ON THE INTERPRETATION OF FORMAL LANGUAGES AND THE ANALYSIS OF LOGICAL PROPERTIES

Josep MACIA*

* Departament de Lògica, Història i Filosofia de la Ciència, Universitat de Barcelona, Baldiri Reixach s/n, 08028 Barcelona. E-mail: josep@alum.mit.edu

BIBLID [0495-4548 (2000) 15: 38; p. 235-258]

ABSTRACT: We can distinguish different senses in which a formal language can be said to have been provided with an interpretation. We focus on two: (i) We provide a model (or structure) and a definition of satisfaction and truth in the standard way (ii) We provide a translation into a natural language. We argue that the sentences of a formal language interpreted as in (i) do not have meaning. A formal language interpreted as in (i) models the way the truth of a sentence would be affected by two factors: the interpretation as in (ii) of the language, and a way the world might be. Viewing in this way the relation between interpreting a formal language as in (i) and as in (ii) allows us to justify the *conceptual* adequacy of the standard model-theoretic definitions of the properties of logical truth and logical consequence.

Keywords: semantics, meaning, formal languages, regimented languages, translation, logical properties, models.

CONTENTS

The work in this paper falls under the general topic of the study of the relationship between natural language and the so-called *formal languages*[1]. Here, we will focus on the question: "what does a sentence of a formal language mean?". In answering this question we will make some distinctions which we will use in the final part when considering some aspects of the relation between natural language and formal languages; in this final part we will also sketch how our findings can be used to defend the conceptual adequacy of the standard model theoretic account of the logical properties.

We can not pretend to be making claims about all formal languages, since there are infinitely many different *kinds* of sign systems that could be

regarded as formal languages, and we would not know even how to approach the task of trying to say anything substantive about all of them. We will restrict our attention to the languages of standard propositional logic and standard first and second order logic. Almost all of the time we will focus our attention specifically to standard first order languages.

1. Formal Languages and What They Mean

I think we can distinguish at least four main senses in which we say that a sentence of a formal language, and specifically, of a first order language, has certain meaning. For our purposes the two important senses will be the ones we will consider in subsections (1) and (3) below. The four senses are the following:

(1) We could say that a sentence of a formal language does not by itself mean anything unless we interpret it, and to interpret it consists in providing in the standard way a so called *model* for it (models are also called *interpretations*, or *structures*).

There are different ways of specifying what a model for a first order language L is. One common way of doing it is to say that a model M is an ordered pair <D,F> such that D is a set, the so called *domain* (or *universe*) of M, and F is a function that assigns an appropriate value to each non logical primitive symbol of L: an element of M to each constant, a subset of D to each 1-place predicate symbol, a subset of n-tuples of elements of D to each n-place predicate symbol ($n \geq 2$), and a subset of n+1 tuples of elements of D (meeting certain conditions) to each n-place function symbol.

A model by itself does not yet endow the formal language L with meaning. If it does so, it is only with respect to a theory that tells us what the interpretation or the value of complex expressions is, and specifically, what the interpretation of the sentences is. There are some differences in the specific form that such a theory can have. We will consider here two slightly different presentations which are both standard. (Several other presentations are possible, including some that are hybrids of the two considered here).

One way to proceed is to provide a truth theory for L and to do this through a definition of satisfaction: we define first that a model M (=<D,F>) and an appropriate *sequence* (or *assignment*) s satisfy a formula α of L. (A sequence is a function whose domain is the set of variables, an appropriate sequence for M is a sequence whose range is a subset of D). Then we can say that a formula α is true in a model M if there is an appropriate

sequence s such that M and s satisfy α (or alternatively, if for any appropriate sequence s, M and s satisfy α). In order to define satisfaction for a first-order language it is common to proceed in the following way:

First, we give a recursive definition of the *denotation* (*designation*, or *value*) of a term t with respect to a model M and a sequence s, which we will write as *M/s(t)*:

if t is a variable then M/s(t)=s(t),
if t is a constant then M/s(t)=F(t),
if t is $ft_1...t_n$ then M/s(t)=F(*f*)(M/s(t_1),...,M/s(t_n))

Then we give a recursive definition of *M and s satisfy formula* α that has the following form (where M=<D,F>):

If $\alpha=t_1 \approx t_2$, where t_1 and t_2 are terms, then M and s satisfy α iff M/s(t_1)=M/s(t_2),
if α= *Pt*, where P is a monadic predicate symbol and *t* is a term, then M and s satisfy α iff M/s(*t*)∈ F(*P*),
if $\alpha=Rt_1...t_n$, where *R* is an n-adic (n≥2) predicate symbol and $t_1...t_n$ are terms, then M and s satisfy α iff <M/s(t_1),...,M/s(t_n)>∈ F(*R*),
if $\alpha=\neg\beta$, where β is a formula, then M and s satisfy α iff M and s do not satisfy β,
if $\alpha=(\beta \wedge \gamma)$, where β and γ are formulas, M and s satisfy α iff M and s satisfy both β and γ,
if $\alpha=\exists x\alpha$, where α is a formula, then M and s satisfy α iff there is an element of D, a, such that M and s_x^a satisfy α, where s_x^a is a sequence that assigns a to *x* and which otherwise is just like s.

An alternative way to proceed in order to provide an interpretation for the sentences of the first-order formal language L is the following: given a model M=<D,F> for L, we recursively define the function *I*, which we might call *the interpretation under M*, that assigns a value to every primitive non logical expression of L, to every closed term and to every sentence of L. To each sentence of L it assigns either the value True or the value False.

