

ANDREA BIANCHI

**NATURALIZANDO LA SEMÁNTICA
Y EL ARGUMENTO DE PUTNAM BASADO EN
LA TEORÍA DE LOS MODELOS**

Resumen: A partir de 1976, en muchas ocasiones Hilary Putnam ha propuesto un argumento, fundado en algunos resultados de la teoría de modelos, según el cual cualquier programa filosófico, cuyo propósito sea naturalizar la semántica, fracasó en dar cuenta de un importante rasgo de todo lenguaje natural: la determinación de la referencia. Aquí, después de haber presentado el argumento, sugiero que no funciona, porque simplemente asume lo que debería probar, esto es, que no debemos extender la metateoría: al parecer Putnam piensa que todo lo que podemos asertivamente afirmar acerca de las relaciones entre las palabras y las entidades en el mundo es lo que la teoría de modelos nos dice, pero nunca ha aportado justificaciones para ello. Al final del artículo, discutiré la aparentemente confiable intuición que me parece yace en la raíz del argumento, esto es, que dada una teoría muy formal, hay un infinito número de modos de conectarla con o proyectarla sobre el mundo. Sugeriré que deberíamos resistir tal intuición, porque descansa sobre una muy dudosa noción del mundo, la cual asume que para cualquier clase de objetos hay una propiedad correspondiente.

Palabras clave: Indeterminación de la referencia, naturalización de la semántica, metateoría.

NATURALIZING SEMANTICS AND PUTNAM'S MODEL-THEORETIC ARGUMENT*

Abstract: Since 1976 Hilary Putnam has on many occasions proposed an argument, founded on some model-theoretic results, to the effect that any philosophical programme whose purpose is to naturalize semantics would fail to account for an important feature of every natural language, the determinacy of reference. Here, after having presented the argument, I will suggest that it does not work, because it simply assumes what it should prove, that is that we cannot extend the metatheory: Putnam appears to think that all we may determinately say about the relations between words and entities in the world is what the model theory tells us, but he has never offered justifications for that. At the end of the article, I will discuss the apparently reliable intuition that seems to me to be at the root of the argument, that is that, given a formal theory, there is an infinite number of ways of connecting it to, or of projecting it onto, the world. I will suggest that we should resist this intuition, because it rests on a very doubtful notion of world, which assumes that for any class of objects there is a corresponding property.

Key words: Indeterminacy of reference, naturalization of semantics, metatheory.

Are what we usually call “natural languages” really *natural*? Put like this, the question is quite vague (what exactly counts as natural?). One discussed way of specifying it is more or less the following. Let us assume that physics is the basic science, the one to which all the other natural sciences could in some way be traced back. Does it have at its disposal, at least in principle, the conceptual resources to account for

* The first time I presented the ideas of this work was at the Seminar on Analytic Philosophy at the University of Padua. I am grateful to all the participants. I would also like to thank Pierdaniele Giaretta, Paolo Leonardi, Enrico Martino and Ernesto Napoli for their comments on earlier written versions.

all interesting properties of a language? That is: will the final physical theory, supposing we reach it, explain every linguistic phenomenon, alongside the chemical, or biological, or meteorological ones? As often happens in philosophy, some think so, others do not. Notoriously, the greatest difficulties are represented by *semantic phenomena*, because of their intentional and normative aspects. If nonetheless we opted for an affirmative answer and turned into the uncertain road of naturalization, we would very soon come across a fork with regard to the traditional semantic notions (especially reference and truth): should we eliminate, perhaps substituting them with surrogates, or reduce them? The latter possibility could gain substance more or less in this way: firstly, we show how the semantic properties of complex expressions of any language are determined by their structure and the semantic properties of their constituents; secondly, we try to identify the primitive expressions and a natural relation, in the sense specified, which holds between each of them and its semantic value (which, in order to avoid circularity, has to be an entity accessible by non linguistic means).¹ Obviously, on

