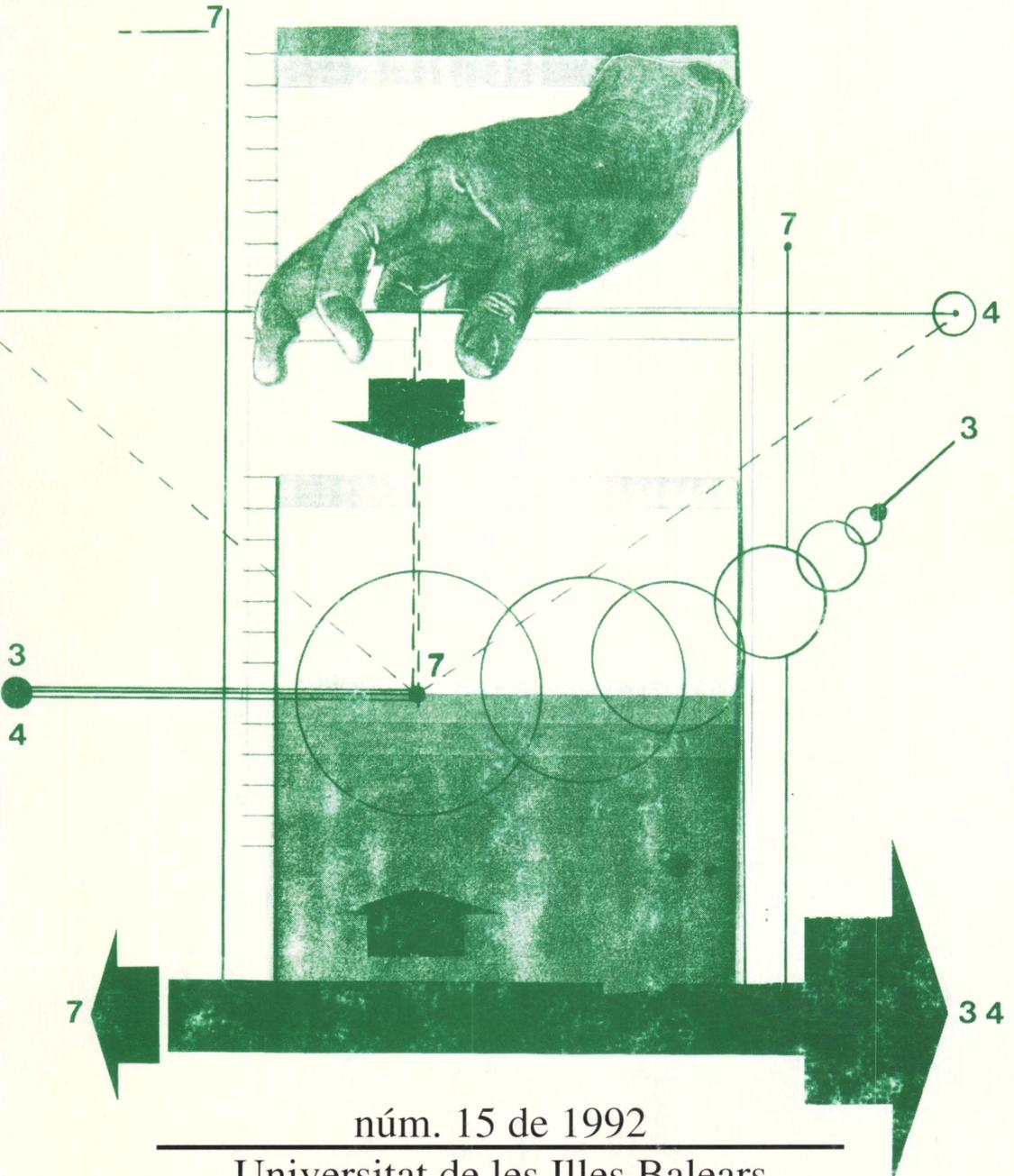


TAULA

quaderns de pensament



núm. 15 de 1992

Universitat de les Illes Balears

Taula

quaderns de pensament

15

Palma 1992

Universitat de les Illes Balears
Departament de Filosofia

UNIVERSITAT DE LES ILLES BALEARS



510152092X

Índex

Presentació.....	5
Antoni Gomila Benejam	

Tema

<i>Fodor's contribution to the study of the cognitive mind</i>	9
José E. García-Albea	
<i>Por qué el Significado (Probablemente) no es el Rol conceptual</i>	19
Jerry A. Fodor, Ernest Lepore	
<i>Contents and contexts: on Fodor's narrow/broad divide</i>	35
Manuel Liz	
<i>La acción, su razón y su circunstancia</i>	49
Fernando Broncano	
<i>Conexionismo y el epifenomenalismo de lo mental</i>	69
Josep E. Corbí	
<i>El problema de la modularidad en el procesamiento auditivo y en el procesamiento del habla</i>	83
Luis Enrique López Bascuas	
<i>The modularity thesis: its implications for interpretations of priming effects</i>	93
Christopher W. Dawis	
<i>Mental Models and Referential processing</i>	101
Manuel de Vega	
<i>Sobre la autonomía de los procesos de formulación en la producción de oraciones</i>	115
José Manuel Igoa	
<i>Confirmación isotrópica-quineana reconsiderada</i>	137
J. Padilla-Gálvez	

Recensions bibliogràfiques

Anton Pacheco, José Antonio. <i>Un libro sobre Swedenborg</i> , Publicaciones de la Universidad de Sevilla, Sevilla 1991 (95 pp.)	149
Gabriel Amengual	
Cordua, Carla, <i>El mundo ético. Ensayos sobre la esfera del hombre en la filosofía de Hegel</i> . Ed. Anthropos, Barcelona 1989 (223 pp.)	150
Gabriel Amengual	

Hegel, G.W.F., <i>Vorlesungen über die Geschichte der Philosophie.</i> Teil 4: Philosophie des Mittelalters und der neuen Zeit. Herausgegeben von Pierre Garniron und Walter Jaeschke. Felix Meiner Verlag, Hamburg 1986 (G.W. F. Hegel, Vorlesungen. Ausgewählte Nachschriften und Manuskripte, Bd. 9) (437 pp.)	152
Gabriel Amengual	
Hegel, G. W. F., <i>Vorlesungen über die Geschichte der Philosophie.</i> Teil 2: Griechische Philosophie 1. Thales bis Kyniker. Hrsg. v. P. Garniron u. W. Jaeschke. Felix Meiner Verlag, Hamburg 1989 (G.W.F. Hegel, Vorlesungen. Ausgewählte Nachschriften und manuskripte, Bd. 7) (442 pp.)	152
Gabriel Amengual	
Lauth, Reinhard, <i>Transzendente Entwicklungslinien von Descartes bis zu Marx und Dostojewski</i> , Meiner, Hamburg 1989 (486 pp.)	155
Gabriel Amengual	
Mayos Solsona, Gonçal. <i>Entre la lògica i empiria. Claus de la filosofia hegeliana de la història</i> . Prefaci de Ramon Valls Plana. Barcelona 1989 (259 pp.)	157
Gabriel Amengual	
Javier Muguerza et al., <i>El fundamento de los derechos humanos.</i> Ed Debate, Madrid 1989	159
Antoni Barber Orfila	

Presentació

De l'11 al 14 de setembre de 1990 van tenir lloc a Maó (Menorca) les Jornades sobre el Model Computacional de la Ment: l'Aportació de Jerry Fodor. Organitzades pel Departament de Filosofia d'aquesta Universitat i per l'IME (Institut Menorquí d'Estudis), i amb el suport del Ministeri d'Educació i Ciència, van permetre aplegar durant quatre dies un grup multidisciplinari d'investigadors al voltant del prestigiós professor Jerry Fodor, de les universitats de Cuny i Rutgers.

Aquí es recullen les aportacions presentades en un intent de donar un abast més ampli al treball d'aquells dies, sabent, però, que la vivacitat i la riquesa de les discussions que seguiren cadascuna de les intervencions aquí aplegades –un dels objectius centrals de l'organització de les Jornades– quedaran necessàriament excloses.

Les aportacions, com es pot veure, s'organitzen segons dos eixos principals, en funció dels aspectes més rellevants del treball de Fodor. L'un, de caire més empíric, tracta de l'arquitectura cognitiva; l'altre, de tradició més filosòfica, és el de la intencionalitat.

Quant al primer, engloba qüestions diverses: la hipòtesi de la modularitat, l'estatus dels processos centrals, el rebuig dels models connexionistes, etc. Els treballs d'Igoa i López Bascuas, de caire experimental, aprofundeixen les possibilitats explicatives de la noció de modularitat, mentre que Davis suggereix com aquesta hipòtesi permet reinterpretar coherentment resultats previs. De Vega, Corbí i Padilla, en canvi, defensen posicions crítiques respecte a Fodor, en qüestions diverses: el paper dels models mentals en el processament cognitiu, la viabilitat del connexionisme, la noció de processos quineans.

Quant al segon, el de la intencionalitat, s'hi inclou tant l'aportació de Fodor que recollim aquí, centrada particularment en el rebuig de l'holisme semàntic, com els treballs de Broncano i Liz, els quals, des de perspectives diverses, ataquen la noció de «*narrow content*», estaló principal de l'antiholisme fodoriana. García-Albea, primer impulsor de la traducció de Fodor al castellà, introdueix les Jornades amb una caracterització dels trets centrals que configuren el pensament de Fodor.

En efecte, s'enganyaria qui pensàs que els dos eixos de discussió esbossats delimiten compartiments independents. Ben al contrari, és precisament la seva interrelació

en un objectiu comú allò que dóna sentit a unes jornades multidisciplinàries com les que es pretenen reflectir aquí. L'objectiu de Fodor és clar: aconseguir trobar per a la psicologia el «camí segur de la ciència», sense renunciar, però, a allò més distintiu dels estats psicològics, el seu contingut intencional. El camí que segueix, per a aquest fi, és el d'un racionalisme renovat, que aspira, de bell nou, a assegurar la correspondència entre subjecte i objecte, entre ment i realitat. Problemàtic, sens dubte, però alhora punt de referència inexcusable en el debat actual.

Esperam, per tant, que aquests treballs resultaran d'interès, que contribuiran a crear les condicions per a la col·laboració interdisciplinària que caracteritza les Ciències Cognitives.

Antoni Gomila Benejam
Universitat de les Illes Balears

Taula (UIB) núm. 15, 1992
Tema

Taula (UIB) núm. 15, 1992

Fodor's contribution to the study of the cognitive mind

José E. García-Albea*

Abstract

This presentation is intended as a general overview of the main issues that Jerry Fodor has addressed concerning the explanation of intelligent behavior. From a mosaic of different pieces of work, which sometimes may look unrelated and even conflicting, our purpose is to extract the global picture, trying to make explicit the links between its parts. Fodor's contribution to the study of the cognitive mind is sketched by means of three antinomies that express the scope and limits of that study. These antinomies are: 1) Functionalism versus Physicalism; 2) Computation versus Association; and 3) Modularity versus Globality.

Resumen

En esta presentación se intenta hacer un repaso general de los principales temas tratados por Jerry Fodor en el contexto explicativo de la conducta inteligente. A partir del mosaico formado por sus diferentes trabajos, que a veces pueden parecer inconexos y hasta contrapuestos, se trata aquí de proporcionar las claves para apreciar su extremada coherencia, destacando así las relaciones que se dan entre ellos. La decisiva aportación de Fodor al estudio de la mente cognitiva se resume a través de tres antinomias que expresan el alcance y los límites de dicho estudio, a saber: 1) Funcionalismo versus Fisicalismo; 2) Computación versus Asociación; y 3) Modularidad versus Globalidad.

* Artículo presentado a la revista TAULA y que corresponde a la conferencia inaugural de las JORNADAS CIENTÍFICAS SOBRE LA TEORÍA COMPUTACIONAL DE LA MENTE: LA APORTACIÓN DE J. FODOR, celebradas en Mahón, del 11 al 14 de Septiembre de 1990.

Dirección del autor: Depto. de Psicología Básica (Procesos Básicos). Facultad de Psicología, U.C.M. Campus de Somosaguas. 28023-Madrid.

It is a great pleasure for me to open this seminar devoted to Jerry Fodor's contribution to the computational model of mind. From the beginning, I want to express my admiration and gratitude to Toni Gomila and to the Department of Philosophy at the University of the Balearic Islands for organizing this meeting, for making possible the presence of Jerry Fodor himself, and for bringing us to this charming island of Menorca, a wonderful scenery for intellectual enterprise, ... among other things.

I'm supposed to make a presentation and to set some background for the oncoming lectures and discussions. For many of you this will be no more than a reminder, but in any case it is what has to be done at the very onset of the game: to put the cards on the table, or if you like this better, to set the frame.

And it is here where the troubles start, in setting the frame to capture what Jerry Fodor has done. (By the way, I suspect that frame problems, in one version or other, are going to be all around during these days). Well, the first difficulty in conceptualizing Jerry Fodor consists in his atypicality. He is not a typical philosopher, even though he was indeed trained as such, got his degrees at Princeton University, has been involved in so many arguments with philosophers, and has actually made so important contributions to the philosophy of mind. Nor is Jerry a typical psychologist, even though he was introduced into experimental psychology by such an authority in the field as Charles Osgood (University of Illinois) and was also the director of the Laboratory of Psycholinguistics at M.I.T. for twenty years, with so many publications in this field. Jerry is not a typical linguist either, and he even declares that his approach to linguistics has always been amateur; but, as it is well known, he was there, among Chomsky's first graduate students and colleagues, giving birth to the Standard Theory, with a major contribution to the formulation of the semantic component of generative grammar.

Jerry is not a typical scholar in any of these fields. In dealing with philosophers, he has not worried about frequently going against the tide, mainly when operationalism, associationism, physicalism, intentional interpretationism, or meaning holism have been on fashion. When he comes into psychology circles, Jerry Fodor is viewed pretty often as a disturbing visitor who forces everyone to think of the sense of their work; in some cases, he is nevertheless acknowledged by psychologists for his contribution to the foundations of the discipline and for keeping them busy with new research problems. And if he shares the table of the linguists, then he sits down in a corner trying to keep track of their arguments, while thinking that what really matters is to understand how the mind works (this is also what he usually does when invited to the Artificial Intelligence meetings).

Perhaps, in order to set Fodor's own frame, it is this very interdisciplinary orientation which provides the first clue. For this is what is precisely at the core of that emerging field called Cognitive Science, and perhaps Howard Gardner is right when, in his book about the history of the cognitive revolution («The mind's new science») devotes an entire paragraph to Jerry Fodor considered as the prototype of the complete cognitivist. Gardner introduces this paragraph with the following words: «... it is

salutary to consider the work of Jerry Fodor, a full-scale cognitivist -one philosopher who seemingly has no reservations whatsoever about the common fate of philosophers and empirical scientists interested in issues of mind» (1985, p. 81).

However, giving a name to an emerging field is not the same as having a clear idea of what are going to be the scope and limits of that field. And here again we have to face another frame problem, the one that refers to the feasibility of Cognitive Science as such, or for that matter, to the possibilities for psychology to count as a proper science. I think that Fodor's approach to this problem might be taken as his basic contribution to the study of the cognitive mind, where all the others stem from and get summarized.

Tracking Fodor's writings in the past 25 years, the to-be and not-to-be of Cognitive Science can be expressed in this three antinomies:

- 1) Functionalism vs. Physicalism
- 2) Computation vs. Association
- 3) Modularity vs. Globality

The left side of each antinomy represents the key value that defines the scope of research in cognition; the terms on the right represent the limits of that enterprise. The first contrast bears on the required level of explanation for intelligent behavior, so it sets up the playing-ground. The second one bears on the type of modelling that best suits the demands of mental operations underlying that behavior, thus providing the explanatory instrument. And the third contrast refers to two major kinds of properties that may be shown by mental operations, which in turn may be more or less accessible to productive research and explanation. Both kinds of properties contribute to the configuration of an architecture for the cognitive mind, with some components that are better understood than others.

Functionalism vs. Physicalism

The possibility of studying mental states and processes depends on the assumption that there is a functional level of explanation wherein those states and processes are involved in the causation of behavior without being reduced to the neurological ones. So it is assumed that behavior is a complex function of internal properties of the system and that for capturing those properties we need a sufficiently abstract level of description, one which is not provided by the more basic sciences (all the way down from neurology). Functionalism preserves psychology (and cognitive science) from two major kinds of reductionism (physicalist and behaviorist) and, at the same time, sheds a new light on the mind-body relation, avoiding ontological dualism: mental types cannot be reduced to neural types, even though mental tokens are physically instantiated by neural tokens (or, for that matter, by tokens of other hardware substances).

On the other hand, functionalism, at least as Jerry Fodor understands it, does not entail that the explanatory power of the so called «special sciences» should be diminished with respect to basic science. The difference between special sciences and

basic science is in the abstraction level they adopt and, correspondingly, in the more or less restrictive character of their respective laws. (The only science which claims to apply to everything is basic physics; as soon as you consider higher order properties, the extension of the laws gets reduced, there are more exceptions.). But, in spite of this difference in extension, there is no reason why the laws discovered in both cases shouldn't have the same explanatory capacity and why the notion of causality can not be applied in the same way (viz., as succession of events governed by a law).

This relates to a further and very important point that has to be made when considering functionalism in the context of behavioral sciences. The attribution of explanatory power to psychology is contingent upon the attribution of causal powers to mental states and processes. Moreover, what distinguishes mental events from other kinds of events is that they are intentional, or referential, or if you like this better, they have informational content. But then we are committed to adopt a strict realistic position with regard to intentionality. Mental events are not just a way of talking (or «a façon de parler») about neural or physical events. There are intentional laws that express causal relations between mental events, just like there are physical laws, biological laws or economical laws that bear on other types of events.

Fodor's insistence on this point has been remarkable, in spite of so much skepticism that he has met around, coming from philosophers like Dennett or Stich (or even from old functionalists like Putnam). This skepticism also comes from many psychologists and A.I. people, that try to keep their scientific enterprise as something «special», but in the precarious sense that it may live without ontological commitments. (By the way, this is very surprising, because the same psychologists that charge Fodor for having reduced the field so much are often the ones that don't care about or are even against intentional realism, the very conditioning factor for the possibility of psychology.)

The intentional character of mental states goes hand in hand with the vindication of propositional attitudes (remember, for example, Fodor's «Three cheers for propositional attitudes» in 1979), and at the same time, reconciles psychological science with commonsense psychology (remember, for example, the first chapter of *PSYCHOSEMANTICS*, 1987). Propositional attitudes are just relations that organisms bear to mental representations. If we thus consider mental states as propositional attitudes, we already have got the way to individuate the objects of our science: mental states are individuated by the kind of relation and by the representational content expressed by propositional attitudes.

So far, so good. But now we have to face the problem of how in the world mental states enter into causal interactions and may finally produce intelligent behavior. It is now when Fodor's computational theory of mind comes onto the stage.

Computation vs. Association

The starting point here is the very notion of mental representation. Mental representations are symbolic entities, and just like all symbols, they have both semantic content and syntactic form. Because they have semantic content, our mental states have

content, that is, they are intentional (or referential). Because they have syntactic form, which makes them susceptible to different combinations and transformations, our mental states can intervene in causal interactions. In this sense, the mental processes that make these interactions possible are computational. Computations, after all, are no more than formal operations defined over representations.

The arguments and supporting evidence for this theory of mind have been developed in many different pieces of work beginning from 1975 landmark *THE LANGUAGE OF THOUGHT*. I am not going to review them all here, but just remind you that if mental causation is computation, it must have at its disposal a medium or vehicle for computation, that is, the system for mental representations that constitutes the language of thought or mentalese. What is known about the human mind and its main achievements (in perception, learning, problem solving, language, reasoning, etc.) suggests that mentalese has to be a pretty abstract language, with a high expressive power, a rich vocabulary and a rich syntax, mostly innate, flexible, and subject to a rational and efficient exploitation by the individual.

To adopt Fodor's view of mind entails certain consequences for setting up the limits of the theories in psychology and cognitive science. First of all, the language of thought story precludes an associationist view of mind, be it in the traditional flavour of the British empiricists or in the more recent versions of the connectionists. The mechanisms of mental causation provided by associationists (contiguity, resemblance, statistical inference, etc.) are just inadequate to account for the properties shown by cognition at large (such as productivity, systematicity, constituency, etc.).

So this is mainly an empirical issue. In this respect, it is important to notice that associationism may recognize the central role of mental representations and, for that matter, be mentalist and honor the constraints of a functional level of explanation. What happens is that it has not got right the mechanism of mental causation.

Notice also that computability as such doesn't do the work. Associationist functions are indeed computable, as connectionists are showing all the time; but it is one thing to deploy a computational apparatus to formalize the theory and to make the model work, and quite another to consider the mind as really computational, at least in the sense in which a Turing machine is computational. Connectionists are not computational in this respect, and that's why, according to the available evidence, their model of the cognitive mind is not the right one. The case made out by Fodor against connectionism is not very different from the case made out by Chomsky against finite state grammars more than thirty years ago (see Fodor and Pylyshyn, 1988, and Fodor and McLaughlin, 1990).

Another unavoidable consequence of adopting Fodor's theory of mind is the exclusion of semantics from the territory of psychology. Remember that the causal properties of our mental states and processes are derived from the formal-syntactic aspects of mental representations. The content can still play a role in the individuation of mental types, but only to the extent that it is subject to that *formality condition*, and not as subject of such semantic properties as truth, reference, meaning, and so on. These semantic notions could be necessary to face the epistemological (or perhaps ontological) problem of how our thoughts match with reality, but they simply are not psychological categories in charge of explaining mental causation.

This point has been one of the main guidelines of Fodor's work in both the empirical side (let's remember «The psychological unreality of semantic representations» in 1975, or «Against definitions» in 1980) and the theoretical side (from his arguments based on the intensionality of mental states -presented in «Methodological solipsism», 1980- to the more sophisticated ones against Meaning Holism, developed in his latest books). At the same time, this point has also been the one that perhaps has raised more controversy and reserve, and I suspect that these will be also raised during these days (so we'll have time to discuss it). What I think has to be clear from the beginning is the coherence of Fodor's view: if cognition is computation, then Cognitive Science (or just Psychology) is not Semantics. (To be or not to be).

Modularity vs. Globality

We can now proceed to examine the third antinomy that configures the study of the cognitive mind. In this case, the issues are more neutral with respect to the status of Cognitive Science. But assuming that human cognition is the prototype to be reproduced by machines, the matter becomes of enormous importance for the actual development of our science. So here the limits, if there are any, will be internal to the field, once we have opted for the first parts of the two former antinomies (a functional level of explanation and a computational theory of mind). And here the underlying questions are entirely open to empirical contrast, and they finally consist in knowing how much we can understand of the human mind. Depending on the answer to this question we'll know to what extent we are going to be able to construct really intelligent machines (at least as intelligent as us).

It is needless to remark Fodor's important contribution to these issues since the publication of *THE MODULARITY OF MIND* in 1983, perhaps one of the most influential and provocative books in the field. For the present purposes, and given that almost half of this seminar is going to be devoted to modularity, I will only make a few observations to prepare the ground.

In the first place, Fodor proposes a *computationally-based taxonomy* of the mind components. According to certain properties that can be shown by mental processes, a fundamental distinction is drawn between modular and non-modular systems; a distinction that has important consequences for the characterization of the functional architecture of the cognitive mind. I think that, in this respect, it is important to emphasize the computational base of the taxonomy. In one sense, because computation is the common feature of all processes, being modular or not (the language of thought story still keeps its force). In another sense, because the differences between systems are differences in their kinds of computations and the critical properties are properties of those computations. In this latter sense, Fodor goes beyond a mere informational modularity (à la Chomsky) where the structure of mind would be based on the structure of knowledge, but not on the nature of its operations.

Secondly, Fodor identifies the bunch of *critical properties* that contribute to make the distinction between modular and non-modular processes and, at the same time, are used for diagnosis. The former ones are domain specific, informationally encapsulated, mandatory, etc., whereas the latter ones are global, isotropic, etc. The former mainly correspond to input systems and the latter to central systems. The available evidence shows that processes like sentence parsing could be considered as prototypical cases of the modular ones, while inductive reasoning could be the paradigm of the non-modular (or central) processes. So far so good. There are many details to be filled in, but this is the general picture of the cognitive mind proposed by Fodor. However, there is still something more to be said.

The third and final observation refers to the prospects for the study of these two classes of systems. According to *Fodor's First Law of the Nonexistence of Cognitive Science*, «the more global a cognitive process is, the less it is understood» (Fodor, 1983, p. 107). So now the internal limits of the field become explicit. The properties used to sort out modular from non-modular systems are critical for assessing progress in psychological research (or in Cognitive Science). As I said before, this is basically an empirical issue open to disconfirmation. Nevertheless, Fodor doesn't resign himself to showing the evidence for his own Law, but goes much further trying to disentangle the possible reasons for the existing limits in understanding non-modular (or, say, unencapsulated) processes. Those reasons can be summarized in just one: the frame problem (but now, leaving rhetoric aside, in its more genuine and technical sense, as it was first raised in the A.I. circles). When a system has a very large data base at its disposal, with so many unconstrained potential relations among them, the question is how to choose the appropriate ones in order to solve a problem, make a decision, plan a course of action, etc. The frame problem is then the problem of how to manage relevance, and it is the problem that, in the first instance, non-modular unencapsulated systems have to solve and, as a matter of fact, do solve very efficiently. Secondly, it is the problem that has to face the cognitive scientist in order to finally understand intelligent behavior and design really intelligent machines. But in this case, given what we have got so far, the prospects are not very optimistic. As Jerry likes to say, there is something here that we just don't understand. Something that seems to depend on the global properties of the computations performed by our central systems.

Well, I think that this is enough for a first encounter with Jerry Fodor. I am not sure if he would put his signature to all I have said, but at least on this occasion we won't get lost in hermeneutic disputes. This meeting gives us the opportunity to hear Fodor's own position about all this (and more) as told by himself. I have just tried to give a very general overview of the main issues that he has addressed concerning the study of the cognitive mind. From a mosaic of different pieces of work, which sometimes may look unrelated and even conflicting, my purpose has been to extract the global picture, trying to make explicit the links between its parts.

However, I don't want to finish without pointing out some questions that still remain open and that could be located near the boundaries we have set for studying the cognitive mind. Or if you prefer, you can take them as a pretext for provoking your discussion from the very beginning.

The first one. According to the former picture, psychology is mainly concerned with cognition (viewed as computation). However, there are still other types of relations between organisms and representations instantiated by propositional attitudes, such as desires, motives, emotions, etc., which may also intervene in the intentional causation of behavior. Question: Are all these mental states susceptible to scientific treatment beyond what is assumed by folk psychology?

Second. The computational theory of mind is based on the syntax of mental representations. In this respect, minds are like computers. But mental representations have also semantic content, and in this respect minds are very different from ordinary computers (remember «Tom swift and his procedural grandmother» from 1978). Even though semantics is not part of psychology (or cognitive science), meaning is still a characteristic of our thoughts. To what extent psychology can move toward a better understanding of the implications of meaning in the explanation of behavior?

Third question. Is there any hope to overcome the problems raised by the frame problem? And another related to this: Why computer technology has made more progress in the domain of their putative central processes (reasoning, problem solving, expert systems, etc.) than in the domain of their putative perceptual capabilities (visual processing, speech perception, natural language understanding, etc.)?

And the last one. The functional architecture of cognition is taken to be mostly innate. Thus it may be relevant for its study to take into account the biological character of this innate endowment, since the software properties of a computer are not entirely independent of its hardware constraints. To what extent, then, can the study of the cognitive mind get rid of its neurobiological substratum and ontogenetic conditions?

I can assure you that these are really unbiased questions, at least in the sense that I don't have a clue of how to answer them. So any help from the audience will be wellcome.

Thank you.

References

- Fodor, J.A. (1975). *The language of thought*. New York: Harper & Row.
- Fodor, J.A. (1978). Tom Swift and his procedural grandmother. *Cognition*, 6. Reprinted in J.A. Fodor, *Representations*, 1981.
- Fodor, J.A. (1979). Three cheers for propositional attitudes. In W. Cooper and E. Walker (eds.), *Sentence processing: Psycholinguistic Studies presented to Merrill Garrett*. New York: LEA.
- Fodor, J.A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *The behavioral and Brain Sciences*, 3. Reprinted in J.A. Fodor, *Representations*, 1981.
- Fodor, J.A. (1981). *Representations*. Cambridge, Ma.: M.I.T. Press.
- Fodor, J.A. (1983). *The modularity of mind*. Cambridge, Ma.: M.I.T. Press.
- Fodor, J.A. (1987). *Psychosemantics*. Cambridge, Ma.: M.I.T. Press.
- Fodor, J.A., Fodor, J.D., and Garrett, M.F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry*, 6, 515-531.
- Fodor, J.A., Garrett, M.F., Walker, E., and Parkes, C. (1980). Against definitions. *Cognition*, 8, 263-367.
- Fodor, J.A. and McLaughlin, B.P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, 35, 183-204.
- Fodor, J.A. and Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Taula (UIB) núm. 15, 1992

Por qué el Significado (Probablemente) no es el Rol Conceptual*

Jerry A. Fodor, Rutgers University y CUNY Graduate Center

Ernest Lepore, Rutgers University

Introducción

Uno de los logros del último par de décadas consiste en que la gente que trabaja en Semántica Lingüística y la gente que trabaja en Filosofía del Language han llegado a un amistoso acuerdo de facto sobre cómo describir sus respectivas tareas. Los términos de ese acuerdo son que los semánticos se ocupan del trabajo y los filósofos de las preocupaciones. Los semánticos intentan construir teorías del significado efectivas (o teorías de la verdad, o teorías de modelos, o lo que sea) para uno u otro tipo de expresión en un lenguaje u otro. Los filósofos, por el contrario, se centran en los grandes problemas fundacionales como: cuál es la relación entre sentido y referencia, o entre pensamiento y lenguaje; si la traducción es determinada o no, y si la vida es como una fuente. Se supone que de tanto en tanto filósofos y semánticos se reúnen para compartir sus respectivos avances. O su falta.

De acuerdo con esta división del trabajo, este artículo no es sobre semántica sino sobre filosofía del language. Hemos estado hurgando en torno a los cimientos de la teoría del significado, y nos parece haber descubierto una grieta grande y desagradable. A nuestro parecer, una de las piedras angulares se está resquebrajando. Pensamos que es mejor notarlo antes de que las cosas empeoren.

* El material de este artículo está adaptado de nuestro próximo libro *Holism: A Shopper's Guide*, Blackwell, 1991. Aparecerá en las Actas de la Reunión de la Chicago Linguistic Society, *CLS 27-I: Papers from the Twenty-Seventh Meeting-CLS 27-II: Papers from the Parasession on Negation*, Dobrin, L., Nichols, L. & Rodríguez. R., eds., a la que agradecemos el permiso para su republicación.

Procederemos de la siguiente manera: primero, intentaremos localizar el problema; después, intentaremos decir de qué se trata; para finalizar, haremos un par de sugerencias sobre qué hacer al respecto. La primera parte de la discusión será muy amplia; la segunda será más específica; la tercera parte será prácticamente inexistente. Allá vamos.

1. Dónde está el problema

Un problema tradicional en la teoría del significado es «¿de dónde vienen las propiedades semánticas?» Lo que esta pregunta da por supuesto es el hecho de que lo que una palabra (o una oración, o lo que sea) significa, no puede ser un hecho bruto. No puede ser un hecho bruto, por ejemplo, que ‘perro’ significa *perro* y no *protón* y que ‘protón’ significa *protón* y no *perro*. Al contrario, ‘perro’ debe tener alguna propiedad *a*-semántica en virtud de la cual significa *perro* y no *protón*. Y ‘protón’ debe tener alguna (distinta) propiedad *a*-semántica en virtud de la cual significa *protón* y no *perro*. Dicho en la jerga filosófica estándar, las propiedades semánticas deben *sobrevenir a* las propiedades *a*-semánticas.

Puede haber ciertas propiedades que *simplemente* las cosas tienen, sin que haya razón alguna para ello. Pero si las hay, son las propiedades de las que habla la física básica (como carga, masa, etc.). Y ciertamente no incluyen los tipos de propiedades de las que hablan los semánticos (como significar *perro* o ser sinónimo de ‘soltero’).

Nótese de pasada que nada de esto debe entenderse como un intento de imponer el fisicalismo. Para nuestros propósitos nos basta que las propiedades semánticas sean, por ejemplo, irreduciblemente intencionales, o irreduciblemente epistemológicas, o irreduciblemente teleológicas. Lo único que es preciso excluir es que sean irreduciblemente semánticas. En resumen, no nos interesa si las propiedades semánticas sobrevienen a algo que es *físico*, en tanto en cuanto sobrevengan a algo distinto de sí mismas.

La cuestión que se plantea entonces es: «¿A qué sobrevienen las propiedades semánticas de los símbolos?» A lo largo de los años los filósofos del lenguaje se han entusiasmado con dos tipos (quizás incompatibles) de respuesta a esta cuestión. Una es lo que llamaremos la respuesta del «Antiguo Testamento», de acuerdo con la cual el significado de una expresión sobreviene a la *relación de la expresión con las cosas del mundo*. Y la otra es la del «Nuevo Testamento», según la cual el significado de una expresión sobreviene al *rol de la expresión en el lenguaje*. El desacuerdo entre estos dos tipos de explicación es venerable, y no nos proponemos entrar en detalles. Un parágrafo o dos, a modo de recuerdo, bastará.

La semántica «Antiguo Testamento» deriva directamente del empirismo británico y tiene entre sus modernos representantes a los psicólogos conductistas, como Watson y Skinner, y a un puñado de filósofos del movimiento de la «semántica naturalizada» (incluyendo a Dretske, Millikan, Barwise y Perry en algunos momentos, y uno o más Fodors). La idea básica es que ‘perro’ significa *perro* debido a cierta relación (*a*-semántica) que se da entre el símbolo y el animal. Si eres un empirista británico (o al menos, si eres Hume), entonces para ti esa relación tiene que ver con la semejanza. ‘Perro’ significa *perro* y no *protón* porque ‘perro’ está asociado a cierta imagen mental que significa *perro* y no *protón*. Y la imagen mental significa *perro* en lugar de *protón*.

porque se parece mucho a los perros pero no tiene nada que ver con los protones. Si eres un psicólogo conductista dirás que se trata de una relación *causal* (típicamente asociativa), entre símbolo y mundo; ‘perro’ significa *perro* y no *protón* porque, a consecuencia de la historia de condicionamientos del hablante los perros provocan emisiones de ‘perro’, a diferencia de los protones. Los desarrollos recientes de la semántica «antiguo testamento» proponen aun nuevas variaciones de este tema, incluyendo, por ejemplo, recurso a las relaciones nomológicas y/o informacionales entre los símbolos y el mundo.

Bien, este enfoque tipo «antiguo testamento» ha recibido muchas críticas, tanto en filosofía como en lingüística. Quizás la principal objeción proviene de Frege: el significado *no puede* ser una relación entre símbolos y el mundo porque la *identidad* de la relación con el mundo es compatible con significados *distintos*. De esta forma, es plausible que expresiones como «la estrella de la mañana» y «la estrella del anochecer» estén ambas ligadas al mismo objeto no lingüístico (al planeta Venus); y es también plausible que tales expresiones *no signifiquen* lo mismo. De lo contrario, no se podría negar que la estrella de la mañana es la estrella del anochecer sin autocontradicción, o afirmarlo sin caer en una tautología. Lo que de hecho, no ocurre.

Así, de acuerdo con este argumento, el significado de una expresión no sobreviene a la forma en que está ligada al mundo. ¿A qué sobreviene entonces?

La respuesta tipo «nuevo testamento» es una elaboración de la siguiente idea: las expresiones «la estrella de la mañana» y «la estrella del anochecer» significan cosas diferentes, a pesar de estar ligadas al mismo objeto (Venus), debido a que desempeñan *roles diferentes en el lenguaje*. «Sólo en el contexto de una oración tiene una palabra significado», afirma Frege, y Wittgenstein añade que «comprender una oración es comprender un lenguaje». El núcleo de la idea es que es la forma en que las expresiones se conectan *entre sí* lo que determina el significado de esas expresiones en el lenguaje. Muy a menudo (de hecho, en todas las teorías semánticas que discutiremos), esta noción de «rol en el lenguaje» recibe un giro epistemológico: aprender una expresión lingüística consiste en estar dispuesto a realizar un conjunto central de *inferencias* que fijan la relación semántica de esta expresión respecto a las demás expresiones del lenguaje. Aprender la expresión ‘perro’ es, inter alia, estar dispuesto a inferir «Rover es un animal» a partir de «Rover es un perro». Aprender la expresión «la estrella de la mañana» es, inter alia, estar dispuesto a inferir «sale por la mañana» de ello. Y así sucesivamente.

Queremos enfatizar que, en este punto, estamos usando nociones como *inferir* y *estar dispuesto a inferir* como cheques en blanco. Por el momento permitiremos que cualquier disposición a que una creencia cause otra cuente como una disposición a inferir la segunda de la primera. De este modo, las inferencias deductivas, inductivas, plausibles y prudenciales, así como meras asociaciones, y Dios sabe qué más, están todas incluidas. Gran parte de este artículo consistirá en señalar los problemas con que la semántica «nuevo testamento» se encuentra cuando intenta decir exactamente cuáles son las inferencias constitutivas del significado de los términos que aparecen en ellas.