If *e* is a constant, a function symbol, or a predicate symbol of L, *I* (*e*)=F(*e*),
if *f* is a n-place function symbol and $t_1,...,t_n$ are closed terms then *I* ($ft_1...t_n$)=*I* (*f*) (<*I* (t_1),...,*I* (t_n)>),
if $\alpha=t_1 \approx t_2$, where t_1 and t_2 are terms, *I* (α)=True if *I* (t_1) = *I* (t_2), and *I* (α) = False if *I* (t_1) ≠ *I* (t_2),

if $\alpha=Pt$, where P is a 1-place predicate symbol and t is a closed term, then $I(\alpha)$=True if $I(t)\in I$ (P), and $I(\alpha)$=False if $I(t)\notin I$ (P),

if $\alpha=Rt_1,\ldots,t_n$, where R is an n-place (n≥2) predicate symbol, and $t_1,\ldots,t_n$ are closed terms, then $I(\alpha)$=True if $<I(t_1),\ldots,I(t_n)>\in I$ (R), and $I(\alpha)$=False if $<I(t_1),\ldots,I(t_n)>\notin I$(R),

if $\alpha=\neg\beta$, where β is a formula, then $I(\alpha)$=True if $I(\beta)$=False, and $I(\alpha)$=False if $I(\beta)$=True,

if $\alpha=(\beta\wedge\gamma)$, where β and γ are formulas, then $I(\alpha)$=True if $I(\beta)$=True and I(γ)=True, and $I(\alpha)$=False if $I(\beta)$=False or I(γ)=False,

if $\alpha=\exists x\beta$, where β is a formula, $I(\alpha)$=True if $I_a^e(\beta_{x/a})$=True for some e in D, $I(\alpha)$=False if $I_a^e(\beta_{x/a})$=False
for all e in the domain of D, where $\beta_{x/a}$ is a sentence obtained from β by replacing all free occurrences of x with a new constant a which does not appear in β, and I_a^e is a function that assigns e to a and which otherwise is just like I.

Now, let's consider some specific formal language, say the language L that has one constant symbol a and one predicate symbol P, and some specific interpretation for the language, i.e. one model for the language, say the model M whose domain is the set of humans, that assigns David Armstrong to a, and assigns the set {x: x philosophizes} to P.

Given this interpretation, does the sentence of L *Pa* mean the same as the English sentence *Armstrong philosophizes*? I think it is clear that it does not. If we consider the second presentation given above we see that the 'value' or 'interpretation' that we assign to a sentence is either True or False. In our specific example, we would have that the value of *Pa* is True. All the other true sentences of P would be assigned the same value as *Pa* by the interpretation function under M. If what determines the interpretation of *Pa* in M, i.e. what determines the meaning of *Pa* according to M, is the value that the sentence gets assigned by I, the interpretation function under M, then certainly *Pa* does not mean the same as *Armstrong philosophizes*, or otherwise we would be equally justified in claiming that, for instance, $\exists xPx$ means that Armstrong philosophizes, since $\exists xPx$ gets assigned by I the same value as *Pa*.

Maybe someone might argue in the following way: it is incorrect to take 'having the value True' as being all that the second approach above says about the interpretation of *Pa*. Given the way the interpretation function is defined it also tells us 'when' *Pa* is true, namely when Armstrong philosophizes. So we would have that given the interpretation function or, at least, given the way it is defined, we can conclude that *Pa* means that Armstrong

philosophizes. This is obscured by the very fact that we use a function and we assign an *object* to each expression. We should understand the claim that $I(\alpha)$=True as just another way of expressing that α is true. Viewing things this way the second approach is just like the first in that it is a way of providing a theory of truth for the language on the basis of a model.

I think that the view expressed in the previous paragraph is not correct. First, the claim that the value of $I(\alpha)$ is True, taken by itself, is a completely different claim from the claim that α is true. True is an object (or so me must assume if the definition of I is to make sense) -an abstract one. So is α. Given any two objects we can always define a function that will make one the value of the other, but this fact by itself does not imply anything about the two objects or their relationship other than we have stipulated that the function we have defined assigns one to the other. We could define another function G that made blueness (if such entity exists) the value for the argument the flag of the People's Republic of China. Then it would be the case that G(China's flag)=Blue, but this does not mean that it would also be the case that China's flag is blue. Analogously, the claim that $I(Pa)$=True is a claim about which objects happen to be related by I, not about whether α is or not true. If we want the second approach to yield a theory of truth we should incorporate a clause such as: if $I(\alpha)$=True then α is true, if $I(\alpha)$=False then α is false. Notice that given that we need a clause such as the one just stated, instead of postulating the range of I to be the set {True, False}, we could postulate it to be the set {1,0} and then have the clause: if $I(\alpha)$=1 then α is true, if $I(\alpha)$=0 then α is false. The only difference between having one or the other set as the range for I is that in the first case is easier to infer on the basis of I (and the fact that I is presented as an *interpretation function*) the clause that would allow us to obtain a truth theory.

Second, even if we left the considerations in the previous paragraph aside and considered the second approach basically as the same as the first one, i.e., as a way of providing a truth theory, it would still not be the case that interpreting *Pa* in accordance with this second approach would make *Pa* to mean that Armstrong philosophizes. This is so because the first approach does not make it the case either. Let's see why it does not:

It is true that, given the model M above, the truth theory that the first approach provides would yield the following biconditional:

(a) *Pa* is true iff Armstrong $\in$ {x: x philosophizes}

But even if the truth theory yields this biconditional, it does not make it the case that *Pa* means that Armstrong philosophizes. We can point two three sort of facts that show that this is so, the most relevant being the first one:

(i) The biconditional in (a) involves only material implication. That is, in order for the biconditional to be true all that is required is that the sentences appearing on the right and on the left of *iff* be both true or both false. Given that the set {x:x philosophizes} does in fact have Armstrong as a member, (a) allows us to conclude that *Pa* is true. We could have obtained this exact same information if instead of (a) we had (a)'

(a)' *Pa* is true iff Lennon was born in Liverpool

The sentence appearing in the right hand side of *iff* does not tell us 'when' *Pa* is true, that is, it does not give us the truth conditions of *Pa*. All it tells us is that *Pa* is actually true if and only if certain fact happens to obtain.