¹ The first clear formulation of this two-stage strategy was in Field, H., "Tarski's Theory of Truth", *Journal of Philosophy*, 69 (1972), pp. 347–75, reprinted in *id.*, *Truth and the Absence of Fact*, Oxford, Clarendon Press, 2001, pp. 3–26. As for the first stage, he wrote that Tarskian semantics, correctly understood, "explains the semantic properties of complex expressions (e.g., truth value for sentences, denotation for complex singular terms) in terms of semantic properties of their primitive components" (p. 6), that is to say it reduces truth and denotation to "primitive denotation". Furthermore, he showed that also truth conditions (for sentences, functions from worlds to truth values) can be explained in the same way, if we take it that a primitive predicate "stands not for a set (its extension) but for a *property* that exists in the actual world" (*id.*, "Mental Representation", *Erkenntnis*, 13 (1978), pp. 9–61, reprinted in *id.*, *Truth and the Absence of Fact*, cit., pp. 30–67, p.41). As for the second stage, he suggested a causal theory of primitive denotation, but without working it out. Just a remark on terminology: hereafter I will use "reference" for the semantic value of primitive expressions (Field's "primitive denotation").

these bases we can articulate very different programmes, depending on the chosen natural relation and on the isolated class of primitive expressions. This second aspect has some consequences on the problems we then have to face: the wider the class, for example up to the inclusion of all or almost all the syntactically simple expressions (roughly, the words), the more difficult to find a natural relation, non linguistically mediated, that each of them would have with its reference; viceversa, the narrower the class, the more worrying Quine's criticisms of the analytic\synthetic distinction.²

All research programmes that have so far moved in this direction have got no further than an embryonic stage, and it is more than legitimate to have doubts and perplexities about their chances of success.³ Here, however, we will suppose that any one of them has reached its goal, having succeeded on the one hand in individuating a naturalistically sound predicate that turns out to be equiextensional to "refer to" when we narrow down to the selected primitives the set on which the latter is defined, on the other in finding the way of specifying, avoiding Quinean criticisms, how non primitives inherit their semantic properties. The question we should discuss is: in this case would naturalization of language have really been achieved?

Since 1976 Hilary Putnam has on many occasions proposed an argument to the effect that the programme in ques-

² The first line is Jerry Fodor's (see Fodor, J.A., *Concepts: Where Cognitive Science Went Wrong*, Oxford, Clarendon Press, 1998). The second has a longer tradition, at the beginning outside the naturalistic framework, starting with Russell's logical atomism and continued by logical positivists, especially Rudolf Carnap. Among naturalists, it is defended by upholders of inferential role semantics. Against them, Fodor repropose Quine's criticisms to the Carnapian project (See Fodor, *Concepts* cit., ch. 4).

³ For a brief critical discussion of them, see Loewer, B., "A Guide to Naturalizing Semantics" in Hale, B. and Wright, C. (eds.), *A Companion to the Philosophy of Language* London, Blackwell, 1997, pp. 108–26.

tion would anyway have missed the mark.⁴ In this article, my purpose is to discuss and criticize the argument. In the next section, I will present a reconstruction of it. In the following one, I will show where and why in my opinion it does not work, and that will lead me to make some general considerations on the relations between model theory and naturalized semantics. To conclude, I will clarify what seems to me to be the alleged intuition that grounds the argument and I will try to explain why I do not find it convincing.

Putnam's argument

We all agree, I think, that the words we use when we speak have a *univocal* and *determined* reference (apart from the ambiguity of some expressions, and perhaps the vagueness of some others): "Aristotle" refers to Aristotle, not to Kant; "cat" to cats, not to dogs (or: to cathood, not to doghood). Things being so, a good theory of language, whether naturalistic or not, should account for it: the reference of the expressions should turn out in some way determined. Unfortunately, according to Putnam, this would not happen with the conceptions of language which move along the lines we have sketched above, which would therefore show themselves to be inadequate.⁵

⁴ The first formulation of the argument was in a 1976 conference, "Realism and Reason", published as the fourth part in Putnam, H., *Meaning and the Moral Sciences*, London, Routledge and Kegan Paul, 1978, where it was acknowledged as a "turn" in his thinking (p. 5). Other, more articulated, versions are in "Models and Reality" (1977), which became the first chapter of *id.*, *Realism and Reason: Philosophical Papers*, Volume 3, Cambridge (Mass.), Cambridge University Press, 1983, and in *id.*, *Reason, Truth and History*, Cambridge (Mass.), Cambridge University Press, 1981, ch. 2. The introduction to the 1983 book (pp. vii-xviii) offers a non technical but clear concise exposition of the argument (inclusive of a "parable", pp. ix-x), and an effective placement of it in the historical context.