El enfoque «nuevo testamento» en filosofía es bastante parecido al enfoque, familiar, de la lingüística estructuralista, según el cual el significado de una expresión es su rol en una «sistema de diferencias». Saber lo que ‘perro’ significa es saber que excluye ‘gato’ y cohabita con ‘animal’. Saber lo que ‘soltero’ significa es saber que

calculó a Juan, y por tanto el ejemplo presente no constituye una objeción a la sistematicidad del español. No obstante, no pretendemos dogmatizar. Si existe la proposición de que la respuesta calculó a Juan, entonces el español *puede* expresarla: por supuesto, es lo que hace la forma de palabras «la respuesta calculó a Juan». El español es sistemático bajo cualquier supuesto.

En conexión con la productividad y la sistematicidad se da otra característica, aparentemente universal, de los lenguajes naturales. La estructura de las oraciones es, en el siguiente sentido, *isomórfica* con respecto a la estructura de las proposiciones que expresa: *si una oración S expresa la proposición que P, entonces los constituyentes sintácticos de S expresan los constituyentes sintácticos de P*. Si, por ejemplo, una oración expresa la proposición que P y Q, entonces la oración incluirá un componente sintáctico correspondiente a P y otro constituyente sintáctico que expresa la proposición que Q. Si una oración expresa la proposición que Juan quiere a María, entonces habrá un componente sintáctico de la oración que se refiera a Juan, otro componente sintáctico de la oración que se refiera a María, y otro constituyente sintáctico de la oración que exprese una relación tal que, x está relacionado con y si y sólo si x quiere a y. Nótese que, aunque todo esto es patente, no se trata de truismos. Los modismos y otras construcciones «holofrásticas» son excepciones, pero del tipo de excepciones que confirman la regla.²

Lo que afirmamos es que estas tres generalizaciones acerca de los lenguajes naturales -productividad, sistematicidad e isomorfismo- están conectadas, y se explican como consecuencias de la composicionalidad. Se siguen del principio según el cual el significado de una oración está compuesto de los significados de sus partes. De otra manera, resultan sorprendentes. Por consiguiente, en lo que sigue insistiremos en la composicionalidad.

Sobre esta base podemos alcanzar la grieta en los cimientos. Consiste en que la composicionalidad resulta *embarazosa* para la semántica tipo «nuevo testamento», que identifica el significado de una expresión con su rol inferencial. En particular, da pie al siguiente tipo de argumento *prima facie*:

1. Los significados son composicionales.
2. Pero los roles inferenciales *no* son composicionales.
3. Por tanto, los significados no pueden ser roles inferenciales.

La segunda premisa, por supuesto, es la que sostiene el argumento. Pero parece obviamente correcta. Considérese el significado del sintagma ‘vaca marrón’; depende de los significados de ‘vaca’, ‘marrón’, así como su sintaxis, tal como la composicionalidad requiere. En primera aproximación, ‘marrón’ significa -o si se prefiere, connota la propiedad- MARRON, ‘vaca’ significa VACA, y la interpretación semántica de la estructura sintáctica [Adjetivo + Sustantivo]_N es la conjunción de propiedades. (Nos

² Nuestra formulación del principio del isomorfismo resulta intencionalmente vaga; importantes problemas surgen cuando se intenta una mayor precisión. Por ejemplo, supóngase que se sostiene que la oración «está lloviendo» expresa la proposición que está lloviendo aquí. Entonces uno debe decir o bien que «está lloviendo» tiene más constituyentes que los que aparecen superficialmente, o bien que el principio del isomorfismo puede ser violado por información transmitida pragmáticamente. Para nuestros propósitos no es preciso entrar en tales cuestiones.

damos cuenta que hay problemas con respecto a patos de reclamo y demás; pero la asunción de la composicionalidad del lenguaje es la asunción que tales problemas pueden resolverse.) Sin embargo, prima facie, el rol inferencial de ‘vaca marrón’ depende no sólo del rol inferencial de ‘vaca’ y del rol inferencial de ‘marrón’, sino también de lo que uno cree acerca de las vacas marrones. De este modo, a diferencia del significado, el rol inferencial no es, en general, composicional.

Supóngase, por ejemplo, que uno cree que las vacas marrones son peligrosas. Entonces parte del rol inferencial de ‘vaca marrón’ en tal dialecto consiste en que aparece, o puede aparecer, en inferencias del tipo ‘vaca marrón → peligroso’. Este hecho acerca del rol inferencial de ‘vaca marrón’, no obstante, al menos a primera vista, no parece derivarse de hechos correspondientes acerca de los roles inferenciales de sus constituyentes. Esto se ve al contrastar el presente caso con, por ejemplo, inferencias como ‘vaca marrón → animal marrón’ o ‘vaca marrón → vaca no verde’. ‘Animal marrón’ se sigue de ‘vaca marrón’ porque ‘vaca’ entraña ‘animal’; ‘vaca no verde’ se sigue de ‘vaca marrón’ porque ‘marrón’ entraña ‘no verde’. Pero no parece que ‘vaca’ o ‘marrón’ entrañen ‘peligroso’. A este respecto, pues, no parece que la inferencia de ‘vaca marrón’ a ‘peligroso’ sea composicional.

En resumen, parece que algunas, pero no todas, de las inferencias potenciales en que ‘vaca marrón’ (parte de su rol en el lenguaje o sistema de creencias), vienen determinadas por los potenciales inferenciales de ‘vaca’ y ‘marrón’ respectivamente, mientras que el resto depende de las creencias de uno acerca de las vacas marrones en el «mundo real». Esto no debería parecer sorprendente o polémico. Se trata sólo de una forma de decir que la expresión ‘las vacas marrones son peligrosas’ (a diferencia de ‘las vacas marrones son animales’ o ‘las vacas marrones no son verdes’ o ‘las vacas marrones son marrones’) es claramente sintética. Es decir, ‘las vacas marrones son peligrosas’ es *contingentemente* verdadero, verdadero en virtud de los hechos acerca de las vacas marrones, y no en virtud de hechos acerca de los significados de los términos (asumiendo que sea verdadero en primer lugar).

Pero, para repetir el argumento, si los significados son composicionales y los roles inferenciales no, entonces se sigue que los significados no pueden ser los roles inferenciales.

A nuestro parecer, el argumento es bastante robusto; en particular, no depende de supuestos de detalle sobre cómo una semántica del rol inferencial construye la noción de rol inferencial. Hay, por ejemplo, un influyente artículo de Hartry Field en el que el rol inferencial es analizado en términos de probabilidades subjetivas; en efecto, el rol inferencial del pensamiento que P se identifica con la probabilidad subjetiva que alguien asignaría a P, respecto al resto de sus posibles pensamientos (Field, 1977). Así, por ejemplo, el rol inferencial del pensamiento que está lloviendo está determinado en parte por la probabilidad subjetiva que alguien le asignaría con respecto al supuesto de que el sol brille; y en parte, por la probabilidad que le asignaría con respecto al supuesto que los elefantes tengan alas, etc. Como otras versiones de la semántica «nuevo testamento», este tratamiento se aplica, de forma clara, tanto a una teoría que asigna roles inferenciales a pensamientos como a una que los asigna a expresiones lingüísticas, o a ambos.

Lo que afirmamos es que la construcción de los roles inferenciales en términos de probabilidades subjetivas, cualesquiera que sean sus virtudes, no ayuda en el caso de

la composicionalidad. Esto se debe a que *las probabilidades subjetivas no son a su vez composicionales*. Por ejemplo, la probabilidad subjetiva que se asigna al pensamiento que *las vacas marrones son peligrosas* no está en función de la probabilidad subjetiva que se asigna al pensamiento que *las vacas son peligrosas*, junto con las probabilidad subjetiva que se asigna al pensamiento que *las cosas marrones son peligrosas*. Si esto no parece obvio, considérese un mundo (o mejor, un mundo creencial, dado que las probabilidades en disputa se supone que son subjetivas) donde hay muchas cosas que son vacas, casi ninguna que sea peligrosa, y un pequeñísimo número de vacas marrones, la mayor parte de las cuales son muy, muy fieras. Sobre tales supuestos, la probabilidad que algo que es marrón sea peligroso es pequeña, y la probabilidad de que algo que es una vaca sea peligroso es pequeña, pero la probabilidad de que una vaca marrón sea peligrosa es tan grande como se desee.

Hasta ahora el argumento puede entenderse como el cuerno de un dilema, y anticipamos la siguiente respuesta: «De acuerdo, si se asume la composicionalidad del significado, entonces los significados no pueden identificarse con el rol inferencial como tal. Pero eso no es realmente embarazoso para la semántica tipo «nuevo testamento» porque los significados pueden aun identificarse con los roles en las inferencias *analíticas*. De este modo, por una parte, la inferencia ‘vaca marrón → animal marrón’ es composicional (se sigue de la inferencia ‘vaca → animal’), y por otra parte, precisamente porque es composicional, ‘vaca marrón → animal marrón’ es analítico. Las inferencias composicionales serán siempre analíticas y las inferencias analíticas serán siempre composicionales; la composicionalidad de una inferencia y su analiticidad *son la misma cosa*.

Se puede mirar de esta forma: si ‘vaca marrón → animal marrón’ es composicional, entonces está garantizada por los roles inferenciales de las expresiones ‘vaca’ y ‘marrón’. En esto consiste la composicionalidad de una inferencia. Pero, de acuerdo con la semántica «nuevo testamento», los roles inferenciales de ‘vaca’ y ‘marrón’ *son sus significados*. Por consiguiente, la inferencia ‘vaca marrón → animal marrón’ está garantizada en virtud de los *significados* de ‘vaca’ y ‘marrón’. Pero una inferencia es analítica *justamente* cuando su validez depende de los significados de sus expresiones constituyentes. En consecuencia, la composicionalidad de ‘vaca marrón → animal marrón’ -o, mutatis mutandis, de cualquier otra inferencia-, entraña su analiticidad. El mismo argumento funciona en el otro sentido. Una inferencia es analítica si depende de los significados de sus constituyentes. Pero, de acuerdo con la semántica «nuevo testamento», los significados son los roles inferenciales. Por tanto, una inferencia analítica resulta ser una inferencia cuya validez depende de los roles inferenciales de sus constituyentes. Y que la validez de una inferencia dependa de los roles inferenciales de sus constituyentes es lo que hace composicional una inferencia. Por tanto, la analiticidad implica la composicionalidad y viceversa. Y por tanto, el significado es composicional, el rol inferencial no, pero el rol en las inferencias analíticas sí es composicional. Lo que se deduce de todo ello es que hace falta una versión revisada de la semántica «nuevo testamento», que identifique el significado con el rol en las inferencias analíticas.»

Lo primero que hay que decir al respecto de esta nueva sugerencia es que la amenaza de circularidad es ahora aparente. Lo que se propone es reconciliar la teoría del

rol inferencial con la composicionalidad del significado mediante la identificación del significado de una expresión con su rol en las inferencias analíticas. Pero la diferencia entre las inferencias analíticas y las inferencias *tout court* es sólo que la validez de las primeras depende de los *significados* de sus expresiones contituyentes. Como resultado, la analiticidad, el significado y la composicionalidad se las arreglan para ir tirando apoyándose mutuamente, y Quine tiene que decir «Ya os lo dije».

Nótese también que esta propuesta pone en peligro el proyecto de naturalización de la semántica del rol inferencial. Gran parte del atractivo de identificar el significado con el rol inferencial consiste en la idea que el rol inferencial de una expresión podría a su vez identificarse con el rol *causal*, suministrando de esta forma una vía naturalista concebible de solución al problema de Brentano. Que las relaciones causales reconstruyen las relaciones inferenciales es uno de los supuestos básicos de las teorías computacionales de los procesos mentales, de manera que quizás se halla aquí el germen de esperanza de una posible unificación de la semántica con la psicología. Pero dejando de lado las propuestas de una teoría causal de la analiticidad, esta táctica no está a disposición del filósofo que identifica el significado con el rol de una expresión de inferencias analíticas.³ La idea de que los procesos mentales son computacionales puede proporcionar la base para una explicación naturalista de la inferencia, pero no ofrece luz alguna respecto a la naturaleza de la analiticidad. Como tampoco lo hace, a nuestro parecer, nada más.

Podemos decir ahora con precisión cuál es nuestro problema: no se puede identificar los significados con los roles inferenciales *tout court*, porque, a diferencia de los significados, los roles inferenciales no son composicionales. *Se puede* identificar los significados con los roles en las inferencias analíticas, no obstante, porque las inferencias analíticas son composicionales. Pero, por supuesto, el coste de identificar los significados con los roles en las inferencias analíticas es el de apelar a la distinción analítico/sintético. Así, el coste de la semántica «nuevo testamento» es el de apelar a la distinción analítico/sintético. Pero en estos días prácticamente todo el mundo cree que la distinción analítico/sintético no se sostiene. Además, se cree que el descubrimiento de que la distinción analítico/sintético carece de fundamento es uno de los dos logros más importantes de la moderna filosofía del lenguaje -el otro sería la teoría del significado como rol inferencial-. Si seguimos asumiendo la composicionalidad como no negociable, se sigue que uno de estos principios fundacionales de la filosofía del lenguaje va a tener que ser abandonado. ¿Cuál? ¿Y con qué vamos a reemplazarlo?

³ Hay diversas teorías causales del significado; por ejemplo, Skinner (1957), Dretske (1981), y Fodor (1991), entre otras muchas. Y como cualquier teoría del significado, cada una de ellas implica una correspondiente noción de analiticidad. Pero todas estas teorías son externalistas y atomistas y por tanto no ofrecen consuelo alguno a las semánticas del rol conceptual o al holismo. Tales teorías causales no legitiman una construcción del significado en términos de inferencia *analítica* porque rechazan la teoría del rol conceptual: no identifican el significado con la inferencia en absoluto.

3. ¿Y ahora qué?

Ahora es cuando deberíamos decir qué hacer para arreglar la grieta en los cimientos. Prometemos que lo haríamos si pudiéramos. Dado que no podemos, nos limitaremos a unas pocas observaciones acerca de las que consideramos opciones abiertas, y de lo que puede suponer adoptar una y otra.

La primera posibilidad a considerar es el intentar resucitar la distinción analítico/sintético. Pensamos que esta sugerencia tiene cierta plausibilidad *prima facie*. Hemos visto que la composicionalidad y la analiticidad vienen a ser lo mismo si se acepta el enfoque «nuevo testamento» según el cual el significado sobreviene al rol conceptual. Pero, de hecho, es plausible pensar que la composicionalidad entraña la analiticidad *tanto si se acepta la semántica «nuevo testamento» como si no*. Así, por ejemplo, es difícil ver cómo alguien podría afirmar que el significado de ‘vaca marrón’ es *composicional* y a la vez negar que la inferencia de ‘vaca marrón’ a ‘marrón’ es *analíticamente válida*. Si es innegable que el significado de ‘vaca marrón’ resulta de los significados de ‘vaca’ y de ‘marrón’. Parece igualmente innegable, y por las mismas razones, que la inferencia de ‘vaca marrón’ a ‘marrón’ está justificada por los principios lingüísticos que permiten esa construcción. Pero una inferencia cuya validez depende de principios lingüísticos *es precisamente* una inferencia analítica. En resumen; las mismas relaciones estructurales entre los constituyentes de una oración que garantizan su composicionalidad parecen engendrar la analiticidad de algunas de las inferencias de las que forman parte sus constituyentes.

De esta forma, parece que la composicionalidad subraya ciertas analiticidades por sí misma, sin necesidad de recurrir al principio de que el significado sobreviene al rol inferencial. Un semántico tipo «nuevo testamento» podría argüir, por tanto, que si la composicionalidad no es negociable entonces tampoco puede serlo la analiticidad, y, en consecuencia, que el rechazo de los argumentos de Quine contra la analiticidad está independientemente motivado. (Vale la pena mencionar a este respecto que Quine dice poco o nada al respecto de ejemplos como ‘vaca marrón → marrón’, donde la analiticidad de una inferencia depende de la *estructura* de sus premisas más que de sus contenidos léxicos.) Y si puede defenderse la recuperación de la distinción analítico/sintético, entonces parece que todo esté bien. Hemos visto que no se puede sostener al mismo tiempo la semántica «nuevo testamento», la composicionalidad y *el rechazo de la distinción a/s*, pero ciertamente se pueden sostener los dos primeros si se abandona el tercero.

Queremos enfatizar, no obstante, que el tipo de distinción *a/s* que la composicionalidad entraña se da sólo entre las expresiones y sus *constituyentes sintácticos*. Sirve para distinguir, digamos, ‘vaca marrón → marrón’ de ‘vaca marrón → peligroso’. Pero no permite distinguir entre, digamos, ‘vaca marrón → animal’ y ‘vaca marrón → peligroso’. Es decir, no permite una distinción *a/s* entre inferencias que dependen del repertorio léxico de las premisas, a diferencia de las que dependen de su estructura lingüística. Sin embargo, son justamente tales inferencias léxicas las que resultan problemáticas para la semántica «nuevo testamento».

Si el significado de ‘vaca marrón’ deriva de los significados de ‘vaca’ y ‘marrón’ (como se desprende de la composicionalidad), y si el significado de ‘vaca’ es

su rol inferencial (si la semántica «nuevo testamento» está en lo cierto), entonces es preciso arreglar las cosas de modo que ‘vaca marrón’ permita la inferencia a ‘animal’, *pero no la inferencia a ‘peligroso’*, debido a los significados de sus constituyentes. Pero eso requiere excluir de la representación semántica de ‘vaca’ información del tipo, por ejemplo, ‘vaca → tipo de x tal que los x marrones son peligrosos’ (y, análogamente, debemos excluir de la representación semántica de ‘marrón’ información del tipo ‘marrón → x tal que las vacas x son peligrosas’).⁴ Pero decir que debemos tratar tales inferencias como excluidas de las representaciones semánticas de ‘vaca’ y ‘marrón’ es decir que tales inferencias *no* son constitutivas de los significados de ‘vaca’ y de ‘marrón’. Y decir eso es presuponer una distinción *a/s* para las inferencias de naturaleza léxica. Precisamente el tipo de distinción *a/s* cuya crítica familiar por parte de Quine es ampliamente aceptada.

Hemos estado acentuando la diferencia entre las analiticidades que resultan de la estructura composicional de una expresión y las analiticidades que se generan por los significados de los ítems de su vocabulario. Esta distinción es, por supuesto, puesta en cuestión si se asume que hay un nivel de representación en el que los ítems léxicos son semánticamente descompuestos: postular tal nivel equivale a afirmar que el significado léxico es a su vez composicional. Por tanto, si hay descomposición léxica, entonces podemos identificar el significado de una expresión con su rol en aquellas inferencias que son determinadas por su estructura composicional, *incluyendo la estructura composicional de sus componentes léxicos*. El resultado es asimilar inferencias del tipo ‘vaca marrón → animal’ a las del tipo ‘vaca marrón → marrón’, dado que, al nivel de la representación semántica, ambas suponen relaciones entre expresiones y sus constituyentes. Y ambas se diferencian de ‘vaca marrón → peligroso’ porque presumiblemente ‘peligroso no es un constituyente de las representaciones de ‘vaca marrón’ a ningún nivel.

Por supuesto, resolver estos problemas dando por descontada la posibilidad de descomposición léxica no respondería las objeciones de Quine. Si el significado léxico fuera composicional, entonces los ítems léxicos serían definibles; y parte del argumento de Quine en contra de la posibilidad de la distinción *a/s* consiste en la falta de una noción de definición válida. Si los argumentos de Quine demuestran algo, es que no hay forma de reconstruir la intuición de que ‘vaca marrón → animal’ es por definición, mientras que ‘vaca marrón → peligroso’ no lo es.

Concluimos, pues, afirmando que aunque puede haber razones para resucitar la distinción *a/s*, la no negociabilidad de la composicionalidad no es una de ellas. La

⁴ Jim Higginbotham nos ha señalado que excluir *peligroso* de las representaciones semánticas de ‘vaca’ y ‘marrón’ no basta para asegurar que ‘las vacas marrones son peligrosas’ sea sintética. Esto se debe a que la composicionalidad requiere sólo que haya una función que determine el significado de ‘vaca marrón’ en base a la sintaxis y a los componentes léxicos de la expresión. Hablando de forma estricta, no requiere que la representación semántica de ‘vaca marrón’ incluya las representaciones semánticas de ‘vaca’ y ‘marrón’ (aunque muchos tratamientos de la analiticidad en la tradición kantiana han hecho tal suposición). Todo esto nos parece bien; lo que afirmamos es que si ‘vaca marrón’ hereda *peligroso* de su estructura composicional (en la forma que sea), entonces ‘las vacas marrones son peligrosas’ resulta analítico. Lo que constituye un resultado inaceptable.

composicionalidad permite una distinción entre ‘vaca marrón → marrón’ y ‘vaca marrón → animal’, pero no entre ‘vaca marrón → peligroso’ y ‘vaca marrón → animal’. Por consiguiente, el problema original sigue en pie: si la composicionalidad no es negociable, entonces o bien hay una distinción a/s para las inferencias léxicas (contra Quine), o bien los significados no sobrevienen a los roles inferenciales (contra la semántica «nuevo testamento»).

Una segunda reflexión es que la situación presente está llena de ironías. Observamos anteriormente que una vez se dice que el significado de una expresión sobreviene a su rol inferencial, es difícil evitar la conclusión que el significado de una expresión sobreviene a la *totalidad* de su rol inferencial. Esto es, es difícil evitar el relativizar el significado de una expresión al lenguaje que la contiene, en su conjunto, con las consecuencias de que las expresiones de lenguajes diferentes son semánticamente inconmensurables. Existe un tipo de idealismo lingüístico que se deleita en estas implicaciones holistas de la semántica «nuevo testamento»; se puede reconocer sin dificultad esta sensibilidad entre las gentes de la ‘lingüística cognitiva’, los conexionistas, y, por supuesto, entre filósofos como Rorty, Putnam, Kuhn y Derrida. Sin embargo, como hemos visto, la semántica del rol inferencial *no tiene* implicaciones holistas después de todo, y por tanto el relativismo sale derrotado cualquiera que sea el rumbo que tome el argumento.

Así es como se suponía que se organizaba el argumento que concluía con el holismo semántico a partir de la semántica del rol inferencial:

1. La Semántica «Nuevo Testamento»: el significado de una expresión se constituye, al menos parcialmente, por las relaciones inferenciales de la expresión.
2. Ausencia de la distinción a/s: carecemos de una distinción de principio entre las inferencias que constituyen el significado de una expresión y las inferencias que no.
3. Conclusión, el Holismo Semántico: el significado de una expresión se constituye por la totalidad de sus relaciones inferenciales, esto es, por su rol total en el lenguaje.

Pensamos que hay multitud de razones para desaprobar esta forma de defender el holismo semántico. No es la menor de ellas el que depende, de forma aparentemente ineliminable, de un argumento de «pendiente resbaladiza», un tipo notorio de argumento falaz por conducir de premisas verdaderas a conclusiones falsas. (Esto es, depende de argüir de ‘no hay una diferencia de principio entre Fs que son Gs y Fs que no lo son’, a ‘o bien ningún F es G o bien todos lo son’). Para nuestros propósitos, no obstante, podemos pasar esto por alto. Lo que queremos destacar es que el argumento no puede ser mejor de lo que lo es su segunda premisa, esto es, *depende de* la negación explícita de la distinción a/s.

De este modo la situación es la siguiente: si la distinción a/s tiene razón de ser, entonces la segunda premisa es falsa y el argumento en favor del holismo semántico incorrecto. Pero si la distinción a/s no se sostiene, entonces la primera premisa es falsa. Esto es así porque, como hemos visto, el respetar la composicionalidad requiere que el significado de una expresión se identifique con su rol en las inferencias *analíticas*; y la composicionalidad no es negociable. Pero el argumento de que no hay distinción a/s es el argumento que niega la existencia de inferencias analíticas. Asumimos la bivalencia:

o hay una distinción a/s o no la hay. De cualquier forma, el argumento que concluye el holismo semántico en base a la semántica del rol inferencial resulta incorrecto. Suponemos que esta demostración debería oscurecer los cielos de los semánticos holistas de la Costa Oeste. Que así sea: necesitan la lluvia.

Para recapitular, el problema fundacional es que no se puede tener todo a la vez: defender una semántica del rol inferencial Y defender la composicionalidad Y abandonar la distinción a/s. Si la composicionalidad no es negociable, nos quedan sólo dos opciones: defender la distinción a/s, o rechazar la idea de que el significado sobreviene al rol conceptual. No podemos decir cuál de estas opciones es la mejor, pero queremos insistir en que el agitarse no nos va a librar del anzuelo. En particular, muchos científicos cognitivos han esperado reconciliarse con los argumentos de Quine optando por una noción de analiticidad gradual, contextual, o desnaturalizada de otro modo. La idea consiste en que, aunque Quine puede haber mostrado que la noción de *identidad* de significado es problemática, queda abierta la posibilidad de desarrollar una noción de *similitud* de significado. Se supone que una distinción a/s gradual nos permitiría tener tanto la composicionalidad como la sobrevenida del significado al rol inferencial. La semántica «nuevo testamento» resultaría así vindicada, aunque en una versión menos afilada.

Nadie ha mostrado que esto es inviable; pero tampoco nadie tiene la menor idea de cómo llevarlo a cabo. Nos parece que los argumentos en contra de una noción de principio de identidad del significado funcionan igualmente bien en contra de una noción de principio de similitud de significado, pero no lo argüiremos aquí.⁵ Baste notar aquí que las conexiones entre la analiticidad y la composicionalidad que hemos estado examinando convierten en poco prometedora la búsqueda de una noción gradualista de la analiticidad. La composicionalidad es, después de todo, un principio que rige las relaciones *entre significados* (rige la relación entre el significado de una expresión compleja y el significado de sus constituyentes). Por tanto, si una teoría semántica reconstruye el significado por medio de la noción de inferencia analítica, y se dispone solamente de una noción gradual de analiticidad, *entonces se tendrá que disponer también de una noción gradual de composicionalidad*. Pero, ¿cómo sería una noción gradual de composicionalidad? En particular, ¿cómo podría esa noción desempeñar la función que se requiere de la composicionalidad, esto es, dar cuenta de la sistematicidad, el isomorfismo y la productividad?

¿Una noción gradual de composicionalidad no implicaría, todo lo más, que un conocimiento finito del lenguaje es adecuado para *casi-entender* expresiones previamente no encontradas? ¿O que si un lenguaje es capaz de expresar la proposición que aRb, entonces es *casi-capaz* de expresar la proposición bRa? ¿O que si la oración S expresa la proposición P entonces los constituyentes de S *casi-expresan* los constituyentes de P? ¿Pero qué sentido tiene afirmar que, por ejemplo, '[«Juan quiere a María» casi-expresa la proposición que Juan quiere a María] sólo si «Juan» casi-denota a Juan'? No parece recomendable tratar de vadear estas pantanosas aguas.

⁵ Cf. los capítulos 1 y 7 de Fodor & Lepore (en prensa).

Aquí, pues, es adónde hemos llegado: la composicionalidad no es negociable; los roles inferenciales son composicionales sólo si se cuenta con una distinción a/s válida para las inferencias que dependen del repertorio léxico. Por tanto, si Quine tiene razón acerca de la distinción a/s, la única alternativa que queda es abandonar la idea de la semántica «nuevo testamento» de que el significado de los ítems léxicos resulta de sus roles inferenciales. Dejando aparte las propuestas en favor de un tipo completamente distinto de semántica, lo que la situación requiere es regresar al enfoque tipo «antiguo testamento» según el cual lo que hace que ‘perro’ signifique perro es algún tipo de conexión entre el símbolo y el mundo; quizás algún tipo de conexión causal, o informacional, o nomológica, entre instancias del tipo de expresión e instancias del tipo de animal. Como hemos señalado anteriormente, existen diversas propuestas de este estilo y, quién sabe, quizá una de ellas puede funcionar.

De todo ello resultan algunas consecuencias de interés para la epistemología, dado que todos los candidatos plausibles para las relaciones símbolos-mundo semánticamente relevantes parecen ser *atomistas*. Si ‘perro’ significa *perro*, por ejemplo, porque los perros causan emisiones de ‘perro’, entonces parece que un lenguaje pudiera tener una palabra que significara lo que ‘perro’ significa en español sin que hiciera falta que tuviera ninguna otra palabra con el mismo significado que las palabras del español. A primera vista, de hecho, parece posible que un lenguaje pudiera tener una palabra con el mismo significado que ‘perro’ sin que tuviera *ninguna otra* palabra. Los argumentos holistas de todo tipo encaminados a demostrar la imposibilidad de traducción (y de forma general, la imposibilidad de conmensurar lenguajes e ideologías), entonces, quedarían probablemente colapsados. Más nubarrones sobre la Costa Oeste.

No estamos, por supuesto, urgiendo a que todo el mundo siga este curso. Por el contrario, muchos lingüistas tendrán buenas razones para no hacerlo. Por ejemplo, si la semántica «antiguo testamento» es esencialmente correcta, ¿qué resulta de estudios lingüísticos del tipo de la ‘semántica léxica’? No se pueden estudiar las relaciones semánticas entre ítems léxicos a menos que *se den* relaciones semánticas entre los ítems léxicos; la semántica «antiguo testamento» debe probar que no se dan. La semántica «antiguo testamento» sostiene que las relaciones semánticas se dan entre los ítems léxicos y *el mundo* y *sólo* entre los ítems léxicos y el mundo.⁶

La semántica «antiguo testamento», por tanto, no tiene tratos con los campos semánticos, descomposiciones léxicas, redes conceptuales y demás. Por supuesto, si lo que se quiere decir mediante «el nivel semántico de descripción lingüística» es el nivel de descripción al que las oraciones que sólo difieren por sus expresiones sinónimas son representadas idénticamente, entonces una forma natural de entender lo que la semántica

⁶ De forma más precisa, lo que la semántica del «antiguo testamento» afirma es que si hay relaciones semánticas entre los ítems léxicos, entonces no se derivan de relaciones entre sus roles en el lenguaje, sino de relaciones entre sus conexiones con el mundo. *Ceteris paribus*, dos expresiones conectadas con el mundo de la misma forma serán sinónimos. Sin embargo, lo que se supone que el nivel semántico de descripción léxica representa son relaciones inferenciales entre ítems léxicos, y no sus relaciones con el mundo.

«antiguo testamento» entraña es que rechaza la existencia de tal nivel de descripción lingüística. En consecuencia, la semántica generativa no tiene cabida, la semántica interpretativa tampoco tiene cabida, y el nivel superior de descripción lingüística es el sintáctico o la forma lógica, esto es, un nivel que preserva el repertorio no-lógico-sintáctico de superficie. El Dios del Antiguo Testamento es un Dios austero.

Como se ha observado anteriormente, nuestro propósito no es convertir a nadie a tan rigurosa teología. Se trata sólo de señalar que parece existir esta grieta en los cimientos de la estructura donde filósofos y lingüistas han estado cohabitando recientemente. Y de incitar a que alguien haga algo antes de que caiga el tejado.

(Trad.: A. Gomila)

Bibliografía

- Dretske, F. (1981): *Knowledge and the Flow of Information* , MIT.
Field, H. (1977): «Logic, Meaning, and Conceptual Role», *Journal of Philosophy* 74: 379-409.
Fodor, J. (1991): *A Theory of Content* , MIT.
Fodor, J. & Lepore, E. (en prensa): *Holism, A Shopper's Guide* , B. Blackwell.
Skinner, B.F. (1957): *Verbal Behavior*, Appleton Century Cross Inc.

Taula (UIB) núm. 15, 1992

Contents and contexts: on Fodor's narrow/broad divide

Manuel Liz

Universidad de La Laguna

Abstract

Folkpsychological knowledge is very ignorant of internal facts of the body. Folkpsychology is really very superficial and very behaviouristic. On the other hand, what happens inside our bodies must be very important and relevant for a scientific psychology. According to Fodor, only mental properties identified by means of causal powers that locally supervene on the body (narrow contents) are the business of a scientific psychology. In this paper, and through the notion of «canonical contexts of attribution», I explore a possible strategy for having also the local supervenience of some folk-contents (broad contents) and to develop causal theories about them.

Resumen

La atribución de actitudes proposicionales se realiza con un completo desconocimiento de cuáles puedan ser los hechos internos que ocurran en nuestros cuerpos. Pero son justamente estos hechos los que deben importar en una psicología científica. Según Fodor, la psicología debe individualizar propiedades mentales sólo en relación a propiedades causales que puedan sobrevenir localmente sobre los cuerpos de los sujetos. En este trabajo y a través de la noción de «contextos canónicos de atribución», se examina una posible estrategia para conseguir que las actitudes proposicionales atribuidas puedan también ser localmente sobrevenientes y tener cierto tipo de eficacia causal.

1. Two points of view: folkpsychology and knowledge of the body

Folkpsychology is the ordinary knowledge we can have of the mental life of whatever psychological subject. Now, what is a psychological subject? Perhaps the best answer would be that a psychological subject is whatever entity that could be known and which behaviour could be predicted and explained by means of folkpsychology. And such a characterization would not be circular or defective because psychological subjects can exist even though we may not have knowledge of their existence. The possibility not the actuality of folkpsychological knowledge is the key characteristic that permits the reality of a psychological subject. This is the reason why sometimes the real existence of a psychological subject —our real existence as psychological subjects for instance— appears to be only a business of counterfactual considerations. People —some animals and perhaps some artifacts too— think, desire, hope, hate and so on, and they act according to the things they think, desire, hope, hate and so on. Folkpsychology is the intuitive and mentalistic psychology of common sense. Its basic rules are well known for almost all people. For example, if someone believes that if A then B, and desires B, then —»*ceteris paribus*»— he/she will try to make A become true. Folkpsychology appeals to a mental life full of beliefs, desires and so forth. Folkpsychology is quite universal, it starts very early in our lives and, for the moment, it has no relevant rivals. Yet, in spite of such pervasiveness, folkpsychological knowledge is very ignorant of the body. We usually describe the mental life of people —roughly, that they think and desire and what they think and desire— with expressions that ignore any consideration to what is happening inside their bodies.

Let's refer to the context as all those facts from which we obtain and justify the folkpsychological descriptions of the mental life of a subject according to a content. In general, the *context of a thing X* will be made up of all the facts of the world —past, present future— but will actually exclude facts —past, present or future— inside X.

Usually we don't need go inside the skin of anyone to know a lot about his/her folkpsychology. People, you yourself, could be made of silicon and cheese, for instance, and people could have the oddest internal structures you can imagine, but from the point of view of folkpsychology nothing of this would matter. The things that matter are in the context. Here, the skin of the body, in a metaphorical sense that includes movements of the body and other such things, belongs to the context as one of its limits. The context starts, so to speak, with the skin and includes or can include all the rest of the world. From the folk perspective all we need —in order to obtain and justify folkpsychological descriptions of the mental life of a psychological subject— is a context. Folkpsychology is really very superficial and very behaviouristic.

On the other hand, from a scientific viewpoint what happens inside our body is very important and relevant and requires consideration if we want to explain and make scientifically clear our psychology. Biology and medical sciences stress the idea that all we are is unavoidably related to our bodies. From such a perspective, facts about the structure and composition of our bodies are crucial for scientifically knowing our psychology.

Let's refer to internal facts as all those facts that belong to our internal functional structure or to our internal stuff. In general, the *internal facts of a thing X* will be made up of all the facts —past, present or future— inside X.

The points of view of folkpsychology and of the knowledge of the body are opposite in at least four respects: 1) in respect to criteria of identification of mental properties, 2) in respect to ways in which mental properties can supervene on another non-mental properties, 3) in respect to how mental causation can be possible, and, 4) in respect to ways of elaborating sound psychological theories.

1) Identification of mental properties

Folkpsychological criteria of identification of the mental have two important characteristics: 1) they are fuzzy and prototypical, that is, they don't offer necessary nor sufficient conditions for the identification of something having a mental property, and, 2) mental properties are individuated, as we have noted, without any concern to internal facts. Let's analyze more the second characteristic.¹ A lot of work about how we attribute mental properties according to a semantical content suggests that, here, mental properties count as only attributed properties in the following sense. A property will be an only attributed property to a thing X—an OA property— iff the assertion that X has that property is completely justified by the knowledge of some peculiar facts of the context of X. If mental properties are OA properties, the only relevant facts in the world in order to make folkpsychological assertions about a subject X will also be facts that belong to the context of X. And the only source of justification for the attributions of these OA properties will be also the context. On the other hand, if the justification of the attribution of a property requires the knowledge of some internal facts, the property will be a not-only attributed one. So, a property will be a *not-only attributed property to a thing X*—an NOA property— iff the assertion that X has that property is justified only by the knowledge of some internal facts of X. The source of justification for the attribution of NOA properties to a thing is some kind of knowledge of its internal facts.