Maybe it could be replied that what makes *Pa* mean that Armstrong philosophizes is not just that the biconditional (a) follows from the truth theory, but the whole interpretation for the language, including the interpretation of the expressions in *Pa*. I do not think this is correct, though. Suppose we interpret the language P with respect to the same model as before, and with the following minor modification to the definition of satisfaction: instead of having the clause in (b) as before, we have (b)'

(b) if $\alpha = Pt$, where P is a monadic predicate symbol and t is a term, then M and s satisfy α iff $M/s(t) \in F(P)$

(b)' if $\alpha = Pt$, where P is a monadic predicate symbol and t is a term and $Pt \neq Pa$, then M and s satisfy α iff $M/s(t) \in F(P)$, if $Pt = Pa$ then M and s satisfy α iff Lennon $\in$ {x:x was born in Liverpool}

Every primitive symbol of P would still be assigned the same value as before. And all the sentences of P would have the exact same truth value. So, there seems to be no reason to claim that under one presentation of the interpretation of P *Pa* means that Armstrong philosophizes, but under the other it means something else -we must keep in mind that a biconditional that follows from the theory does not say anything about any connection between what the two sentences on each side of the *iff* express; otherwise put: from the truth of a biconditional and what one of the two sentences expresses, we can not conclude anything about what the other sentences expresses, other than it expresses something that determines the same truth value as the one determined by what the former sentence expresses.

(ii) A second way of realizing that the fact that the truth theory yields (a) does not make it the case that *Pa* means that Armstrong philosophizes is by noticing that sets are extensional. The set

{x: x philosophizes}

is presumably the same as

{x: x philosophizes and x is a rational being}

or as

{x: x philosophizes and x is not a new born}

We would still have the same model *M* if we had specified the value of *P* as being the set

{x:x is not a new born and x philosophizes}

We would then say that the truth theory would have as a consequence (c):

(c) *Pa* is true iff Armstrong $\in$ {x:x is not a new born and x philosophizes}

If having (a) as a consequence made it the case that under the interpretation induced by *M* *Pa* meant that Armstrong philosophizes, then if the theory yields (c) we would have to say that *Pa* means that Armstrong philosophizes and is not a new born. This is absurd since, as we pointed out, the model is the same no matter how we specify the set that is the value of *P*.

(iii) A third way of realizing that the fact that the interpretation induced by *M* yields the biconditional (a) does not suffice to make it the case that *Pa* means the same as the English sentence *Armstrong philosophizes* is by noting that the English sentence does not say anything about sets or the membership relation, whereas the sentence in the right hand side of (a) is about the membership of an object in a set[2]. And even if we think that (a) is not by itself what determines what the meaning of *Pa* is, we should note that the value assigned to *P* is a set, and whatever we might want to say about how the value of *P* contributes to what *Pa* means, it is this set and no something else that will play a role.

There seems to be very good reasons, then, for thinking that *Pa* when interpreted in the standard way on the basis of *M* does not mean the same as the English sentence *Armstrong philosophizes*. Does *Pa* so interpreted mean the same as any English sentence? Well, which English sentence could it be? It seems that the most plausible candidate would be *Armstrong is a member of {x:x philosophizes}*. It seems clear, though, that this English sen-

tence will not do either. First, there is the fact that, as noted in (ii) above, whatever *Pa* might mean is not sensible to the different ways of specifying the set {x:x philosophizes}, whereas this is not true of the English sentence under consideration. Moreover: even if we were interested only in a notion of 'sameness of meaning' according to which the sentences *{x:x is author of* Carrie*} has one member* and *{Stephen King} has one member* would have the same meaning, the sort of difficulty raised in (i) above would also apply to the candidate English sentence we are now considering: Given the same model *M* we can provide an alternative formulation of the truth theory that provides the same interpretation for each expression in the language, but which does not give any condition involving reference to any set when specifying the truth condition of *Pa* (we can use, for instance, a clause such as (b)").

It seems, then, that *Pa* does not mean the same as any English sentence. I believe that this fact makes it very plausible to think that it does not mean anything at all, if for a sentence to mean something requires not just that it possesses some semantic property or other like, for instance, to include some expression that refers to some specific individual, but also that the sentence does 'the same sort of thing' that natural language sentences do.

If *Pa*, when evaluated with respect to the model for L M, does not mean anything, what do we do when we provide in the standard way a so called *interpretation* for a first order formal language? Do the expressions of the language have any sort of semantic property? We will try to say something about this later on, in section 3.

(2) Sometimes we might say things such as: *sentence (d) says that R is transitive*; or: *sentence (e) says that there are infinitely many things*; or: *sentence (f) says that nothing is P.*

(d) $\forall x \forall y \forall z(Rxy \wedge Ryz \rightarrow Rxz)$

(e) $\exists X(\forall x \exists y \forall z(Xxz \leftrightarrow z=y) \wedge \forall x \forall y(\exists z(Xxz \wedge Xyz) \rightarrow x=y) \wedge$
$\wedge \exists x \forall y \neg Xyx)$

(f) $\neg \exists x Px$

These claims exemplify another sense of what a sentence of a formal language means. Here the claims about what a sentence α means have to be understood as claims about what will be the case in all and only the models in which α is true. Furthermore, what we pretend to be claiming about some primitive symbol of P appearing in α ('R is transitive'), is actually what will be true *of the interpretation* of that primitive symbol in each model where α is true. So, for instance, we say that (d) means that R is

transitive because (d) is true in all and only the models where *the interpretation of R* is a transitive relation; or we say that (e) means that there are infinitely many objects because in each model in which (e) is true the domain will be an infinite set and, furthermore, (e) is true in all models with an infinite domain.

Maybe there is also a looser use of this sense of 'the meaning of α' where a sentence of a formal language is said to mean that p if α being true in some model is enough to guarantee that p is the case with respect to that model. That is, α is said to mean that p if in all the models in which α is true it is the case that p (without requiring as well that α be true in only those models with respect to which it is the case that p). For instance, in this looser sense we could say that (g) means that there are infinitely many things

(g) $\forall x \forall y(fx=fy \rightarrow x=y) \wedge \exists x \forall y \neg fy=x$

Any model in which (g) is true has a domain with infinitely many objects. Nevertheless there might be models with an infinite domain but where (g) is false.

Be it as it may, these two senses of a sentence of a formal language meaning something that we have considered in this section (2) are not the senses that interest us the most here. We have considered them just not to confuse them with the ones we do have a primarily interest in.

(3) It can not be denied that sometimes the sentences of a formal language are used so that they do mean the same as certain natural language sentences. For instance a mathematician might express the thought that every number is the sum of two primes by using the English sentence E*very number is the sum of two primes*, but also by using the first order formal language sentence:

$\forall x \exists y \exists z(Py \wedge Pz \wedge x \approx y+z)$

Or he can express that there is at least one prime number by using the sentence *there is at least one prime number* but also by using the sentence: $\exists xPx$.