⁵ Putnam thinks that the only way of saving the determinacy of reference is to adopt the perspective of the *internal realism*, which ceases to consider the re-

Reconstructing Putnam's argument in detail is not easy, for a number of reasons. First of all, he has proposed versions quite unlike each other, and not always compatible, as various commentators have noted.⁶ Moreover, he makes use in turn of two different, though related, aspects of the classical metatheory: on the one hand, the Löwenheim–Skolem theorem (with connected Skolem 'paradox'), on the other, the incapacity of any formal theory to discriminate between isomorphic models. Finally, Putnam's target is not directly the above sketched naturalistic line of thought, but more in general the *metaphysical realism* (even though he focuses particularly the one that does not turn on supernatural powers of the human mind, like that of grasping platonic forms),⁷ which he equates with the thesis that "the theory that is 'ideal' from the point of view of operational utility, inner beauty and elegance, 'plausibility', simplicity, 'conservatism', etc., *might be false*".⁸ With his argument, he aims to show the "unintelligibility" (*ibid.*, p. 126) of this thesis. In

lation of reference as the naturalist being discussed does, namely as a relation which the expressions would bear to objects in a world that is independent from our representations. Instead, the latter would always be the objects of a theory, and that would make the determinacy of reference almost tautological. Here I am not going to discuss Putnam's proposal, which I find difficult to understand. Instead, I will try to show that it is by no means the only way of saving determinacy.

⁶ See for example Lewis, D., "Putnam's Paradox", *Australasian Journal of Philosophy*, 62 (1984), pp. 221–36, p. 222, and Hale, B. and Wright, C., "Putnam's Model-theoretic Argument against Metaphysical Realism", in Hale and Wright (eds.), *A Companion to the Philosophy of Language*, cit., pp. 427–57, p. 428.

⁷ See Putnam, *Realism and Reason*, cit., p.1. In his "Model Theory and the 'Factuality' of Semantics" (in George, A. (ed.), *Reflections on Chomsky*, Oxford, Blackwell, 1989, pp. 213–32), he wrote: "The upshot of *this* model-theoretic argument is simply that semantic physicalism does not work. (However, physicalism seems to be the only sort of metaphysical realism that our time can take seriously.)", p. 220.

⁸ Putnam, *Meaning and the Moral Sciences*, cit., p. 125.

fact, if consistent, as it should be, the ideal theory will have a model, whose domain is the infinite set of objects in the world, namely an interpretation onto objects of the world that makes it true.⁹ Obviously, the metaphysical realist would react by pointing out that this interpretation might not be the *intended* one, and on the intended one the theory might still turn out to be false. But if we are not able to make sense of the notion of intended interpretation, as Putnam argues, the supposition that the theory might really be false “appears to collapse into *unintelligibility*”: the truth in that model seems to be sufficient for the truth *tout court*.¹⁰

However, in spite of these complications, the core of the argument is, in its structure, quite simple. As Putnam himself recognizes, what he is doing is just extending “in a very strong way”¹¹ some Quinean results, by adding a few corollaries. After having proved that establishing the truth value of every sentence of a language is not sufficient to determine the reference of its subsentential expressions, and that establishing their truth conditions (their truth value in any possible world) does not improve the situation, he disposes of some alleged ways of solving the puzzle, including the naturalist one, and concludes that we should look upon reference from a very different perspective, metaphysically innocuous, by achieving the so-called “internal realism”, according to which truth does not transcend our final criteria of justifica-

⁹ Strictly speaking, that would require two independent arguments in order to show that the theory would have at least one infinite model and that there is an infinite number of objects in the world. If it was not so, in fact, nothing would guarantee that the theory has a model with the right cardinality. On this point, Putnam cheats: “If THE WORLD is finite, let the theory be compatible with there being only N individuals (where N is the cardinality of THE WORLD)” (*ibid.*, p. 139).

¹⁰ Of course, that is valid just for ideal theories, because only in this case nothing could ever falsify them.

¹¹ Putnam, *Reason, Truth and History*; cit., p. 33.

tion.¹²

Let us suppose we have a first-order formalization of our language, and that “the operational and theoretical constraints (the ones rational inquirers would accept in some sort of ideal limit of inquiry) determine which sentences in the language are *true*”.¹³ It seems that in this way we possess absolutely everything it makes sense to require. What else, apart from all the observational data available in principle and what in some way marks our being rational, could ever be relevant? In any case, we are faced with a particular first-order theory. Hence, in spite of the ‘ideality’ of the theory, the language in which it is formulated, as any first-order lan-