Let's suppose, for example, the property P. «Px» is true iff Px. So far, so good. Now, if the knowledge of some facts of the context of x is enough to justify the assertion Px, P will be an OA property. If, on the other hand, it is necessary to know some internal facts of x in order to justify the assertion Px, P will be a NOA property. «To be number one in a race», «to have a lot of homework», «to look happy» and so on, could be examples of OA properties. Properties like «to be green» or «to have a certain shape» are also OA properties that refer to the particular appearance, to the «skin», of the objects. We are justified in their attribution if we know a lot of facts about how we are looking at the objects, facts about our eyes, the light, the perspective and so on. And we don't need to know anything about the internal facts of those things with those OA properties. By contrast, «To be water», «to be a lion», «to be made of cheese», «to have five components», etc, are examples of NOA properties. In order to attribute these properties with enough justification, we need the knowledge of some internal facts of the things that have them. This is the point of view of the knowledge of the body with respect to mental properties. Its criteria of identification of the mental intend to offer, in the ideal case, necessary and sufficient conditions. They intend also to individuate

¹ Let's remember, for instance, the works of Burge, Davidson, dennett, Putnam and Stich about this topic.

mental properties according to internal facts of the bodies. Hence, mental properties must count as NOA properties with characteristic causal powers associated mainly via our nervous systems. These NOA properties could have, not only an intrinsic character, but also a relational one. If this is so, then there would have to exist some causal laws reflecting this relational character. Here, the distinction between relational and intrinsic—not relational—properties is, in a logical and metaphysical sense, irrelevant. It has only an heuristic and pragmatic interest. The important thing is the notion of NOA properties associated with genuine causal powers. These causal powers must be capable of being represented with certain, monadic or polyadic, predicates in a way that can support genuine and lawful causal connections. These causal connections are considered the only useful and definitive criteria of identification of the mental.

2) *Supervenience*

What about supervenience? Roughly, let's say that a set of properties *supervenes on* another set of properties if any distinction that can be put with regard to the having of the former properties entails some difference with regard to the having of the latter ones, the properties of the so called basis of the supervenience relation. If the properties of this basis are internal properties of a thing, we'll say that the supervenience is *local*. Now, if mental properties are OA properties, as they are from the folkpsychological viewpoint, it seems unlikely that they can locally supervene on the internal facts of the subjects. If, on the other hand, mental properties are NOA properties, their local supervenience on more basic properties of the subjects appears much more likely.

3) *Mental causation*

Mental causation must be covered by causal laws. This is the only—not metaphorical—sense in which we can understand mental causation as a part of the order of the world. But, how could an OA property be the antecedent or consequent part of the description of a genuine causal law? Local supervenience and mental causation seem to need NOA properties. Folkpsychology poses a very hard problem here. Really, we don't know how to identify mental properties without the folk criteria of identification. If we leave the knowledge of the body alone, we don't have any psychological knowledge at all. But, mental properties only are capable of local supervenience and they only can have causal efficiency through the internal facts of the subject that has them. It cannot be otherwise unless to have mental properties is a mere truism or a shorter way of speaking of another thing, of another thing that belongs to the context.

4) *Psychological theories*

Finally, in order to elaborate psychological theories, the adoption of the folkpsychological perspective seems to entail our looking for them inside sociolinguistic theories of attribution or something like that. It is supposed that the psychological properties described by these theories can't locally supervene on internal properties of the psychological subjects. They supervene on other properties of the world. And, perhaps, mental causation could be constructed with these bricks. But, the important

point is that all this leads to a very odd project. To adopt this sort of perspective means to give up a lot of plausible intuitions about the supervenience of the mental, about mental causation and about what must be relevant for a scientific psychological theory in which the body of the psychological subject plays a central role.

At a first look and through these four areas, the opposition between the point of view of folkpsychology and the point of view of the knowledge of the body is clear. Of course, you can say that this is not at all true. As far as folkpsychology accepts some general constrictions about what sort of bodies can have mental life, it is not so indifferent to internal facts of the body. Nowadays, this might be the case for our folkpsychology. If the old folkpsychology was linked to magic and religion in respect to the internal facts of our bodies, our folkpsychology is linked to science in a lot of folk ways. But, folkpsychology can work very well without these commitments. Moreover, it must do it in most of its working situations. This is an important point that we must not forget. We want to establish some links between the point of view of folkpsychology and the point of view of the knowledge of the body. But, the links set up by means of constrictions on the folk attributions are only contingent and accidental. They exist, but they are an irrelevant part of the folk. The scientifically important links between folkpsychology and the knowledge of the body must be found in another place.

2. Fodor: methodological individualism and the narrow/broad divide

Which is Fodor's view about these problems?. According to Fodor,² *individualism* is a general methodological position in science. Having a property is something individuated with respect to its causal powers. Only properties individuated this way count as scientifically relevant properties. Most of scientific taxonomies respect individualism. Psychology is individualistic too. Psychological states must be individuated with respect to their causal powers. Identity of psychological states must entail identity of causal consequences across contexts. Individualism doesn't prohibit the fact that some properties can be relational. But, if they are, then there must be some causal laws reflecting this.

Twin's stories —and related cases— try to show that some psychological properties are relative to the context and that they can't locally supervene on the actual body of the subject. Fodor considers this unproblematic. Twin's stories only show that the semantical evaluation is relative to a context. This relativity doesn't support lawful generalizations across contexts. Therefore, semantical evaluation of mental states is relative to each context, but it has no psychological relevance for elaborating respectable psychological theories.

Semantical properties are like the property of being an *H-(T-)particle*. Let's say, with Fodor, that whatever particle in the universe is an H-(T-)particle at time *t* iff at *t* my dime is heads-up. And that whatever particle in the universe is at time *t* a T-

² Mainly, FODOR (1980 and 1987).

particle iff at t my dime is tails-up. Whether something is an H or an T particle is completely irrelevant to its causal powers. Similarly for the property of being a mental state of a person who lives in a world where there is XYZ rather than H₂O in the Twin cases.

Local supervenience of the mental on the nervous system of the subject must be preserved. So if we don't have *mechanisms* that link our semantical evaluations of the mental with the causal powers of the nervous systems of the subjects, semantical evaluations cannot count as relevant indicators of psychological properties.

Methodological individualism, local supervenience of mental properties on neural properties —perhaps functional ones— and the lack of such a mechanism lead to *methodological solipsism*: the doctrine that psychological properties must be individuated without respect to their semantical evaluation.

Psychological states so individuated have *narrow contents*. Narrow contents are complex functions from thoughts and contexts to semantical properties. These kinds of functions are instantiated in our nervous systems. The particular values of these functions in each context constitute the *broad contents* of our folkpsychological speech in relation to that context. How could we pick out narrow contents? Well, Fodor's answer involves procedures of mention and counterfactual use of our broad contents, but narrow contents are strictly unexpressible. Their direct expression would require not only a mention but also a use of our broad contents, and broad contents aren't the contents a scientific psychology demands.

I have always had serious doubts about this notion of narrow content. One of my main problems might be put roughly as follows. Narrow contents are, we have said, complex functions between thoughts, contexts and semantical values. In some way, these functions are instantiated in our bodies. The *only guide* we have with which to posit and differentiate narrow contents is this: in whatever context, narrow contents related with a thought must yield the semantical values we would ascribe to that thought in that context. But, this is not very much. *A lot of different functional and causal networks instantiated in our nervous systems could do the work*. We have at hand a lot of these plausible functional and causal networks for each thought instantiated in our body. If we think about other sorts of bodies, the problem becomes much more perplexing: too many networks and too many kinds of networks. They may differ only in very small details, but how to discriminate the importance or not of these details? How to select the adequate, functional³ and causal, networks? *How to cut* into the adequate portions the functional and causal cake (in portions that we can scientifically eat)? The problem is that if you come back to your «only guide», and you claim that all these networks and kinds of networks constitute an unique narrow content, then *to be this specific narrow content instead of some other becomes a property only identified by means of the identification of OA properties*. But, we must, I am sure, make the body matter more.

³ Here, you can remember the arguments Putnam (1988) uses about the logical proliferation of functional explanations. Unless we adopt a perspective similar to the one I will sketch, folk-identified mental properties will be properties so open with respect to their functional specification as functional properties are with respect to their physical and causal implementation.

3) Change of perspective: canonical contexts of attribution and only-attributed but locally supervenient and causally efficacious properties

Now, I want to explore a possible way out of this problem. *What we have always desired to have* is the local supervenience of the mental properties identified according to our folk criteria and to develop causal theories about these mental properties. This is, I believe, a warranted wish.

All the so-called *secondary properties* of things (colours, sounds, etc.) share with our folk-identified mental properties the following characteristics: 1) they are OA properties in the sense defined, 2) their supervenience is generally supposed non local, and, 3) they don't belong to our deep microphysical theories of the world. But, this is not all the story. Although the first characteristic remains unchanged, we have *controlled situations* in which secondary properties are locally supervenient and belong to our, not-so-deep, theories of the world. Very often we consider colours as locally supervenient properties and, in this way, we use colour terms in our biological and chemical theories. Whether colours are or not metaphysically dispensable, in a deep scientific image of the world, is a very interesting philosophical problem. But, the fact is they are not dispensable in the ordinary scientific practice. This seems crucial to psychological theory. A psychological theory about the mental is not a philosophical one.

According to Fodor, we must distinguish strictly between narrow and broad mental contents and to base/construe scientific theories only on/about the first ones. But it is *like saying* physical objects must have some colours, narrow colours, that in no way are the colours, the broad colours, we see and describe from our context and that those narrow colours are the topic of a scientific theory of colour. If we don't consider in any way the colours with which we see the objects, objects couldn't have any colour at all. They could have physical, chemical properties and so forth, but not colours. *To speak about mental contents is like speaking about colours*. There must be some *canonical context of attribution*, some appropriate circumstances, in which the broad can be locally supervenient and can have lawful causal efficiency. To find out these particular circumstances for the mental is not only a conceptual problem but also, and mainly, an empirical one. A problem of the same variety is the problem of finding out the circumstances in which some peculiar internal structures are responsible for the emission of the light waves we associate with determinate colours.

Now, let's call a property *intrinsic* if it is representable with a certain monadic predicate capable of supporting genuine and lawful causal connections. A property will be *relational* if it is representable by a non monadic predicate capable of being incorporated in genuine and lawful causal connections. It is clear that properties can exist that are neither intrinsic nor relational. Some OA properties are of this kind, for instance, Fodor's property of being an H-(T-)particle. Often it is supposed that 1) OA properties can only be, if they are something, relational properties with a non-local supervenience, and that 2) only NOA properties can be intrinsic and locally supervenient. But, I consider both suppositions erroneous. First, as we have already mentioned, usually the distinction between intrinsic and relational properties depends, only, on

heuristic and pragmatic decisions. That distinction is relative to our interests and very often it doesn't entail any relevant logical or metaphysical difference. Second and more important, the way in which we have characterized our concepts, there isn't any reason why an OA property cannot be, in some cases, intrinsic and locally supervenient. Colours are a paradigmatic example of properties of this kind.

The statement that something has a determinate colour is completely justified by the knowledge of some facts of its context. But, in some special conditions of the context of attribution, the colours of an object are really intrinsic and locally supervenient properties, part and parcel of the object's microphysical structure and composition. In normal conditions of vision, colours are intrinsic and locally supervenient but also OA properties. If the conditions of vision are less than normal, if we have on our «sunglasses», for instance, colours are no longer either intrinsic or locally supervenient properties. To determine, precisely, normal conditions of vision is, as we have said, a difficult empirical problem; and the hypothesis that normal conditions exist involves numerous commitments. There is nothing trivial about them, and the same questions must be confronted in relation to mental contents.

If we have our «sunglasses» on, we lose the intrinsic and locally supervenient character of our colours, of the colours seen. But, this loss does not alter the intrinsic and locally supervenient character of OA properties. It only entails the relativity of this character to some fixed canonical context of attribution. In all empirical sciences the notion of normal conditions, those defining a canonical context of attribution, has a very clear function. And why not in psychology as well? Why must mental contents, broad mental contents, be more like H-(T-) particles than like colours?. To be in normal conditions is an extremely complex predicate and one that surely doesn't refer merely to OA properties, but it does, indeed, imply the intrinsic and locally supervenient character of some OA properties.

4. Only one kind of content

Fodor⁴ says:

«Your thought is content-identical to mine only if in every context in which you thought has truth condition T, mine has truth condition T and vice versa»

This is, contents instantiate the same functions from contexts to truth conditions. These are the mental contents relevant to our scientific theories of the mental. They are the narrow contents. On the other hand:⁵

«The 'broad' content of a thought is what you can semantically evaluate; it's what you get when you specify a narrow content and fix a context»

In the perspective I am sketching, there are not two sorts of mental contents. As we have seen, if there are two —broad and narrow contents as they are defined by Fodor—, and we try to specify the peculiar way narrow contents are instantiated in our body, then there would be *too many functional and causal candidates*. And the class to

⁴ FODOR (1987: Chap. 2).

⁵ Also FODOR (1987: Cap. 2).

which they all belong can only be characterized by means of OA properties; that is by means of contextualized semantical properties like having —the thoughts in question— such and such truth conditions in such and such contexts. Really, the body must matter more. There must, thus, be only one sort of mental content. One sort of mental content with a condition of identity similar to this:

Your thought is content-identical to mine if 1) in every context in which your thought has truth condition T, mine has truth condition T and vice versa, and 2) your internal structure and composition when you have your thought is sufficiently similar to mine when I have my thought in a canonical context of (self-)attribution

Let's explain. Subcondition 2 would reduce the proliferation of structural and compositional internal facts capable of fulfilling subcondition 1. The identity of contents is defined over the folkpsychological contents even though, in principle, they were ignorant of internal facts of the subjects. They are OA properties. If it is possible, a psychological theory related with the knowledge of the body must look for the local supervenience of these properties and for their causal relationships. To do this, the idea of canonical contexts of attribution is rich and suggestive. It is the standard, most travelled road in all of science. Science by nature covers OA properties with NOA ones through the fixing of canonical contexts of attribution. I am afraid that if we don't try something similar, we will find ourselves relegated to the field of *peripheral cognitive sciences*: that is, sociolinguistics, artificial intelligence, neurophysiology, reflexive folkpsychology and so on, far from a scientific cognitive psychology. But, I am also convinced that, in fact, something like this actually occurs in the everyday efforts of most psychologists.

In Fodor's view, narrow contents are complex functions from thoughts and contexts to the semantical properties, truth conditions, for instance, that would constitute the broad contents of these thoughts in these contexts. Those functions are instantiated in our bodies, mainly in our nervous systems. What makes a thought have the semantical values it has —according to folk attributions in whatever context—, is something instantiated in the body. But, if folkpsychology is as ignorant of internal facts of the bodies as we have claimed, no NOA property of the body can do this work. Your *classical Twins* and you share the same narrow contents since your thoughts in the same contexts *must* be semantically evaluated in the same way. But, this is not really a very difficult case: the same contexts plus the same internal facts *must originate* the same worlds. And broad contents, I am sure, supervene on the world. Now, *imagine another case of Twins*. You have a Twin who in no way shares with you any internal fact —structural neither compositional— but whose thoughts have, in every possible context, the same semantical values, the same truth conditions for instance, that your thoughts have. How could this be possible?. Well, your Twin is a very superficial Twin: he/she merely behaves like you, he/she merely reacts like you and so on. According to Fodor, your thoughts and the superficial Twin's thoughts share the same narrow mental contents because the thoughts of both have the same semantical evaluations in the same contexts. But, by hypothesis, you do not share any internal fact with your superficial Twin!. The only kind of properties that you share with him/her are OA properties. Hence, narrow mental contents are OA properties too. Of

course, extensionally, a narrow content is an equivalence class of internal facts. The class of all relevant internal facts of the subjects whose thoughts have in the same contexts the same semantical properties. But, again, belonging to this class of internal facts -functional and causal internal facts- is also an OA property.

Fodor's strategy doesn't help to select interesting NOA properties that can be associated with the folk mental life; that is, with the OA properties that constitute the folk contents. Classical Twins or superficial Twins don't help either. The former are exactly like ourselves in all internal facts. And the latter are completely different. The right solution must be found between them, in some notion of similarity between us, or our classical Twins, and whatever possible superficial Twin. In other words, we don't know what Fodor's ineffable narrow contents are. We only have the broad contents of our folkpsychological attributions, the various —real or counterfactuals— contexts in which we —really or counterfactually— attribute them and some knowledge of the internal facts of the psychological subjects to which we attribute these contents. These constitute the pieces of the puzzle. And the main options —in order to select in the subjects NOA properties related to these contents— are:

- 1) To consider that there will be no definite NOA properties of this kind unless —perhaps future— self-restrictions of the folk attributions of mental properties reduce this indefiniteness
- 2) To choose a privileged class of contexts of attribution —the sort of contexts we have called «canonical»—, to describe the relevant internal facts that covary with our attributions of mental contents and to construe, by means of similarity relationships, a class of internal facts on which the attributed mental contents can be locally supervenient on the body and can have the causal efficacy we want for mental properties as mental properties embodied in a subject.

The first is not really an option. Folkpsychology is as predictive, successful and universal as it is because of the autonomy of its attributions: the supposed self-restrictions are either an irrelevant part of folkpsychology, as we have said already, or, if relevant, they only spell suicide. But, the second option needs much more refinement.

5. Some refinements

We need to make clear a lot of points. First of all, what are, more precisely, those canonical contexts of attributions and how might they be discovered?. Intuitively, canonical contexts of attribution for mental properties, like whatever canonical context, are contexts in which all goes well. We cannot discover in them anything that converts our attributions into errors. The mental properties attributed by folkpsychology are OA properties, but the attributions may be fallible. There are sound, correct attributions and unsound, incorrect attributions. Let's suppose, as I assumed earlier, that colours are OA properties too. This doesn't mean that the colours we see and describe when we have our «sunglasses» on and when we know this fact —fact that belongs to the context— are colours we are disposed to attribute to the objects. In relation to mental properties we are in a very similar situation.

To be an H-(T-)particle is an OA property too. But, it is a property that doesn't have any causal relevance. We don't have any causal law in which these properties appear. On the other hand, to be of a certain colour or to have certain mental content are properties causally relevant in certain conditions: in those conditions that convert a determinate context of attribution of these properties into a canonical one. In these conditions, we can know what internal facts are happening —inside the skin— and we can try relevant covariations between these internal facts and our attributions, retaining the canonical character of the contexts in which we make the attributions. We don't need any other metaphysical background about meaning, causality and the order of the world. We need only to apply to mental life the *same epistemological strategy* that we apply in natural science to properties like colours, sounds, macroscopic objects and so on when we think they are interesting properties from the scientific point of view.

Once again, in some cases colours are locally supervenient on classes of internal facts of the things. Supervenience epistemologized is no more than an affair of lawless covariation. The only metaphysics we need is inside the limits of science, its epistemology and its methodology. So, colours, as colours, could have some sort of causal efficacy because of some sort of internal facts. Of course, a lot of attributions of colours are out of play: for instance, all the attributions we can make when we have on our «sunglasses» —and, it can be sunglasses of a very wide physical and philosophical variety—. Moreover, there are not two kinds of colours, narrow and broad colours. There is only one kind. Some special members of this kind are locally supervenient and have a causal efficacy according to our science.

In reference to folk contents, we could say exactly the same. Not all folk contents but only some special ones must be locally supervenient on the body. Perhaps not in the precise format of «beliefs-that» and «desires-that» but in the format of «beliefs-about» and «desires-about».⁶ These contents must be easily transportable with the bodies to other accesible and near-possible worlds. To other worlds in which we could be equipped with an epistemic and linguistic apparatus very close to the one we have in our actual world. Folkpsychology is very ignorant of the internal facts of the psychological subjects, but it is equally ignorant of any internal facts of things within the context. So, the following counterfactual thesis must be true:

IF 1) there would exist a world with internal facts of the same type as those in our world —in relation to a canonical context of attribution of a certain mental content— that are the basis of a supervenient relation of this content,
 AND 2) in that world we were equipped with our same epistemic and linguistic powers,
 THEN 3) in such a world the subjects that have these internal facts will have that mental content too.

In these conditions, we could have very different worlds but the same mental contents, the same folk, broad contents. So, these contents could have a local supervenience

⁶ Or in the format of «core beliefs-about» and «core desires-about». I don't think the core-strategy is a dead way for some --let's say «simple minded»-- psychological subjects like non-human organims, children, vetry handicaped people or AI engines.

and a causal efficiency. Of course, this would be applicable only to some folk contents. The other cases of folk contents, contents to which all of this is inapplicable, must belong more to the field of sociolinguistic theory and practice than to the field of scientific psychology dealing with facts happening inside our body.

A final and rather puzzling remark. Perhaps the sentence «the snow is white» doesn't need to be true in every possible world in which the sentence «X believes that the snow is white» can adequately *describe and refer to* a true belief of X in that world. Perhaps the sentence «the snow is white» and the internal facts we could associate, from a canonical context of attribution and for subjects like you and me, with «to believe that the snow is white» —the token of the mental sentence we have when we believe that the snow is white, if you prefer— don't belong to the same semantical type. Perhaps, they don't have the same truth conditions. Why suppose they do?. *In order to refer and describe a thing with semantical properties, we don't need to do it with an expression that has the same semantical properties.* We can, but we are not forced to do so. And, I think that in relation to the *basic and pre-folkpsychological semanticity of the mental*, the things are so complicated to suspect that here this may be not the case it is best, as a strategy. When we identify mental contents —folk, broad contents— and we try to establish their local supervenience and their causal powers through a supposed canonical context of attribution, we are in the same situation as we are when we interpret a text, a text in which the symbols —the tokens of these symbols— are causally connected among themselves and with some pieces of the world. This is, for instance, the situation we are in when we are interpreting an E-mail in our computer. We describe the message with expressions like «it says that ...», «if now it says that ..., I am sure that in the next page it will say that —», and so on. Let's say that this is the folkreading of the message. It is possible to identify unambiguously an expression of the computer with the description «it says that ...» even though that expression —the computer expression— doesn't have the same semantical properties that the expression «...» can have. Also, it is possible to look for the local supervenience in my computer of content «...» in spite of the fact that «to say that ...» is an OA property to my computer. Also, we can try to describe the causal relationships associated in my computer with content «...» as such a content. Furthermore, all of this could be made with mental contents and could serve as grounds for a scientific psychology in which folk contents are related with our knowledge of the body. But a problem remains. The expressions in my body —or in my computer— can have other basic and pre-folk semantical properties that the ones «...» in «I believe that ...» —or in «it says that ...»— can have. The problem is more complicated, because there is *nothing in my body* that is the trade-mark of a basic and original semanticity⁷. My mental word «cat» could be causally linked to cats in an

⁷ I remember Dennett exploring recently this problem in a lecture given at the «II Seminario interuniversitario de ciencia cognitiva» (Salamanca, March-1990. The title of his lecture was «The Interpretation of Texts, People and other Artifacts»). Even if we are realists about mental properties, as Dennett seems to be now, we will have the very hard problem of finding out the natural --non artificial--, original and pre-folkpsychological mechanisms that are the source on any mental property. But, perhaps, the divide between the natural and the

artificial way. In the same *artificial way*, the word «cat» in my computer can be causally linked to cats, but with, as well, the same *natural appearance* with which our mental word «cat» is causally linked to cats according to Fodor⁸. What is the difference between natural causal chains—natural causal chains that originate the desired basic semanticity—and the artificial ones?. Perhaps, this difference is only relative to our conceptual schemes. Moreover, this difference cannot entail any difference in our bodies or in their near contexts. Of course, we can look for some basic, original and pre-folkpsychological semanticity although it depends on our conceptual schemes as a lot of very important things do! But, in this case, it must be clear that *we will not be looking for another kind of content but for the source of whatever content*⁹. From the point of view of psychological models guided by Fodor's representational theory of mind, all of this is ignored. But, I think that it could be an important starting point in order to coordinate such models with other new models, with connectionist models for instance. *Perhaps scientific psychology needs to have more than one sort of models and the models for folk describable cognitive processes don't need to be the models for the acquisition of basic and pre-folkpsychological semanticity*. This is a very important distinction, because it is quite possible to have theories about the former without any clear idea about how to describe and explain completely the latter, even with the help of new connectionist models.

References

- FIELD H. (1978). «Mental Representation», *Erkenntnis*, 13.
 FODOR J. (1980). «Methodological Solipsism Considered as a Research Strategy in Cognitive Science», *Behavioral and Brain Sciences*, 3.
 (1987). *Psychosemantics*, Cambridge, MIT Press.
 (1990). *A Theory of Content and Other Essays*, Cambridge, MIT Press.
 PUTNAM H. (1988). *Representation and Reality*, Cambridge, MIT Press.

artificial depends too much of the various, scientific and folk, humans conceptual schemes and points of view. «Cat» can be caused by cats also in a lot of artificial ways. And the ways dogs could cause in some circumstances «cat», for instance, also could be asymmetrically dependent on these artificial causal paths! --see FODOR (1990: Chap. 2 and 3)--. Now, our problem is not only the problem of distinguishing situations of type I from situations of type II, but also the problem of distinguishing the natural from the artificial. And, how to do it?

⁸ For instance, FODOR (1990: Chap. 3 and 4).

⁹ It is obvious that, for subjects like you and me, the «natural» learning of the public word «cat», for instance, must be an important part of the source of internal facts -mental tokens- that have the semantical content we describe by the word «cat». See last remarks of FIELD (1978). But, we must not be confused. One thing is the source of a mental content and other thing is the way we describe and identify this content. The basic semanticity that a source provides can be different from the semantical properties by means of which we are able to identify mental causal efficacy, inside of psychological theories sensible to internal facts of the body.

Taula (UIB) núm. 15, 1992

La acción, su razón y su circunstancia

Fernando Broncano

1. La intencionalidad en la explicación de las acciones

Los avatares que ha sufrido la teoría de la acción en el último siglo y medio, en el jardín de los senderos que se bifurcan del cientificismo y el humanismo espiritualista, han promovido un cierto acuerdo en que el modelo de explicación intencional, el modelo de creencias- deseos- acciones, constituye un punto de partida (para algunos es también el punto de llegada). Pero al tiempo, el desarrollo de las ciencias sociales, cognitivas y psicobiológicas ha generado una imagen científica que se enfrenta a la imagen del sentido común que nos hacemos de nosotros mismos, presente y presupuesta de manera necesaria en este modelo de explicación.¹

La explicación intencional de la conducta humana, en tanto que algo diferente de la mera acción física, tiene una historia tan larga como la propia Filosofía y, según la gran mayoría de los filósofos, es un modo de explicación que no ha sido superado ni será fácilmente superable. La noción de intencionalidad connota todos aquellos rasgos que constituyen nuestras señas de identidad, de forma que quienes ponen en cuestión la capacidad explicativa de las intenciones atentan presuntamente contra los principios fundamentales que hacen posible todo pensamiento legítimo: pertenecen, en palabras de N. Goodman,² a las tribus de nuevos bárbaros que en nombre del cientificismo están destruyendo nuestra cultura.

No se trata de poner en cuestión el modelo de explicación intencional teniendo sobre nuestras cabezas la amenaza de una acusación tan grave. Aceptamos que el modelo intencional es el que mejor explica qué es lo que tenemos de especial los seres

¹ Stich (1983) y Dennet (1988).

² Goodman (1978) pg.

inteligentes, a saber, la capacidad de dar sentido o asignar contenido a las cosas que, en virtud de ello, se convierten en cosas para nosotros o, como afirmarían los antirrealistas, simplemente en cosas. No se nos escapa que la aceptación de este modelo entraña el compromiso con ciertas entidades del sentido común, las actitudes proposicionales, como entidades teóricas de las que nos servimos para explicar el contenido de los estados mentales. Para no bajar a mayores profundidades consideramos que los estados mentales no son sino estados cerebrales que tienen la virtud de contener información, así como la capacidad de actuar causalmente sobre los órganos encargados de la acción.

Pero la verdadera cuestión está en si el modelo de acción intencional es capaz de situarnos de forma adecuada en el mundo como seres naturales, por muy especiales que seamos. La idea de que formamos parte continua de la naturaleza, que somos producto de la evolución natural y que vivimos en dependencia permanente de nuestro medio ambiente, mucho más dependientemente de lo que parece implicarse en las imágenes dualistas, ha llegado a ser un componente de nuestra cultura que ya constituye un elemento regulador de cualquier teoría de la acción. Cuando situamos al agente intencional en un medio ambiente más complejo que el que sugiere el esquema de una mente en un cuerpo que le obedece, la acción intencional, sobre todo la más valiosa, aquélla que se articula en proyectos y programas, desborda el estrecho reducto en que ha sido confinada por los filósofos intencionalistas.

Ni el conjunto de nuestras acciones forma una clase homogénea ni los objetivos de las distintas teorías de la acción son coincidentes: hay, por citar un caso relevante, una perspectiva de la acción humana de carácter básicamente individualista que tiene interés para la Psicología, pero, también es cierto, ésta no es la única perspectiva, ni tiene por qué ser la perspectiva fundante de las demás, ni siquiera está claro que sea una perspectiva homogénea en sí misma. El tipo de acciones que nos interesa aquí es el que agrupa a las acciones que, además de *racionales*, tienen la virtud de ser *colectivas, cooperativas e históricas*. Son acciones **paradigmáticamente racionales** en el sentido de que constituyen un modelo de nuestras formas más perfeccionadas de transformar el medio y la parte del medio que somos nosotros mismos. Este tipo de acciones, por las características que acabamos de enunciar, pertenecen al terreno de lo que los filósofos han denominado *praxis*: se realizan a través de la interacción social y siguiendo reglas, normas y proyectos. Son acciones que, para castigo de escépticos, en no pocas ocasiones tienen éxito y, para desolación de dogmáticos, también a veces, provocan mayores males que los que pretendían resolver.

En la explicación de estas acciones de carácter social hay algunas preguntas que están en la base de todo deseo explicativo: ¿cómo es que a veces conseguimos hacer lo que queremos hacer?, ¿cómo conseguimos que los demás entiendan lo que queremos hacer?, ¿cómo conseguimos entender lo que están haciendo los demás?, ¿cómo es posible la acción cooperativa?, ¿cómo es posible la interacción?. Todas estas preguntas deben ser contestadas en una teoría de la acción que tenga algo que ver con lo que todos consideramos que son acciones paradigmáticamente inteligentes: pelear, trabajar, investigar, jugar, conversar, hacer el amor... Aceptemos que la intencionalidad es un presupuesto para la explicación de estas acciones, la cuestión es si el modelo de intención-acción es suficiente para dar cuenta del complejo de sucesos que ocurre en cualquiera de las acciones anteriores.

Desde Aristóteles hasta Davidson se ha seguido una estrategia similar: comenzar elaborando un modelo de acción simple que posteriormente se va haciendo más complejo. En este modelo tenemos un objeto o intención de la acción *O*, un deseo de conseguirlo *D*, moralmente aceptable o reprochable, (en algunos modelos se añade también una voluntad explícita de acción), la creencia verdadera o falsa de que para conseguir *O* hay que hacer *A* y acción propiamente *A*, determinada causalmente por lo anterior. El resto consiste en complicar mediante algún mecanismo de composición las acciones simples para formar las complejas. Hay razones para pensar que esta estrategia difícilmente consigue evitar el solipsismo o al menos el individualismo metodológicos en la explicación de la acción². No hay objeciones de principio a la adopción de estrategias solipsistas o individualistas. Es posible que ciertas ramas de la investigación, por ejemplo la teoría formal de la decisión, o algunas formas de la investigación psicológica, puedan emplear ocasionalmente estas estrategias, pero estoy convencido de que no son adecuadas para la explicación de las acciones que consideramos como sustrato paradigmático de la racionalidad práctica: las acciones sociales en general y las técnicas en particular.

Frente a la estrategia solipsista, una explicación de la acción racional debería tener en cuenta en primer lugar el carácter circunstancial y contingente de las acciones, en segundo lugar, la existencia de procesos previos de socialización, en consecuencia, pues, la doble dependencia contextual y lingüística (cultural, más precisamente) de la explicación de la acción⁴ y, en tercer lugar, la naturaleza *sui generis* de la interacción social que subyace a todos los procesos de acción social cooperativa. Sólo desde esta triple perspectiva nuestra teoría de las acciones resultará ser una teoría de lo que nosotros somos capaces de hacer.

2. Las explicaciones de las acciones como racionalizaciones de la conducta: Los patrones de conducta *normal*

Las acciones inteligentes pertenecen a una clase muy especial de hechos naturales. Son procesos que comienzan en un estado mental del agente y terminan en

³ La distinción entre solipsismo e individualismo se establece en Fodor(1988). Personalmente no creo encontrar la diferencia: según Fodor «el individualismo metodológico es la doctrina de que los estados psicológicos son individuados con respecto a sus capacidades causales, el solipsismo metodológico es la doctrina de que los estados psicológicos son individuados sin tener en cuenta su evaluación semántica». No es el momento de entrar en polémica con esta distinción, aunque Fodor parece haber dado marcha atrás en su programa: si las capacidades causales, desde el punto de vista funcionalista, no están determinadas por, y sólo por, el conjunto de conexiones sintácticas que determinan un ítem dado, -en el caso contrario no existe ninguna diferencia entre individualismo y solipsismo- el término de «individualismo» es completamente arbitrario y depende solamente de cómo se definan las acciones. Por supuesto que los agentes actúan individualmente, pero eso no implica que la causalidad sea individual. La causalidad no es menos sensible a la descripción contextual que las propias acciones.

⁴ El individualismo es sólo uno de los vicios de la psicología actual: el cientificismo, causalismo a toda costa y el parroquianismo que confunde la descripción psicológica en la cultura propia con la esencia de la mente humana son otros tantos pecados que merecen espacios separados. Véase Harré, R & D. Clarke & N. de Carlo (1989) para un análisis más pormenorizado.

una modificación de algún estado de cosas del mundo⁵. Algunos procesos biológicos son clasificados como acciones inteligentes en el caso de que este proceso suceda de *una cierta manera*: debe ocurrir de tal forma que *el estado mental del agente, en sus varios aspectos cognitivos, motivacionales y volitivos sea la causa* que desencadene el proceso⁶. En caso de que ésto no ocurra la acción puede tener un tipo de explicación diferente, no será explicada o reconstruida como una acción inteligente. Algunas orugas cuando salen del capullo dirigen sus movimientos siempre hacia arriba⁷; podríamos interpretar el hecho de manera pseudo-intencional afirmando que «buscan» las hojas más tiernas de los árboles que las alimentan, que «intentan» protegerse de sus depredadores... La realidad es que sus aparatos sensorio-motrices están dispuestos de forma que si la luz no incide simétricamente sobre los dos ojos, las patas del lado menos iluminado se mueven más rápido. Una consecuencia es que generalmente se mueven hacia arriba, a menos que coloquemos una fuerte luz en la base del tronco. Si sustituimos las primeras explicaciones por las segundas no ocurre nada, con la ventaja añadida de que ahora disponemos de una auténtica explicación causal. Pero ésta es la sustitución que no parece posible hacer en el caso de la conducta inteligente porque es el hecho de que la causa sea un estado representacional, y no otra cosa, lo que hace de las acciones intencionales tales acciones.

La atribución de intencionalidad con la consiguiente atribución de sentido a una acción resulta de un arriesgado juicio que realizamos sobre las transformaciones en el medio a que da lugar el movimiento físico de un agente. La preferencia de palabras, los saludos o las amenazas son hechos físicos que transportan la información de las intenciones de su autor. Este es el juicio que hacemos, aunque podemos equivocarnos, y de hecho lo hacemos muy a menudo, dado que los procesos naturales que constituyen las acciones intencionales son de un carácter ontológicamente muy extraño: una parte del proceso es, para decirlo en términos quineanos, *referencialmente abierta* y otra parte es *referencialmente opaca*. La parte abierta es aquélla que admite una descripción física o fisico-química del proceso: es el conjunto de sucesos que contribuyen mediante transformaciones físicas al resultado de la acción. La parte opaca corresponde al estado mental, motivacional y desiderativo.

Al adscribir sentido a la acción de un agente podemos cometer varios tipos de errores: puede haber ocurrido que la acción no haya sido intencional sino, por ejemplo, producto de un acto reflejo, puede que no tengamos familiaridad con los objetivos de las acciones, como ocurre con las acciones técnicas para los no especialistas, o puede que no estemos familiarizados con la circunstancia y cultura en la que se produce la acción, como ocurre cuando tratamos de reconstruir hechos del pasado a partir de sus

⁵ Véase Dretske(1988).

⁶ Dretske (1988) hace una defensa muy convincente, desde el punto de vista naturalista, de la causalidad de los estados cognitivos. Este punto del trabajo, como podrá comprobar quien conozca el libro, tiene una gran deuda con él contraída en su gozosa y apasionada lectura.

⁷ El ejemplo aparece en Dretske (1988) pg. 93.

productos. Incluso en el marco de una acción intencional interpretable, por ejemplo, la que uno mismo realiza⁸, actúan muchos mecanismos de carácter pre o subintencional que pueden ocultarnos el verdadero sentido de nuestras propias acciones, tal es el caso del pensamiento desiderativo o de otros varios efectos del marco sobre la representación.

Estos casos cuasi-patológicos, pero corrientes, nos indican que la explicación de la acción humana es mucho más compleja de lo que nos permiten suponer nuestros hábitos cotidianos, por dos razones fundamentales: en primer lugar porque las acciones humanas, como la categoría de hechos biológicos a la que pertenecen, se describen de manera *funcional* que se abstrae del sistema que las realiza (la digestión es un proceso biológico que consiste en la ruptura de las cadenas químicas de los alimentos para convertirlas en unidades asimilables, función que puede ser realizada, en principio, por sistemas diferentes a los órganos correspondientes). Pues bien, la atribución de funciones a órganos, como a máquinas y artefactos, todos lo sabemos, es más dificultosa de lo que parece y sólo se consigue tras laboriosas investigaciones. Cuando los sistemas que vamos a investigar son muy complejos, o bien simulan o recuerdan conductas nuestras, el deslizamiento hacia el lenguaje intencional es casi inevitable, como ocurre con los biólogos cuando divulgan sus conocimientos de la conducta animal o, como gusta Dennet de recordar⁹, cuando los programadores de programas de ordenador complejos, los jugadores de ajedrez por ejemplo, hablan de las «estrategias» del programa.