So we have that sometimes we take a formal language to be just like a regimented version of a part of natural language.

To use another example: we might, for instance, regard *a* just as another name for David Armstrong, and to take *P* to make the same contribution to the meaning of the sentences where it appears as *philosophizes* makes to the meaning of the English sentences where it appears, and to take *Pa* just as another way (in addition to *Armstrong philosophizes*, *Armstrong filosofa*, and

many others) of expressing that Armstrong philosophizes; and we might take $\exists xPx$ just as one alternative way of expressing that there is a thing that philosophizes.

Under this view of formal languages a formal language is similar to Esperanto: a language artificially created, with which one expresses thoughts that can also be expressed with the natural languages.

Understood in this way then the sentence *Pa* can mean the same as the English sentence *Armstrong philosophizes*. The question now is, how do we manage to make a particular formal language, understood in the sense we are describing here, to mean what it means?

When considering the sense in subsection (1) of a sentence of a formal language meaning something, we saw that providing a model and a truth theory in the standard way was not enough to have a formal language whose sentences would possess the characteristic that we are considering here: to mean the same as some sentences of a natural language. It might be thought, though, that we can obtain a language with such a characteristic if we amend the truth theory we were considering in (1) so as to avoid the features that were the basis for our argumentation that the sentences of P did not mean the same as any English sentence.

We could avoid having the value of a predicate symbol to be a set by not assigning a value to it through the model but rather having one clause in the definition of satisfaction for each predicate symbol, this clause being of the same sort as the one we offer here for *P*:

> if $\alpha = Pt$, where *t* is a term, then M and s satisfy α iff M/s(t) philosophizes

Then we would have as a consequence:

(h) *Pa* is true iff Armstrong philosophizes

We could as well decide to use a stronger biconditional, instead of the one involving only material implication. There are several possibilities here, since conditionals can be postulated to be more or less strong[3]. For the sake of the argument let's suppose we chose the strongest possibility and make the biconditional to be metaphysically necessary equivalence (we can think of it as placing a necessary operator in front of the whole biconditional sentence in (h)). Nevertheless, this strong biconditional would still be too weak[4] to avoid the difficulty (i) pointed out in subsection (1): we could still have another theory with respect to the same model such that all expressions would be assigned the same value and all sentences declared as

true in exactly the same possibilities but that yields a biconditional where *Pa is true* is not paired with *Armstrong philosophizes* but with another sentence that intuitively has another meaning. For instance, this alternative theory could yield (j)

(j) *Pa* is true iff the square of 11 is 121 and Armstrong philosophizes

Both the theory that would have (h) as a consequence and the theory that would have (j) as a consequence would assign the same value to all expressions and declare all sentences true or false in exactly the same circumstances. They are indistinguishable with respect to what they say about the language P. So if we claim that according to one of the theories *Pa* means that p then we should also maintain that according to the other *Pa* means that p. On the other hand, though, we would like to claim that according to the first theory, and given (h), *Pa* means that Armstrong philosophizes, but then, given (j), we should claim that according to the second theory *Pa* means that the square of 11 is 121 and Armstrong philosophizes. We are led, then, to the contradiction that the sentence *Pa* does and does not mean the same according to the two theories. The contradiction seems to arise from supposing of each of the theories that it suffices to endow *Pa* with certain specific meaning, in the same way that English sentences have meaning.

Even if the kind of theories considered so far do not suffice to make the sentences of L to mean something in the sense that we are considering in this section (3), there is another way of accomplish it, which seems to be what we, one way or other, in fact do when introducing a formal language which is used and understood in the way we are considering here. This other way is to provide a translation from the formal language into a natural language. Unlike what was the situation in subsection (1) here there is no standard way to proceed, since the translation procedure is not usually presented in an explicit way. One possible way of interpreting the expressions of L by explicitly indicating how to translate them into English would be the following[5]:

(If α translates as β, we will also write $tr(\alpha)=\beta$)

"a" translates as "Armstrong"
if v is a variable, v translates as v
"P" translates as "philosophizes"
if t_1 and t_2 are terms, $\lceil t_1 \approx t_2 \rceil$ translates as $tr(t_1)$ ^ "is identical with" ^ ^ $tr(t_2)$

if P is a predicate and t is a term, ⌜Pt⌝ translates as tr(t) ^ tr(P)
if α is a formula, then ⌜¬α⌝ translates as "it is not the case that" ^ tr(α)
if α and β are formulas, ⌜α∧β⌝ translates as "it is both the case that" ^ ^ tr(α) ^ "and that" ^ tr(β)
if α is a formula and v is a variable, ⌜∃v⌝ translates as "there is an object" ^ v ^ "such that" ^ tr(α)

So we have, for instance, that the sentence of L *∃x(¬Px ∧ x=a)* translates into English as *there is an object x such that it is both the case that it is not the case that x philosophizes and that x is identical with a.* And, of course, *Pa* translates as *Armstrong philosophizes.* Since we understand the English sentences we understand what the meaning that we postulate for the sentences of L is.

One comment about how formal languages are sometimes, and maybe even often, taught in introductory courses to logic which I believe has a significance beyond the pedagogy of logic: when formal languages and, in particular, first-order formal languages are first introduced, it is not uncommon to begin by explaining what sort of things can be expressed with these languages. So, for instance, students are taught that *John loves Mary* can be expressed as *Ljm*, or that *there is something that loves everything* is expressed in a first order language as *∃x∀yLxy*. That is: the semantics for the formal language is presented in the sense of (3): students are told what the expressions and the sentences of particular formal languages mean by giving English equivalents, and they are trained in translating from English into a formal language and from the formal language into English. Then they are told something of the following kind: 'Now we are going to do in a rigorous way what we have done so far in an intuitive way'. And then they are introduced to models, interpretation functions, assignments and the recursive definition of satisfaction. That is, they learn how to interpret a formal language in the sense of (1). Notice, though, that, whatever reasons there might be for presenting the topic in this way, the teacher who proceeds in this way is in some respect fooling her students: to interpret a language as in (1) is not a rigorous way of doing what we do when we interpret it as in (3). It is to do something else. We can see this in the fact that the sentences do not mean the same. To use our example once more: in the case of P, the sentence *Pa* can mean that Armstrong philosophizes when interpret as in (3) but not when interpreted as in (1).