¹² Almost all philosophers who as far as I know have discussed the matter, with the partial exception of David L. Anderson (“What is the Model-theoretic Argument”, *Journal of Philosophy*, 90 (1993), pp. 311–22) that I will briefly consider later on, seem to me to agree in regarding what I am presenting as the core of the argument. Some, like Timothy Bays (“On Putnam and His Models”, *Journal of Philosophy*, 98 (2001), pp. 331–50), more interested in the mathematical aspects of the argument, favour the version that makes use of the Löwenheim–Skolem theorem, on which Putnam insists particularly in *Realism and Reason*, cit. In any case, the difference is not substantial. I prefer to introduce here (following Hale and Wright, “Putnam’s Model-theoretic Argument against the Metaphysical Realism”, cit.) the version that makes use of permutations mainly for two reasons (for another one, see Hansen, C., “Putnam’s Indeterminacy Argument: the Skolemization of Absolutely Everything”, *Philosophical Studies* 51 (1987), pp. 77–99, pp. 82–7). Firstly, it has a more general range, since it is valid even if we assume that the domain of interpretation is in some way independently fixed, and it does not require a first-order formulation. Secondly, it has more disruptive effects as regards indeterminacy of reference: every expression of the language, with the exception of the logical apparatus, is affected by it in an uncontrollable manner. On the contrary, the Löwenheim–Skolem theorem, alone, on the one hand does not have consequences on the reference of singular terms, on the other does not bring about a complete inscrutability of the extension of the predicates. To conclude, Ernie Lepore and Barry Loewer, in their “A Putnam’s Progress” (*Midwest Studies in Philosophy*, 12 (1988), pp. 459–73) offered a (sympathetic) reconstruction of Putnam’s argument that focuses on the differences between metaphysical and internal realism.

¹³ Putnam, *Reason, Truth and History*, cit., p. 33.

guage, will be liable to an infinite number of different interpretations. For example, it is easy to show that, given a language and an interpretation of it, for any non identical permutation of the domain (namely, for any one-to-one function that maps any element of it onto another), there exists another interpretation in the same domain that agrees with the first one for the truth value of the sentences but assigns a different reference to the primitive non logical expressions.¹⁴ Let us suppose then that the language contains the proper name "Aristotle", which appears in some sentences of our 'ideal theory'. According to the intended interpretation, all these sentences are about Aristotle, and attribute to him all the properties that he possesses, or rather that we are justified in attributing. Yet, if in the domain of the interpretation there is, as there has to be, Kant as well, then by properly adjusting the reference of other expressions we can interpret the same sentences as being about him without their truth values changing. What is more, we can interpret them as being about any other object in the domain, and anything seems to go: we are left with a *total indeterminacy of reference*.

It could be noted that I have incidentally said that the interpretation that assigns Aristotle, and not Kant, to "Aristotle", is the *intended* one. Putnam's crucial point, however, is about

¹⁴ For simple proof, see Hale and Wright, "Putnam's Model-theoretic Argument against Metaphysical Realism", cit., pp. 448–9. We can also prove something stronger, that is that, given a language and an interpretation of it, for any non identical permutation of the domain there exists another interpretation in the same domain that agrees with the first one for the *truth conditions* of the sentences (functions from worlds to truth values) but assigns a different reference to the primitive non logical expressions (*ibid.*, pp. 449–51). In any case, it is unclear to me how in Putnam's perspective we could ever get these truth conditions, since observational and theoretical constraints do not seem to determine them. To conclude, we should add that first-order formalization is not essential. Analogous results can be proved for second-order formalizations (*ibid.*, p. 451) and for those in languages with modal operators (*ibid.*, pp. 451–2).

exactly that. Both interpretations, we have seen, make all the sentences of the ideal theory, and only them, true, that is to say they satisfy any possible operational and theoretical constraint. But if they are both compatible with all the *evidence* we might ever have, it seems that absolutely nothing, short of magical mental powers, allows us to single out one as being intended. Appealing to *mental states* and their *intentionality*, for example, does not help. If we conceive them as “pure”, that is individuated only by what goes on inside us (as sentences of the mentalese, or mental images, or functional states, or neural configurations, or something else), they do not solve the problem (what in the world are they related to?). If we conceive them as “impure”, that is individuated by what they are about, they presuppose our ability to refer to objects and properties of the external world instead of explaining it.¹⁵ In short, nothing seems to allow us to state that “Aristotle”, in our speech, refers really to Aristotle rather than to Kant, or to any other individual.