Surge una segunda dificultad en la explicación de la conducta, no ya de la complejidad funcional del sistema que estamos considerando, sino del *holismo* que caracteriza a los estados mentales. Nuestras representaciones, sean lo que sean, se caracterizan por estar unas con otras en relaciones lógicas y semánticas¹⁰ en diversos niveles de lenguaje formando una red cuya estructura contribuye de una forma u otra, aun no sabemos muy bien cómo, a la producción causal de la acción. Cuando atribuimos una causa intencional a un agente estamos indicando la existencia dentro de sus sistemas cognitivos y motivacionales de un estado mental relacionado de forma determinada con otros estados mentales. Y esta atribución es siempre arriesgada porque la hacemos *siempre* desde *nuestro* sistema conceptual y desde *nuestro* lenguaje¹¹.

Estas dos dificultades hacen que la adscripción de sentido a las acciones sea un proceso de una sensibilidad extrema al contexto del intérprete. Las acciones adquieren sentido intencional en la medida en que se integran en otras redes de acciones dentro de culturas y comunidades de agentes. Claro que ésto no es decir mucho, no es el punto final de nuestra historia, como lo es en la tradición wittgensteiniana, sino el punto de partida. Es posible incluso que tenga razón Umberto Eco¹² cuando señala que las cosas han ido demasiado lejos con el receptor y tal vez haya que volver a hablar de la *intentio operis* junto a la *intentio auctoris* y la *intentio lectoris*. Tal vez; tengámoslo en cuenta y no lo olvidemos en ningún momento, pero observemos que la triple dimensión de la que habla

⁸ Véase Elster (1988).

⁹ Dennet (1978).

¹⁰ Stich (1983).

¹¹ Liz (1988). Véase también la discusión entre Putnam (1988) y Fodor (1988).

¹² Eco (1997).

Eco involucrada en la adscripción de sentido indica que la explicación de las acciones intencionales tiene elementos *sui generis* de carácter metodológico: por una parte no nos está permitido explicar las acciones humanas siguiendo el modelo de leyes que se aplican al caso concreto, ésto es algo que sabemos desde los famosos escritos de Von Wright¹³, ya que no sabemos muy bien qué podrían ser leyes que rigieran la conducta intencional, más, por otra parte, cuando reconocemos una acción como acción intencional queremos decir también que un agente *R* en circunstancias *C* similares se habría comportado de manera similar a como lo ha hecho el agente cuya conducta estamos interpretando.

¿Existe algún modo de concebir la explicación de las acciones intencionales de manera que no abramos la puerta al dualismo ontológico por la vía del dualismo metodológico?, ¿existe alguna forma de reconciliar al intérprete en su contexto cultural con el agente en su estado individual y privado?, ¿hay forma de hacer compatible el holismo de los estados mentales con la determinación de una acción concreta?. Mi propuesta, siguiendo la de R. Garret Millikan, es concebir la explicación de las acciones intencionales no como aplicación de leyes sino como aplicación de *patrones de comportamiento normal*¹⁴.

Hay muchos sistemas en la naturaleza que describimos bajo patrones de comportamiento normal: aquéllos en los que nos interesa analizar los procesos que ocurren en ellos bajo el prisma de su adecuación a una determinada función. Es lo que hacemos al describir las funciones de los órganos: no nos interesa solamente la descripción físico-química de los procesos sino la medida en que éstos se desarrollan de forma «normal» o «anormal». Tal condición de normalidad no puede determinarse observando solamente el órgano, sino atendiendo al tiempo a la aportación que hace su conducta al funcionamiento general del organismo. Lo que decimos de los órganos podemos decir igualmente de la conducta del organismo en su entorno. La conducta de las orugas de las que hablábamos puede que sea la misma desde el punto de vista físico cuando le hemos colocado una luz en la base del tronco, pero desde el punto de vista de la adecuación de su conducta a las necesidades de supervivencia, no lo es. En un sentido muy similar hablamos de los instrumentos y herramientas: pueden realizarse de muchos materiales y emplearse para muchas tareas, pero a nosotros nos interesa solamente su *función normal*: no se suelen matar moscas a cañonazos. Casi nadie lo hace.

La idea de patrón de conducta normal no se expresa, ni es expresable, en términos de leyes naturales, entre otras razones por la *dependencia de las circunstancias*¹⁵ de la que hacen gala los patrones de conducta normal. Se puede, bien es cierto, establecer un patrón de conducta normal de acuerdo a propósitos muy variados: hablamos de normalidad del funcionamiento de un órgano respecto a los procesos generales del organismo, de normalidad de la conducta respecto al medio ambiente, etc. Por otra parte la elección del contexto determina también el modo de explicación que elegimos. Pues

¹³ Von Wright (1971).

¹⁴ La propuesta sigue de lejos las de Millikan (1984) (1986) y Pettit (1986).

¹⁵ Hay muchas teorías acerca de qué son las leyes naturales, pero todas ellas admiten la necesidad de un cierto grado de universalidad e independencia de las circunstancias. No parece que haya otro modo de distinguir las leyes de la naturaleza de las regularidades casuales.

bien, cuando analizamos una acción desde la perspectiva de la racionalidad definimos la *normalidad de las acciones* por referencia a un doble contexto de *circunstancias medio-ambientales* y de *agentes racionales* que conforman la cultura del agente cuya acción explicamos.

La subsunción bajo patrones de conducta normal, cuando tratamos de agentes inteligentes, tiene el carácter de una *racionalización* de la acción. Se trata de una reconstrucción normativa de un proceso causal iniciado en y por un estado mental que, en virtud de esta descripción, se considera una *razón* para la acción. Las acciones no racionalizables bajo patrones normales se convierten inmediatamente en candidatas a ser *explicadas* causalmente de otras formas en tanto que hechos que le suceden al agente, y no como acciones causadas por sus intenciones. Hay que tener en cuenta además que la racionalización de las acciones entraña una valoración y, por consiguiente, la aplicación de valores de muy diverso carácter que, como ocurre con los significados, no dependen solamente de lo que ocurre en la cabeza o en el sistema nervioso del agente, sino que en su mayoría dependen de convenciones o regularidades presentes en la cultura en la que ese sistema intencional se ha llegado a formar como tal.

Racionalizar una acción es situarla, como acabamos de proponer, bajo el alcance de un patrón de conducta normal, lo que significa algo más que aducir una razón para su producción; al menos en cuanto entendamos por razón algo que está en la creencia del agente: vamos a defender la idea de que lo que está en la cabeza del agente no es suficiente para determinar ni explicar las acciones. La noción de patrón involucra una necesaria atención a las circunstancias en las que se realiza que introduce en los patrones de acción un elemento radicalmente contingente como lo es toda circunstancia¹⁶.

3. La contingencia de la acción

Cuando subsumimos acciones bajo patrones de normalidad estamos reconociendo acciones como ejemplares de *tipos ideales*. Para hacerlo debemos ser capaces de clasificar acciones en tipos, lo que entraña la disponibilidad de algún criterio de similaridad entre acciones. Permítasenos suponer que este criterio existe y que puede ser definido con cierta precisión. Supongamos por el momento que este criterio es el que incorporan nuestras lenguas naturales en forma de los verbos de acción, lo que nos proporciona una cierta taxonomía de acciones que podemos enriquecer, si tal es nuestro deseo, con los vocabularios técnicos que aparecen en las diversas culturas. Pues bien, el problema con el que nos encontramos ahora es el contrario, la individualización de acciones concretas que ocurren en circunstancias diversas, en particular, el hecho de que

¹⁶ Los patrones pueden concebirse como regularidades fundamentadas en hábitos y habilidades que se producen en circunstancias similares y que son poseídos por agentes de la misma cultura. No se trata de reglas de acción porque con las reglas sucede algo similar a lo que sucede con las leyes: para que sean tales reglas necesitan eliminar la referencia a la circunstancia concreta en la que se realizan. Pero pueden estar involucradas reglas de acción en ciertos casos, particularmente en aquéllos en los que el agente dispone de un conocimiento y capacidad práctica suficiente como para formular planes de acción. Sin embargo, como veremos más adelante, este criterio no es necesario ni constitutivo de la acción inteligente.

los sistemas inteligentes sean capaces de realizar acciones concretas del mismo tipo en situaciones diferentes.

Consideraremos que las acciones inteligentes tienen la propiedad que **John Perry (1986)** ha denominado *eficiencia*: una acción es eficiente si es capaz de alcanzar su objetivo en circunstancias distintas. La capacidad de supervivencia de los organismos en circunstancias cambiantes depende de que adquieran habilidades en la producción de acciones eficientes puesto que sus capacidades de producción son limitadas por principio. Podemos pensar la eficiencia de las acciones como una propiedad similar a la que tienen las oraciones de los lenguajes naturales: una misma oración puede ser empleada en circunstancias diferentes para comunicar información también diferente¹⁷. El argumento consiste en mostrar que la eficiencia de las acciones solamente puede ser descrita en forma de una relación entre el estado mental del agente y el contexto:

Una acción es adecuada si contribuye a la satisfacción de un objetivo O dada la *verdad* de una representación o creencia R ¹⁸. Para que se produzca la adecuación no solamente son necesarias la verdad de nuestras creencias y la corrección de la dirección causal que las une a los efectos, sino también, siguiendo aún la terminología de J.Perry, la *benevolencia* del medio para promover dicha satisfacción. La benevolencia es una propiedad circunstancial que depende de cómo sean las cosas; es el correlato de la eficiencia: las acciones pueden ser eficientes a causa de que el medio es benevolente en el sentido de que circunstancias distintas, pero similares, contribuyen a la satisfacción de los deseos mediante la acción. Esta característica obliga a considerar la relación circunstancial con el medio como una parte sustancial de la determinación de la acción que tratamos de explicar.

John Perry (1986) ha observado una dificultad que surge al clasificar las acciones en tipos o individualizarlas como acciones específicas. Según cómo individualicemos dos acciones serán o no la misma acción. Si usted y yo estamos frente a un vaso de agua a cuarenta centímetros de él, extendemos la mano y lo llevamos a los labios, las acciones son la misma o del mismo tipo si las individualizamos mediante un criterio de similaridad física de los movimientos, pero no si las individualizamos por los resultados. No es la misma acción el que usted beba agua o el que sea yo quien la beba: podría haber ocurrido que yo acercase el vaso a sus labios¹⁹. Lo que está en juego aquí es la referencia a la circunstancia y al agente como elementos determinantes de la acción; y esta referencia es independiente de que la acción se determine aludiendo al estado mental que la causa, al modo de Davidson, o a un lenguaje público intencional, al modo de Von Wright²⁰. No estamos diciendo simplemente que, tal como ocurre en la descripción de todo sistema²¹, sea necesario especificar un marco de referencia espacio-temporal, lo que estamos afirmando es que la situación en la que se realiza la acción

¹⁷ Barwise & Perry (1983).

¹⁸ Perry(1986).

¹⁹ Obsérvese que el problema que plantea Perry es independiente de cómo describamos la acción, que es el problema al que se enfrenta Davidson (1967).

²⁰ Soutland (1983) ofrece una buena reconstrucción de la polémica entre ambas líneas de trabajo.

²¹ Véase Bunge (1979).

contribuye decisivamente a su descripción interna, lo mismo que ocurre con el estado interno del agente.

Si estas observaciones parecen perogrulladas es a causa de que las lenguas naturales contienen mecanismos que de modo inmediato sitúan la acción en su sujeto y en su lugar, pero esta articulación del lenguaje se olvida en las descripciones lógicas o filosóficas de la acción a causa de una mala semántica que tiende a situar los significados en la cabeza de los hablantes y no en las relaciones que mantienen aquéllos con el entorno en el que son proferidas las expresiones. Las lenguas naturales, al igual que las acciones, y posiblemente por las mismas razones epigenéticas, también son *eficientes* y permiten cambiar de significados con una asombrosa economía de medios. La deixis es, como sabemos, el principal mecanismo responsable de la eficiencia²². Una oración como «te lo digo claramente, amigo, estás equivocado» ancla su significado en tantas situaciones como sea usada y en cada una de ellas, si el hablante o el oyente son distintos, el significado también lo es.

La deixis es una propiedad lingüística que corresponde a una habilidad de los hablantes, la de abstraer y anclar información de y en situaciones concretas. Sin embargo la eficiencia de la acción humana no puede explicarse mediante un recurso semántico como éste que, por su propia naturaleza, se limita al transporte o tratamiento de la información pero no a la modificación de los estados y situaciones que la generan²³. Dado que la acción es un proceso que transforma, y no simplemente transporta, información, la eficiencia práctica debe serlo en el cambio de los estados de cosas, lo que exige una capacidad interna del agente para transformar situaciones concretas. Los tipos de acciones que realizamos, y por consiguiente, los tipos de acciones que conseguimos reconocer o conseguimos que nos reconozcan es necesariamente limitado. Las lenguas naturales, más los diversos vocabularios técnicos, a través de los verbos de acción, constituyen un sistema de clasificación natural de las acciones. Ciertamente hacemos más cosas que las que nuestros verbos son capaces de expresar y, por supuesto, ningún agente individual es capaz de hacer todo lo que el lenguaje expresa -la división del trabajo lingüístico refleja la división social del trabajo-, pero, a pesar de ello, podemos considerar esta clasificación como un catálogo de nuestros recursos de acción que, pese a su limitación, permiten una ilimitada capacidad de transformaciones creativas.

Propongo que analicemos la eficiencia práctica mediante la noción de *habilidades adquiridas*, aunque reconozco que las habilidades son capacidades de muy difícil análisis tanto filosófico como psicológico. **Quintanilla (1989)** distingue en el *conocimiento práctico* un componente teórico expresable en creencias del agente, un componente operacional, expresable en reglas del tipo «si quieres conseguir el objetivo *O* en las circunstancias *C*, debes hacer *A*», y un saber hacer adquirido por entrenamiento.

²² Este y otros mecanismos similares, como son las frases nominales se estudian en la Semántica Situacional desarrollada por Barwise & Perry (1983).

²³ Salvo cuando las preferencias son actos lingüísticos que modifican las situaciones, por supuesto, pero tales actos pertenecen a la categoría de acciones con todo el derecho y, por consiguiente, caen bajo el alcance de nuestra observación. Lo único que ocurre es que en ciertas ocasiones, como cuando el juez o el sacerdote dice «yo os declaro marido y mujer», es la información el instrumento que usamos para modificar las situaciones. Estos casos no presentan ningún problema especial.

Estoy de acuerdo en la propuesta, aunque el problema interesante desde el punto de vista del psicólogo, pero también del filósofo, es dilucidar en qué puede consistir la capacidad de hacer lo ordenado por la regla operacional. Hasta ahora han sido la percepción y el lenguaje los únicos campos en los que se han elaborado modelos computacionales de competencia -otro nombre para la habilidad-, pero eso no impide que podamos hablar de habilidades prácticas en campos más extensos de la acción, particularmente en aquéllos en los que nos encontramos con esquemas de acción aprendidos. En este sentido cabe esperar que el desarrollo de modelos conexionistas de diversas actividades inteligentes pueda ser muy útil para desentrañar no ya la arquitectura, sino también y sobre todo los mecanismos por los que el cerebro es capaz de organizar un esquema sensorio-motriz de acción semi o cuasi-reflejo²⁴. El problema interesante que presentan las habilidades es el de cómo pueden realizarse en las más variadas circunstancias, incluidas las que no pueden haber sido previstas por novedosas, lo que parece implicar un tipo de procesamiento más allá de las reglas de producción del tipo formulado anteriormente: lo que me distingue de mi fontanero, para mi desgracia, es mi falta de habilidad, no son seguramente las capacidades de realizar acciones particulares, tomadas éstas una a una, sino la capacidad para salir de forma inmediata de problemas prácticos que presentan las tuberías y que a mí se me antojan irresolubles.

En cualquier caso, independientemente de cómo sean los mecanismos internos al sujeto, hay una característica de la eficiencia que no podemos olvidar: no puede ser definida de forma internalista: la eficiencia de una acción es un predicado, al menos, diádico, en el que la circunstancia contingente ocupa uno de los lugares de la relación.

La cuestión se complica cuando consideramos la eficiencia de un tipo de acciones muy especial: las acciones cooperativas. Individuo a individuo, bien lo sabemos, nuestra eficiencia es muy limitada, pero también es cierto que una de nuestras habilidades no menores, que también poseen en cierto grado varias especies de depredadores que cazan en manada, es la capacidad de trabajo cooperativo: la resituación de nuestras acciones en función de las acciones e intenciones de los otros. Esta es la cuestión que vamos a tratar en la segunda parte de este trabajo. La noción de racionalidad está estrechamente ligada a esta capacidad en un sentido en el que no lo está necesariamente respecto a la acción intencional elemental. Mientras que es necesario que esta acción sea adecuada en cuanto a que la relación con la circunstancia sea la apropiada, en el caso de las acciones cooperativas, el requisito de racionalidad de los agentes se impone como relación necesaria para que sus acciones sean comprendidas.

4. La acción cooperativa y sus representaciones: planes y esquemas

En la acción cooperativa la dimensión temporal no puede reducirse a un marco externo de referencia en el que sucede la acción. Ya hemos establecido que también el espacio es algo más que un marco de referencia, pero el tiempo añade una nueva característica. Cada uno de las etapas del proceso se produce, por decirlo de algún modo, *contingentemente* respecto al anterior,

²⁴ Véase Rumelhart, D.E., Smolensky, P., J.L. McClelland & G.E. Hinton (1986) y Churchland, Patricia (1985).

resulta de singulares acumulaciones de condiciones iniciales que constituyen el marco para la siguiente acción. No hay interacción, en el sentido en el que estamos usando el término, sino entre sistemas dotados de capacidades de realimentación informativa y ajuste de las expectativas a las circunstancias de cada momento.

Pero no es suficiente un sistema de realimentación: se produce un proceso interactivo de carácter histórico cuando *cada uno de los agentes es capaz de comprender o conjeturar las intenciones de los demás y, sobre todo, es capaz de manifestar correctamente las suyas* (incluido en el caso de que pretenda mentir). Es decir, el agente considera como parte de la información obetida en el escenario de la acción los cursos de acción posibles de los participantes en el proceso.

Berliner (1980) describe las partidas de backgammon celebradas en el Casino de Montecarlo entre Sergio Villa, campeón del mundo en esa época, y BKG, un programa de juego por ordenador diseñado por el autor que consiguió ganar al campeón. BKG es un sistema que ante cada configuración del tablero responde efectuando una exploración de posibles configuraciones que se derivarían de ella dadas las posibles tiradas de dados. BKG, ciertamente, no explora todas las posibles salidas, unas cuatrocientas para cada configuración, y mucho menos todas las posibles en todos los cursos posibles. Por eso es un sistema «inteligente»: BKG asigna valores a las más interesantes y explora las consecuencias, siguiendo su sistema de valores hasta cierta profundidad y después elige la más conveniente. En la jerga de la Inteligencia Artificial, BKG dispone de un sistema de reglas que constriñe su «motor de inferencias», en la jerga de la metodología, BKG emplea reglas heurísticas. BKG puede considerarse, desde mi punto de vista, como un sistema que interactúa con la situación pero no un sistema que interactúa o actúa cooperativamente (aunque sea en un enfrentamiento) con su oponente, por lo tanto sólo *vicariamente* puede calificarse de racional: en primer lugar, aunque no habría que darle demasiada importancia, sólo en apariencia, o sólo trivialmente, puede afirmarse que el proceso es temporal e histórico, en el sentido en que lo es una secuencia de situaciones medioambientales. El espacio de posibilidades que define cada situación está artificial y convenientemente cerrado, a diferencia de lo que ocurre en una situación de juego con dos jugadores humanos. En segundo y más importante lugar, la evolución del sistema no es interactiva puesto que el programa no hace conjeturas sobre sobre la conducta del oponente. El programa no «ajusta» sus proyecciones futuras al curso de acción que conjetura tomará su oponente.

No estoy adoptando ninguna posición a favor o en contra de la Inteligencia Artificial. Por el contrario, tiendo a pensar que los programas inteligentes son modelos, en el sentido metodológico, de prototeorías psicológicas. De hecho programas de juego como el descrito por **Findler (1978)** se aproximan mucho más a una situación interactiva, pero en la medida en que necesitan un «modelo de usuario», necesitan también un «programa» de comprensión de acciones del oponente.

La ciencia cognitiva actual, cuyo dominio pretendido es la clase de sistemas que procesan información, dejando a un lado cuál es su realización física, ha comenzado a seguir una estrategia de aproximación a la arquitectura de un sistema que es capaz de interactuar con el medio. Se basa en el supuesto metodológico de que comprender una acción cotidiana, en cuanto causada por las razones de un agente, es generar un plan en

las circunstancias actuales que conduciría al propio sistema a la realización de la acción²⁵. La investigación se desarrolla sobre todo alrededor de la comprensión de historias en las que el sujeto debe reconstruir las cadenas causales ocultas o defectuosas. Una oración como

«Juan lloró cuando María le dijo que estaba enamorada de Andrés» es inmediatamente comprendida por los oyentes porque éstos reconstruyen las cadenas causales implícitas en la frase, con las que están familiarizados a causa de que, en este caso, quizá, conozcan los efectos típicos que produce la información en las relaciones amorosas asimétricas.

Shank y Abelson persiguen la construcción de un sistema que pudiera representar la operación simétrica de entender o contar una historia perforando el túnel desde dos partes: formalmente siguen la estrategia formulada por los pioneros de la Inteligencia Artificial A. Newell y H. Simon de sustituir los mecanismos deductivos por mecanismos heurísticos mucho más rápidos. Un mecanismo heurístico consiste en una restricción del número de inferencias posibles, lo que nos permite desechar un nutrido grupo de situaciones que en otro caso habría que haber contrastado. La restricción se realiza haciendo interactuar el motor de inferencias con el conocimiento del que dispone la memoria del sistema. Para ello almacenamos ciertas reglas inferenciales que establecen que en ciertas circunstancias dadas -que el programador experto debe conocer por experiencia propia- hay que comportarse de tal y tal forma. Un motor de inferencias «inteligente» aplica, cuando reconoce situaciones típicas, acciones que ya han tenido éxito en casos similares.

El problema que plantea la elaboración de un sistema heurístico con limitaciones inferenciales de esta clase es cómo estructurar el conocimiento depositado en la memoria del sistema. S. Pappert y M. Minsky, expertos en I.A. de inspiración piagetiana, desarrollaron un sistema de almacenamiento del conocimiento que, aunque no es el único y ha sido discutido por muchos psicólogos (**Anderson (1983)**, sí ha devenido en la forma más conocida de representar la memoria de los sistemas inteligentes: se trata de un almacenamiento estructurado en microsistemas que se caracterizan por relacionar mediante reglas de varios grados de abstracción grupos y tipos de circunstancias concretas. La articulación o estructuración de las reglas se realiza en una o varias pirámides de progresivas abstracciones. La base de la pirámide esta compuesta por esquemas de situaciones que simplemente deben ser rellenados con datos tomados de la realidad. El mecanismo heurístico consiste en llamar *ordenadamente* a estos microprogramas cada vez que se encuentran situaciones características. Esto es lo que se denomina a grandes rasgos en I.A. un sistema experto.

La otra dirección de la investigación consiste, en el caso definido por Shank y Abelson, en construir una base de conocimiento que sea empíricamente similar, o simule serlo, al conocimiento activo en la cabeza del hablante de un lenguaje natural.

²⁵ El trabajo clásico es Shank & Abelson (1977). Un modelo más sofisticado es el de Wilensky (1983) aunque la estrategia fue diseñada por Miller, Galanter & Pribram (1960). Winograd (1972) consigue un primer programa de ordenador que en un mundo reducido de objetos geométricos entiende fragmentos del lenguaje.

Se definen situaciones o micromundos construidos desde una perspectiva humana en los que se producen secuencias de acciones con «causalidades» características en la cultura que se supone al usuario del sistema o al productor del texto que el sistema debe comprender. Se presupone en esta estrategia que el lenguaje natural contiene entre sus reglas semánticas una clasificación de objetos, acciones y situaciones posibles y dentro de las acciones, de relaciones de causalidad. Esta estrategia no es diferente al proyecto de construcción de una psicosemántica: las acciones se reconocen dentro de un sistema de clasificación natural de acciones.

Una vez que está construida la base de conocimiento y en esta hemos almacenado la taxonomía de acciones y situaciones suficientes para la comprensión de la historia, el programador debe construir un modelo de planificador, y por consiguiente un modelo de intérprete. Se trata, ahora ya, de un sistema simulador de inteligencia que emplea los mecanismos heurísticos de los que hemos hablado antes y contiene varios subsistemas en el que corresponden a funciones que se supone que el agente realiza. Los siguientes componentes son los establecidos por **Wilensky(1983)**, que puede considerarse como una versión más sofisticada del modelo de Schank y Abelson:

- *Detector de objetivos*: dada una situación, al menos dada la información que nuestro sistema inteligente es capaz de extraer de ella, el detector propone objetivos que deben alcanzarse. Para ello se generan estados posibles relevantes y accesibles desde esa situación y se les concede un valor determinado dados los deseos y necesidades del sistema planificador.

- *Proponente de planes*: este subsistema es el encargado de elaborar uno o varios planes de acción que conduzcan el objetivo. Para ello debe buscar y encontrar en la memoria algún plan aplicable a la situación dada o, mediante razonamientos difusos²⁶, algún plan que haya sido usado en situaciones análogas.

- *Proyectista*: este subsistema puede ser ya considerado como un subsistema de control. Su función es construir mundos posibles en los que se somete el plan propuesto a un test de funcionamiento: el subsistema se encarga de desarrollar el plan en la memoria o, si se quiere, de calcular las consecuencias. Para ello puede tomar las condiciones iniciales de la situación real o «imaginar» situaciones diferentes en las que el plan se ve sometido a otras constricciones. la evaluación de las consecuencias en la adopción del plan y la consiguiente decisión sobre su validez es un primer mecanismo de racionalidad o control.

- *Sistema metaplanificador*: Wilensky (83) introduce una corrección muy sensata en el modo habitual de computar planes que se venía siguiendo desde Shank y sus colaboradores. Se trata de un mecanismo o subsistema en el que es posible evaluar los objetivos en su relación con otros objetivos posibles y/o deseables: mientras que la consecución de un objetivo puede reforzar la consecución de otro y, por consiguiente, hacer menos importante o urgente su persecución, e incluso eliminarlo como objetivo

²⁶ El sistema de almacenamiento de información en forma de patrones cuyas circunstancias concretas pueden quedar abiertas permite esta búsqueda de p en situaciones análogas puesto que la analogía la produce el que dos situaciones compartan varios aspectos estructurales relevantes al objetivo que tratamos de conseguir.

que hay que conseguir, en otro caso puede ocurrir lo contrario, que dos objetivos compitan de manera que la consecución de uno inhiba la producción de otro. En este caso el sistema debe decidir que objetivo se persigue o cual es el punto de equilibrio entre los dos. Para decirlo en términos antropomórficos, el sistema debe estar provisto de algún conjunto de valores regulativos o de una «ética» no meramente consecuencialista.

- *Ejecutor*: Si el sistema pretende ser o simular ser un agente intencional debe estar provisto de un mecanismo causal que sea capaz de llevar a cabo los planes, es decir, que sea capaz de transformar el medio en función de los contenidos establecidos en el estado mental o computacional del sistema.

El planificador e intérprete de historias es, como ya puede suponerse, algo más que un trabajo de mayor o menor interés tecnológico en el campo de la inteligencia artificial, que lo tiene, por cierto; su principal interés filosófico está en que se trata de un programa que integra de manera comprensiva elementos de sentido común incorporados en el lenguaje natural, una fenomenología de la vida cotidiana, versión rigurosa y crítica del sentido común y la simulación cuasi empírica que proporciona la inteligencia artificial del modelo conceptual. Puede considerarse un modelo acerca de cómo se adscribe racionalidad a sistemas de acciones más complejos que los que suelen considerar los filósofos cuando hacen teoría analítica de la acción: la racionalidad de los sistemas de acciones es una propiedad que ha de adscribirse a los planes o proyectos que el sujeto agente *tiene en la cabeza*.

Aunque existen diversos modos de analizar la racionalidad de las acciones, todas ellas se reducirían en último extremo al establecimiento de condiciones sobre las decisiones que toma el agente (el sistema en el caso de los simuladores) dada la información de la que dispone: los sistemas heurísticos, en el mismo sentido en que los seres humanos, son más o menos racionales en virtud de que aplican estrategias consistentes de decisión que pueden ser comparadas por su eficiencia: organizan sus objetivos en preferencias, establecen una distribución de probabilidades y en base a ambas asignaciones deciden. El resultado debería ser un plan que fuese el mejor de los disponibles. Una vez que el proyecto se pone en marcha la racionalidad consiste en la capacidad del sujeto para restaurar la marcha del plan ante perturbaciones posibles.

5. Modelos conversacionales versus modelos narrativos

La comprensión de narraciones es el modelo de comprensión de acciones que ha postulado la ciencia cognitiva de manera casi incuestionada, falta de crítica a la que no es ajena el *bouquet* racionalista que despide el programa de la inteligencia artificial. Digo casi incuestionada porque recientemente L.A. Suchman²⁷ ha publicado una crítica a este modelo y una contrapropuesta que me parece de una importancia transcendental, entre otras razones porque se trata de una crítica *desde dentro*. La propuesta de L. Suchman está basada en experiencias empíricas llevadas a cabo en sujetos que interactuaban con una máquina de cierta complicación para conseguir un resultado deseado. Se trataba en

²⁷ Suchman (1986).

concreto de una fotocopiadora de funcionamiento un poco complejo a la que se había añadido un programa tutorial de aprendizaje en un ordenador acoplado. El problema que interesaba a Suchman era el de los problemas de la comunicación entre hombre y máquina en procesos interactivos. En este caso se trataba de que la máquina realizase las intenciones del usuario para lo que debería pasar por varios estados en función de las órdenes emitidas por un usuario inexperto. Al igual que ocurría en las programaciones anteriores de sistemas capaces de comprender historias, el problema tenía, tiene, un innegable interés técnico, pero lo más importante desde la perspectiva conceptual era que la estrategia de programación refleja modelos y teorías implícitos acerca de la naturaleza de la acción social y la acción interactiva. El análisis que Suchman hace del problema presentado es que, al contrario de las metodologías habituales, el proceso de interacción entre el usuario y la máquina no puede entenderse como un mero proceso de acción y reacción que está basado en dinámicas mentales solipsistas de los sujetos que participan, en este caso el usuario y el programa de instrucciones, dinámicas que, a su vez, son entendidas como instancias de procesos de computación del tipo descrito anteriormente. La interacción hombre-máquina es un proceso social de interacción con su propia sustantividad que no puede separarse de la circunstancia concreta en la que tiene lugar. Los dos peligros que se tratan de evitar aquí son el individualismo metodológico y el solipsismo: la concepción del nivel social como un nivel autónomo nos lleva a buscar reglas y normas implícitas en quienes ya son expertos en este nivel social, los humanos, como modelos para ser implementados en programas de menor nivel de interacción; la concepción situada de la interacción nos precave del optimismo de pensar que un proceso interno, mental o computacional es suficiente para determinar de forma unívoca qué sea una acción o cuál es su sentido.

El estudio de Suchman parte de la constatación empírica de que el problema más arduo con el que se enfrenta el programador de sistemas tutoriales -el caso más complejo, por cierto, de los que pretende resolver actualmente la IA, para escarnio de quienes piensan que enseñar es más sencillo que calcular- no es tanto el de saber el estado cognitivo del interlocutor, axioma intocable de la psicología educativa de orientación cognitiva, lo que se traduce computacionalmente en dotar a la base de datos del sistema de un modelo de usuario, sino en ser capaces de que el usuario entienda lo que el hablante quiere decir: los diseñadores de esos sistemas, al observar las expresiones verbales de usuarios de la máquina, encontraban que se producía un problema de opacidad referencial. Los usuarios no entendían o malinterpretaban las instrucciones de la máquina, cuyo programa, independientemente de que estuviese «pensando» en un usuario inexperto, estaba diseñado siguiendo modelos ideales de conducta racional al margen de las situaciones concretas de interacción.

El problema es simétrico para la máquina, pues si a una orden malinterpretada el sistema responde con una señal de error o una parada, la interacción colapsa también; de poco nos servirá entonces nuestro programa tutorial para usuarios inexpertos. La solución es que el programa aprenda a extraer información, no de él mismo o de su memoria como ocurre en los modelos solipsistas, sino de la acción que se ha realizado en circunstancias variables y contingentes. La estrategia que debemos usar entonces para diseñar nuestro sistema de instrucciones no puede ser entonces el modelo narrativo

basado en planes de conducta, porque lo que no puede ser previsto en los planes es la reacción del usuario ante la situación de la que forma parte la acción que vamos a formular, pero también otros factores en que no dependen sólo, como se deduce de lo que venimos indicando, del estado o conocimientos del receptor sino del modo en que se presenta la acción. La propuesta de Suchman es que observemos acciones sociales en las que la cooperación de los agentes no puede entenderse solamente como un proceso de acción-reacción sino como un proceso en el que el resultado final está siempre presente pero que no puede ser planificado. La conversación es, en este sentido, un modelo de lo que puede considerarse como conducta cooperativa. En una conducta verbal se producen curiosos y muy significativos procesos de conducta cooperativa como, por ejemplo los sistemas que usa el oyente para asignar sentido a las palabras del interlocutor aunque no parezcan tenerlo, ni en los más finos análisis de las implicaturas conversacionales que pueda hacerse. Lo mismo ocurre con los malentendidos que frecuentemente se producen en el discurrir de la conversación²⁸: si la conducta del oyente no es la esperada el hablante puede inducir o bien que no se le ha entendido o bien que la actitud del oyente es diferente de la que él pensaba. Tanto en el caso de las aparentes inconsistencias como en el caso de los malentendidos los hablantes no suspenden el curso de la conversación para aclarar la situación, sino que inician caminos marginales de aclaración que se van autoajustando en el curso de nuevas situaciones conversacionales. La cuestión importante es que una conversación no sólo no podemos describirla como un proceso de mera acción- reacción basada en la comprensión mutua de planes, que de hecho no se produce, al menos eso es lo que indican los estudios empíricos, sino que ni siquiera podemos entenderla como un proceso guiado por alguna extraña forma de consenso comunicativo, como algunas teorías neotranscendentales pretenden afirmar. Una conversación es un proceso de cooperación esencialmente situada en el que cada uno de los interlocutores pone en la situación sus propias expectativas, que corrige instante a instante en la medida en que está convencido de que el proceso de interacción continúa correctamente, pero al tiempo usa la situación concreta conversacional, no como mera instancia de planes que tiene en la cabeza, sino, como fuente esencial de información.

El libro de Suchman comienza con una analogía etnográfica de lo que podría entenderse como dos modelos alternativos de racionalidad: los navegantes occidentales del siglo XVIII y los nativos trukeses de las islas de la Melanesia, quienes viajan asiduamente de unas islas a otras en pequeñas canoas. Los navegantes occidentales conciben el proceso técnico de la navegación como un proceso guiado por planes de navegación que se establecen al comienzo del viaje en un mapa. La función de los navegantes será restaurar el plan de navegación cada vez que se produce una perturbación que altera el curso previsto y la información que obtengan en el viaje será una información seleccionada por su carácter instrumental en el plan. Los nativos trukeses, por el contrario, cuando salen a navegar no solamente no disponen de cartas de

²⁸ Belder (1983) saca unas conclusiones muy inteligentes del problema de las dificultades de comprensión, dificultades que, como hacían los empiristas con los errores, se achacan achacan ingenuamente a criterios externos y sencillos de resolver

navegación, sino que ni siquiera tienen la noción de plan. Si se les pregunta por su objetivo en cualquier momento y circunstancia pueden indicar el punto al que se dirigen, pero se sorprenden ante una pregunta por su plan de navegación. Los trukeses explotan de forma habilísima la información disponible en cada situación particular: el color del agua y el viento, la velocidad y dirección de la corriente, las algas y el vuelo de las aves,...; con la información obtenida se resitúan permanentemente sin perder de «vista» la finalidad de sus actos, pero sin que ésta se convierta necesariamente en plan.