At this point we can introduce some terminology that will distinguish among different senses of what we have so far ambiguously called *a formal*

language, or simply *a language*. We will refer to what sometimes is called *uninterpreted language* (which, if it does not have any semantic property would seem not to deserve the name *language* at all) as *a system of forms*; the only systems of forms we will be concerned with here are those of standard propositional logic, first-order and second-order logic, so that when we say 'a system of forms' we mean one the those three types; we will refer to a system of forms with an interpretation in the sense of (1) as a *formal language*, and we will refer to a system of forms with an interpretation in the sense of (3) as *a regimented language*.

(4) The sentences of a formal language or of a regimented language sometimes are said to mean something through *encodement*. Whatever they might mean through encodement is something they mean *in addition* to what they 'mean' given the interpretation that they have. There is encodement when the objects the language talks about have other objects associated with them, and some formulas of the language can be seen as codifying or playing the role of predicates about these other objects.

The most significant kind of encodement is the so called *Gödelization* where we codify the primitive symbols of the *language* (system of forms) of arithmetic by means of natural numbers. One way of doing it is, for example, to associate the numbers 1,3,5,7,9,11,13 and 15 to, respectively the constant *0*, the monadic function symbol *s*, the 2-place function symbols + and ., and the primitive logical symbols $\exists$, $\neg$, $\wedge$ and =. To the variable x_i we assign the number 2i+17. Furthermore we can assign an (even) number to each sequence of primitive signs of the *language* of arithmetic, and we also assign an (even) number to each finite sequence of finite sequences of primitive symbols of the *language*[6]. Then when we interpret the *language* of arithmetic in the usual way as about natural numbers, we can see the sentences of the language of arithmetic as also encoding claims about the symbols and sequences of symbols of the *language* of arithmetic, and to see formulas of arithmetic with n-free variables as encoding predicates about the *language* of arithmetic. A very simple and uninteresting example would be the formula $\neg x=sss0$ which because in its usual interpretation is true when the value of x is not 3 (and because the number that corresponds to the function symbol *s* is 3), can be seen as *encoding* the predicate 'it is not the symbol *s*'.

When all the objects that are the interpretation of the *language* are encoding some object or other (in the example of the previous paragraph: if every natural number encodes some primitive symbol of the *language* of arithmetic, or some sequence or sequence of sequences of primitive symbols) then

instead of 'encodement' we could directly talk of just 're-interpretation' of the *language*: the encoding provides us with another way of giving an interpretation for the *language*. In the case of the *language* of arithmetic and the encoding of expressions and sequences of expressions of the very *language* of arithmetic by means of natural numbers, we could understand the codification as providing another interpretation of the *language* of arithmetic: the domain of this alternative interpretation has as individuals the primitive symbols of the *language* of arithmetic, sequences of those, and sequences of those sequences; and we would give an interpretation for the different primitive non-logical symbols making use of the correspondence we have between the elements of the domain of the usual interpretation (natural numbers) and the elements of the domain of this alternative interpretation (symbols, sequences of symbols, and sequences of sequences of symbols). For instance, according to the alternative interpretation we would interpret the function symbol *s* as the function that assigns to an element of the domain *e* (a symbol of the *language* of arithmetic, or a sequence of symbols, or a sequence of sequences of symbols), the symbol or sequence that is associated with a number which is the successor of the number associated with *e* (i.e., if *e* is associated with the number *n*, n^* is the successor of *n*, and e^* is the symbol or sequence associated with n^*, then the interpretation of *s* assigns e^* to *e*).

We could distinguish different levels in how a formula or sentence that has certain interpretation about certain objects 'says' something or codifies some claim about some other objects. For instance we can say of a certain formula of the *language* of arithmetic with one free variable that codifies the predicate 'to be a formula provable in the theory Z'[7] (or that it says that the value of the variable x is a formula provable in Z), only because on the usual interpretation of the *language* of arithmetic the formula is true for exactly those values of x that are numbers that are associated by the codification with sequences of primitive symbols of the *language* of arithmetic that are formulas that can be proved in Z. We could have stronger reasons, though, for saying of a formula $\alpha(x)$ of the language of arithmetic that it *expresses* or *corresponds to* the predicate 'to be a formula provable in Z'. It could be that the formula $\alpha(x)$ not only is true for exactly those values of x that are numbers that correspond to formulas that are provable in Z, but also that $\alpha(x)$ is built up from subformulas that are coding the predicates 'to be a formula of the *language* of arithmetic', 'to be an axiom of Z', 'to be a sequence', 'to be a member of a sequence', 'to be earlier in a sequence than', 'to be the result of applying Modus Ponens to', and $\alpha(x)$ can be seen as say-

ing that 'there is something that is a derivation of x from the axioms of Z' or more explicitly 'there is something that is a sequence of formulas such that each one is either an axiom of Z or the result of applying Modus Ponens to two earlier formulas in the sequence, and the last formula of the sequence is x'. The subformulas of $\alpha(x)$ can, in turn, be built up from other subformulas that are also coding predicates about the *language* of arithmetic (for instance the formula $\beta(x)$ corresponding to 'to be a formula' can contain subformulas that correspond to the predicates 'to be an atomic formula', 'to be the negation of', 'to be the conjunction of', 'to be an existential quantification of', and $\beta(x)$ can be seen as saying that 'there is a sequence such that each member is either an atomic formula, or the negation of an earlier member, or the conjunction of two earlier members, or an existential quantification of an earlier member, and x is the last member of the sequence'). It seems clear that one such formula $\alpha(x)$ can be said in a more proper or fuller sense that expresses the predicate 'to be provable in Z' than one formula that simply is true for the right values of x.

There would be a lot more to say about encodement and Gödelization (in particular, it would be interesting to clarify what exactly the distinction between 'expressing in a more/less full sense' consists in). We leave it here, though, since the sense of a sentence having certain meaning that we have considered in this section is not one we want to focus on in this article, and we have included it just for the sake of completeness and to distinguish it from the other senses.