At this point, the naturalist comes into play and tries to make his move. Of course there is something that makes “Aristotle” the name of Aristotle, he says. It is Aristotle, and not Kant or some thing else, that has the natural relation with “Aristotle” (let us call it “*R*”) to which he has reduced the semantic one – I remind you that we are assuming that he has in some way reached this. So, even if other interpretations of the theory are logically possible, the actual one is that which associates every primitive of the language with the

¹⁵ The Twin Earth case helps in making clear the distinction between pure and impure mental states (introduced by Putnam in *Reason, Truth and History*, cit., pp. 42–3): Oscar and his twin have the same pure mental states, but they differ with regard to some impure mental states (water-states). Another way of explaining the distinction is to draw a parallel with the case of language: syntactically individuated sentences are pure, interpreted sentences are impure.

individual or the class of n -tuples of individuals with which it has the relation R . After all, a logical possibility is not a fact: it is true that we can in some way balance the books even associating "Aristotle" with Kant, but, in fact, it is only Aristotle who the expression refers to. In so many words, it would not be the mind but *nature* that 'selects' the 'right' interpretation.

According to Putnam, however, this move is not a good one. After all, what our naturalist is doing is just supplementing our ideal theory by adding to it a formula like "x refers to y if and only if $R(x, y)$ ", or something similar.¹⁶ What does he achieve in that way? Nothing more than another theory, which obviously will be affected by the same problems as the former one: it will have a plurality of models –even if perhaps not the same–, so the reference of its expressions will not come out more determined, in spite of the intentions of the naturalist. Ironically, to the same predicate "R" different interpretations, coinciding in the truth values and truth conditions that induce on the sentences of the theory, will assign different extensions. Insult added to injury...¹⁷

¹⁶ More correctly, he is telling us that the theory will contain the formula in question.

¹⁷ The remark that what the naturalist is doing "is just adding more *theory*" (Putnam, *Realism and Reason*, cit., p. 18) is in some way present in all the three pieces of writings that I am discussing (compare with *Meaning and the Moral Sciences*, cit., pp. 126–7, and *Truth, Reason and History*, cit., pp. 45–8). In the face of the many criticisms that have been moved against this strategy of reasoning, Anderson (see Anderson, "What is the Model-theoretic Argument", cit.) tried to maintain –without in fact being very convincing– that the argument should not be interpreted in that way. Beside the model-theoretic part there should instead appear the specific criticisms that Putnam formulates in various places against some projects of naturalization, and in particular against those which appeal to the notion of *cause*. Even leaving aside the fact that philologically this interpretation does not work, this defence of the argument seems to me quite weak, because on the one hand the specific criticisms to those projects are slight and on the other they are unable to rule out the possibility that different ones have success.

Putnam's moral is simple. We should give up the hope of finding a correspondence between words and objects from an external and objective point of view, and be content with the little that we can say about semantics within the theory: that "Aristotle" refers to Aristotle, "Kant" to Kant, and so on. Whatever they mean, these assertions are true, and that is sufficient.¹⁸ In short, semantical facts cannot have any deep explanation, and thankfully they do not need it.

Theory and metatheory: what is wrong in the argument

Unfortunately for Putnam, but fortunately for those that find his internal realism unintelligible, there is a very simple objection that we can make against the argument.¹⁹ As we have seen, the problem is raised beginning from some metatheoretical considerations, or, to be more explicit, from some model-theoretic results. Now, the crucial thing that we have to note is that there is no argument unless it is assumed that the interpretation of the metatheoretical language is determined. Only if the words that occur in the formulation of the

¹⁸ The assertions are "analytic relative to the theory" (Putnam, *Meaning and the Moral Sciences*, cit., p. 136), because they follow "immediately from the definition of 'refers'" (*ibid.*), where the definition is a mere list, in a Tarskian style.

¹⁹ It is an objection in some way current, even though never quite explicitly, in literature. For example, it seems to me to be at the root of Michael Devitt's criticisms (see Devitt, M., *Realism and Truth*, Second Edition, Princeton, Princeton University Press, 1991, pp. 225–9). Some recent eye-opening remarks by Bays ("On Putnam and his Models", cit., pp. 347–50) go in the same direction. See also Loux, M.J., *Metaphysics: a Contemporary Introduction*, Second Edition, London, Routledge, 2002, pp. 281–3. In my opinion, the objection is clearer and stronger than the well known one that David Lewis formulated (Lewis, "Putnam Paradox", cit., pp. 225–6), which is however related. Anthony L. Brueckner ("Putnam's Model-theoretic Argument against Metaphysical Realism", *Analysis* 44 (1984), pp. 134–40), Mark Heller ("Putnam, Reference, and Realism", *Midwest Studies in Philosophy*, 12 (1988), pp. 113–27) and James Van Cleve ("Semantic Supervenience and Referential Indeterminacy", *Journal of Philosophy*, 89 (1992), pp. 344–61) move more or less along the same lines as Lewis.

