprinted in L. Wittgenstein, *Tractatus Logico–Philosophicus*, Routledge and Kegan Paul, London 1961).

WITTGENSTEIN L. (1921), *Logisch–philosophische Abhandlung*, "Annalen der Naturphilosophie", 14: pp. 185–262 (Engl. trans. *Tractatus Logico–Philosophicus*, Routledge and Kegan Paul, London 1961).