2. Some issues regarding the relation natural language-regimented languages-formal languages

In the previous section we have distinguished between what we called *formal languages* and *regimented languages*. We content that this distinction is useful in the study of the relationship between natural language and formal languages at least for the following reason: we can break the question 'what is the relationship between natural language and a certain formal language?' in two parts: there is first the question of what the relation is between natural language and the regimented language that shares the same system of forms with the formal language under consideration, and then there is the question of what the relationship is between this regimented language and the formal language. Different issues arise depending on which of the two kinds of question we are considering. We content that is beneficial to keep the different issues separate and not to confuse them as we would easily do

if we where to directly discuss 'the relation between natural language and formal languages' without making any further distinction.

2.1. Natural language and regimented languages

What is the relationship and the differences between natural language and regimented languages? In this section we are going to briefly look at three or four aspects of this relationship. Some of our comments will be tentative or inconclusive. This is a difficult topic. The main reason for considering here the relationship between natural language and regimented languages is simply to see what sort of issues arise regarding this relationship and to be able to separate them from the issues that arise about the relationship between formal languages and regimented languages.

(1) As we saw, a regimented language is interpreted by using part of natural language. A regimented language, though, has some differences even with that part of natural language that is used to interpret it. One of them has to do with the domain of quantification. When we introduced one particular regimented language in section 1-(3) we did not provide any specific domain of quantification -we took our quantifiers to range over everything there is. We already mentioned that, unlike what is the case for formal languages, there is no standard way of characterising regimented languages. When they are provided, though, a domain of quantification is usually specified, i.e., there are some things that are stipulated to be the ones the language will talk about. So if the domain of quantification consists of the things that are p, then an existential quantification $\exists x$ will be translated as *There is a thing x which is p such that.* The domain of quantification is the same for all the sentences of a particular regimented language. This is not the case with respect to natural language, where the domain of quantification can change from one sentence to another, or even from one part of a sentence to another, as in (k) and (l):

(k) I entered the room. Everything was in order. I looked into the fridge. Everything had been stolen.

(l) After the attack on Ganymede, someone was happy to learn that everyone was dead.

In (k) what has been stolen is not what was claimed to be in order, and in (l) whoever was happy after the attack is not someone who was dead.

Having a fixed domain for all uses of quantifiers is, then, a feature that distinguishes a regimented language form natural language. This feature of

regimented languages is one aspect of two related general characteristics of these languages: not to be subject to context dependency and to approximate the ideal of displaying in an explicit way all the features that are relevant for the meaning of the sentence.

(2) Another such feature is that in a regimented language different syntactic categories correspond to different semantic categories. We will not go here into the very interesting and very difficult topic of characterising what a semantic category is. We will make just one comment regarding predicate symbols. Predicate symbols are translated by means of expressions whose meaning is such that either applies or does not apply to an individual (in the case of a monadic predicate) or to n individuals (for an n-place predicate). We might, for instance, interpret the 1-place predicate symbol *R* by indicating that it translates as *runs*. Then we would have that it applies to those individuals that run and does not apply to those individuals that do not run. Now: we could also interpret the monadic predicate symbol *Q* by indicating that it translates as *runs quickly*. If we translate the constant *a* as *Armstrong*, do *Ra* and *Qa* mean, respectively, the same as *Armstrong runs* and *Armstrong runs quickly*?

Our concepts of 'meaning' and 'meaning the same' are probably not precise enough as to allow us to go into too fine-grained distinctions, but I would just want to point out that it is not obvious that the answer to the question is 'yes'. *Armstrong runs* follows logically from *Armstrong runs quickly*, whereas it would seem that *Ra* does not follow logically from *Qa* (we are here appealing to the intuitive, pre-theoretic notion of 'following logically from' or 'being a logical consequence of'). If we believe that logical properties depend only on the meaning of the expressions (and not, for instance, on their spelling or pronunciation), then the two pairs of sentences must differ in meaning.

We might wonder: given the interpretation that *Ra* and *Qa* have, is it really the case that *Ra* is not a logical consequence of *Qa*? I think that it's clear that *Ra* is an analytic consequence of *Qa* (or that $\neg(Qa \wedge \neg Ra)$ is an analytic truth). Nevertheless, it seems reasonable to believe that *Ra* is not a logical consequence of *Qa*, in the same way that we believe that the English sentence *John is an unmarried man* is not a logical consequence of *John is a bachelor*.

Notice that the situation would have been different if the sentences of a regimented language were regarded as *notational simplifications* or *abbreviations* of English sentences. We could have introduced such type of language by means of some clauses that would look very much like the ones

we gave to introduce our sample regimented language in 1-(3): all we would need to do would be to substitute "is an abbreviation of" for "translates as". If "Ra" abbreviates "Armstrong runs", "Ra" means something only through its connection with "Armstrong runs". This relation of 'being an abbreviation of' can not simply be analysed in terms of those of 'being a name of', 'being a token of' and 'being a type of'. If "Ra" is an abbreviation of "Armstrong runs" then it is not the case that "Ra" names "Armstrong runs", since to use "Ra" is not to mention "Armstrong runs" but rather to express something about Armstrong; it is not the case either, though, that when we make a particular use of "Ra" we have used a token of the sentence "Armstrong runs": we have only used a token of the expression "Ra". If *Ra* and *Qa* are just abbreviations of *Armstrong runs* and *Armstrong runs quickly* then certainly the same logical relations must hold between *Ra* and *Qa*, and the fully expanded English sentences.

(3) It might be thought that another difference between regimented languages and a natural language like English arises because the use of regimented languages is not subject to the conversational norms that in the case of natural languages have an influence on what is communicated with the use of some sentence. For instance, if I say *There is a man waiting for John* it will be understood that there is only one man who is waiting, whereas the sentence $\exists x Wxj$ (with the appropriate interpretation of W and a, and quantifying over men) can not be taken to communicate that there is only one man waiting for John. This is not because the existential quantifier in the regimented language has a different interpretation from the existential quantifier in natural language (this could hardly be the case given that we have interpreted existential quantification in the regimented language by means of existential quantification in English and problems such as the ones regarding the structure of meaning that we mentioned above do not arise here). The use of the English sentence *There is a man waiting for John* is subject to the effect of the Cooperative Principle and, in particular, to the maxim of quantity: the maxim requires to give as much information as might be needed; whether there is only one man or more would usually be relevant information in a context where the sentence is used; lacking information to the contrary it will usually be assumed that the person using the sentence knows whether there is only one man or there are more; that there are two men, that there are three men, that there are several men or that there are many men can be expressed with as much brevity and simplicity as that there is a man; the speaker is abiding by the Cooperative Principle and chose to say that there is a man rather than any of the other stronger

claims, this must be because he knows the stronger claims not to be true and, so, there must be only one man. A reasoning of this kind is what makes it the case that when we use *There is a man waiting for John* we usually communicate that there is only one man waiting for John.