Löwenheim–Skolem theorem, or in the one that we have presented, have the ‘intended’ interpretation, could the latter be used to deduce what Putnam deduces. It does not matter that, in the case of the ideal theory, the metatheory has to be part of the theory, and hence in no better condition. After all, Putnam, unlike Quine, does not mean to prove the indeterminacy or inscrutability of reference *tout court*, but only to show that the metaphysical realist, and, in particular, our naturalist, cannot avoid it. Therefore, replying to Putnam as Aristotle replies to the skeptical that denies the principle of non contradiction, that is maintaining that he simply contradicts himself (in denying, he would assume what he is denying) is missing the mark: he has the total right to assume the determinacy of his talking, and, particularly, of the metatheory that he uses in his reasoning, in order to criticize people who are not able to account for this determinacy (of the theory, but finally also of the metatheory).

At this point, however, those who want to question Putnam’s conclusions have at their disposal, in my opinion, a winning move, which consists of questioning his metatheory: why that one exactly?

Obviously, we are not to deny the truth of the metatheorems that Putnam uses. But why should we assume that the metatheory is exhausted by them? In particular, why could we not see the move of the naturalist we considered in the previous section as a move at this level, namely as the proposal of supplementing the metatheory with the formula “ x refers to y if and only if $R(x, y)$ ”, or something similar? If we maintain that what our naturalist is doing “is just adding more *theory*”, we not only misunderstand, but also cheat him.²⁰ After all, the formula contains a variable that ranges

²⁰ As Bays write, “Putnam wants to limit his opponents to working *within* a particular theory, while he himself steps outside this theory to talk about its

over expressions of the language and a predicate as “referring to”: it therefore seems to be an assertion *about* the theory, and not an assertion *of* the theory. If so, it should have the same status as the metatheorems that Putnam uses in his reasoning. Hence, as in these, the expressions that occur in it (“*R*”, in particular) would have, contrary to what Putnam claims, a determined reference. Since we are assuming that the formula is true (see first section), which he never questions, the reference of the expressions of the object language would turn out to be equally determined: they would precisely stand for whatever bears the relation *R* to them.

In the end, we just have to choose between two different metatheoretical proposals: Putnam’s and the naturalist’s. The latter includes the important theorems of the first, thus recognizing that there are many (infinite, if the domain is infinite) interpretations of the expressions of a language which are logically possible, but supplements them with another simple assertion which allows us to pick up, from these interpretations, the ‘intended’ one.²¹ We are in no position to establish now which of these two proposals is right, also because, as I have written in the opening section, the naturalist’s in fact still does not exist. Yet certainly Putnam with his argument does not help one against the other, because, as we have just seen, he moves precisely from the assumption of the first, and he claims, without any justification, that the second cannot be assumed in the same way: that is, he is simply assuming what he should prove.

semantics. [...] Putnam’s arguments depend on allowing himself just a little more than he allows those against whom he is arguing” (Bays, “On Putnam and His Models”, cit., p. 350).

²¹ Moreover, the opportunity of a similar integration is suggested also from partly different perspectives. For example, model-theoretic semantics, according to Diego Marconi, is not able to account for our lexical competence (see Marconi, D., *Lexical Competence*, Boston, MIT Press, 1997, ch. 1).

Besides these, could there be other metatheoretical proposals? For sure, for example, the classical metatheory could be supplemented with an assertion more or less of this type: “*x* refers to *y* if and only if *y* is conventionally associated with *x*”, with the necessary specifications. In fact, even this easy integration would allow us to save determinacy of reference, though it could not be appreciated by naturalists because of its use of the intentional notion of convention. During a seminar on this topic, another option was suggested to me, that of simply fixing in the metalanguage the interpretation of the language of the theory. In my opinion, however, this is misunderstanding the problem: Putnam’s point is not that we cannot ever single out an interpretation, but rather that we cannot ever establish whether this is the right, intended, one. Suppose for example you and I observe some people speaking. If we possessed all the data we needed, we could offer an interpretation of their talk. Yet, unfortunately, our interpretations could differ in the way we have sketched above. The crucial questions are: Who is right, you or I? By using the name “Aristotle”, are they speaking about Aristotle or Kant? What is more, is there a fact of the matter in this case? Putnam denies it, whereas the naturalist’s and conventionalist’s proposals appeal respectively to natural and to social facts in order to solve the puzzle. Instead, the ‘simple metalinguistic’ proposal saves determinacy in a completely vacuous manner: it is true that I can ‘rightly’ determine, from my own perspective, the reference of the words people are uttering, yet you can ‘rightly’ determine it in a different way from yours. Even worse, when you and I talk to each other perhaps we are using the same words with a different reference. So, the indeterminacy reappears at a deeper level. In the end, the proposal turns out to be almost equivalent to Putnam’s rather than a real alternative.