Las tentaciones relativista o dogmática son tan cercanas al tipo de argumentación que acabo de esgrimir que no podemos acabar sin un comentario sobre el significado de esta presunta dicotomía de modelos de racionalidad: nos sentimos tentados a pensar que existe algo así como una racionalidad práctica del salvaje y una racionalidad práctica occidental, al modo de Levi-Strauss, o una racionalidad pre-tecnológica y una racionalidad tecnológica. Pero no es así. Con esta respuesta se tranquilizan tanto los relativistas etnológicos como los dogmáticos científicistas: lo que estamos intentando insinuar es que los modelos situados, contingentes y cooperativos de acción no son alternativos a los modelos racionales de planificación sino que, por el contrario, son su condición de posibilidad. La acción situada no es un momento de la acción planificada, sino al contrario: aquélla es constitutiva de ésta y no al contrario. Podemos planificar en situaciones normales y normalizadas en las que nuestras habilidades son ayudadas por un control rígido de procesos que los convierte en predecibles, pero ésa no es nuestra única ni nuestra más importante habilidad racional práctica. Al contrario: porque somos capaces de ser racionales en situaciones no planificadas ni planificables es por lo que podemos buscar universos y procesos en los que la teoría de la decisión se aplique de forma estándar. Tal vez podemos hablar de tres momentos de desarrollo de nuestras capacidades racionales: un momento pretecnológico en el que simplemente aplicamos nuestras habilidades cotidianas, un momento tecnológico en el que estas habilidades están normalizadas en sistemas técnicos de reglas nomopragmáticas y un momento posttecnológico en el que debemos controlar la adecuación de los propios procesos técnicos. En este último caso, el más interesante desde el punto de vista de la racionalidad, no nos encontramos en mejor situación que en el primero: los resultados de los procesos sociales de cooperación técnica no son previsible cuando los observamos a una escala algo mayor que la que usa el ingeniero contratado por la empresa, pero eso mismo no los hace racionales ni irracionales. Lo que ocurre entonces es que ya no podemos seguir actuando siguiendo políticas consecuencialistas ni siguiendo planes que no existen más que en los manuales de racionalidad. Es el momento en el que debemos de aprender de nosotros mismos, cuando somos racionales en las situaciones más complejas que presenta la vida social y cotidiana.

7. Moraleja

Dice un proverbio chino: «si el sabio se empeña en aprender de la actuación del tonto, con un poco de esfuerzo llegará a ser tan tonto como él; aprender de sí mismo le llevará varias vidas pero sólo entonces será sabio».

Referencias

- ANDERSON, JOHN (1983). *The Architecture of Cognition* Cambridge. (Mass): Harvard University Press.
- BARWISE, J. & J. PERRY (1983). *Situations and Attitudes*. Cambridge (Mass): MIT Press.
- BELDER, BEATRICE (1983). «I know you but if only I understood you...» en PARRETT & BOUVERESSE (eds). *Meaning and Understanding*. Berlin: Gruyter.
- BERLINER (1980). «Chaquete por ordenador» *Investigación y ciencia*. Agosto 1980.
- BUNGE, M. (1979). *A world of systems. Treatise on Basic Philosophy*, vol. IV Dordrecht: Reidel
- CHURCHLAND, PATRICIA (1985). «Epistemology in the age of Neuroscience» *Mind*.
- DAVIDSON, D (1967). «The logical form of action sentences» en *Essays on Actions and Events*. Oxford: O.U.P.
- DENNET, D. (1978). «Intentional Systems» en *Brainstorms*. Cambridge (Mass) MIT Press.
- (1988). «Making sense of ourselves' en *The Intentional Stance* Cambridge (Mass) MIT Press.
- DRETSKE, F. (1988). *Explaining Behavior. Reasons in a world of causes*. Cambridge (Mass) MIT Press.
- ECO, U. (1987). «El extraño caso de la intentio lectoris» *Revista de Occidente*, 69.
- ELSTER, J. (1988). *Uvas amargas. Ensayo sobre las subversiones de la racionalidad*. Barcelona: Península.
- FODOR, J. (1987). *Psychosemantics: the problem of meaning in the Philosophy of Mind*. Cambridge (Mass): MIT Press.
- FINDLER, N. (1978). «Poker computerizado» *Investigación y ciencia*. Septiembre.
- GOODMAN, N. (1978). *Ways of World-Making*. Indianapolis: Hackett.
- LIZ, MANUEL. «Ways to obtain contents» *Actas del II Simposio Iberoamericano sobre Información y Contenido*. México: SOFIA.
- MILLER, G.H., GALANTER, E. & PRIBRAM, K.H. (1960). *Plans and the Structure of Behavior*. N.York: Holt.
- MILLIKAN, R.G. (1984). *Language, Thought and others biological categories: New foundations for realism*. Cambridge (Mass) MIT Press.
- (1989) «In Defense of Proper Functions». *Philosophy of Science* 56 228-302.
- MINSKY M. (1975). «Aframework for representing knowledge» en P.H. WINSTON *The Psychology of Computer vision*. N. York: McGraw Hill.
- (1987). «Líneas-K: una teoría de la memoria» en D.A. NORMAN *Perspectivas en ciencia cognitiva*. Barcelona: Paidós.
- PERRY, J. (1986). «Circumstancial attitudes and benevolent cognition» en BUTTERFIELD, J. (ed) *Language, mind and logic*. Cambridge: C.U.P.
- PETTIT, P. (1986). «Broad-minded explanation and Psychology» en PETTIT & McDOWELL (eds) *Subject, thought and context*. Oxford: O.U.P.
- PUTNAM, H. (1988). «Racionalidad en teoría de la decisión y en la ética» en L. OLIVE ed) *Racionalidad en ética, política, ciencia y tecnología*. México: F.C.E.
- QUINTANILLA. M.A. (1989). *Tecnología: un enfoque filosófico* Madrid: FUNDESCO.
- RUMELHART, D.E., SMOLENSKY, P., McCLELLAND & HINTON, G.E. (1986). «Schemata and secuential thought in PDP models» en McCLELLAND & RUMELHART; *Parallel Distributed Processing*, vol 2 Cambridge (Mass): M.I.T.Press.
- SCHANK, R. & ABELSON, R. (1977). *Scripts, plans, goals and Understanding*. Hillsdale (N J.): Lawrence Erlbaum.

Taula 15

- SITCH, S. (1985). «Could man be an irrational animal?. Some notes on epistemology of rationality». *Synthese* 69 .
- SOUTLAND, F. (1983). «Davidson, Von Wright and the debate over causation» en FLOISTAD, G. (ed) *Contemporary Philosophy. Philosophy of action*. Amsterdam: M. Nijhoff.
- SUCHMAN, L.A (1987). *Plans and situated actions. The problem of human-machine communication*. Cambridge: C.U.P.
- VON WRIGHT, G. (1971). *Explicación y comprensión*. Madrid: Alianza.
- WILENSKY, R. (1983). *Planning and understanding. A computational approach to human reasoning*. Reading (Mass): Addison-Wesley.
- WINOGRAD, T.A. (1972). *Understanding natural language*. N.York: Academic Press.

Taula (UIB) núm. 15, 1992

Conexionismo y el epifenomenalismo de lo mental

Josep E. Corbí

Universitat de València

Resumen

En este artículo discuto una serie de argumentos presentados por J. Fodor, Z. Pylyshyn y B. McLaughlin donde se defiende que el surgimiento de los modelos conexionistas no representa realmente una amenaza para la hipótesis del lenguaje del pensamiento, dado que esos modelos no pueden dar cuenta de la sistematicidad del pensamiento ni de la eficacia causal de las propiedades mentales. A lo largo del artículo intento mostrar, sin embargo, que los argumentos mencionados no aciertan a excluir la posibilidad de dar cuenta, dentro de un modelo conexionista, de la sistematicidad y la eficacia causal de lo mental. Este hecho afectaría ciertamente a los mejores argumentos en defensa de la hipótesis del lenguaje del pensamiento, dado que tales argumentos tiene la estructura de una inferencia en favor de la mejor explicación.

Abstract

I discuss some arguments presented by J. Fodor, Z. Pylyshyn and B. McLaughlin where they defend that the emergence of connectionist models does not constitute a threat to the language of thought hypothesis, since they cannot account for the systematicity of thoughts and the causal efficacy of mental properties. I argue, however, that the arguments supplied do not rule out the possibility of accounting, within connectionist models, for the systematicity and the causal efficacy of the mental; and this fact represents a major challenge to the best arguments for the language of thought hypothesis, which have the structure of inferences to the best explanation.

1. Introducción

En este artículo, se discuten algunos argumentos de J. Fodor y B. McLaughlin en los que se defiende que el surgimiento de los modelos conexionistas no constituye una amenaza para la hipótesis del lenguaje del pensamiento (HLP). Intentaré mostrar, por el contrario, que, si bien los modelos clásicos y conexionistas no son necesariamente incompatibles, el surgimiento de estos últimos socava los mejores argumentos en favor de HLP. En este sentido, argüiré que los argumentos de Fodor y McLaughlin fracasan en su intento de excluir la posibilidad de construir una teoría representacional de la cognición que, además de incorporar los modelos conexionistas, prescindiera de HLP.

Así, en la primera sección, ofrezco una caracterización de HLP y presento los argumentos que Fodor propone en su defensa y que, como veremos, descansan básicamente en la carencia de explicaciones alternativas. En la segunda sección, describo la crítica de Fodor y McLaughlin al modelo de Smolensky en tanto que posible alternativa a HLP. Su crítica se centra en la imposibilidad de que este modelo dé cuenta de la eficacia causal de las propiedades mentales, por lo que la discusión de este tópico nos obliga a analizar las condiciones que, según Fodor, debe reunir una propiedad para que sea reconocida como causalmente eficaz. En la tercera sección, se desarrolla una crítica a las condiciones que Fodor establece para la eficacia causal y se presenta una propuesta alternativa, que, entre otras cosas, permite hacer compatible el modelo de Smolensky con la eficacia causal de las propiedades mentales. Este hecho me lleva, finalmente, a concluir que el surgimiento de los modelos conexionistas parece socavar los mejores argumentos en favor de HLP.

1. La Hipótesis del Lenguaje del Pensamiento

La Hipótesis del Lenguaje del Pensamiento va más allá del Realismo Intencional. Así, un realista intencional asume: (i) «...Que hay creencias y deseos y que hay una cuestión de hecho acerca de sus contenidos intencionales.» (ii) «Podría ocurrir que creer y desear fuesen estados del cerebro.» (iii) «...Que las creencias y los deseos tienen roles causales y que la conducta pública es, típicamente, el efecto de complejas interacciones entre estos estados mentales.»¹

HLP añade, sin embargo, «que *los estados mentales* - y no sólo sus objetos proposicionales- *tienen típicamente estructura constituyente*».² Un estado tiene estructura constituyente si y sólo si tiene partes constituyentes. La noción de 'parte constituyente' puede, a su vez, analizarse en términos de transportabilidad en el seno de una semántica combinatoria. Diremos, en este sentido, que una parte de un estado mental es una parte constituyente sólo si es transportable a otro estado mental, es decir, si realiza una aportación estable al contenido de los estados mentales de los que forma parte. Por tanto,

¹ Todas las notas en este párrafo pertenecen a Fodor (1987), p. 135.

² Fodor (1987), p. 136.

lo que HLP defiende es que los estados mentales (ya sean tipos o casos particulares), y no sólo sus objetos proposicionales, tienen partes constituyentes y, en definitiva, estructura constituyente.³ Por otro lado, si los estados mentales particulares son, como sugiere la tesis (ii), estados cerebrales particulares, entonces podemos añadir que los estados cerebrales particulares tendrán también partes constituyentes.

Los argumentos en favor de HLP descansan, según Fodor, en el siguiente prejuicio metafísico:

«No creo que haya mecanismos intencionales. Es decir, no creo que los contenidos por sí mismos determinen roles causales. En consecuencia, tiene que ser posible contar la historia completa de la causación mental (la historia completa de la implementación de las generalizaciones que las psicologías del deseo/creencia articulan) *sin referirse a las propiedades intencionales que se subsumen bajo tales generalizaciones.*» (Fodor (1987), p. 139)

Podemos, ciertamente, distinguir entre una interpretación débil y una interpretación fuerte de la tesis que en esta cita se propone. Según la interpretación débil, podríamos decir que todo lo que pretende afirmar Fodor cuando dice que no hay mecanismos intencionales es que la historia causal completa acerca de los procesos mentales podría contarse sin apelar a sus propiedades intencionales.

Hay razones para sospechar, sin embargo, que esta es una interpretación demasiado débil; pues, en primer lugar, el prejuicio metafísico se reduciría a un principio general, y más bien inocuo, de supervinencia de lo mental respecto a lo físico; y, en segundo lugar, la referencia a la inexistencia de mecanismos intencionales aparecería como inadecuada e irrelevante en un contexto en el que se pretende fundamentar HLP mediante una inferencia a la mejor explicación, ya que el hecho de que la historia causal completa pueda contarse sin apelar a propiedades intencionales no excluye por sí mismo la posibilidad de que surjan historias causales alternativas en las que intervengan propiedades intencionales.

De acuerdo con esto, podríamos proponer una interpretación más interesante, y también más fuerte, del texto de Fodor y, para ello, podríamos dejarnos guiar por el significado de los términos que forman la expresión 'mecanismo intencional'. La idea de mecanismo es, en general, la idea de un proceso causal regular desarrollado en el seno de un sistema relativamente cerrado. Mas, en este contexto, ¿qué añade el adjetivo 'intencional' al sustantivo 'mecanismo'? La segunda parte de la cita sugiere la siguiente respuesta: Un mecanismo intencional es un proceso causal regular que se desarrolla en el seno de un sistema relativamente cerrado en el que intervienen propiedades intencionales que condicionan causalmente su desarrollo. He de reconocer que, a mi entender, Fodor se compromete con el prejuicio metafísico en su versión fuerte, pues en él descansan los

³ Así, dice Fodor:

«De manera similar, 'mutatis mutandis', el 'P' que expresa la proposición P en la fórmula 'P' es un caso del mismo tipo que el 'P' que expresa la proposición P en la fórmula 'P & Q'. Si quisiéramos ser un poco más precisos, podríamos decir que la historia de HPL equivale a decir que (i) (algunas) fórmulas mentales tienen fórmulas mentales como partes; y (ii) las partes son 'transportables': las mismas partes que aparecen en muchas fórmulas.» (Fodor (1987), p. 137)

⁴ Cf. Fodor (1987), appendix.

tres argumentos que desarrolla Fodor en apoyo de HLP.⁴ Los tres argumentos tienen la estructura de una afirmación del consecuente o, dicho de otro modo, la estructura de una inferencia a la mejor explicación. Esta estructura argumentativa es, sin embargo, bastante frágil, pues sólo resulta válida en la medida en que no se descubra otro antecedente que, respaldado por evidencia independiente, dé lugar al mismo consecuente. En concreto, lo que vamos a discutir es si los modelos conexionistas nos proporcionan elementos adecuados para generar esa historia alternativa. Sin embargo, con el fin de saber lo que esta historia alternativa se supone que debe contarnos, necesitamos describir, previamente, los rasgos de nuestras capacidades cognitivas para cuya explicación se postula HLP.

En este sentido, Fodor considera que una teoría satisfactoria de la cognición debe ser capaz de dar cuenta de la responsabilidad causal de lo mental, así como de la productividad y sistematicidad de nuestros pensamientos, sin violar nuestro prejuicio metafísico fuerte. En este contexto, Fodor entiende, como cabría esperar, que HLP constituye en la actualidad la mejor hipótesis disponible para dar cuenta de estos rasgos de lo mental.⁵

En mi opinión, existen al menos dos estrategias diferentes para poner en cuestión la inferencia a la mejor explicación en la que descansa HLP. En primer lugar, podríamos poner en duda que HLP pueda proporcionarnos una buena explicación de los rasgos básicos de nuestras capacidades cognitivas. Esta línea de argumentación no intentaría atacar la posibilidad de dar cuenta de la sistematicidad y productividad del pensamiento gracias a HLP; sino que subrayaría la existencia de rasgos de nuestro pensamiento para los que HLP difícilmente podría constituir un modelo realista. Pensemos, por ejemplo, en el problema del esquema y en la explosión computacional que a la que conducen todos los intentos de solución de este problema en el seno de un modelo clásico. Esta explosión computacional es especialmente importante porque entra en conflicto con el número de neuronas y la velocidad a la que se producen las transmisiones en nuestro cerebro; de donde parece seguirse que lo que sabemos acerca de nuestro cerebro parece minar la plausibilidad HLP en tanto que modelo realista de nuestra mente, dado que no puede dar cuenta de algunos de los rasgos de nuestro pensamiento dentro de condiciones neurológicamente plausibles.

Una segunda estrategia, frente al argumento de Fodor, intentaría proponer una historia alternativa acerca de la eficacia causal de los estados mentales, así como de su productividad y sistematicidad. La posibilidad de narrar esta historia alternativa se ha hecho más plausible desde el surgimiento de los modelos conexionistas. En lo que sigue me limitaré a discutir este último punto, es decir, si hay alguna razón de principio en contra de la posibilidad de que los modelos conexionistas puedan constituir la base para esa historia alternativa, es decir, para un tratamiento de la productividad y sistematicidad del pensamiento, así como de la responsabilidad causal de las propiedades intencionales, que prescindiera de HLP.

⁵ Fodor afirma en *Psychosemantics* que, apoyándonos en la productividad y sistematicidad de los pensamientos, puede construirse el siguiente argumento en apoyo de HLP: (a) Las capacidades lingüísticas son productivas y sistemáticas en virtud de tener partes constituyentes; (b) los pensamientos son también productivos y sistemáticos; (c) por tanto, una inferencia a la mejor explicación nos permite afirmar que los pensamientos también tienen partes constituyentes.

2. Fodor y McLaughlin: el epifenomenalismo de lo mental en el modelo de Smolensk

En esta sección, consideraremos, en primer lugar, la crítica de Fodor y McLaughlin al modelo de Smolensky como un modelo que pueda dar cuenta de la eficacia causal de las propiedades intencionales, lo que nos llevará, en segundo término, la necesidad de ofrecer una caracterización más cuidadosa de las condiciones que propone Fodor para determinar la eficacia causal de una propiedad. Será precisamente la revisión de estas condiciones (que se llevará a cabo en la sección tercera) la que nos permitirá reivindicar el modelo de Smolensky frente a las objeciones de Fodor y McLaughlin.

2.1. Fodor y McLaughlin: Crítica del modelo de Smolensky

Conviene advertir, para empezar, que, en este artículo, no entraré a discutir en detalle el intento de Fodor y Pylyshyn (1988) de presentar los modelos clásicos y conexionistas como complementarios, si bien situando los modelos conexionistas a un nivel implementacional, y no cognitivo, de manera que estos modelos constituirían, como mucho, una de las implementaciones posibles que pueden recibir los modelos clásicos. Brevemente, diría que, incluso si asumimos la tesis de Fodor y Pylyshyn acerca de que los modelos conexionistas no se ubican a un nivel cognitivo, este hecho por sí mismo no excluye la posibilidad de elaborar una teoría representacional de la cognición cuyos estados básicos estén definidos a un nivel conexionista y donde las propiedades cognitivas se entiendan, sin postular HLP, como surgiendo de estados de un nivel inferior. En otras palabras, lo que parece crucial es determinar no si los estados conexionistas están a un nivel cognitivo, sino si hay alguna razón de principio en contra de la posibilidad de elaborar una teoría conexionista de la cognición que, a pesar de prescindir de HLP, dé cuenta de los rasgos básicos de nuestras capacidades cognitivas.⁶

De hecho, Fodor y McLaughlin intentan presentar esa razón de principio en su escrito «Connectionism and the Problem of Systematicity; Why Smolensky's Solution Doesn't Work». En este texto, se indica que HLP se postula con el fin de dar cuenta de la eficacia causal de los estados mentales en tanto que estados mentales, sin violar el principio metafísico fuerte que excluye la existencia de mecanismos intencionales; y se argumenta, por el contrario, que los modelos conexionistas no pueden preservar la eficacia causal de los contenidos mentales, una vez que se acepta el prejuicio metafísico fuerte.

En concreto, Fodor y McLaughlin centran sus críticas en la propuesta de Smolensky para dar cuenta de la sistematicidad de nuestros pensamientos dentro de un modelo conexionista. En general, podemos decir que Smolensky elabora su aparato conceptual en términos de vectores de productos de tensores y de superposición de vectores, si bien no es necesario entrar en los detalles. En lo que ahora nos interesa, basta con asumir, siguiendo a Fodor y McLaughlin, «...Que cada árbol clásico puede codificarse mediante un vector complejo de tal manera que los constituyentes del árbol se corresponden de algún modo regular con los componentes del vector.»⁷

⁶ Para una discusión detallada de este punto, cf. Corbí (en prensa).

⁷ Fodor y McLaughlin (1990), p. 197.

Otro punto importante que aceptan las dos partes implicadas es que, aunque los vectores complejos deben corresponder a pautas efectivas de actividad a lo largo de las unidades de la máquina, no hay ninguna necesidad de que los componentes del vector queden reflejados efectivamente en tales pautas de actividad, pues, entre otras razones, un mismo vector complejo puede descomponerse de modos diferentes; si bien algunos modos de descomposición pueden resultar más fructíferos que otros, desde el punto de vista de la simulación de árboles clásicos:

«... Un punto sobre el que el propio Smolensky es explícito y resulta vital en este respecto: Los componentes de un vector complejo no necesitan responder a pautas de actividad en la unidades efectivas de la máquina. En términos del propio Smolensky, los estados de actividad de los rellenadores y las unidades de rol pueden ser ‘imaginarios’, si bien los vectores de actividad últimos -los que no sirven ellos mismos como rellenadores o como componentes de rol de estructuras más complejas- deben de ser pautas de actividad que se den efectivamente a lo largo de las unidades de la máquina.» (Fodor and McLaughlin (1990), p. 197)

De todo esto se sigue, según Fodor y McLaughlin, que los componentes de los vectores complejos no son causalmente responsables del procesamiento que lleva a cabo la máquina conexionista, pues sería realmente milagroso que algo que no se da físicamente tuviese poderes causales, al menos si aceptamos nuestro prejuicio metafísico en cualquiera de sus versiones. Este hecho lleva a Fodor y McLaughlin a concluir que si, como parece seguirse del análisis de Smolensky, los componentes vectoriales son causalmente inertes y, por tanto, no pueden explicar la evolución de un sistema, entonces todavía no se ha mostrado cómo un modelo conexionista puede dar cuenta, en términos no-clásicos, de la sistematicidad y de la eficacia causal del pensamiento:

«[Smolensky] reconoce que los constituyentes de las representaciones mentales complejas no juegan ningún papel en la determinación de lo que ocurre cuando se instancian las representaciones... Parece, de este modo, que Smolensky debe de esconder en su manga una solución *no*-clásica para el problema de la sistematicidad; alguna solución que *no* dependa de asumir que los procesos mentales sean sensibles a su estructura constituyente.... No queda claro como los constituyentes (‘componentes’) de los vectores complejos pueden explicar su evolución [de la máquina] cuando, ‘*ex hypothesi*’, son como mucho causalmente inertes si no meramente imaginarios.» (Fodor and McLaughlin (1990), p. 200-1)

Me gustaría descartar, para empezar, una interpretación de la posición de Smolensky que parece resultar algo atractiva para Fodor y McLaughlin, y según la cual Smolensky estaría afirmando que los componentes vectoriales son entidades imaginarias.

Es cierto, como hemos visto, que Smolensky afirma explícitamente que los componentes vectoriales son entidades ‘imaginarias’ (si bien haciendo uso de las comillas), y esta indicación lleva a Fodor y McLaughlin a decir que los componentes vectoriales son «... como mucho causalmente inertes si no meramente imaginarios.» (aquí las comillas ya han desaparecido).

Estoy, por supuesto, de acuerdo con Fodor y McLaughlin en que entidades o propiedades meramente imaginarias, como ‘ser una sirena’, son causalmente inertes, dado que nunca se han instanciado. Sin embargo, no estoy seguro de que los componentes vectoriales sean imaginarios en este mismo sentido.

A primera vista, parece que el hecho de que puedan elaborarse diferentes análisis composicionales para el mismo vector complejo no impide que tengamos criterios para determinar, en el seno de cada análisis composicional, si un cierto componente vectorial está o no instanciado. Obviamente, los componentes vectoriales sólo serían causalmente responsables cuando estuviesen instanciados, mas, en contra de lo que ocurre con ‘ser una sirena’, podemos afirmar sin reparo que hay instanciaciones de componentes vectoriales.

Fodor y McLaughlin podrían replicar, sin duda, que se necesita un poco más para garantizar que una propiedad sea causalmente eficaz, y que estas condiciones adicionales no quedan satisfechas por la propuesta de Smolensky. Esta es precisamente la cuestión que pretendo examinar en las páginas que siguen.

2.2 Fodor: Condiciones suficientes para la eficacia causal

Una discusión detallada del punto anterior requiere, sin duda, una caracterización previa de cuándo se puede decir que una propiedad es causalmente responsable, y cuándo podemos excluir una propiedad como causalmente irrelevante. En este sentido, podemos empezar considerando la propuesta de Fodor en «Making Mind Matter More», donde se establece explícitamente una condición suficiente para que una propiedad sea causalmente responsable:

Condición I: «P es causalmente responsable si es una propiedad en virtud de la cual los individuos se subsumen bajo leyes causales.» (Fodor (1990b), p. 143)

Fodor entiende que esta condición suficiente no es circular a pesar del uso de la palabra ‘causal’ en el ‘explanans’, pues, en su opinión, puede que, al fin y al cabo, todas las leyes sean causales o, en cualquier caso, siempre podría ofrecerse una caracterización no-circular de ley causal en términos, por ejemplo, de sucesión temporal, asimetría, etc.

De todos modos, la condición I necesita complementarse con una segunda condición que tenga en cuenta la distinción entre las leyes de la ciencia básica y las leyes de las ciencias especiales. En este sentido, habría que decir que, mientras que en la ciencia básica no tiene sentido plantearse ninguna cuestión adicional acerca de cómo se implementan sus leyes, las ciencias especiales están siempre abiertas a la cuestión acerca de los mecanismos de la ciencia básica que explican cómo el antecedente de una ley no-básica causa su consecuente. Esta situación se deriva directamente de nuestro compromiso con el prejuicio metafísico fuerte, y daría lugar a la siguiente condición para la responsabilidad causal de una propiedad:

Condición II: «Pero una diferencia metafísicamente interesante entre leyes básicas y no-básicas es que, en el caso de las últimas pero no de las primeras, es necesario que haya *un mecanismo en virtud del cual* la satisfacción de su antecedente cause la satisfacción de su consecuente. Si ‘Fs causan Gs’ es básica, entonces no hay ninguna respuesta a la cuestión de *cómo* Fs causan Gs; simplemente lo hacen, y el hecho de que

lo hagan forma parte de los hechos que-no-requieren-explicación-adicional acerca del modo en el que el mundo está organizado.» (Fodor (1990b), p. 144)

En mi opinión, sin embargo, el análisis de Smolensky concerniente a la conducta de un sistema conexionista en términos de vectores de productos tensoriales y de superposición de vectores, así como su descomposición reflejando la estructura de los árboles clásicos, satisface la condición suficiente para la eficacia causal indicada por Fodor. En primer lugar, cabe indicar que las propiedades de los componentes vectoriales se subsumen bajo leyes causales si las propiedades de las partes constituyentes clásicas lo hacen, pues hemos asumido que puede establecerse una correspondencia estable entre ambas. Podemos, por tanto, concluir que las propiedades de los componentes vectoriales satisfacen la condición I.

En segundo término, el procesamiento al nivel de las pautas de conectividad entre las unidades puede verse como el mecanismo que garantiza la validez de las leyes que vinculan las propiedades de los componentes vectoriales, dado que, según el modelo de Smolensky, los vectores complejos encuentran su instanciación en las pautas de conectividad entre unidades. En definitiva, podemos concluir que las propiedades de los componentes vectoriales satisfacen también la condición II para la existencia de eficacia causal.

Se hace necesaria, en consecuencia, una lectura más restringida de la condición II si Fodor y McLaughlin quieren negar la eficacia causal de las propiedades de los componentes vectoriales. Nos referiremos a esta versión fuerte de la condición II como 'Condición II*'. Así, podríamos decir que la condición II* requiere la instanciación física de las propiedades causalmente responsables, de manera que pueda establecerse una relación sistemática entre los componentes de los tipos de estado no-físico y físico que se instancian. Parece claro que el análisis de Smolensky no satisface esta condición instanciadora fuerte pues, como vimos, no se puede establecer ninguna relación sistemática entre los componentes vectoriales y los componentes de las pautas de actividad.

Asumo, en definitiva, que Fodor y McLaughlin se comprometen con la condición II*, dado que este compromiso les permite concluir: (i) que Smolensky sólo consigue dar cuenta de la sistematicidad en un modelo conexionista al precio de la inercia de las propiedades intencionales; y (ii) que HLP es la mejor explicación disponible tanto para la sistematicidad como para la eficacia causal de los contenidos mentales, pues HLP no sólo satisface la condición I, sino que se compromete directamente con la condición II*, en la medida en que reivindica que los estados mentales mismos, y no sólo sus objetos intencionales, tienen partes constituyentes.

3. Un tratamiento alternativo de la responsabilidad causal

En esta sección, intentaré mostrar que la condición II* es demasiado fuerte como condición necesaria para la responsabilidad causal. En segundo término, defenderé un tratamiento alternativo de la responsabilidad causal que, aparte de concordar con nuestras intuiciones, parece disolver el conflicto entre los modelos conexionistas y la responsabilidad causal de las propiedades intencionales.

3.1. Una objeción a la condición II*

Una propiedad no-física es causalmente responsable, según la condición II*, sólo si puede establecerse, al menos en principio, una relación sistemática entre la instanciación de esta propiedad y la instanciación de ciertas propiedades físicas⁸. Pienso, sin embargo, que una condición necesaria tan fuerte entra en conflicto con nuestras intuiciones acerca de la responsabilidad causal.

Para empezar, podemos indicar que nosotros adscribimos responsabilidad causal a todo tipo de propiedades en las ciencias especiales sin tener en cuenta si realmente puede establecerse esa relación sistemática. Se pueden señalar, además, razones de principio para dudar acerca de la posibilidad de establecer tales relaciones tipo-tipo, pues, en múltiples ocasiones, tales relaciones exigirán la introducción de disyunciones indefinidamente abiertas de propiedades físicas, y este hecho nos llevará a algo muy cercano a una mera correlación caso-caso.

Recordemos, por ejemplo, cuando la selección nacional de Alemania Occidental marcó el gol de la victoria en la final de la Copa del Mundo de 1990. Ríos de alegría recorrieron las calles de Alemania, los jugadores se abrazaban entusiasmados. Pero las consecuencias de este gol irán mucho más lejos: incrementaría el orgullo nacional de los alemanes, triplicaría la prima que recibirían jugadores y entrenadores, aumentaría la cotización de los jugadores alemanes en general, se organizarían recepciones multitudinarias. Nadie en su sano juicio dudaría de la eficacia causal de ese gol y, en general, de los goles que dan la victoria en la final de una Copa del Mundo.

Y, sin embargo, este hecho tan trivial parece poner en apuros al fiscalista fuerte, a la condición II*. Pensemos, ¿podemos, acaso, especificar un conjunto de propiedades físicas que constituyan condición suficiente de la realización de un gol? A primera vista, la respuesta es sencilla: basta con que el balón traspase la raya de la portería. Aunque habría que ser un poco más cautelosos: no basta con traspasar la raya, tendrían que concurrir otras circunstancias, por ejemplo, que el hecho en cuestión ocurriese durante un partido de fútbol y no durante un entrenamiento. Hay que reconocer, con todo, que también hay entrenamientos tan serios que se llaman partidos de entrenamiento y donde los goles sí que se computan. Lo que esta ambigüedad parece sugerir es que también podemos encontrarnos con dificultades a la hora de indicar un conjunto de propiedades físicas que constituyan condición suficiente de la realización de un partido de fútbol. De hecho, habría que distinguir entre partidos de entrenamiento, partidos oficiales, partidos amistosos, partidos de liga, partidos de copa etc. Todas estas distinciones tienen su relevancia causal: la prima no es la misma si se gana la Liga que si se gana la Copa, si se vence en un partido amistoso que si el partido es oficial; las emociones que agitan a las masas en los estadios y el aspecto que presentan sus gradas son bastante diferentes. Distinguimos con facilidad entre todos estos tipos de partido atendiendo a los decretos de los diferentes órganos rectores del fútbol nacional e internacional, pero ¿cómo indicar condiciones físicas suficientes para hacer tales distingos?

⁸ Las propiedades físicas que estamos considerando están constituidas muy a menudo por conjuntos disyuntivos de propiedades físicas.

Habría que recordar, por otra parte, que no estábamos hablando de un gol cualquiera ni de un partido cualquiera, estábamos hablando del gol que daba la victoria en la gran final. Los goles que han dado la victoria en la final de las diferentes ediciones de la Copa del Mundo de Fútbol han sido de muy distinta factura, se han visto rodeados de condiciones muy diferentes, los organismos organizadores de la Copa y del partido han ido variando, también las normas del fútbol que obligan a conceder o invalidar ciertos goles etc. Parece imposible encontrar un conjunto de propiedades físicas que sea condición suficiente de la instanciación de la propiedad 'ser el gol de la victoria de una final de la Copa del Mundo de Fútbol' y, sin embargo, esta propiedad tiene una exagerada eficacia causal. Esto no quiere decir que para que algo sea un gol de la gran final no sea *necesario* que ocurran ciertos eventos físicos. Ese no es el problema que aquí se discute. Lo que más bien se sugiere es que no hay un conjunto de eventos físicos que sea condición *suficiente* para la instanciación de tales goles.

En términos generales, podríamos decir que lo que en principio se presentaba como una relación tipo-tipo entre un conjunto de condiciones físicas y la propiedad de 'ser un gol' o 'ser el gol que da la victoria en la final de la Copa del Mundo de Fútbol', se aproxima cada vez más a un relación caso-caso. Así, para determinar que algo sea un gol no basta con indicar que una forma esférica con ciertas características físicas ha traspasado una raya entre dos postes y un largero, hace falta observar la completa distribución de organismos en un rectángulo, hay que extender nuestras consideraciones más allá del estadio y describir los elementos físicos implicados en la multitud de decisiones institucionales que preceden a un partido, a la organización de una competición, etc; de manera que pequeñas variaciones en lugares muy alejados del estadio pueden hacer que algo no sea un gol o que no sea un gol válido. Además, lo que en un momento era un gol válido podría dejar de serlo por la alteración del reglamento del fútbol y, en consecuencia, lo que un momento del tiempo eran condiciones físicas suficientes pueden dejar de serlo, a no ser, por supuesto, que en la enumeración de esas condiciones se incluyan más y más elementos del mundo, con lo que nos iremos acercando peligrosamente al establecimiento de una relación caso-caso en la medida en que las condiciones físicas de instanciación de propiedades como 'ser un gol' parecen incluir aspectos muy extensos del mundo si no el universo entero, pues pequeñas alteraciones físicas en algún lugar del mundo pueden provocar que lo que era un gol válido deje de serlo. Al final, lo único que parece que podemos decir es que en un mundo físicamente idéntico a este en todos sus aspectos coincidiría con este en todas sus propiedades futbolísticas. Pero esto no es más que el principio general de supervinencia de las propiedades no-físicas del mundo respecto de sus propiedades físicas.

Se podría replicar, con todo, que no está claro que esta circunstancia se dé en ciencias naturales especiales como la geología o la biología. Pienso, sin embargo, que estas ciencias especiales se verían afectadas también por mi argumento, aunque en menor grado. Así, en la medida en que mutaciones en las propiedades físicas de objetos individuados no-físicamente son compatibles con la estabilidad de sus propiedades no-físicas, puede haber casos en los que cambios físicos en otras áreas del universo hagan que un conjunto de condiciones físicas deje de ser suficiente para la posesión de ciertas propiedades no-físicas. Este hecho introduce, en definitiva, los dos factores que, en el

caso de las categorías económicas, parecían socavar el carácter tipo-tipo de la relación que la condición II* postula, a saber: la relatividad a un cierto tiempo y la necesidad de incorporar más y más rasgos del mundo.

De todos modos, el hecho de que teorías distintas parcelen el mundo de modo diferente es compatible, a mi entender: (i) con el carácter básico de las taxonomías físicas y, por tanto, con la posibilidad de establecer diversos tipos de conexiones entre las diferentes taxonomías, incluso si ya no podemos esperar que estas conexiones sean tan fuertes como se sugiere en la condición II*; y (ii) con la responsabilidad causal de las propiedades no-físicas. Así, en la próxima sección, presentaré un tratamiento de la responsabilidad causal que aspira a integrar todos estos elementos.

3.2 *Un tratamiento alternativo*

Me gustaría proponer el siguiente conjunto de requisitos como una condición suficiente para la responsabilidad causal de una propiedad:

- (a) que la propiedad forme parte de una ley causal
- (b) que la propiedad esté físicamente instanciada o, al menos, sea superviniente respecto a las propiedades físicas del mundo, y
- (c) que no exista hay ningún conjunto alternativo de leyes causales más generales y rigurosas que explique aproximadamente el mismo tipo de fenómenos, y en los que la propiedad en cuestión no aparezca.