Similar observations could be made with respect to other contrasts between sentences of a regimented language and natural language sentences that are used to interpret them. (For instance: the effect of the maxim of manner and the difference between *John got rich and he took up philosophy* and *John took up philosophy and he got rich*, but the equivalence between $(Rj \wedge Tj)$ and $(Tj \wedge Rj)$ [with their obvious interpretation]).

I believe, though, that this particular distinction between natural language and regimented languages is not so much about the languages themselves but about their use. The sentences of a regimented language do not communicate anything other than what constitutes their meaning because they are not evaluated within the framework that is assumed when we consider the use of a sentence as part of the cooperative effort of a conversation. But this is not something essential to regimented languages themselves. They could in principle also be used to engage in a conversation, and then they would also be subjected to the Cooperative Principle.

2.2. Regimented languages and formal languages

In this final section we will make some comments on the relationship between formal languages and regimented languages. We will mainly focus on the issue of the conceptual adequacy of the standard analysis of the logical properties (subsection (2) and (3)). We will briefly explain how having distinguished between regimented languages and formal languages is useful in order to justify the adequacy of the standard analysis.

(1) One fact that makes it easy to overlook that formal languages are not regimented languages, but just modelations of them, is that it is often very easy to go from a formal language to a 'corresponding' regimented language and vice versa. For instance given a formal language where, say, the interpretation of the predicate symbol P in the model $M=<D,F>$ is given with a clause of the form in (s)

(s) $F(P)=\{x: x \text{ so-and-so}\}$

we have a corresponding way of interpreting the predicate symbol P in a regimented language, namely with a clause of the form in (t)

(t) P translates as 'so-and-so'

And conversely, given a clause of the form in (t) that allows us to interpret the predicate *P* in a regimented language, we can think of the corresponding way of interpreting *P* in a formal language by means of a clause like (s). The same could be said for the other kinds of expressions.

Even if in many cases an interpretation of a system of forms as a regimented language already suggests an specific way of interpreting the system of forms as a formal language, and also vice versa, the two sorts of 'language' are very different. Remember that as we saw at some length in section 1 a sentence of a formal language does not really mean anything in the way that sentences of natural language or of a regimented language do.

George Boolos writes[8]

> When we say that + denotes plus in *N*, using "plus" or a synonym to say so, we allow it to be understood that + is to have the sense of "plus", whatever that might be (and not, say, that of "plus the cube root of the square root of the cube of the square of"). Similarly for the other symbols of the language, including the variables, the manner of specification of whose range, i.e., *as* over the natural numbers, contributes in large measure to the determination of the meanings of quantified sentences of PA.

In this passage professor Boolos seems to be aware of the difference between having a formal language (whose sentences would not really have any meaning) and having a regimented language (whose sentences can have the sort of meanings that we intuitively attribute, for instance, to the sentences of the 'language' of arithmetic), and of the tension that arises between interpreting the 'language' (system of forms) of arithmetic as a formal language while pretending at the same time that the sentences have meaning and say things about the natural numbers in the way that English sentences say things about the natural numbers. He seems, though, to pretend to be having both a formal language and a regimented language when he introduces the standard formal language for the 'language' of arithmetic. He suggests that when introducing a formal language we are also introducing a corresponding regimented language. Other authors also seem to assume something like this, even if they do not make it explicit in the way Boolos does. Doing so without saying anything else, though, is unjustified. If nothing else is added the definition of a formal language does not provide by itself anything else other than the formal language itself, and this sort of 'language', as we have argued, is not a regimented language.

(2) What is the relationship between formal languages and regimented languages? I want to claim that formal languages are *models* of regimented languages. Here by *model* we do not mean what is meant in model theoretic semantics, i.e., an structure or interpretation, but rather what we usu-

ally mean when we say that we construct a model of something: something else that has *some* of the same properties as the original object, and that it is usually made in order to facilitate studying those properties that the two objects have in common. To avoid terminological confusions we will call a model in this (when not talking about logic) most common sense a *modelation.*

In what sense is a formal language a modelation of a regimented language? At least in the following sense: a formal language models the way the truth of the sentences of a regimented language is affected by the combined effect of the meaning of the expressions of the regimented language *and* the way the world is. Here by 'the way the world is' we do not just mean 'the way the world actually is', rather we mean that the formal language models how *one* way the world *might be* and the meaning that the expressions of *some* regimented language have would affect the truth of the sentences of the regimented language. With respect to each specific expression of a regimented language, a formal language models how the meaning of the expression in the regimented language and a way the world is affect the contribution that the expression makes to the truth value of the sentences where the expression occurs.

(3) That a formal language is a model of a regimented language in the sense we have just pointed out can be used to justify the adequacy of the standard definitions of logical truth and logical consequence in the following way[9]:

Logical truth and logical consequence are the two fundamental logical notions. Our intuitions about these two fundamental properties (which because they are fundamental we are particularly interested in clarifying) seem to be roughly the following: a sentence (of a natural or a regimented language) is a logical truth if it is true just in virtue of the meaning of certain expressions, the so called *logical expressions*, and the 'form' of the sentence. 'Form' here does not mean 'grammatical form' but rather 'the semantic category of the different expressions of the sentence and the way the expressions are combined'. As for the logical expressions they are expressions characterised by having a meaning of a particularly general kind; unlike the other expressions, it is meaningful to apply logical expressions to all sorts of discourse. Besides this vague and general idea, our intuitions about logical expressions seem to include enough as to allow us to recognise one when we see it, with the limitations imposed by the vagueness that the concept of 'logical expression', as most others, surely has: 'or' and 'every' are logical expressions, but 'John' and 'dog' are not, we might not be com-

pletely sure regarding expressions such as 'is one of them' or 'exactly five'. A sentence is a logical consequence of some sentences if: the sentence is true if all these other sentences are and this is so just in virtue of the meaning of the so called logical expressions and the 'form' of all the sentences involved. A sentence is a logical truth if it is true just in virtue of the meaning of the logical expressions it contains and of the 'form' of the sentence. For simplicity we will from now on focus just on the notion of logical truth.