A final remark. It has been pointed out to me that Putnam

could reply by noting that natural relations (like R) do not have a metatheoretical nature and this would make a great difference between his proposals and the naturalist's. I find this objection wrong. The point is that no notion is metatheoretical *by essence*, and thus those used in the classical metatheory (e.g. reference and truth) are in a no better position than R . In our theories about the world we can speak of reference and truth, for example in studying other languages. Only when we investigate our own language do we have in some way to step outside it in order to talk about its semantics. It is this investigation, and not the notions that we use in doing it, that is metatheoretical. Thus, if our language belongs to the natural order, as naturalists believe, the notions which we use in investigating the natural world are the notions which we should use in investigating the language, namely in doing metatheory. If we deny this we assume, once again, what should be proved.

What world?

If I am right, Putnam's argument does not work. People who are sympathetic to the perspective of internal realism could however keep on appealing to the model-theoretic theorems in order to safeguard the apparently reliable intuition that, given a formal theory, there is an infinite number of ways of 'connecting' it to, or of 'projecting' it onto, the world. I wish to conclude this article by saying something about why in my opinion we should resist this intuition.

Is it true that given a theory there is an infinite number of ways of 'connecting' it to, or of 'projecting' it onto, the world, obviously leaving aside the fact that if our naturalist is right, there would be only one way in which the theory is actually connected to it? Yes and no, I am tempted to answer. Yes, if we conceive the 'connecting' as the interpreting in a domain

that contains all the individuals that there are in the world and this domain is infinite: the metatheorems show clearly that we can do it in an infinite number of ways. *No*, if we take the notion of *world* a bit more seriously.

What do I mean, exactly, by that? Simply, that the world is not a sheer set of individuals. On the contrary, within it there are individuals with properties, and in relations with each other (“The world is the totality of facts, not of things”, Wittgenstein wrote in the *Tractatus*). In order to meet the requirements of formal semantics, one can reformulate this by saying that in the world there are individuals which belong to classes that are subsets of the domain, and which are members of n -tuples of individuals that belong to subsets of the n th Cartesian product of the domain with itself. Yet, it is important to always remember that the notion of class is introduced to this end, and is dependent in that use on those of property and of relation (for any property or relation there is a corresponding class), and not viceversa.²²

Now, the point is that the interpretations we can construct from permutations do assign ‘strange’ classes to the predicates. ‘Strange’, obviously, not from a set-theoretic point of view (every subset of the domain is legitimate), but from the point of view of the conception of world that the use of set-theoretic notions within semantics claims to capture. A class which contains only male dogs and female cats, for example, is ‘strange’. We can certainly isolate it, but perhaps only because we have in our language primitive predicates that stand for more ‘natural’ properties and relations, whatever they are. There is no reason, I think, for admitting strange properties and relations in our ontology. Thus, there

²² If you think there are paradoxical properties or relations, you can reformulate: for any non paradoxical property or relation there is a corresponding class.

is not a corresponding property or relation for any class. And, to come back to our subject, if we maintain that the world is constituted by individuals with ‘natural’ properties, and in ‘natural’ relations with each other, an interpretation that connects our ideal theory to the world will consequently be an interpretation that maps as far as possible its primitive predicates onto classes that correspond to these natural properties and relations.²³ Nothing, contrary to the alleged intuition from which we started, leads us to believe that there is an infinite number of these interpretations, or even more than one.

Obviously, a lot remains to be clarified. What is, for example, a ‘natural’ property? Putnam, who briefly takes this possibility into consideration, accuses it of “medieval essentialism” (*ibid.*, p. xii), and of being tied up with a completely discredited psychology, according to which we would be equipped with a mysterious capacity to grasp platonic forms.²⁴ The latter charge seems to me to be groundless: the belief in the existence of natural properties does not involve the belief in the existence of a mental power of immediately and precisely determining the boundaries of the corresponding classes. On the contrary, we can see in scientific activity exactly that whose purpose is to individuate them by trial and

²³ Lewis seems to be more or less of this opinion (see Lewis, “Putnam Paradox”, cit., pp. 227–30). However, we have to note that we cannot use these metaphysical considerations against Putnam’s argument, because they presuppose the solution (see Putnam, *Reason, Truth and History*, cit., pp. 35–8).