El requisito (a) reproduce literalmente la condición I establecida por Fodor. A primera vista, este requisito parece excluir la posibilidad de que las propiedades intencionales sean causalmente eficaces, dado que se tiende a pensar que no hay leyes estrictas en psicología ni tampoco leyes psicofísicas, pues todas las generalizaciones que se puedan proponer en estos campos han de incluir cláusulas ‘*ceteris paribus*’. Donald Davidson es, como sabemos, el filósofo que ha planteado con más rigor esta objeción. No obstante, Fodor ha propuesto una respuesta a esta objeción que encuentro satisfactoria: el argumento de Davidson es tan potente que afectaría no sólo a la psicología sino a todas las ciencias especiales, dado que todas ellas incluyen inevitablemente ese tipo de cláusulas, las cuales, por otra parte, no pueden eliminarse, al igual que ocurre en psicología, manteniéndose dentro del vocabulario de cada una de esas ciencias. Por tanto, si la estrategia de Davidson fuese correcta no sólo las propiedades intencionales dejarían de satisfacer este requisito, sino también las propiedades definidas por la geología, la biología, etc. En estas condiciones, me parece más apropiado ampliar nuestro concepto de ley causal y entender que la presencia de las cláusulas ‘*ceteris paribus*’ no descarta la posibilidad de que las propiedades intencionales, o las propiedades de cualquier ciencia especial, satisfagan este primer requisito.⁹

El requisito (c) nos ayuda a excluir del ámbito de las propiedades causalmente eficaces algunas propiedades que, a pesar de aparecer en algunas generalizaciones

⁹ Con esto no pretendo que se hayan resuelto todos los problemas en torno al carácter legaliforme de las generalizaciones psicológicas, sino únicamente indicar cómo se podría evitar una de las objeciones más inmediatas que se suele plantear (Cf. Fodor (1988), pp. 15 ff).

verdaderas, no nos sentimos inclinados a reconocerlas como causalmente eficaces. A mi entender, eso ocurre cuando disponemos de una teoría alternativa que da cuenta de esas generalizaciones en virtud de un sistema de principios más generales y rigurosos. Consideremos, por ejemplo, la siguiente generalización: «Si un niño tiene fiebre alta y le duele la garganta, probablemente alguno de sus hermanos se encontrará en la misma situación unos días después». Concedamos que esta generalización, o alguna semejante, es verdadera, con lo que las propiedades ‘tener fiebre alta’ o ‘doler la garganta’ cumplen el requisito (a) y, sin embargo, no me parece nada correcto decir que la causa de que un niño tenga fiebre alta y le duela la cabeza el viernes sea que su hermano estaba en esas circunstancias a principios de semana. Entiendo, por el contrario, que la propiedad causalmente eficaz, en este caso, es la transmisión de ciertos gérmenes bacteriológicos de un hermano a otro. Considero que lo que me induce a pensar que esta última es la verdadera causa es que nos encontramos con una teoría de la transmisión de enfermedades infecciosas que no sólo es más general y rigurosa que la recogida en la generalización con la que iniciábamos el ejemplo, sino que además da cuenta de aproximadamente el mismo tipo de fenómenos: el contagio de enfermedades.

Es importante resaltar, sin embargo, que el requisito (c) no disminuye las expectativas de que las propiedades intencionales tengan eficacia causal porque estamos muy lejos de disponer de una teoría alternativa que explique aproximadamente el mismo tipo de fenómenos que la psicología intencional.

Respecto al requisito (b), cabría empezar advirtiendo que, tras la discusión anterior acerca de los modos en que algunas propiedades de las ciencias especiales se relacionan con las propiedades físicas, debemos estar preparados para reconocer la existencia de diferentes grados de instanciación física compatible con nuestro prejuicio metafísico acerca de la inexistencia de mecanismos intencionales. En algunos casos encontraremos una correlación tipo-tipo entre conjuntos disyuntivos cerrados de propiedades físicas y propiedades no-físicas, pero, en otros casos, ese conjunto disyuntivo se hará más extenso y menos cerrado.¹⁰

¿Cuál es, no obstante, el límite en la flexibilidad de las posibles conexiones entre las propiedades físicas y no-físicas que estoy dispuesto a tolerar? La respuesta que, en mi opinión, se sigue del debate anterior es que la supervivencia de las propiedades no-físicas respecto a las propiedades físicas define los límites del requisito (b). De hecho, este requisito define la versión débil del prejuicio metafísico, que, a su vez, caracteriza a la posición fisicalista más endeble y la única versión del prejuicio metafísico que realmente considero digna de tal nombre. En realidad, este último principio forma parte del suelo rocoso de mi visión del mundo; en cierto modo, no puedo concebir ningún tipo de evidencia que pudiese convencerme de su falsedad. En cambio, no sólo me resulta plenamente comprensible que la tesis fisicalista fuerte sea falsa, sino que además he intentado mostrar por qué pienso que lo es. Otra cuestión que no discuto

¹⁰ En este contexto, el carácter cerrado de un sistema es claramente una cuestión de grado: por ejemplo, podríamos decir que las taxonomías biológicas están más cerradas desde un punto de vista físico que las categorías económicas.

aquí es si la condición fisicalista fuerte puede interpretarse no tanto como un requisito, sino como una estrategia de investigación tendente a incrementar el nivel de cohesión entre las diferentes teorías que componen nuestra concepción materialista del mundo.

La satisfacción de estos tres requisitos es, por supuesto, una cuestión de grado. Se puede esperar, por tanto, que la cuestión acerca de la responsabilidad causal de una cierta clase de propiedades no sea en todos los casos un asunto de todo o nada. En este sentido, es posible encontrar combinaciones diferentes en la satisfacción de los requisitos (a)-(c), de manera que cuanto menos quede satisfecho cada uno de ellos, menos nos sentiremos tentados a adscribir responsabilidad causal a la clase de propiedades que en cada caso se considere.

Finalmente, parece obvio que, si atendemos a mi tratamiento de la responsabilidad causal, los modelos conexionistas no son incompatibles con la responsabilidad causal de las propiedades intencionales, pues la propuesta de Smolensky no sólo satisface los requisitos (a) y (c), sino también el requisito (b) en una de sus versiones más restrictivas, pues, si bien no se concede la versión más fuerte de este requisito, al menos se defiende la existencia de relaciones tipo-tipo entre los vectores complejos y las pautas de actividad.

4. Conclusiones

Hemos indicado, al inicio, que el surgimiento de los modelos conexionistas puede representar una amenaza seria para HPL, pues todos los argumentos en defensa de HPL tienen la estructura de una inferencia a la mejor explicación y, en consecuencia, el nacimiento de un tratamiento alternativo, basado en evidencia independiente, tiende ciertamente a invalidar tales argumentos.

Fodor y McLaughlin replican, sin embargo, que una teoría conexionista de la cognición que abandone HPL no puede constituir tal planteamiento alternativo; pues tales modelos no son compatibles con la responsabilidad causal de las propiedades intencionales, dado que sólo HPL puede garantizar la eficacia causal de los estados mentales en tanto que estados mentales.

He intentado mostrar, sin embargo, que el argumento de Fodor y McLaughlin descansa en una concepción demasiado restrictiva de la responsabilidad causal, que conduce a negar la responsabilidad causal de muchas propiedades no-físicas definidas en el seno de ciencias especiales tales como la economía y la biología, a pesar de nuestra tendencia intuitiva a reconocerlas como causalmente relevantes.

Este hecho me llevó a proponer un conjunto alternativo de requisitos que definen una condición suficiente para la responsabilidad causal. Según esta propuesta, (i) se respeta el prejuicio metafísico débil concerniente a la inexistencia de mecanismos intencionales; (ii) quedan excluidas, en cambio, las propiedades que no resultan interesantes desde un punto de vista explicativo o que forman parte de teorías no-empíricas o metafísicas; (iii) las propiedades postuladas por la economía y la biología serán igualmente causalmente responsables; (iv) los modelos conexionistas son, en principio, compatibles con la responsabilidad causal de las propiedades intencionales; por lo que podemos concluir (v) que los argumentos de Fodor y McLaughlin no

consiguen hacer frente a la amenaza que la emergencia de los modelos conexionistas supone para la hipótesis del lenguaje del pensamiento.

Bibliografía

- CORBI, J.E. (en prensa), «Classical and Connectionist Models: Levels of Description», *Synthese*.
- FODOR, J. (1987), *Psychosemantics*, Cambridge, Mass., The MIT Press.
- FODOR, J. (1990), «Making Mind Matter More» en: Fodor, J., *Theory of content and other essays*, Cambridge, Mass., The MIT Press, pp. 137-159.
- FODOR, J. & PYLYSHYN, Z.W. (1988), «Connectionism and Cognitive Architecture: A Critical Analysis» in *Cognition*, v. 28, pp. 3-71.
- FODOR, J. & MCLAUGHLIN, B. (1990), «Connectionism and The Problem of Systematicity; Why Smolensky's Solution Doesn't Work» in *Cognition*, v. 35, pp. 183-204.
- SMOLENSKY, P. (1987), «The Constituent Structure of Mental States: A Reply to Fodor and Pylyshyn» en: *Southern Journal of Philosophy*, 26, 137-160.
- SMOLENSKY, P. (1988 b), «Connectionism, Constituency and the Language of Thought.» en: Loewer, B. & Rey, G. (eds.), *Meaning in Mind: Fodor and his Critics*, Oxford, Basil Blackwell.

Taula (UIB) núm. 15, 1992

El problema de la modularidad en el procesamiento auditivo y en el procesamiento del habla

Luis Enrique López Bascuas*

Universidad Complutense de Madrid

Resumen

En este trabajo se comentan algunas aportaciones experimentales que han resultado ser relevantes para dilucidar hasta qué punto el procesamiento temprano del habla puede ser considerado un módulo. No se pretende ser exhaustivo a la hora de presentar la evidencia disponible, sino desarrollar una línea argumental coherente que permita establecer algunas conclusiones que ayuden, al menos, a centrar la polémica.

Abstract

In this work relevant experiments are commented concerning the existence of a speech module. The paper is not intended to provide an exhaustive review of the evidence. Instead, it tries to develop a coherent line of argument that may lead to some relevant conclusions in order to set the debate in proper terms.

* Este trabajo ha sido financiado, parcialmente, por la Dirección General de Investigación Científica y Técnica, proyecto PB87-0531. Quiero agradecer a R.P. Fahey y B.S. Rosner (Universidad de Oxford) y a J.E. García-Albea (Universidad Complutense) toda la ayuda prestada en el curso del mismo.

Evidencia primitiva a favor de la existencia de un módulo de procesamiento de habla

El debate en torno al problema de la modularidad (Fodor, 1983) ha tenido una larga tradición en el campo del procesamiento del habla. Ya en los años 50, la singularidad del habla fue defendida enfáticamente en los laboratorios Haskins. Sin embargo no es menos cierto que precisamente en este dominio la hipótesis de la modularidad ha sido duramente criticada.

La evidencia psicológica primitiva que apoyaba la idea de que el habla era «especial» provino de estudios realizados con habla sintética en los ya mencionados laboratorios Haskins. En 1957, Liberman, Harris, Hoffman y Griffith sintetizaron un conjunto de estímulos iguales en todas sus características físicas excepto en la frecuencia de inicio de la transición del segundo formante. Variando en pasos iguales ese valor frecuencial, construyeron un continuo físico que, desde el punto de vista perceptivo, cubría los diferentes lugares de articulación de las oclusivas sonoras (/b/ /d/ /g/). Las frecuencias centrales de las partes estacionarias de los formantes aproximaban la vocal final al sonido /e/. La tarea de los sujetos consistía en identificar los estímulos. Sorprendentemente, aunque los estímulos formaban un continuo físico, los sujetos los clasificaban conforme a tres categorías bien definidas (esto es, las probabilidades de identificación variaban bruscamente en vez de gradualmente). Más aun, Liberman *et al.* trataron de determinar si los sujetos eran capaces de detectar diferencias entre estímulos que pertenecían a la misma categoría. Para ello utilizaron una tarea de discriminación ABX. A y B son dos estímulos diferentes del continuo y X puede ser o A o B. La tarea del sujeto es decidir si X es A o B. Los resultados mostraron que los sujetos sólo podían discriminar entre estímulos que habían sido extraídos de distintas categorías percibidas pero no podían discriminar (al menos, difícilmente podían) entre estímulos que pertenecían a la misma categoría percibida. Este patrón básico de resultados es lo que se conoce como «Percepción Categórica».

Estos hallazgos concernientes al lugar de articulación fueron extendidos a otros contrastes fonológicos. Entre otros, el contraste sordo/sonoro ha sido estudiado extensamente. Existen muchas claves acústicas que potencialmente pueden marcar la distinción sorda/sonora pero una de las más usadas en la investigación sobre habla es el VOT (voice onset time). El VOT es el intervalo que transcurre entre la relajación de los articuladores tras el periodo de oclusión y el inicio de la vibración glotal. Ambos eventos pueden ser localizados en una representación espectrográfica de la señal del habla. La relajación de los articuladores está marcada por un estallido breve de energía en la región de frecuencias medias o altas, en tanto que la actividad glotal se manifiesta en el inicio de la periodicidad en bajas frecuencias. Lisker y Abramson (1970) obtuvieron para esta dimensión el patrón clásico de Percepción Categórica.

Dado que para otro tipo de estímulos los sujetos suelen mostrar un rendimiento mucho mayor en tareas de discriminación que en tareas de identificación y que, para ciertas señales no lingüísticas (por ejemplo Liberman *et al.*, 1961), se encontró percepción continua, la percepción categórica se tomó como evidencia de que algún tipo de procesador especial computaba las señales de habla. Aquí yacen los orígenes de la Teoría Motora de la Percepción del Habla, teoría que en su formulación más reciente (Liberman y Mattingly, 1985) apoya una organización modular de los sistemas de entrada.

Evidencia en contra de un módulo específico: los trabajos con continuos de no habla

Sin embargo, Miller *et al.* (1976) trataron de demostrar que la percepción categórica podría ser una propiedad general de la conducta sensorial, poniendo así en crisis la afirmación de que existe un modo de percepción único para el habla. Para ello construyeron un continuo de no habla con ruido termal filtrado y un zumbido (una función de onda cuadrada) variando en pasos iguales el inicio relativo del ruido y el zumbido. En los experimentos de identificación se presentaron a los sujetos sonidos aislados con asincronías de -10 a +80 ms. Los sujetos tenían que utilizar las etiquetas «no-ruido» o «ruido» para identificar los estímulos. En la tarea de discriminación se presentaba a los sujetos dos estímulos iguales y uno distinto. El sujeto debía indicar si el sonido diferente era el primero, el segundo o el tercero. Los resultados mostraron funciones de identificación con pendientes muy pronunciadas y funciones de discriminación con máximos locales en torno a la frontera entre categorías (en tanto que ejecución cercana al azar en el resto de los puntos).

Pisoni (1977) siguió la misma línea argumental en una serie de cuatro importantes experimentos. Su continuo de no habla trataba de simular otra clave acústica que marca el rasgo de sonoridad, a saber, la frecuencia de inicio del primer formante relativa a la frecuencia de inicio del segundo. Para ello utilizó dos funciones sinusoidales con frecuencias de 500 y 1500 Hz. El continuo se construía variando, en pasos de 10 ms., el inicio del tono grave en relación con el tono agudo. Los puntos extremos del continuo consistían en el tono grave precediendo al agudo 50 ms. y en el tono agudo precediendo al grave 50 ms.

La primera parte del experimento 1 consistió en dos sesiones de entrenamiento donde los sujetos asociaban a cada sonido un botón de respuesta. Acto seguido los sujetos se sometían a una tarea de identificación, usando las respuestas que habían aprendido en la sesión de entrenamiento, y a una tarea ABX de discriminación.

Pisoni obtuvo funciones de identificación como las pronosticadas si se asume percepción categórica al menos para 5 de sus 8 sujetos. Hay que resaltar que la frontera entre categorías no ocurría a los 0 ms. de asincronía sino que estaba desplazada hacia la derecha, esto es, hacia la categoría en la que el tono grave se retarda con relación al agudo. En este punto existen dos posibles explicaciones. Una en términos de enmascaramiento (mayor enmascaramiento de las altas frecuencias por las bajas frecuencias -la amplitud del tono grave era 12 dB mayor que la del agudo-) o una limitación en el procesamiento del orden de eventos temporales. Un experimento piloto no mostró cambios en la frontera entre categorías cuando se variaba la amplitud relativa de los tonos, sugiriendo que la hipótesis en términos de enmascaramiento podía ser descartada.

En las funciones de discriminación aparecieron los picos habituales; así pues se encontró un nuevo ejemplo de percepción categórica con señales de no habla. Sin embargo la cuestión que se suscita es si ese patrón de resultados podría ser el resultado de algún proceso de etiquetado adquirido en las sesiones de entrenamiento. Este es un punto importante porque, de acuerdo con el argumento, los picos deben tener un fundamento puramente psicofísico. Por ello en el experimento 2 los sujetos realizaron la tarea de discriminación ABX sin sesiones de entrenamiento previas.

Los resultados muestran el pico ya encontrado en torno a los +20 ms. Esto significa que los picos no aparecen debido exclusivamente a los procedimientos de entrenamiento, sino que existen áreas en el continuo psicofísico para las cuales el sistema es más sensible. Los datos mostraron también un pico secundario en torno a los -20 ms. Parecería haber tres regiones diferentes: el tono agudo precede al grave 20 ms. o más, el tono grave precede al agudo 20 ms. o más y una región de simultaneidad entre -20 y +20 ms.

Si esto es así es posible avanzar dos predicciones adicionales. En primer lugar, si ofrecemos a los sujetos la posibilidad de utilizar tres etiquetas en vez de dos, deberíamos obtener tres categorías con las fronteras localizadas en -20 y +20 ms. En segundo lugar, si la tarea de los sujetos es decidir si al inicio oyen uno o dos eventos, deberán contestar «dos» sólo cuando los estímulos estén fuera del rango (-20,+20). Ambas predicciones fueron comprobadas en los experimentos 3 y 4.

Todos estos resultados sugieren que la distinción sorda/sonora podría estar basada en las capacidades generales del sistema auditivo, desestimando así cualquier explicación basada en módulos específicos de procesamiento del habla.

Así, Miller *et al.* propusieron una explicación del fenómeno de la percepción categórica en términos de un umbral de enmascaramiento. Si se toman dos estímulos que están por debajo del umbral, no podrán ser distinguidos y los sujetos contestarán al azar, si ambos estímulos están por encima del umbral, entonces podrán ser detectados y las discriminaciones de los sujetos seguirán la ley de Weber. El pico en la función discriminación aparece cuando se toma un estímulo que está por debajo del umbral y otro que está por encima. Esto marcaría una discontinuidad natural que haría sencillo el establecer etiquetas (Cf. Miller *et al.*, 1976).

Por su parte, Pisoni sugirió que el fenómeno de la percepción categórica refleja una limitación básica del sistema auditivo para la estimación del orden en que se producen eventos temporales. Habría una suerte de umbral para este tipo de tareas en torno a los 20 ms. Esto es consistente con los hallazgos de Hirsh (1959) (aunque véase Rosen y Howell, 1987). Los distintos sistemas fonológicos, por tanto, habrían aprovechado esta situación para situar las categorías fonológicas en distintos lados de la discontinuidad natural, con lo que las harían claramente discriminables (Cf. Pisoni, 1977).

Con independencia de los detalles, todas estas explicaciones de la distinción sorda/sonora apoyan fuertemente la idea de una facultad horizontal encargada de tratar todo tipo de señales acústicas.

Nueva evidencia a favor de un módulo específico de procesamiento del habla

Sin embargo, existe también un buen cúmulo de información en el que no encaja la asunción de que barreras psicoacústicas naturales explican las fronteras de VOT. Para empezar por lo más obvio hay que señalar que tales fronteras no son fijas sino altamente flexibles. Varían con el contexto que proporcionaría otro estímulo presentado con el estímulo test (efectos de contraste, de rango, de adaptación selectiva) y también varían con las peculiaridades de la estructura interna del estímulo (efectos de la velocidad de habla y relaciones de intercambio entre claves acústicas) (Repp y

Liberman, 1987). Por ejemplo, es sabido que se da una relación de intercambio entre VOT y la frecuencia de inicio del primer formante. Así la frontera de VOT se alarga según se disminuye la frecuencia de inicio del primer formante. Este resultado no se ha obtenido con análogos de no habla (Summerfield, 1982; aunque véase Diehl, 1987).

Otra fuente de evidencia importante que permite dilucidar si el efecto de la frontera fonémica se debe a una facultad horizontal o vertical proviene del campo de la investigación translingüística. Parece claro que, si algún tipo de umbral psicoacústico está involucrado en el efecto, entonces la posible localización de la frontera fonémica estaría altamente constreñida para los distintos idiomas. Así, un primer paso en la investigación consistió en determinar dónde de hecho sitúan las fronteras los hablantes de diversas lenguas para los distintos contrastes fonológicos. Las diferencias entre el inglés y el español son particularmente interesantes para el problema que nos ocupa. En contraste con su contrapartida inglesa, la oclusiva sonora española se produce con actividad glotal antes del periodo de oclusión de los articuladores, esto es, la /b/ española es presonora. Williams (1977), estudió tanto los aspectos de producción como los de percepción del contraste sordo/sonoro en español (dialectos latinoamericanos) y encontró que, primero, se daba una estrecha correspondencia entre los aspectos de percepción y producción y, segundo, que las fronteras entre categorías y los picos de discriminación diferían de los que se habían descrito en hablantes ingleses.

En particular los ingleses exhibían las fronteras y picos en torno a un VOT de 20 ms. Sin embargo, para el grupo de español ambos valores estaban en torno a -5 ms. Estos resultados no son consistentes con los argumentos presentados anteriormente en tanto que, en ese rango de valores de VOT, se espera una ejecución en discriminación cercana al azar. No obstante, los hablantes de español mostraron un pico secundario en torno a los 20 ms. Parece, por tanto, que pudieron discriminar entre oclusivas aspiradas y no aspiradas en términos puramente auditivos, en tanto que la discriminación correspondiente a los -5 ms. de VOT podría ser de naturaleza fonémica.

Un avance importante para esclarecer estas cuestiones se puede encontrar en un trabajo reciente de Kewley-Port, Watson y Foyle, 1988. Estos autores se propusieron mostrar que el fenómeno de la percepción categórica y las capacidades últimas de resolución del sistema auditivo son aspectos independientes en buena medida. Esto es así tanto para señales de habla como para señales de no habla.

Según estos y otros autores conviene distinguir entre dos tipos de tareas: tareas de alta incertidumbre y tareas de mínima incertidumbre psicofísica. Las primeras suponen la presentación de una gran variedad de estímulos en orden aleatorio, en las segundas un mismo par es presentado dentro de una serie de ensayos. En tanto que las primeras serían recomendables cuando estamos interesados en el modo habitual de percibir sonidos complejos, las segundas serían más convenientes si nuestro objetivo es determinar las capacidades básicas (el poder de resolución) del sistema auditivo. Sachs y Grant (1976) ya habían mostrado cómo aumenta el rendimiento usando tareas de mínima incertidumbre en un continuo de VOT con consonantes velares.

Así pues, el argumento es bastante directo: si la ejecución de los sujetos cuando detectan cambios en VOT sigue la ley de Weber (esto es, necesitamos un incremento mayor de estimulación para notar un cambio según aumentamos en VOT) entonces la

percepción categórica no puede explicarse en términos puramente psicoacústicos (por ejemplo, aludiendo a las capacidades de resolución temporal del sistema auditivo).

Cuando usaron las tareas habituales de identificación y ABX obtuvieron los resultados convencionales. Sin embargo cuando usaron una tarea de incertidumbre mínima encontraron una mejora general en las discriminaciones y, lo que es más interesante, una discriminación cada vez más fina según se acortaba la duración del VOT. Este resultado es consistente con la ley de Weber.

Siguiendo los mismos procedimientos repitieron el experimento con un continuo de ruido-zumbido similar al usado por Miller *et al.* (1976). Una vez más se obtuvieron amplias mejoras en ejecución y las funciones eran monótonas decrecientes.

Estos resultados muestran claramente que no hay evidencia fuerte que permita sostener la idea acerca de un umbral psicoacústico que limita la discriminación del orden temporal y, por tanto, otras explicaciones han de ofrecerse para dar cuenta del fenómeno de frontera entre categorías.

Algunas consideraciones en torno al módulo auditivo

Sin embargo, ahora surge una pregunta inesperada: ¿qué determina las fronteras y los picos que se obtienen con sujetos no entrenados en continuos de no habla? De acuerdo con los resultados comentados de Kewley-Port *et al.*, las fronteras podrían haber aparecido en un rango amplio de valores. Dado que de hecho han aparecido en posiciones semejantes a las que se obtienen con habla, una posible hipótesis es que la experiencia con el idioma determina las fronteras y picos en el no habla. Pero si esto es así, información específica del módulo lingüístico está siendo utilizada por el módulo auditivo con lo cual la supuesta modularidad queda violada asumiendo que las computaciones que originan las fronteras son internas a los módulos.

Si tenemos en cuenta los resultados de los estudios translingüísticos citados, tenemos un modo bastante directo de comprobar si la hipótesis de que la experiencia con el idioma determina las fronteras del no habla es sostenible. Como dijimos, las fronteras entre categorías para hablantes de español se sitúan en un VOT menor que las de los hablantes de inglés. Así pues, si la hipótesis es cierta hay que esperar que las fronteras en continuos de no habla sean también menores en hablantes españoles. Fahey, López-Bascuas, García-Albea y Rosner (en preparación) pusieron a prueba esta hipótesis. Aunque el trabajo no está completado, merece la pena comentar los resultados provisionales obtenidos.

En este trabajo se utilizaron dos grupos de sujetos (monolingües españoles y monolingües ingleses) y tres continuos estímulares (uno de habla y dos de no habla).

El continuo de habla estaba constituido por 10 estímulos que perceptivamente variaban de /ba/ a /pa/. El continuo fue generado con una versión modificada del sintetizador de Klatt (1980) desarrollada en el I.B.M. UK Science Center. El punto extremo correspondiente al estímulo /ba/ tenía 35 ms. de presonoridad, el punto extremo correspondiente a /pa/ tenía un VOT de 55 ms. Las transiciones de los tres primeros formantes no eran lineales con el objeto de obtener habla más natural. F4 y F5 se mantuvieron constantes.

Los continuos de no habla trataron de simular los de ruido-zumbido y tonos ya comentados. El continuo de ruido-zumbido fue también creado con el sintetizador. Para obtener el ruido se excitó un filtro paso-todo con ruido de fricación; el zumbido se consiguió con una onda cuadrada de 100 hz. En un extremo del continuo el zumbido precedía al ruido en 35 ms., en el otro extremo el ruido precedía al zumbido en 55 ms.

Los tonos fueron creados con un programa en Pascal. El programa permitía definir la asincronía entre dos funciones sinusoidales, una de 500 Hz. y la otra de 1500 Hz. De nuevo los puntos extremos eran de 35 ms. (cuando el tono grave precedía al agudo) y 55 ms. (cuando el tono agudo precedía la grave).

Los sujetos pasaron por 200 ensayos de identificación y 360 ensayos de discriminación bajo el paradigma 2IAX (se presentaban dos estímulos del continuo y el sujeto tenía que decidir si eran iguales o diferentes). Sin entrar en detalles, los resultados hasta ahora disponibles indican que: 1) Existen diferencias significativas entre las fronteras de habla del grupo de españoles y el grupo de ingleses, por un lado, y entre el continuo de ruido-zumbido y el continuo de los tonos en el grupo de españoles, por otro lado. El resto de las comparaciones posibles arrojan diferencias que no son significativas. 2) Ninguna de las seis posibles correlaciones entre continuos (tres dentro de cada grupo de sujetos) resulta ser significativa.

La discusión de estos resultados conviene integrarla en el siguiente epígrafe de consideraciones generales.

Consideraciones finales

El trabajo anterior trataba de decidir si era plausible que la experiencia con un idioma determinara las fronteras que se establecen en continuos de no habla. Si así fuera no debería haber diferencias entre fronteras pertenecientes a distintos continuos dentro de cada idioma. Como hemos señalado el trabajo citado pone de manifiesto un contraejemplo a esta condición: los tonos y el ruido-zumbido difieren significativamente en el grupo de español.

Además, teniendo en cuenta que las fronteras de habla difieren significativamente, habría que encontrar diferencias también significativas entre los continuos de no habla a través de los idiomas. De nuevo encontramos un contraejemplo de esta condición: las fronteras en el ruido-zumbido no son diferentes a través de los idiomas.

Así pues, por un lado parece implausible que el efecto de frontera entre categorías pueda ser debido a propiedades generales del sistema auditivo y, por otro lado, no se ha encontrado evidencia convincente que apoye la idea de que las fronteras de no habla vienen determinadas por la experiencia con el lenguaje. Por lo tanto, la evidencia que aquí se presenta es compatible con una organización modular de los sistemas de entrada.

No obstante, sería poco sensato obviar la evidencia que apunta en la otra dirección así como algunas cuestiones que quedan pendientes de respuesta. Al menos dos pasos deben darse para tratar de resolver el conflicto que plantea la evidencia disponible.

En primer lugar, habría que retomar con rigor la idea de que en los resultados experimentales se mezclan tanto factores sensoriales como factores decisionales. Es

este un aspecto central que no puede obviarse al interpretar los datos. En este sentido nótese que el efecto de frontera entre categorías podría deberse a cualquiera de los dos factores aludidos. Por ejemplo, podría ocurrir que las densidades de la señal que corresponden a las categorías del habla estuvieran claramente separadas o que dichas densidades tuvieran una varianza pequeña. No obstante, también podría ocurrir que las categorías de habla generaran un criterio de respuesta de varianza pequeña en la frontera (Rosner, en prensa).

Los métodos convencionales que siguen el modelo de la teoría de detección de señales tienen una seria limitación: asumen que la varianza del criterio es uniforme y habitualmente cero. Por tanto, es necesario un nuevo modelo que permita varianzas no uniformes del criterio de respuesta. Recientemente se ha hecho un avance importante en esa dirección (Rosner, en prensa). Este tipo de análisis podría revelar diferencias en el procesamiento de señales de habla y no habla.

En segundo lugar, habría que afrontar el problema de encontrar una representación temprana adecuada para las señales de habla y de no habla. Este paso es esencial para desarrollar modelos fuertes de procesamiento auditivo y lingüístico. Los procesos que hayan de ser ejecutados en estadios ulteriores de procesamiento podrían depender críticamente del formato y la información especificados en esa primera representación del estímulo físico. Además, algunos problemas importantes como el de la variabilidad de la señal, podrían resolverse, al menos parcialmente, si se desarrollaran algoritmos que proporcionaran la representación requerida. Más aún, diferentes sistemas de entrada podrían ser caracterizados por los diferentes tipos de representaciones que computan con el fin de resolver el problema perceptivo con el que se enfrentan. Por supuesto, este punto es especialmente relevante para el caso de las señales acústicas y de habla, en tanto que potencialmente podrían compartir muchas características físicas. Así, para identificar a qué dominio pertenece una señal, se debería examinar el tipo de representación inicial que computa.

La naturaleza específica de la representación tiene que determinarse, por un lado, haciendo referencia a la tarea computacional que el sistema ejecuta (esto es, por referencia al problema que el organismo resuelve) y, por otro lado, refiriéndose al conjunto de constricciones que operan sobre los procesos posibles (Marr, 1982). Así existen constricciones matemáticas (cierto teorema podría avisarnos de que una determinada transformación formal provoca un resultado no deseado), constricciones físicas (relativas a la naturaleza del estímulo que ha de ser procesado) y constricciones informativas (qué información puede consultar el proceso).

Por ejemplo, es bien sabido que las claves acústicas que marcan contrastes relevantes en el habla están distribuidos tanto en el dominio del tiempo como en el de la frecuencia. Por tanto, cualquier representación del habla, para ser adecuada, ha de ser máximamente eficiente en ambos dominios. Si se prueba que el procesamiento auditivo general requiere un tipo distinto de representación, tendremos un argumento fuerte para postular dos módulos diferentes.

Bibliografia

- DIEHL, R.L. (1987). Auditory Constraints on Speech Perception. En M.E.H. SCHOUTEN (Ed.), *The Psychophysics of Speech Perception*. Martinus Nijhoff.
- FAHEY, R.P., LOPEZ-BASCUAS, L.E., GARCIA-ALBEA, J.E. & ROSNER, B.S. (en preparación). Temporal lag perception in Spanish and English speakers.
- FODOR, J.A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- HIRSH, I. (1959). Auditory perception of temporal order. *JASA*, 31, 759-766.
- KEWLEY-PORT, D., WATSON, CH. & FOYLED. (1988). Auditory temporal acuity in relation to category boundaries: speech and nonspeech stimuli. *JASA*, 83 (3). 1133-1145.
- KLATT, D. (1980). Software for a cascade-parallel formant synthesizer. *JASA*, 67, 971-995.
- LIBERMAN, A.M., HARRIS, K.S., HOFFMAN, H.S. & GRIFFITH, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358-368.
- LIBERMAN, A.M., HARRIS, K.S., KINNEY, J.A. & LANE, H. (1961). The discrimination of relative onset times of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 61, 379-388.
- LIBERMAN, A.M. & MATTINGLY, I.G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- LISKER, L. & ABRAMSON, A. (1970). The voicing dimension: some experiments in comparative phonetics. Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 563-567.
- MARR, D. (1982). *Vision*. Freeman.
- MILLER, J.D., WIER, C., PASTORE, R. KELLY, W. y DOOLING, R. (1976). Discrimination and labeling of noise-buzz sequences with varying NLT: an example of categorical perception. *JASA*, 60, 410-417.
- PISONI, D.B. (1977). Identification and discrimination of the relative onset times of two component tones: implications for voicing perception in stops. *JASA*, 61, 1352-1361.
- REPP, B.H. & LIBERMAN, A.M. (1987). Phonetic category boundaries are flexible. En S. HARNAD (Ed.), *Categorical Perception*. Cambridge, U.P.
- ROSEN, S. & HOWELL, P. (1987). Is there a natural sensitivity at 20 ms. in relative tone-onset-time continua? A reanalysis of Hirsh's (1959) data. En M.E.H. SCHOUTEN (Ed.) *The Psychophysics of speech perception*. Martinus Nijhoff.
- ROSNER, B.S. (en preparación). Thurstonian models for changes in perceptual performance.
- SACHS, R.M. & GRANT, K.W. (1976). Stimulus correlates in the perception of voice onset time (VOT): II Discrimination of speech with high and low stimulus uncertainty. *JASA*. Supplement 1, 60, S91.
- SUMMERFIELD, Q. (1982). Differences between spectral dependencies in auditory and phonetic temporal processing: relevance to the perception of voicing in initial stops. *JASA*, 72, 51-61.
- WILLIAMS, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics*, 5, 169-184.

Taula (UIB) núm. 15, 1992

The modularity thesis: its implications for interpretations of priming effects

Christopher W. Davis

Universidad Complutense de Madrid

Abstract

The aim of this paper was to show that Fodor's modularity thesis provides a framework for interpreting data thought to be relevant to the structure of the language processor. In setting out a background to this, the basic research enterprise of isolating processor properties was considered, with an emphasis on the logic underlying the use of a particular experimental technique (repetition priming) as a way of determining processing architecture within the language module. This enterprise was considered with respect to Fodor's account of mental architecture; the suggestion being made is that in its traditional form, the priming technique will give ambiguous results. In order to overcome this, a modification to the priming technique was investigated.

Resumen

El objetivo de este artículo es mostrar que la tesis de la modularidad de Fodor proporciona un marco de trabajo desde el que se puede interpretar datos relevantes sobre la estructura del procesador del lenguaje. Para ello, se propone aislar propiedades de dicho procesador, utilizando la lógica subyacente a la técnica experimental del «priming de repetición» como medio para determinar la arquitectura de procesamiento dentro del módulo del lenguaje. Esta línea de investigación se examina considerando la propuesta de Fodor sobre la arquitectura mental. Finalmente, se sugiere que la técnica de «priming» en la forma que se ha utilizado tradicionalmente proporciona resultados ambiguos, y se propone una modificación de la misma.

The primary interest of this paper concerns the possible implications of Fodor's modularity thesis for experimental inquiry in psycholinguistics. Specifically, the assumptions implicit in the use of a particular experimental task thought to index lexical structure will be examined in order to show how theories of mental architecture and the interpretation of data interact. In order to see how these two are linked it is necessary to provide some background into research concerning visual word recognition.

The priming paradigm has become one of the most extensively used techniques in the study of visual word recognition. The term *priming* has had broad application in the word recognition literature. It seems to have been used to describe virtually any situation where a facilitation in performance for one stimulus occurs, given an encounter with another. The idea, simply put, is that an initial stimulus input, the *prime*, benefits the processing of a subsequent input, the *target*. Priming effects have been shown to occur over a range of prime-target relationships: in *repetition* priming, the prime and target stimuli are the same (Scarborough, Cortese and Scarborough, 1977); in *semantic* and *associative* priming, there is either a meaning-based or an experiential link between the prime and target (Meyer and Schvaneveldt, 1971); and in *form* priming, the prime and target share properties of form (Hillinger, 1980).

Effects of this kind have been used to make inferences about the properties of the information processing systems underlying word recognition performance. The particular interest of this paper centers on repetition priming and the variables influencing its occurrence and magnitude in experiments which typically use the lexical decision task. In these experiments, the dominant dependent variable is response latency, a measure which is assumed to reflect the duration of processes involved in word recognition.

When items are repeated in a lexical decision experiment, with first encounters acting as primes for second encounter targets, response latencies for target words in general decrease markedly. This holds true even when a considerable time (say, 20 minutes or more) elapses between successive presentations (Forbach, Stanners, and Hochhaus, 1974; Kirsner and Smith, 1974; Scarborough, Cortese and Scarborough, 1977).

This finding, henceforth referred to as the *repetition priming effect*, has been interpreted by many investigators in terms of a reduction in the processing time required for the *recognition* of the target word. The mechanism proposed to account for this reduction varies with the theoretical position adopted to explain word recognition, in general. For example, Morton (1979) interprets the finding within his system of logogens, the evidence-gathering detectors whose activation is the basis of word recognition in this theoretical framework. Where the same logogen can be activated by both prime and target, the raised activation level which is a product of the prime presentation is said to persist, to permit a reduction in the evidence, and hence time, required for the target to raise activation above threshold.