We are certainly interested in making these intuitions more precise. One way of doing so is by realizing that we capture these intuitions if we say that a sentence is a logical truth if it is true whatever way the world might be and whatever the meanings of the non-logical expressions might be, provided that they have a meaning that keeps them in the same semantic category. This formulation does not appeal to the notion 'in virtue of' which was certainly in need of clarification.

Having this formulation of what it is for a sentence to be a logical truth we might want to restrict our attention to regimented languages since they seem to be rich enough so that what we say about the logical properties as applied to them is extendable to a language in general but at the same time, since they are regimented, they avoid some of the unwelcome complications of natural language (some having to do with facts that we already mentioned in section 2, like the effect of Gricean maxims, others having to do with some facts about natural language such as the vagueness about which expressions are logical expressions -in a regimented language we list the logical expressions, and all of them will be among the ones for which we have no doubt that they are logical).

If (i) each formal language is a model of how the meaning of a regimented language and a way the world might be affect the truth of the sentences of the regimented language, and (ii) for each regimented language and a way the world might be we have a corresponding formal language that models it in the sense we just mentioned[10], then we can say that a sentence is a logical truth iff it is 'true' in all formal languages that share the same system of forms as the regimented language. Of course, once we talk about formal languages it does not really matter if the property that the sentences have is that of being true or simply, say, that of being assigned the value 1. All that matters is that it is a property that can model the property of being true that the sentences of the regimented language do indeed have. For this it is enough that it be one of two properties that will be assigned to the sentences depending in the right way on the values that

the expressions in the sentence have. (In fact, given that the sentences of a formal language do not really mean anything, probably it can not be said in a proper sense that they are true or false).

Now, given the last rendering of what it is for a sentence to be a logical truth, and given that which formal languages we have is determined by which models or structures we have, we can say that a sentence of a formal language is logically true iff it is true (has value 1, etc) in all models. This is, of course, the standard formulation.

One question that would require further discussion but which we will not examine any further here is this: the standard model-theoretic analysis of the logical properties can be seen as having two parts. One the one hand, we have what is properly the conceptual analysis of the logical concept, where we say, for instance regarding the concept of logical truth, that a sentence is logically true if it is true in all models that correspond to one combination of a way the world might be and a meaning the sentences could have. Only with this first part we are not yet able to apply the analysis to any specific sentence. We need to know which all the models are that correspond to a combination of a way the world could be and a meaning the expressions could have (in order to see whether the sentence is 'true' in all of them). In the second part we throw in our substantive metaphysical assumptions about how the world could be, and what sets we actually have to model them, in order to determine which are all the models that should be considered when applying the analysis that we arrived at in the first part[11].

Notes

[1] We will be making three distinct uses of italics: (i) for emphasis; (ii) to talk about an expression type of which we are exhibiting a token, i.e. we might use italics instead of quotation marks; (iii) to talk about expressions type while exhibiting tokens of some of them and/or using metavariables, i.e. we might use italics instead of corners. We do this for simplicity. Only when being precise becomes essential we will have recourse to quotation marks or corners.

[2] The position according to which asserting some predication is the same as asserting some membership relation does not agree with intuition and the burden of justification is on the side of the one that wants to hold that view. Still here are two reasons, in addition to its conflict with basic intuitions, for not holding it: (a) someone can believe that Armstrong philosophizes but not believe that Armstrong$\in$ {x: x philosophizes}; (b) if we accept the principle that assertion of predication is assertion of certain membership relation then we have to accept that the principle also applies to the

assertion that Armstrong$\in${x:x philosophizes}, and so that this assertion is the same as: <Armstrong, {x:x philosophizes}>$\in${<x,y>: x$\in$D & y$\in$P(D) & x$\in$y}.

But then the principle would also apply to this latter predication of membership, etc.; it becomes intuitively less and less plausible that all these other assertions of membership are the same as the original assertion that Armstrong philosophizes.

3 We could require that the equivalence holds in all worlds where Edinburgh is the capital of Scotland, or in all physically possible worlds, in all worlds where the interpretation of language P is the same, etc.

4 Notice, though, that, in another respect, this biconditional would also be intuitively too strong since would require *Pa* to be true in any world where Armstrong philosophizes, even in those where *Pa* does not have its actual meaning, and means something which is not the case in that world.

5 We use the sign "^" to express concatenation of expressions. So, for instance: "John" ^ ^ "runs" = "Johnruns".

6 We could do this using a 'pairing function' and Gödel's β-function. See, for instance, George Boolos' *The Logic of Provability*, 1993, Cambridge University Press, pp. 17 ff.

7 Z is the theory of Elementary Peano arithmetic. For a list of its axioms see, for instance, Boolos, op. cit., pp. 18-19.

8 Op. cit., p. 33.

9 For a more complete exposition of what follows see Macià (1997).

10 For a justification of the truth of (i) and (ii) see Macià (1997, pp. 170-171).

11 I am grateful to George Boolos, Manuel García-Carpintero, Irene Heim, Ignasi Jané, Vann McGee, Robert Stalnaker and Gabriel Uzquiano for comments and discussion. Financial support was provided by the research project PB96-1091-C03-3, funded by the DGES, Spanish Department of Education.

Josep Macià is particularly interested in semantics, the relation syntax-semantics-pragmatics, philosophy of logic and philosophy of language. He got his Ph.D. from MIT with the dissertation 'Natural Language and Formal Languages' (1997). He is the author of 'On Concepts and Conceptions' (*Philosophical Issues* 9, 1998) and 'Does *Naming and Necessity* Refute Descriptivism' (*Theoria* 13:33, 1998). He is assistant professor at the Department of Logic, Philosophy and History of Science, University of Barcelona.