²⁴ Carsten Hansen (“Putnam’s Indeterminacy Argument...”, cit.) criticized Putnam’s argument against metaphysical realism because it would rest on the very strong epistemological claim that “we have access only to our own representations” (p. 92). In my opinion the premise which Putnam appeals to is weaker and less disputable, that is that we don’t have any *direct access to non individual* entities (properties or relations). Obviously, by denying it one could save metaphysical realism from indeterminacy. However, I think that the metaphysical realists can accept Putnam’s epistemological premise without any serious risk.

error. After all, now we know: whales are not fish. Perhaps, we could even overturn the criticism on Putnam. By assuming the existence of a property for any class, no matter how 'strange' the latter is, we would make science itself an understandable practice. However, I do not wish to discuss this point. My limited goal here was to show the fallacy of Putnam's argument and investigate the intuition that is at its basis. I hope I have succeeded.

Università di Bologna and
Università del Piemonte Orientale

REFERENCES

- Anderson, D.L., "What is the Model-theoretic Argument", *Journal of Philosophy*, 90 (1993), pp. 311–22.
- Bays, T., "On Putnam and His Models", *Journal of Philosophy*, 98 (2001), pp. 331–50.
- Brueckner, A.L., "Putnam's Model-theoretic Argument against Metaphysical Realism", *Analysis*, 44 (1984), pp. 134–40.
- Devitt, M., *Realism and Truth*, Second Edition, Princeton, Princeton University Press, 1991.
- Field, H., "Tarski's Theory of Truth", *Journal of Philosophy*, 69 (1972), pp. 347–75, reprinted in *id.*, *Truth and the Absence of Fact*, Oxford, Clarendon Press, 2001, pp. 3–26.
- Field, H., "Mental Representation", *Erkenntnis*, 13 (1978), pp. 9–61, reprinted in *id.*, *Truth and Truth and the Absence of Fact*, Oxford, Clarendon Press, 2001, pp. 30–67.
- Fodor, J.A., *Concepts: Where Cognitive Science Went Wrong*, Oxford, Clarendon Press, 1998.
- Hale, B. and Wright, C., "Putnam's Model-theoretic Argument against Metaphysical Realism", in Hale, B. and Wright, C. (eds.), *A Companion to the Philosophy of Language*, London, Blackwell, 1997, pp. 427–57.
- Hansen, C., "Putnam's Indeterminacy Argument: the Skolemization of Absolutely Everything", *Philosophical Studies*, 51 (1987), pp. 77–99.
- Heller, M., "Putnam, Reference, and Realism", *Midwest Studies in Philosophy*, 12 (1988), pp. 113–27.
- Lepore, E. and Loewer, B., "A Putnam's Progress", *Midwest Studies in Philosophy*, 12 (1988), pp. 459–73.
- Lewis, D., "Putnam's Paradox", *Australasian Journal of Philosophy*, 62 (1984), pp. 221–36.
- Loewer, B., "A Guide to Naturalizing Semantics" in Hale, B. and Wright, C. (eds.), *A Companion to the Philosophy of Language*, London, Blackwell, 1997, pp. 108–26.
- Loux, M.J., *Metaphysics: a Contemporary Introduction*, Second Edition, London, Routledge, 2002.
- Marconi, D., *Lexical Competence*, Boston, MIT Press, 1997.
- Putnam, H., *Meaning and the Moral Sciences*, London, Routledge and Kegan Paul, 1978.
- Putnam, H., *Reason, Truth and History*, Cambridge (Mass.), Cambridge University Press, 1981.
- Putnam, H., *Realism and Reason: Philosophical Papers, Volume 3* Cam-

HUMÁNITAS. Portal temático en Humanidades

NATURALIZING SEMANTICS AND PUTNAM'S ...

21

bridge (Mass.), Cambridge University Press, 1983.

Putnam, H., "Model Theory and the 'Factuality' of Semantics", in George, A. (ed.), *Reflections on Chomsky*, Oxford, Blackwell, 1989, pp. 213–32.

Van Cleve, J., "Semantic Supervenience and Referential Indeterminacy", *Journal of Philosophy*, 89 (1992), pp. 344–61.