Regardless of any particular theoretical stance, the assumption that repetition priming reflects some change in the processes underlying word recognition has important consequences. In particular, it leads the investigator to interpret variables which modulate the size of the repetition priming effect as yielding information about lexical organization. In the end, the interest is not in repetition priming, per se, but in its use as a means to probe the lexical component of the language processor.

To illustrate the way the arguments about lexical structure are constructed on the basis of facts about repetition priming, consider Morton's (1979, 1980) use of his finding that the effect is eliminated when primes and targets are presented in different modalities. Because Morton treats repetition priming as a purely lexical effect, he is obliged to postulate that each word has a separate representation (logogen) for each input modality. A modality dissociation in the repetition priming effect translates to a modality dissociation in lexical architecture: auditory presentation of a word leads to residual activation in its phonological logogen, but not in its orthographic logogen.

Monsell (1987) has adopted a similar strategy for carving up the language processor. Treating repetition priming as the product of temporary modifications to the processes of lexical access, he argues from the results of a range of experiments that the orthographic lexicon should itself be divided; writing a word, it seems, does not necessarily activate the orthographic lexicon used in reading. The priming technique has also been used as a research tool to study the basis of bilingual lexical representation, with the failure of priming from one language to another being regarded as evidence for separate, non-connected lexicons. The unstated idea in all these studies has been that this task gives a *theory free* look at mental structure. In order to understand why this might not be the case, we need to consider why the modularity thesis cautions against making such inferences.

Fodor's modularity thesis is a proposal about mental organization and the division of mental labor between processing domains. The primary claim is that there *is* a mental architecture in which the information exchange between processors is limited. This restriction thereby sets a boundary on those confirmation relations needed to determine perceptual belief. Fodor (1983) has proposed that the fixation of any perceptual belief involves the operation of three functionally distinct systems. At the interface between the organism and the environment, *transducers* convert sensory stimulation into some form of low-level neural code. Computations are then performed on these transducer outputs by *input systems*, which yield hypotheses about the nature of the distal object underlying the proximal stimulus configurations at the transducer. Finally, *central systems* arrive at a belief about the external object by correcting the outputs of an input system in the light of background knowledge and of the output of other input systems.

For Fodor, the language faculty is a strong candidate for an input system. In common with all input systems, a language processor is *modular* in possessing a number of properties. For example, it is domain specific (its processes operate only on inputs of a specified kind, and its outputs are similarly limited); its operation is mandatory (given an appropriate input, its processes automatically run their course); and its working products are unlikely to be accessible to central systems, and hence to consciousness (see Fodor, 1983, pp. 47-101).

The idea of a modular language processor is reflected in theories of lexical access which regard the lexicon as a data-base in the service of sentence processing devices. What follows naturally from this is that those vocabulary features which are crucially involved in language processing are recorded in the database (e.g., spelling, pronunciation, syntactic class, and so on), while other incidental facts are not. To put it

succinctly, the lexicon is placed within an input module (see Fodor, 1983; Forster, 1979; Seidenberg, 1985).

Now if the modular view is taken seriously, a problem now arises in the interpretation of priming effects. For these effects could be generated in a number of ways and at markedly different locations in the perceptual-cognitive apparatus. At lower levels storage may be simply be a change in state in the operating characteristics of a transducer. At the highest level, the fact that a priming event has occurred may be stored as a record of a perceptual belief, fully elaborated. This seems to be an inevitable problem in the study of mental processes, for «overt, observable behavior is an interaction effect par excellence» (Fodor, 1983, p. 1). That is, a researcher's dilemma is that observed behavior is the joint product of contributing sub-systems which may differ from each other in important ways. The individuation of the separate contributions of different sub-systems is likely to be difficult. In particular, data which are presumed to yield information about the structure of an input module (e.g., data about repetition priming effects with word materials) are likely to be contaminated by the influence of systems outside this module (e.g., central memory effects).

The task of eliminating competing explanations of putatively lexical phenomena is not a newly arising one. Coltheart (1978) makes a cautionary point with respect to the suitability of methods for examining lexical processing:

If we wish to discover how lexical access is achieved, it seems advisable to use experimental tasks which cannot be performed in the absence of lexical access, but require little else. (p. 170)

The problem is, if anything, exacerbated by the fact that most available language tasks and performance measures involve the operations of the highest levels of the cognitive apparatus (Forster, 1985, p. 9). If the properties of a *particular* processing system are to be investigated, experimental ingenuity is called for to ensure that variations in the performance measure reflect only processes within that system, with the contribution from other processors minimized.

In sum, what cannot be disregarded is the possibility that although a task includes language module processes, it may also include something more. Since behavior in a language task is determined jointly by the products of the language module and higher inferential systems, «lexical decision times will reflect lexical access times only to the extent that irrelevant influences have been eliminated» (Forster, 1985, p. 10). Forster's observation is not of course particular to lexical decision latencies, or even to latencies more generally, though it has real force in the case of repetition priming. If an initial encounter with a word is to have an effect at a later stage, it must be stored in memory. The critical question is *which* memory or system: a specialized lexical memory, or central system - a general memory for episodes? For the interpretation of patterns of repetition priming, it is perhaps not so much a question of controlling for «irrelevant influences», but of deciding upon the system to which the effect is relevant.

Thus, from the modularity perspective, the problem becomes to limit central processing involvement, while still allowing the input system (the lexicon) to complete its processing. How might this be done? The modularity view itself offers a clue, for Fodor's (1983) view of the architectural arrangements governing exchanges of informa-

tion between input systems and mechanism of cognition permits a plausible account of «cognitive masking» of the central processing system.

The relevant properties, here, concern the way input systems interface with central ones. As outlined earlier, input systems are modular and rapid, while the fixation of perceptual belief handled by central systems is informationally promiscuous and relatively slow. Moreover, belief fixation is sensitive to the utilities of the organism, e.g., the attentional demands present in any given situation. The interplay of these properties, as it pertains to a possible masking situation, has been neatly summarized by Holender (1986):

What is fascinating in [Fodor's] theoretical formulation is precisely that a good (supraliminal) sensory input (1) can undergo a fair amount of processing into a modular input system, including lexical access, without necessitating any intervention from the central processor and (2) that the central processor can be ignorant of the fact that such processing has occurred (p. 60).

This approach then, suggests that a prime stimulus could affect the state of a representation within the lexical input module, but its effect upon subsequent perceptual belief might be negligible because of the action of an appropriately timed mask. It should also be stressed that in adopting an approach more in terms of the allocation of attentional resources, the nature of unconscious capacities are not at issue.

As it turns out, experiments have shown that this may be achievable in practice (e.g., Forster and Davis, 1984). In these experiments, the display characteristics of masks and prime are set so that the subject has no *explicit* memories of the prime event, even though a robust repetition priming effect is observed. Thus masking the prime apparently provides a straightforward way to restrict the possibility of truly high level memories.

In order to be confident that the technique of *masked* priming overcomes the interpretation problem, it is necessary to show that it is sensitive solely to aspects of the input system. Yet if these aspects could be specified in advance, research would be made unnecessary; in practice we are limited to an indirect approach, showing that masked priming is *not* influenced by either pre- or post-input system properties.

One possibility is that masking the priming stimulus has created unusual conditions in early perceptual or even «neural» territory, that is, in the process of sensory transduction. For this to occur, the neural signals of prime and target would need to interact; or if it were sited at the level of preliminary perceptual analysis, the processes identifying features of the stimulus shape might be conflated across successive inputs. But this does not seem to be the case, for an experiment (Davis and Forster, 1991) which used prime-target pairs specifically designed to be either legible or illegible when fused together show no differential effects. Further, other experiments (Davis, 1990) have shown that the amount of priming drops off very sharply as the prime differs from the target. This suggests that stimulus registration of the prime is very good, and once again implies that the prime and target are not visually degraded.

At the other extreme, higher level decision processes (central systems) have also to be ruled out, as experiments (Davis, 1990) suggest that information from the prime does not affect the decision processes which are mediated in part by central

inferential systems. So for instance, there was no indication that the lexical status of the prime influenced the judgment of the lexical status of the target, or that judgments concerning the identity of a pre-specified target were influenced by the nature of the prime.

From these results a picture of masked priming can be fashioned: When the prime is presented, it accesses (and opens) its appropriate entry in the mental lexicon, which then remains open for a short period of time. However, because of the action of the masks, the fact that the priming word occurred does not become a perceptual belief and so does not become an item considered by decision systems. When the target arrives, it accesses its appropriate entry, and if this is the same one as the prime has opened, then a benefit will accrue.

It is beyond the scope of this article to go into detail about the various research findings made using the masked priming technique, suffice to say that not only is the pattern of data obtained different from that obtained with the standard long-term unmasked version, (see Forster and Davis, 1984; Forster et al., 1987; Davis, Sánchez-Casas and García-Albea, 1991), but masked priming also appears to provide a new method of examining issues connected with the modularity debate. For instance, recently Forster and Davis (1991) have shown that the lexical decision and naming task (where the subject must pronounce the target string) seem to be differentially sensitive to an aspect of the masked stimulus. That is, unlike lexical decision, the naming response can be affected by the initial letter of the masked stimulus. One way of describing this difference in task sensitivity is in terms of the availability of information. It is argued that pronouncing a word may be a process sensitive to within language module properties, whereas the lexical decision task (which involves central system decisions) is not.

In conclusion, the modularity thesis, with its notion of specialized processors feeding into more general ones, has provided a basis for questioning the interpretation of previous results based on repetition priming, and has also furnished the rationale for a new investigative technique.

References

- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of Information Processing* (pp. 151 -216) London: Academic Press.
- Davis, C.W. (1990). Masked priming effects in visual word recognition. Unpublished Doctoral thesis, Monash University, Australia.
- Davis, C. W., Sánchez-Casas, R.M. and García-Albea, J.E. (1991). Bilingual lexical representation as revealed using a masked priming procedure. Manuscript submitted for publication.
- Davis, C.W. and Forster, K.I. (1991). Masked orthographic priming: The effect of prime-target legibility. Manuscript submitted for publication.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: The M.I.T. Press.
- Forbach, G. B., Stanners, R. F., and Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory and Cognition*, 2, 337-339.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W.E. Cooper and E. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum and Associates Ltd.
- Forster, K. I. (1985). Comments on the précis of Modularity of Mind. *The Behavioral and Brain Sciences*, 8, 9-10.

- Forster, K.I. and Davis, C.W. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 680-698.
- Forster, K.I., Davis, C.W., Schoknecht, C. and Carter, R. (1987). Masked priming with graphemically related forms: Repetition or Activation? *Quarterly Journal of Experimental Psychology*, 39A, 211-251.
- Forster, K.I. and Davis, C.W. (1991). The Density Constraint on Form-priming in the Naming Task: Interference Effects from a Masked Prime. Manuscript submitted for publication.
- Hillinger, M.L. (1980). Priming effects with phonemically similar words: The encoding bias hypothesis reconsidered. *Memory and Cognition*, 8, 115 - 123.
- Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *The Behavioral and Brain Sciences*, 9, 1-66.
- Kirsner, K., and Smith, M.C. (1974). Modality effects in word identification. *Memory and Cognition*, 2, 637 - 640.
- Meyer, D. E., Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence in retrieval operations. *Journal of Experimental Psychology*, 90, 227 - 234.
- Monsell, S. (1987). Nonvisual orthographic processing and the orthographic input lexicon. In M. Coltheart (Ed.), *Attention and Performance XII* (pp. 299 - 323). London: Lawrence Erlbaum and Associates Ltd.
- Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P.A. Kolers, M.E. Wrolstad, and M. Bouma (Eds.), *Processing of Visible Language*. New York: Plenum Press.
- Morton, J. (1980). The logogen model and orthographic structure. In U. Frith (Ed.), *Cognitive Processes in Spelling*. London: Academic Press.
- Scarborough, D. L., Gerard, L., and Cortese, C. (1977). Frequency and Repetition effects in Lexical Memory *Journal of Experimental Psychology: Human Perception and Performance*, 7, 3-12.
- Seidenberg, M.S. (1985). Comments on the précis of Modularity of Mind. *The Behavioral and Brain Sciences*, 8, 15.

Taula (UIB) núm. 15, 1992

Mental Models and Referential Processing

Manuel de Vega
University of La Laguna

Abstract

This paper proposes that the representation of the referent of discourse might be characterized as the construction and updating of a non-linguistic mental model. Mental models of the referent allow the reader to set up some implicit parameters such as temporal, spatial, and behavioral scales, magnitude of quantifiers, etc. The set up of these parameters is critical to appropriately understand and integrate new pieces of information from the incoming text.

A second function of modelling is to compute the consequences of the events described by the text. There is psychological evidence demonstrating that comprehenders are very efficient in calculating on-line the consequences of described events or changes. However, it is difficult for a purely deductive processor (e.g., a declarative data base plus some rules) to compute appropriately the effects of change. This is the so called «frame problem». Such problem might be overcome by means of a mental model account. Mental models are sensitive to the implicit knowledge of world's regularities and, as a consequence, some courses of modeling are more likely than others. This implies a functional advantage: The number of potential models for a given linguistic message can be drastically reduced.

Traditionally the mental models framework has focused on the representation of spatial contents. However, I suggest that mental models involve also the representation of psycho-social and self traits. Comprehension, is a knowledge state that potentially involves an ecological, a psycho-social, and a self model. At a given moment certain level of modelling is more prominent (foreground) than the others (background).

Some computational principles are proposed in order to guide an eventual implementation of mental models. First, the building blocks that underlies models are primitives. Primitives are variable rather than discrete and fixed semantic values. Secondly, the building and updating of models demand cognitive resources, therefore mental models are sensitive to the structural limits of the system (e.g., working memory, attentional resources). Thirdly, the updating of mental models is cyclic rather than continuous, and it takes place at the end of the main text constituents (clauses, sentences and paragraphs). Fourthly, some models' parameters are structurally isomorphic (some spatial relationships, the temporal structure of events, etc.). Finally, the computation of primitives in the model is based in a multiple constraints principle.

According to the modular approach to cognition the only scientifically describable events of human mind are those related to the encapsulated «modules» (e.g., Fodor, 1983). On its side, the conceptual or «central processor» is, by definition, too flexible and indeterminate as to be properly analyzed by conventional scientific procedures. Therefore, if we study language comprehension we must focus on relatively specialized or modular components such as phonological and graphemic processing, lexical access, syntactic parsing and so forth.

It is likely that the functional architecture of language processing involves some modular or encapsulated mechanisms. However, at some stage of processing, language might have close interfaces with the conceptual system. Thus, comprehension of discourse requires that the listener or reader builds up a non-linguistic representation of the referent of discourse and this conceptual activity is almost simultaneous to the more linguistic levels of discourse processing. In the same vein, speakers encode their linguistic messages starting with a pre-verbal representation elaborated by the conceptual system or the «conceptualizer» (Levelt, 1989). The mechanisms responsible for the representation of reference are still unclear. However, the notion of «mental models» is a promising framework which have generate a remarkable amount of relevant research.

The critical insight of the mental models framework (also named «situation models», «possible worlds», «scenarios», etc.) is that the comprehender builds a representation of the referent of the text (the objects, events, objects and processes described or implicit in the text) rather than -or in addition to- a representation of the text itself (Sanford & Garrod, 1982; Johnson-Laird, 1980, 1983; van Dijk and Kintsch, 1983; Garnham, 1987; Glenberg et al., 1987). An illustration of this idea is the experiment of Glenberg, Meyer and Linden (1987). Subjects were given small text like this:

- (1) John was preparing for a marathon in August (Setting sentence)
- (2a) After doing a few warm-up exercises, he put on his sweatshirt and went jogging (Critical associated)
- (2b) After doing a few warm-up exercises, he took off his sweatshirt and went jogging (Critical dissociated)
- (3) He jogged halfway around the lake without too much difficulty (Filler)
- (4) Further along his route, however, John's muscles began to ache (Filler)

Target: SWEATSHIRT

Subjects read a given version of the text with instructions of comprehension, and immediately they received the target word. They were asked to decide as soon as possible whether or not the target had been included in the text. The critical manipulation was that in the second sentence there were two alternative versions. In the critical associate version (2a) the meaning of the target word was «attached» to the main character by means of the action described by the verb. In the critical dissociated version (2b) the target was dissociated as a consequence of the character's action.

As was expected the decision time for targets was shorter in the associated than in the dissociated version of the texts. The association between the target object and the main character determined that the critical concept were more accessible. This acces-

sibility relies on the situation model rather than on the surface features of the text. In fact, the words between the critical word («put on» or «take off») and the target were exactly the same under both versions of the texts.

A propositional account of text's representation cannot explain either the observed differences, as both the associated and the dissociated versions of the text have virtually the same propositional structure. Thus, the propositional structure for the critical associated sentence and for the critical dissociated one only differ in a single concept.

(5) ... he put on his sweatshirt...

(5') <JOHN, SWEATSHIRT, PUT ON>

(6) ... he took off his sweatshirt...

(6') <JOHN, SWEATSHIRT, TOOK OFF>

There is no apparent reason for becoming more closely associated the nodes <JOHN> and <SWEATSHIRT> in (5') than in (6'). Notice that a propositional system involves computations upon uninterpreted symbols. Namely, the content or meaning of the symbols might be irrelevant in terms of the resulting strength of connections among them. Therefore, both predicates (PUT ON and TOOK OFF) should take the same associative power in order to connect the corresponding concepts. The empirical results (better accessibility of the target SWEATSHIRT under the associated condition) suggest that the computation of text referent relies on fine-grained details of our world knowledge.

Approaching comprehension from a mental model framework has demonstrated to be useful. Mental models put the emphasis in a different theoretical perspective and, consequently, new hypothesis are available. Thus, the emphasis of the mental model approach is on:

- The representation of referents triggered by the text (including thematic inferences) rather than on the structural or formal features of the text.
- Post-lexical processing rather than on lexical processing.
- On-line or incremental changes in representations rather than the final (off-line) products of comprehension.
- Isomorphic dimensions of text representation rather than on linguistic or propositional codes.

Some function of mental models in comprehension

The notion of mental model has been used as an explanatory notion for several psycholinguistic phenomena such as text structure and co-reference, thematic inferences and expectations, anaphoric reference, narrative perspective, focus and background in comprehension, etc. I will focus on two less mentioned phenomena that in my opinion have to do with our modeling capabilities: the implicit scale set up and the modelling of change.

Scale set up

When the reader or listener processes a linguistic message she has to set up the appropriate scale for several dimensions such as time, spatial coordinates, speed, action

level, etc. These particular scales determine a dimensional field that allows the reader to appropriately interpret and integrate the incoming text. Dimensional scales frequently are not explicit in the linguistic message but should be figured out in a top-down manner from some semantic clues provided by the text.

For instance, spatial distance referred by the verb: TO APPROACH is different in (7) and (8). The size of the corresponding goals determines a different «approaching distance» in the reader's mental model.

(7) We were approaching Chicago

(8) We were approaching the door

(Morrow and Clark, 1988)

The temporal scale established from a given temporal adverbial depends on the semantic context provided by the sentence (van Jaarsveld & Schreuder, 1986). For instance, the present denoted by NOW involves different magnitudes in (9) and (10), and the time interval denoted by JUST differs in (11) and (12).

(9) The sky's color is changing now

(10) Europe is changing now

(11) John has just smoked a cigarette

(12) John has just got married

(van Jaarsveld & Schreuder, 1986)

A clear demonstration that parameters are set up is that any violation in the implicit scales is often perceived as an error, a joke or a bizarre expression. For instance (13) is a violation of both the temporal and the behavioral scales:

(13) Q: Well... now that you have finished your PhD what are you going to do?

A: I am going to have a coffee

The clues used by the comprehender to establish the scales of a dimensional field are varied (the sizes of objects, places or agents, the level of the character's goals and intentions, etc.). Hypothetically, dimension scales correlate among them, therefore, the reader might be able to compute the value of implicit parameters from the explicit cues provided by the text. For instance, given a very small moving agent (e.g. an spider) we can «infer» a dimensional field involving small scales for time, space, behavioral level, etc.

Notice that the elusive phenomena of «connotative meaning» might be partially explained as an implicit set up of mental model scales. Actually, connotative meaning cannot be properly understood as a lexical phenomenon. The meanings denoted by some words in different linguistic or communication contexts are not fixed entries «pre-stored» nor «accessed» from a lexicon. For instance, there are not pre-stored magnitudes for «few», «now», «approaching distance», etc. Instead, these magnitudes are computed from a dimensional field or mental model.

The modeling of change

Mental models are dynamic. In other words, the representations of one or several of its entities are transformed in the course of time. However, the question of change is rather complex, as a given transformation in a single entity can determine an effect on several parameters of the model. Let us see a short text that describes a situation model or scenario and some events:

- (14) John is sitting in front of his office table
- (15) At his right there is a cup of coffee on its plate
- (16) In front of him there is an article on the table
- (17) At his left there is a bookshelf
- (18) John takes the cup and gives a small drink of coffee
- (19) John took a new book from the bookshelf

Sentences (14) to (17) allow the reader to build a representation of a static model of a physical environment. Sentence (18) conveys information about a simple action. This action (an object manipulation) produces several changes in the parameters of the model. Not only the cup changes its location but, as a consequence of this change, the relative positions and states of a few things became different while others remain equal. Thus, the action modifies the relative position between: the article and the cup, the cup and the plate, etc. However, another spatial relations remain unchanged: the plate and the book, the book and John, etc.

The principle underlying the readjustment of positional parameters that follows the action seems straightforward in the above example. The moving object approaches a landmark (in this case John's lips) and, consequently, the moving object takes the positional parameters of the landmark. However, things are more complicated than that: Why does the coffee follows the cup in its displacement? and why the plate does not?. Initially the triad «coffee, cup, plate» belonged to the same spatial cluster in the model, however a given action causes that two entities became detached from the third one. Furthermore, a small modification in the action can produce a different pattern of consequences. For instance, if the action (18) substitutes: «John got the plate and gave a small drink of coffee» then the three entities remain together in their movement. In summary, the readjustment of positional parameters in a model involves some subtle knowledge about the behavior of supporters, containers, gravity, and so forth.

Even more complex is the pattern of change during and after the action (19) because the initial scene had been described from the character's perspective. Therefore, if the character moves many situational parameters should be readjusted to keep his perspective. Moreover, often the character's actions determine that some entities came out of focus while others enter the focus and became more accessible (Anderson, Garrod and Sanford, 1984; Glenberg et al., 1987).

A possible argument is to say that the reader just does not make all these implicit computations because it is not necessary. That is probably true in many cases of superficial reading in which the subject does not bother to elaborate a complete and coherent representation of the referent (Oakhill et al., 1989). However, there is some evidence that the reader is able to some extent to make parameters adjustment after a given change has been described by the text. For instance, Duffy (1986) found that subjects generated expectations about the consequences of a physical phenomenon described in short texts. This is an example of an item:

- (20) John was eating his first meal ever in the dining car of a train
- (21) The waiter brought him a large bowl of bean soup
- (22) John tasted the hot soup carefully
- (23a) Suddenly the train screeched to a stop (high expectation)

- (23b) The train slowed down entering a station (low expectation)
- (24a) The soup spilled all over John's clean shirt and pants (key event)
- (24b) The waiter came to offer John some coffee (script related)

There were two critical experimental manipulations: the expectation value of the fourth sentence (high or low) and the content of the target sentence (either describing the inferred event or an script based action). The reading time of the target sentences was the dependent measure. The results showed that under the high expectation condition (23a) reading time was faster for the sentence describing the key event (24a) than the script related one (24b). Conversely, under the low expectation conditions (23b) there were no differences. This experiment suggest that the subject is able to make some rather sophisticated computations on the implicit parameters of the situation model. Furthermore, the outcome of these computations determines inferences or expectations more salient than the script-based default values.

How a given change (e.g., an action or an event) affects the whole model state?. What should be up-dated and what would remain equal?. These are non-trivial questions. Many of these changes are not explicit in the text because the text producer assumes that the reader will infer them by means of her implicit world knowledge. However, there are not simple rules that allow to predict what will change and what will remain equal. In order to appropriately address these questions we need a more elaborate notion of the underlying mechanisms and constraints of mental models.

The problem of modelling change is familiar to the Artificial Intelligence researchers who have labeled it the «frame problem» (Fodor, 1983, 1988; Janlert, 1988). The problem emerge in robots design. An efficient robot that performs goal-oriented actions in a complex environment has to be able to build a dynamic mental model. The robot's mental model must keep track of the state of the robot's goals as well as the consequences of any environment change produced either by any external agent or by the robot itself. The problem of doing that is both quantitative and qualitative. The quantitative problem is that an unlimited number of predicates should be checked continuously to see whether their truth values are appropriate after the change. This is hardly compatible with the time pressure under which a robot (or a human) has to make decisions. The qualitative problem is that there are not simple «logical» principles governing change and its consequences. Therefore, it is unsuitable to store in the robot's memory some simple deductive rules that enable it to adapt its representations to the changing world.

However, the frame problem is probably derived from the technical decision of using a declarative format of representation and trying to model change according to a deductive logic (Janlert, 1988). In my opinion, the use of some isomorphic parameters in the mental model (such as mimicking the temporal structure of events, representing structurally spatial relations and quantifiers, and some kinemati transformations) could reduce considerably the problem (see section 2.8).

Mental models are sensitive to world knowledge

Some authors, notably Pylyshyn (1986), have criticized the notions of cognitive science that are cognitively penetrable, namely that are influenced by our tacit world knowledge, intentions, wishes and so forth. The only scientifically acceptable mechanisms are those non-penetrable that operate isolated or encapsulated from the influence of the «central system». However, mental models are by definition cognitively penetrable. Therefore, there must be features of the models functioning partially derived from our implicit knowledge of the world. For instance, the set up of implicit dimensions, the representation of change (e.g., mental kinematic), the computation of the consequences of change, the changing of focus or perspective. In fact, a mental model is a simulation device that operates upon representations of world entities (tokens), and analogically represent their transformations, temporal structure, etc. This simulation in order to be efficient must be influenced by our knowledge of world regularities. The sensitivity of mental models to the knowledge of world regularities probably avoids computational overloads, inasmuch as some modelling courses are more permissible than others. We usually do not waste our time trying to model impossible or implausible world-states, such as «five-dimension cats».

Consider the well known phenomena of mental kinematic. Our mental representation of moving objects (e.g., rotation or scanning) usually mimics our implicit knowledge of the ecological environment. In other words, for a living organism mental kinematic is anisotropic. First, there is a vertical anisotropy. Although geometrically the three dimensions of space are isotropic, from an ecological point of view this is not true. Specifically, the vertical dimension is privileged because the bias introduced by gravity and/or the head-feet axis of human body (Franklin & Tversky, 1990). Furthermore, the vertical axis is asymmetric as objects fall from top to down, and in order to produce the opposite movement some source of power is required. This determines a lot of predictable events and relations in our mental model without need to mention them in the surface text.

Secondly, the environment is horizontally anisotropic. This anisotropy results from the design of our body, specifically its motor and perceptual functions. Depth dimension is more central because it is associated with the direction of march. In addition, the depth's frontal pole is more salient because it is related to the goals and to the perceptual field (Clark, Franklin & Tversky, 1990).

Furthermore, our ecological mental models are probably constrained by the implicit knowledge of the reversibility or non-reversibility of particular events (e.g. rotation is usually reversible, whereas breaking is non-reversible); and our psychosocial models might be constrained by social norms and implicit theories that make some modelling courses more permissible than others; finally, our self model is constrained both by the social determinants and the self-concept.

However, there must be something non-penetrable in our mental models. The segmentation of reality in some primitive components (Miller & Johnson-Laird, 1976; Johnson-Laird, 1983), and the laws governing kinematic transformations (Shepard, 1984, Shepard & Cooper, 1982) might be universal, innate functions, or built-in «a priories» that underlay mental models.

Mental model's contents

Mental models involve at least three levels of contents:

The ecological (or Gibsonian) model. A three-dimensional frame, unfolding in time, and containing physical entities spatially and dynamically related, and a particular observer's perspective. Actions and events are represented as «mental transformations» of objects or objects' parts (e.g., displacement, rotation, physical deformation, etc.). These transformations not only affect the target object, but they also determine a readjustment of several parameters of the ecological model, such as the relative position of the objects, the new distance, the coming in or out of focus, the updating of causal links (either their activation or deactivation), the computation of some predictable consequences of events, and so forth.

The psycho-social model. A frame involving characters (actors or agents) with individual features, goals and intentions, interpersonal links (emotional links, professional roles, social status, etc.), and person-to-objects links (ownership, instrument, transference). Notice that the psycho-social model, unlike the ecological model, does not correspond to physically «objective» features of the world. To be true, there are physical events and objects in our psycho-social models, but these events and objects are only cues for a higher order interpretation (e.g., in terms of between-personal links, status, intentions, etc). For instance, certain facial expressions and gestures are physical actions, but they are interpreted as signals that convey an interpersonal meaning, such as intentions, feelings, etc.

The self model. The receiver of a verbal message is herself a psycho-social character with a very articulate representation of the self and its states (e.g., goals, intentions, status, emotional and affective links, ownership, and so on). In many occasions, specially if the receiver is involved in a conversational setting, she has to monitor the self model as a function of the events and actions that are incorporated in the model generated by the conversation. The monitoring of the self model relies on the fact that the listener belongs herself to the ongoing psycho-social model. However, even if the listener attends some impersonal source of discourse (a fiction narrative or an expository text) the self model might be monitored and readjusted to some extent. Thus the reader can vicariously shift her mood (feeling intrigued, amazed, surprised, etc., see Brewer, 1985) as a function of her unfolding model of text's characters and situations.

There is another aspect in which the self model is actively engaged in the comprehension processes. The self operates as a metacognitive monitoring device. Thus the reader or the listener is usually aware of her understanding level, the plausibility, permissibility or truth of the world state referred by the text, etc. This metacognitive monitoring has functional consequences in terms of processing strategies. The subject can regulate her cognitive activity according to the comprehension state monitored by the self model. This regulation is based on several possible strategies as asking questions (in conversational settings), going back to parts of the text previously read, trying to elaborate more the current model according to her world knowledge, etc.

Knowledge is a three-way model. The ecological model, the psycho-social model, and the self-model converge to produce a knowledge state in a given moment.

In fact, the three models are not simply additive. They compete in strength and prominence and they constraint each other and the resulting knowledge state is a compromise among them. This compromise involves a definition of the prominent level that must be the most accessible to subject's awareness. In most situations (including comprehension of narratives) the more activated model might correspond to the psycho-social parameters, whereas the ecological and the self model are in the background.

During the comprehension of discourse our knowledge state gradually changes. This change can be interpreted as the unfolding of the three-way model. All the relevant parameters -belonging to the ecological model, the psycho-social model and the self model- are updated as the incoming text is processed. This can produce breaks or transitions in the knowledge state (the three-way model state). Thus, in a given moment, the text describes that the character has reached a goal and the corresponding intention might be deactivated in the reader's model, perhaps new intentions and goals became activated, and the new actions involve a different temporal and/or spatial scale than the previous one, and so on. Gersnbacher (1985) calls «processing shifts» to these breaks in the representational continuity. According to her hypothesis an ongoing «substructure» is kept while the incoming information is coherent with it. When information is less coherent (changes in topic, point of view, place, time, etc.) then a new substructure starts. These breaks of shifts in our representation of reference have several empirical effects that I will describe in the next section.

Computational principles

Mental models must be computable (Johnson-Laird, 1983, 1988). However, up to now the notion of mental models has been difficult to implement and it remains as an intuitive notion. However, some empirical and theoretical constraints can be established on an eventual computational theory of mental models.

Building blocks. Our modeling capabilities must require a limited set of elementary units or «primitives». These primitives might be universal entities such as «time», «space», «cause», «event», «object», «person», «intention», etc. (e.g., Miller & Johnson-Laird, 1976; Johnson-Laird, 1983). Primitives should be understood as variable entities rather than as fixed and discrete semantic values. A primitive is actualized in different ways according to the value of the other primitives concurrently activated in the model. We already have described some examples of this interactive actualization of primitives: the contextual set up of dimensions such as time, size, quantity, etc. In the same vein, cause is not a fixed, and abstract relationship but a specific content-dependent link (compare: «the collision of a stone causes the breakdown of a glass», «her love causes his happiness»).

Structural limitations. Mental models presumably are resource demanding, as they likely consume working memory space. There is some evidence of the cost of mental models in terms of cognitive resources. Johnson-Laird and Bara (1984) demonstrated that the larger the number of mental models required by deductive problems the more time was spent by subjects and the larger was the amount of errors. On the other hand, to build a mental model from a text segment takes longer than a more superficial

processing of the text. Thus, subjects who were presented with predictive tasks, spent more reading time on the scenario inducing source (a single case or individual information) than in the less salient quantitative source (Rodrigo, de Vega & Castañeda 1990). The high resource-demand of modeling was also illustrated by a study of de Vega & Díaz (1990). Subjects who read short texts, spent more time in reading the first sentence, as well as the next filler sentences, when the starting sentence included an scenario inducing word rather than a generic or indeterminate term.

Intermittent up-dating. As we have mentioned above the foundation and updating of a mental model is not a continuous process. Instead, there are breaks or shifts generally located at the end of the minor or major constituents of the text (clauses, sentences, and paragraphs). This is an important difference with other text processes, such as phonological processing, lexical access, and syntactic parsing, that take place immediately or on-line (e.g., Marslen-Wilson, 1975).

There are some empirical findings that support the end-of-constituents locus of integration or mental models updating. First, there is a memory loss of the words' surface codes immediately after the reader has crossed a clause boundary (Jarvella, 1979; Gernsbacher, 1985). This suggests that the reader, at the end of clauses, shifts from the lexical codes to some high-order representation. Secondly, studies with on-line techniques have shown that the reader makes long pauses in the last word of each constituent (Just & Carpenter, 1980; Haberlandt & Graesser, 1985; Daneman & Carpenter, 1983; de Vega et al., 1990). These last-word pauses are functional as they are related to several text parameters. Thus, the pauses duration is a function of the number of «new ideas» in the constituent (Haberlandt & Graesser, 1985; de Vega et al., 1990); pauses are also longer when subjects receive instruction stressing memory rather than comprehension (Haberlandt et al., 1986); finally, pauses increase when the sentence includes a desambiguating «garden-path» word, suggesting that, at the end of constituents, subjects make a semantic assessment of the whole constituent (Daneman & Carpenter, 1983).

There are some alternative interpretations of end of constituent pauses. Particularly, Kintsch and van Dijk maintain that at these positions the reader actualizes the macrostructural representation of the text (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983; Kintsch, 1988). According to Kintsch & van Dijk, the macrostructure is not a mental model but a propositional code. They predict and found out that sentences reading times increase as a function of the number of underlying propositions (Kintsch & Keenan, 1973). However, the number of propositions can be confused with other variables such as the number of new concepts. Some recent studies that rule out the effect of other variables did not find any significant effect of the number of propositions neither on the sentence reading times nor on the end of constituent pauses (Haberlandt & Graesser, 1985; de Vega et al., 1990).

Isomorphism. Several authors assume some perceptual-like properties of mental models but usually they do not elaborate much about it. In principle, it can be proposed that mental models format is a conventional propositional one (e.g., van Dijk & Kintsch, 1983). However, rather than conceiving the encoding of the referent in terms of any sort of *lingua mentis*, I propose that some of the referent properties are encoded in an analogical or isomorphic format.

Of course I do not assume a first order isomorphism, nor any sort of «picture in the head» metaphor. Rather I am thinking on a second-order or functional isomorphism: Some processing components are shared by direct perceptual experience and the modeling of a linguistic source (see similar notions in Shepard & Chipman, 1970; Finke, 1980; de Vega, 1988). These are some proposed isomorphic dimensions and properties of mental models: (a) Static spatial properties like position and distance among represented entities; (b) Temporal structure. For instance, the serial order of events in the model has a one-to-one correspondence with the perceived order of world's events, or the relative duration of events can be also mimicked in the model, (c) Kinematic patterns such as mental rotation and mental scanning (Shepard & Cooper, 1982; Kosslyn, 1980; Denis, 1988); (d) Some episodic associative links between entities mimic the world structure of phenomena. For instance the attachment between objects and characters in the mentioned experiment of Glenberg et al., 1987).

The advantages of an analogical processing of some parameters is that it avoids the practical shortcomings of the frame problem, typical of a purely deductive system. The updating of change might be unsolvable if we assume that the subject represents in a declarative format all the ongoing knowledge and he applies some predicate calculus to this massive data base. Even if this were computationally possible (a propositional data base plus some deductive algorithms can be a virtual Turing machine) it would be unsuitable for practical reasons. A psychological theory of change representation has to be able to make the appropriate computations at the appropriate times, namely it has to make computations on-line in order to fit adaptive and efficiency requirements.

On its side, a system that is able to represent analogically some entities and their transformations on-line can figure out many consequences of change without much effort, by simply «inspecting» the outcome of the dynamic transformations. These analogical transformations have been demonstrated in the research field of mental imagery (Shepard & Cooper, 1982; Kosslyn et al., 1978; de Vega 1988), and it has been suggested that they play an important role in understanding text with spatial descriptions (Denis, 1988).

Multiple constraints. Mental models emerge when the conceptual pieces conveyed by a sentence or text are integrated. Integration is holistic as every concept (either explicit or implicit) modifies its semantic value depending on the whole conceptual matrix implicit in the model. The end product is not a simple addition of lexical traces or semantic features but an interactive composed trace that is the outcome of some kind of «goodness» function. The notion of «variable constraints» proposed by Rumelhart & Ortony (1977) in their schema theory is very similar. They proposes that the instantiation of a given schema component is context-dependent, as its particular value is modulated by the other components currently activated.

The interactive nature of integration is supported by several facts that I have described: the attributes instantiation phenomena (context activates a subset of conceptual attributes), the computation of implicit parameters (spatial and temporal scales, level of intentions and actions, etc.), and the computation of some remote consequences of changes.

References

- Anderson, R.C. & Ortony, A. (1975). On putting apples into bottles. A problem of polysemy. *Cognitive Psychology*, 7, 167-180.
- Anderson, A.; Garrod, S.C., & Sanford, A.J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly J. of Experimental Psychology*, 35A, 427-440.
- Brewer, W.F. (1985). The story schema: Universal and culture specific properties. In D.R. Olson, N. Torrance y A. Hildyard (Eds.). *Literacy, language, and learning*. Cambridge, England. Cambridge University Press.
- Daneman, M. y Carpenter, P.A. (1983). Individual differences in integrating information between and within sentences. *J. of Experimental Psychology: Learning, memory and Cognition*, 9, 4, 561-584.
- Denis, M (1988). Imagery and prose processing. In M. Denis, J. Engelkamp y J.T.E. Richardson (Eds.). *Cognitive and Neuropsychological Approaches to Mental Imagery*. NATO ASI Series. Boston.
- van Dijk, T.A. y Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, Academic Press.
- Duffy, S.A. (1986). Role of expectations in sentence integration. *J. of Experimental Psychology: Learning, Memory and Cognition*, 12, 2, 208-219.
- Finke, R.A. (1980). Levels of equivalence in imagery and perception. *Psychological Review*, 87,2, 113-132.
- Fodor, J.A. (1983). *The modularity of Mind. An essay on faculty psychology*. Cambridge, Massachusetts. Mit Press.
- Fodor, J.A. (1988). Modules, Frames, Fridgeons, Sleeping Dogs, and the Music of the Spheres. In Zenon W. Pylyshyn (Ed.): *The Robot's Dilemma. The Frame Problem in Artificial Intelligence*. Norwood, N.J.: Ablex.
- Garnham, A. (1987). *Mental Models as Representations of discourse and text*. Wiley & Sons. N.Y.
- Garrod, S. y Sanford, A. (1983). Topic dependent effects in language processing. In G.B. Flores d'Arcais y R.J. Jarvella (Eds.). *The process of language understanding*, N.J. Wiley & Sons.
- Gernsbacher, M.A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, 17, 3, 324-363.
- Glenberg, A.M., Meyer, M. y Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *J. of Memory and Language*, 26, 69-83.
- van Jaarsveld, H.J. & Schreuder, R. (1986). Implicit quantification of temporal adverbials. *Journal of Semantics*, 4, 327-339.
- Haberlandt, K. y Graesser, A.C. (1985). Component processes in text comprehension and some of their interactions. *J. of Experimental Psychology: General*, 114, 3, 357-374.
- Haberlandt, K.F., Graesser, A.C., Schneider, N.J. y Kiely, J. (1986). Effects of tasks and new arguments on word reading times. *J. of Memory and Language*, 25, 314-322.
- Janlert, Lars-Erik (1988). Modeling change-The frame problem. In Zenon W. Pylyshyn (Ed.): *The Robot's Dilemma. The Frame Problem in Artificial Intelligence*. Norwood, N.J.: Ablex.
- Jarvella, R.J. (1979). Immediate memory and discourse processing. In G.H. Bower (Ed.). *The psychology of learning and motivation*. New York, Academic Press.
- Johnson-Laird, P.N. (1983). *Mental models*. Cambridge, Inglaterra. Cambridge University Press.
- Johnson-Laird, P.N. (1988). A computational analysis of consciousness. In A.J. Marcel & E. Bisiach (Eds.), *Consciousness in contemporary science*. Oxford: Clarendon Press.

- Johnson-Laird, P.N. & Bara, B.G. (1984). Syllogistic inference. *Cognition*, 16, 1-61.
- Just, M.A. y Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 2, 163-182.
- Kintsch, W. y Keenan, J.M. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 5, 257-274.
- Kintsch, W. y Van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- Kosslyn, S.M. (1980). *Image and mind*. Cambridge, Mass.: Harvard University Press.
- Kosslyn, S.M., Ball, T.M. & Reiser, B.J. (1978). Visual images preserve metric spatial information. Evidence from studies of image scanning. *J. of Experimental Psychology: Human Perception and Performance*, 4, 1, 47-60.
- Levelt, J.M. (1989). *Speaking. From intention to articulation*. Cambridge, Mass.: The MIT Press.
- Marslen-Wilson, W.D. (1975). Sentence perception as an interactive parallel process. *Science*, 189, 226-227.
- Miller, G.A. & Johnson-Laird, P.N. (1976). *Language and perception*. Cambridge: Cambridge University Press.
- Morrow, D.G. & Clark, H.H. (1988). Interpreting words in spatial descriptions. *Language and Cognitive Processes*, 3, 4, 275-291.
- Oakhill, J., Garnham, A. & Vonk, W. (1989). The on-line construction of discourse models. Technical Report (non published).
- Polyshyn, Z. (1986). *Computation and cognition. Toward a foundation for cognitive science*. Cambridge, Mass.: The MIT Press.
- Rumelhart, D. y Ortony, A. (1979). The representation of knowledge in memory. In R.C. Anderson, R. Spiro y W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J., Lawrence Erlbaum Associates.
- Sanford, A & Garrod, S. (1981). *Understanding written language: Exploration of comprehension beyond the sentence*. Chichester: Wiley, 1981.
- Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, 91, 4, 417-447.
- Shepard, R.N., & Cooper, L.A. (1982). *Mental images and their transformations*. Cambridge, Mass.: The MIT Press.
- Shepard, R.N. & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1, 1-17.
- de Vega, M. (1984). *Introducción a la psicología cognitiva*. Madrid: Alianza Editorial.
- de Vega, M. (1988). Mental imagery and perception: Modularity or functional equivalence?. In M. Denis, J. Engelkamp & J.T.E. Richardson (Eds.): *Cognitive and neuropsychological approaches to mental imagery*. Dordrecht: Martinus Nijhoff Publishers.
- De Vega, M., Carreiras, M., Gutiérrez-Calvo, M. & Alonso-Quecuty, M.L. (1990). *Lectura y comprensión. Una perspectiva cognitiva*. Madrid: Alianza Editorial.
- De Vega, M. & Díaz, J.M. (1990). Mental models as an integrative device in discourse comprehension. Paper presented in the *Third European Workshop on Imagery and Cognition*. Aberdeen.

Taula (UIB) núm. 15, 1992

Sobre la autonomía de los procesos de formulación en la producción de oraciones*

José Manuel Igoa
Universidad Autónoma de Madrid

Resumen

La mayoría de los modelos de producción del lenguaje asumen que el proceso en virtud del cual el hablante formula sus intenciones comunicativas en forma de enunciados lingüísticos debe concebirse como una compleja interacción de subprocesos algunos de los cuales operan de forma relativamente autónoma, esto es, como módulos o sistemas cognitivos de propósito específico (Fodor, 1983). El objetivo de este trabajo es someter a prueba la hipótesis de la modularidad aplicada al sistema de salida responsable de la producción de oraciones, a fin de determinar en qué grado y de qué modo la formulación de enunciados lingüísticos es influida por las propiedades semánticas o conceptuales de los mensajes. El trabajo consta de dos partes. En la primera, un análisis de errores espontáneos del habla en castellano revela que el acceso de los procesos de formulación a las propiedades conceptuales del mensaje se halla considerablemente limitado. En la segunda, los resultados de un experimento de producción de oraciones en castellano en el que se empleó una tarea de «priming» léxico (Bock 1986, 1987b), muestran que existe una relación sistemática en el nivel funcional entre los procesos de selección léxica y planificación sintáctica, y que esta relación se halla fuertemente constreñida por características conceptuales de mensaje tales como la «animación» y la «agentividad».

Abstract

Most models of language production assume that the process whereby speakers formulate their communicative intentions in terms of linguistic utterances should be conceived of as a complex interplay of subprocesses, some of which operate in a relatively autonomous fashion, i. e., as cognitive, special-purpose modules (Fodor, 1983). The aim of this paper is to put to an empirical test the modularity hypothesis as applied to the sentence production system, so as to find out in what sense and to what extent the formulation of linguistic utterances is influenced by the semantic or conceptual properties of messages. The paper has two parts. In the first part, an analysis of spontaneous slips of the tongue in Spanish shows that the access of the formulation processes to the conceptual properties of the message is considerably limited. In the second part, the results of an experiment on sentence production in Spanish using a lexical priming procedure developed by Bock (1986, 1987b) suggest that there is a principled relationship between the processes of lexical selection and syntactic planning, and that this relationship is heavily constrained by conceptual characteristics of the message, such as animacy and agency.

* El presente trabajo ha sido financiado en parte con una ayuda de investigación de la Dirección General de Investigación Científica y Técnica (DGICYT) del Ministerio de Educación y Ciencia (Proyecto nº PB87-0531).

1. Introducción

Las teorías actuales de procesamiento del lenguaje sostienen que la generación de planes para la producción de oraciones comprende un conjunto relativamente amplio y, a la vez, considerablemente específico de tareas cognitivas. Estas tareas presentan características y exigencias peculiares claramente diferenciadas de las que plantean, por un lado, el pensamiento y la comunicación y, por otro, el control motor del habla. Esto significa que, más allá de las intenciones comunicativas que el hablante concibe, de las restricciones ambientales y de otra índole que pesan sobre los mensajes que han de ser formulados, y de las demandas articulatorias de las emisiones que el hablante ha de producir, queda una extensa gama de problemas que cualquier teoría del procesamiento del lenguaje tiene que resolver. Entre ellas cabe citar la recuperación de palabras, la combinación de las mismas en secuencias de constituyentes gramaticalmente bien formadas y la especificación de la estructura fonética y de entonación de las secuencias a emitir, por limitarnos a los procesos más generales que supuestamente se hallan implicados en la formulación de planes lingüísticos. La postulación de estos procesos por los modelos de procesamiento lingüístico tiene como fin dar cuenta de un sinnúmero de fenómenos surgidos de la observación del comportamiento lingüístico de los hablantes, tales como la estructura rítmica y melódica del habla, las interrupciones y disfluencias que se producen de forma habitual en monólogos y conversaciones, la variedad de formas oracionales empleadas por los hablantes de una lengua, los lapsus y errores que se cometen al hablar y las perturbaciones de las habilidades lingüísticas y comunicativas que siguen a lesiones cerebrales o a otras condiciones patológicas.

Todo esto no implica, empero, que las características conceptuales y pragmáticas de los mensajes no desempeñen papel alguno en los procesos de formulación. Pero sí sugiere que puede haber un dominio de procesamiento, llámesele mecanismo de formulación de oraciones, que opere con relativa autonomía con respecto a los objetivos de comunicación de los hablantes, a su conocimiento enciclopédico y a los contextos social, físico y cognitivo en que tienen lugar sus intercambios comunicativos. Por consiguiente, dos de los intereses primordiales de los modelos de procesamiento de la información lingüística deben ser, en primer lugar, definir las tareas efectuadas y los recursos de procesamiento empleados por este supuesto mecanismo de formulación de oraciones, y en segundo lugar, restringir y aclarar su interacción con otros sistemas involucrados en la producción de mensajes verbales.

La idea de un sistema de procesamiento funcionalmente autónomo que opera en un dominio de información altamente restringido, como son las representaciones lingüísticas en su totalidad, y que permite tan sólo un acceso limitado a otros componentes de procesamiento da pie a pensar que el sistema de formulación de oraciones debe concebirse como un «módulo cognitivo» (Fodor, 1983). No obstante, esta suposición requiere justificación si se ha de tomar realmente en serio. En su formulación inicial, el concepto fodoriano de módulo se aplica a sistemas de entrada como la percepción auditiva del habla, el reconocimiento de rostros y la percepción visual, entre otros, y no a los sistemas de salida o responsables de la acción. Por otra parte, para extrapolar la noción de sistema modular según se entiende actualmente a sistemas de salida como el

control motor o la producción del lenguaje, probablemente habría que modificar en cierto grado algunas de las propiedades definitorias de los sistemas de entrada o incluso limitar su aplicación en el caso particular de los sistemas de salida. Una de las razones de ello es que los criterios de modularidad definidos por Fodor ponen más énfasis en contraponer los sistemas de entrada, en tanto que clase natural de procesos cognitivos, con respecto a sistemas no cognitivos de transducción sensorial y a sistemas cognitivos destinados a la fijación de creencias o a la realización de inferencias no demostrativas, los llamados sistemas centrales.

En otro orden de cosas, una posibilidad que Fodor no hace explícita en su monografía sobre la modularidad de la mente es que los módulos cognitivos exhiban también una arquitectura modular: a saber, que los procesadores específicos de dominio, informativamente encapsulados y de propósito específico se hallen internamente organizados en términos de componentes funcionalmente autónomos que compartan, en un plano microestructural, las mismas propiedades modulares que las facultades verticales de orden superior. De hecho, muchos investigadores en psicolingüística han mantenido de forma implícita este supuesto durante los últimos veinte años. Esta estrategia de investigación incluso se ha articulado explícitamente en forma de postulados teóricos relativos al carácter modular de los procesos subyacentes al uso del lenguaje. Así ha ocurrido con el reconocimiento de palabras (Forster, 1976; Swinney, 1979; Seidenberg y Tanenhaus, 1986), o el análisis sintáctico (Ferreira y Clifton, 1986; Frazier, 1987), en el terreno de la percepción del lenguaje, y con la planificación sintáctica (Garrett, 1976; Bock, 1987a), la recuperación léxica (Kempen y Huijbers, 1983), o la codificación fonológica (Shattuck-Hufnagel, 1979), en el de la producción del lenguaje.

La mayor parte de las teorías al uso sobre producción del lenguaje comparten el supuesto de que el sistema de procesamiento destinado a la producción del lenguaje está organizado en tres componentes. Según Levelt (1989), cabe distinguir entre la *conceptualización*, la *formulación* y la *articulación*. La conceptualización comporta la concepción de una intención comunicativa, la selección de la información no lingüística adecuada que se desea transmitir y el seguimiento del modelo del discurso en desarrollo, así como de los aspectos más relevantes del contexto no lingüístico en que tiene lugar el acto de comunicación. La conceptualización debe considerarse como un caso paradigmático de actividad de «selección de creencias», dado que la tarea a realizar por el hablante es la de seleccionar un subconjunto de sus creencias o conocimientos para plasmarlos en forma lingüística. Resulta razonable suponer que esta actividad es vulnerable al problema del marco (Fodor, 1983; 1987), dado que en principio cualquier información sobre el conocimiento del mundo del hablante puede entrar en relación con la intención comunicativa que va a transmitir. El resultado de este proceso es un «mensaje preverbal», representación codificada en un formato no lingüístico cuyos elementos se hallan organizados de forma sincrética, es decir, no lineal o sin referencia alguna al orden serial de sus constituyentes.

En el extremo opuesto se halla el articulador, cuyo cometido es planificar un conjunto de órdenes motoras para ser ejecutadas por los órganos del tracto vocal a fin de producir habla manifiesta. Parte de este mecanismo ha de ocuparse de traducir la

representación abstracta de la secuencia de sonidos que integran el enunciado verbal (el plan o representación fonética) a una representación de la secuencia de comandos motores que habrá que ejecutar (el llamado «plan motor»). Otra parte de él se ha de encargar de la «transducción motora», es decir, de acoplar entidades simbólicas, codificadas en un formato lingüístico, como son los planes motores, con entidades físicas, pertenecientes a un dominio no cognitivo, como son los movimientos neuromusculares.

Por último, el proceso de formulación tiene la misión de traducir el mensaje preverbal al plan fonético, representación codificada en términos lingüísticos en la que se especifica el orden serial de constituyentes. Esta traducción se efectúa mediante el acoplamiento de información conceptual con información lingüística y la recuperación y organización de distintas clases de elementos lingüísticos (sintagmas, palabras, morfemas, sílabas y fonemas) en niveles jerárquicos de representación.

En este artículo me propongo cumplir dos objetivos: primero, evaluar la autonomía del proceso de formulación en su conjunto con respecto a la conceptualización, a la luz de la evidencia procedente del análisis de errores y correcciones espontáneas del habla en castellano. Segundo, exponer y comentar un experimento sobre producción de oraciones llevado cabo recientemente a fin de examinar la relación entre, y con ello la hipotética autonomía de, dos procesos que operan dentro del formulador: la recuperación léxica, o proceso por el cual el sistema accede a las representaciones almacenadas del significado del enunciado, o sea, palabras y morfemas; y la planificación sintáctica, o proceso que permite la recodificación de representaciones conceptuales a una secuencia lineal de constituyentes lingüísticos en la que se especifican las relaciones y dependencias estructurales entre los mismos. En suma, el objetivo de este trabajo es arrojar luz sobre la cuestión de la modularidad del componente de formulación lingüística en la producción de oraciones, vista tanto desde fuera como desde dentro de este componente de procesamiento.

2. La modularidad del proceso de formulación: análisis de errores espontáneos del habla y sus correcciones¹

Buena parte de la investigación actual en producción del lenguaje basada en el análisis de errores espontáneos del habla asume que la relación entre la conceptualización y la formulación procede estrictamente de arriba hacia abajo. Esto supone que las decisiones que se toman en el nivel del mensaje (conceptualización) no dependen de propiedades o restricciones de los procesos del nivel de la oración (formulación). Una prueba indirecta de ello es que la mayoría de los errores del habla vienen dados por la sustitución, trasposición o supresión de elementos del enunciado pretendidos por el hablante, o sea, de unidades de información pertenecientes a un único mensaje preverbal

¹ Los análisis de errores expuestos en este trabajo se han efectuado sobre un *Corpus de Errores en Castellano* integrado por 3611 errores (Viso, Igoa y García-Albea, 1987). Los datos cuantitativos aparecen en el prólogo del corpus, así como en Viso (1990). Asimismo, gran parte de los ejemplos presentados en este apartado proceden de este corpus.

que el hablante desea proferir. En cambio, muy pocos errores surgen de la interacción de elementos pretendidos y no pretendidos en el plan lingüístico, esto es, se generan por el solapamiento de diferentes mensajes dentro de un mismo plan. Por consiguiente, casi todos los errores suelen explicarse en virtud de un transitorio funcionamiento anómalo del proceso de formulación, considerándose «internos al plan».

Sin embargo, se puede argüir que muchos errores clasificados como internos a un plan surgen en realidad como resultado del despliegue de «planes lingüísticos alternativos» que expresan un único «mensaje» (Butterworth, 1982). El caso más patente en nuestro corpus se da en las fusiones de palabras (ejemplo 1a) o sintagmas (ejemplo 1b), aunque puede aplicarse también a ciertos casos de trasposición de palabras (ejemplos 2 y 3). Esta nueva interpretación, pues, atañe más a la hipotética causa del error que a sus efectos lingüísticos.

- (1) a. Hay un buen *trozo* (trocho /trecho)
 b. Es que vas con *la lengua al cuello* (la lengua fuera /el agua al cuello)
- (2) Intercambios
 a. ¿Quieres ir al *coche en banco*? (al banco en coche /en coche al banco)
 b. ¿Has hablado *con esto de* tu maestro? (de esto con tu maestro /con tu maestro de esto)
- (3) Desplazamientos
 a. ¿Te sienta.....[pausa] te molesta *mal*? (te sienta mal /te molesta)
 b. Merece *ser* la penaescuchado (merece ser escuchado / merece la pena ser escuchado)

Por otra parte, también se ha apuntado la posibilidad de que ciertos errores surjan de la competencia entre dos (o más) mensajes distintos. En tales casos, algunos elementos lingüísticos pertenecientes a un mensaje (M2) diferente al pretendido por el hablante (M1) se mezclan con, o sustituyen a, elementos del mensaje pretendido (M1) originando el error. Butterworth (1982) y Harley (1984) aportan abundante evidencia de estos errores de «planes en competencia», también denominados «intrusiones cognitivas» dado que el elemento «intruso» procede del contexto cognitivo en que tiene lugar el acto locutivo, sea éste el entorno perceptivo del hablante (ejemplo 4a), el contexto de la conversación (4b), e incluso, a falta de otra evidencia, la memoria de trabajo del hablante (4c).

- (4) a. A y veintisiete salía el *perro* (tren) de la estación
 [Había un perro acostado al lado del hablante al producirse el error]
 b. El estómago de las *uñas* (los rumiantes)
 [El hablante acababa de regañar a una niña por morderse las uñas]
 c. Se me cayó una *ensalada* (zapatilla) en la cueva de Nerja
 [causa desconocida]

La diferencia entre errores internos al plan, de planes alternativos y de planes en competencia presenta gran interés para la evaluación de la autonomía del conceptualizador frente al formulador. A diferencia de los errores internos al plan, los de planes alternativos y planes en competencia suponen el despliegue y la coexistencia al menos de dos planes distintos simultáneamente formulados y que eventualmente interactúan para dar origen al error. Con todo, una diferencia capital entre errores de planes alternativos y de planes en competencia es que sólo en estos últimos se selecciona

más de un mensaje en el conceptualizador. Cuando dos planes entran en competencia, el formulador acoge elementos de mensajes diferentes suministrados por el conceptualizador y genera un enunciado contaminado por elementos de un mensaje ajeno al pretendido. Así pues, los errores de planes en competencia se originan en el nivel del mensaje (conceptualización) y penetran el formulador hasta cierto grado, arrastrando las consecuencias lingüísticas de la formulación hasta la articulación. En cambio, en los errores de planes alternativos el conceptualizador suministra un solo mensaje al formulador, el cual genera dos planes lingüísticos alternativos para codificar dicho mensaje.

Los modelos interactivos de producción del lenguaje (Dell y Reich, 1981; Harley, 1984; Stemberger, 1985) sostienen que los errores de palabras se ven facilitados por propiedades de «bajo nivel» de las representaciones lingüísticas en que se hallan reflejados, lo que ocurre con independencia de las causas que los originan, afectando por igual a los errores internos al plan y a los de planes alternativos y de planes en competencia. En concreto, la semejanza fonológica entre los elementos pretendido y emitido incrementa la probabilidad de que se produzca un error. Este fenómeno ya ha sido constatado empíricamente en algunos trabajos de análisis de errores espontáneos en inglés (Dell y Reich, 1981; Harley, 1984; Dell, 1988; Levelt, 1989), aunque no ha podido ser confirmado en los errores internos al plan en castellano (Viso, Igoa y García-Albea, 1991). No obstante, si esto resultara cierto, sobre todo en errores de planes en competencia, apoyaría la idea de que además del reconocido flujo de arriba-abajo del conceptualizador al formulador, las representaciones del nivel del mensaje también pueden ser activadas por medio de retroinformación procedente de representaciones fonológicas en el nivel de la oración, lo que supondría que la conceptualización no es autónoma con respecto a la formulación. En cambio, si la influencia fonológica sólo afecta a los planes alternativos, el efecto de retroinformación quedará confinado al formulador, preservándose la autonomía del conceptualizador con respecto a propiedades lingüísticas relativamente locales de la planificación de oraciones.

En el análisis que voy a exponer a continuación, se evaluó la semejanza fonológica entre el error y la intención en una muestra de errores léxicos en castellano originados por la intrusión bien de un plan alternativo o bien de un plan en competencia en el proceso de formulación. Se emplearon cuatro criterios de semejanza fonológica: primero, el porcentaje de casos en que el error y la intención compartían los mismos segmentos fonémicos en el mismo orden serial; segundo, el porcentaje de casos en que ambas palabras tenían la misma estructura vocálica (también en el mismo orden serial); tercero, número de sílabas, y cuarto, patrón de acento de los elementos que entraban en interacción en el error.

A efectos de mayor fiabilidad de los resultados, se compararon los porcentajes empíricos obtenidos con las probabilidades, previamente calculadas, de que un emparejamiento al azar de dos palabras cualesquiera arrojara una coincidencia en cada uno de los parámetros fonológicos citados. (Este procedimiento fue empleado ya por Dell y Reich, 1981; los cálculos sobre los materiales en castellano aparecen en Viso, 1990.)

Los análisis se efectuaron sobre un total de 105 errores de planes alternativos y 102 de planes en competencia. Los resultados relativos al primer criterio de semejanza fonológica, la correspondencia en los segmentos fonémicos del error y del «target» en

Taula 15

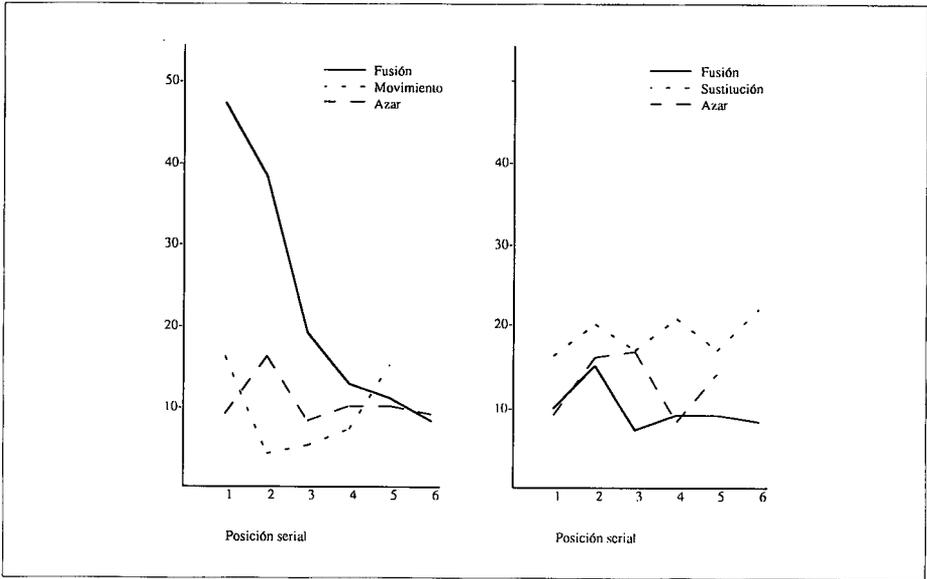


Figura 1. IZQUIERDA: Porcentaje de correspondencia entre error y target en los fonemas situados en las posiciones seriales 1ª a 6ª, en errores de movimiento y fusión de palabras originados por planes alternativos. **DERECHA:** Porcentaje de correspondencia entre error y target en los fonemas situados en las posiciones seriales 1ª a 6ª, en errores de sustitución y fusión de palabras originados por planes en competencia.

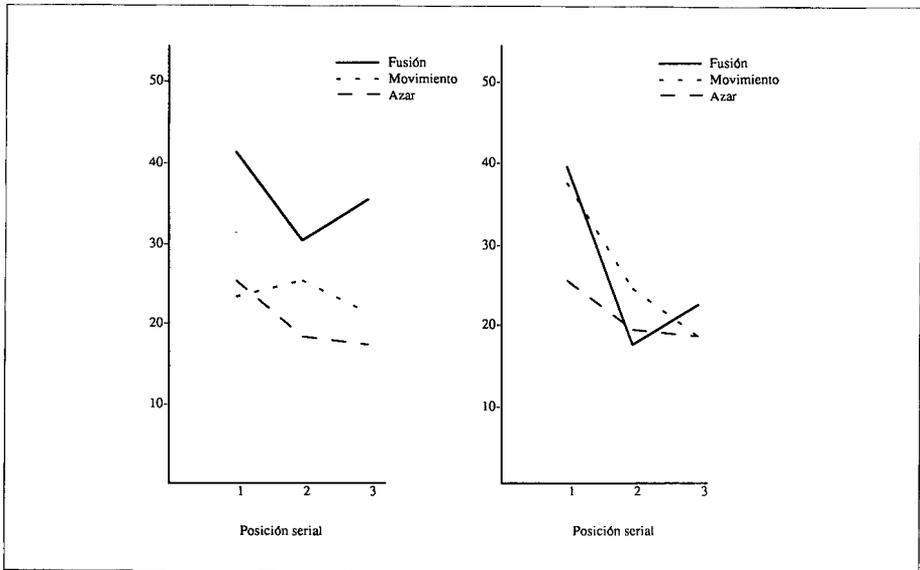


Figura 2. IZQUIERDA: Porcentaje de correspondencia error-target en estructura vocálica en errores de movimiento y fusión de palabras por planes alternativos. **DERECHA:** Porcentaje de correspondencia error-target en estructura vocálica en errores de sustitución y fusión de palabras por planes en competencia.

Orden serial, indican una correspondencia fonológica significativa en fusiones de palabras originadas por planes alternativos, en comparación tanto con la estimación al azar como con los errores de movimiento de palabras por planes alternativos. Esto implica que en las fusiones de palabras debidas a planes alternativos, las unidades error y «target» muestran una considerable semejanza fonológica. Por su parte, en los errores de planes en competencia no se aprecian diferencias significativas en la correspondencia de segmentos fonémicos en orden serial en ninguna de las categorías de error sometidas a análisis en comparación con el azar (Figura 1).

La diferencia observada en los errores de planes alternativos entre las fusiones y los errores de movimiento no es sorprendente, habida cuenta de que las palabras que son objeto de transposición en errores causados por planes alternativos no compiten, a diferencia de lo que ocurre en las fusiones, por un único «target» léxico del enunciado lingüístico, sino que ambas forman parte de los dos planes que entran en interacción y, por consiguiente, disponen de un lugar asignado en la estructura superficial del enunciado. Así pues, no hay razón alguna por la que deban compartir propiedades segmentales (compárense los errores de (1) con los de (2) y (3).

En lo que atañe al segundo criterio, la estructura vocálica, se obtuvieron idénticos resultados a los registrados en el primer índice (Figura 2): se dio una correspondencia fonológica significativa en las fusiones de palabras por planes alternativos, aunque no en los errores de movimiento. En cambio, no hubo diferencias con respecto al azar en el caso de los planes en competencia.

Por lo que a los criterios 3º y 4º se refiere, las fusiones de palabras registraron una correspondencia estadísticamente significativa en comparación con el azar en el número de sílabas, aunque no en el patrón de acento. Por su parte, los errores de movimiento no mostraron coincidencia fonológica alguna en ninguno de los dos índices. Similarmente, las dos categorías de errores de planes en competencia (fusiones y sustituciones) mostraron una correspondencia superior a la esperada por azar en el número de sílabas, pero no en el patrón de acento (Figuras 3 y 4)

La correspondencia exhibida por los errores de planes en competencia en número de sílabas es un resultado inconveniente, ya que no hay razón alguna por la que una propiedad en principio ajena a la estructura conceptual de un mensaje, como es la estructura silábica de las palabras que lo integran, deba aumentar la probabilidad de que se cometa un error lingüístico de origen conceptual (planes en competencia). Pese a ello, cabría argumentar que la longitud en sílabas o el patrón de acento de las palabras no son parámetros tan determinantes de su identidad fonológica como su estructura segmental o vocálica, dado que aquéllos están más asociados a los marcos sintagmáticos en los que se insertan las palabras de contenido durante la planificación de la oración por lo que se considera que operan en etapas previas a la codificación fonológica (Cf. Cutler, 1982, para una discusión sobre este asunto). La información sobre la estructura silábica de los constituyentes léxicos podrá facilitar la ocasional selección errónea de elementos léxicos previamente activados por la concepción de un plan alternativo o incluso en competencia introducido en el formulador. Con todo, esto no deja de ser una especulación que no permite descartar la posibilidad de que los procesos de formulación impongan restricciones débiles sobre la selección errónea de palabras en la planificación de oraciones.

Taula 15

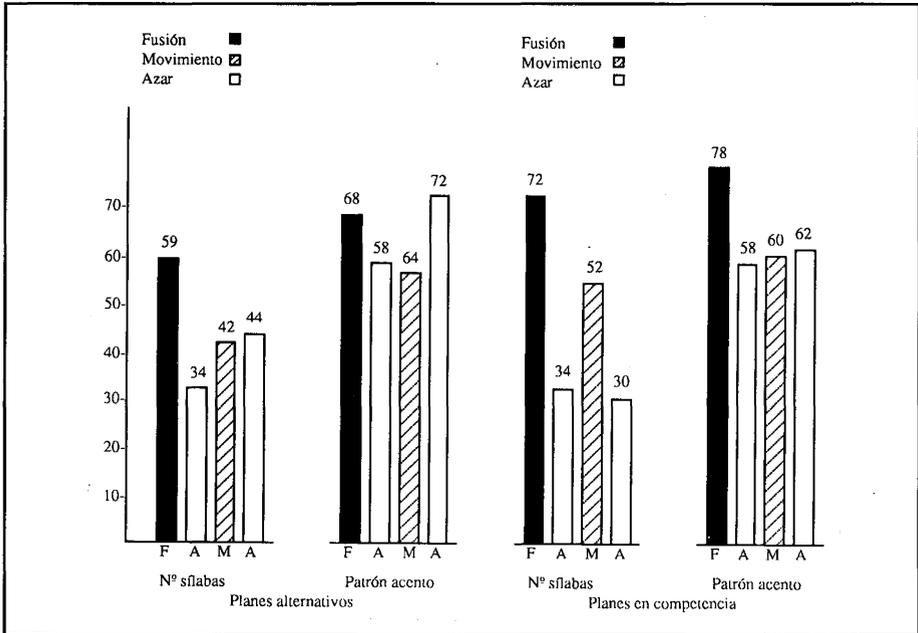


Figura 3. IZQUIERDA: Porcentaje de correspondencia error-target en número de sílabas (izquierda) y patrón de acento (derecha), en errores de movimiento y fusión de palabras originados por planes alternativos. **DERECHA:** Porcentaje de correspondencia error-target en número de sílabas (izquierda) y patrón de acento (derecha), en errores de sustitución y fusión de palabras originados por planes alternativos.

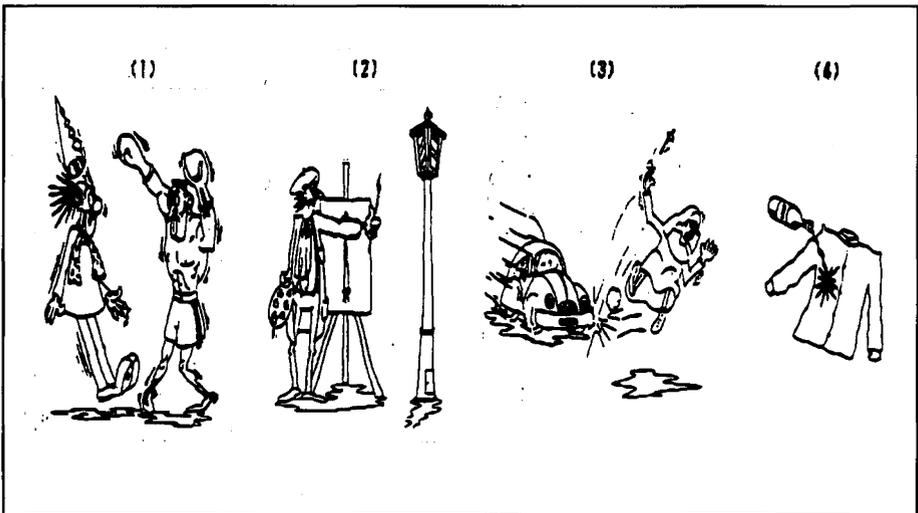


Figura 4. Ejemplos de dibujos empleados en cada uno de los cuatro subexperimentos; (1) agente y paciente animados; (2) agente animado y paciente inanimado; (3) agente/instrumento animado y paciente animado; y (4) agente/instrumento y paciente/meta inanimados.

Estos resultados sugieren que ciertos errores hipotéticamente derivados de la intromisión de planes lingüísticos alternativos concebidos para expresar un solo mensaje se ven facilitados por las propiedades fonológicas de sus constituyentes. Dado que tales propiedades se recuperan en un nivel de procesamiento relativamente inferior, cabe concluir que la información fonológica es retroalimentada hacia el nivel de la selección léxica, o bien que el formulador genera planes lingüísticos alternativos que se amalgaman en la codificación fonológica para dar lugar a errores. Por el contrario, los errores inducidos por planes en competencia, originados por el conceptualizador, no parecen ser sensibles a cualquier tipo de restricción fonológica, lo que supone que la retroinformación fonológica está confinada a los procesos de formulación de oraciones, sin que llegue a afectar a la concepción de mensajes. Dicho en términos modulares, los análisis efectuados revelan el encapsulamiento del conceptualizador con respecto a información representada en el formulador, a la vez que el funcionamiento interactivo de al menos algunos de los procesos internos a este último componente.

Pasemos ahora a considerar en qué sentido el examen de las correcciones de errores puede aportar evidencia sobre la modularidad de la formulación lingüística. Es sabido que los hablantes controlan e inspeccionan casi de continuo sus emisiones lingüísticas y a menudo corrigen los errores que cometen. Según Levelt (1983, 1989), la actividad de inspección y corrección de errores se puede aplicar de modo independiente a las diversas estructuras de información implicadas en la producción del lenguaje y, por tanto, puede acontecer en diferentes etapas de procesamiento. Parece comprobado que los tres sistemas de procesamiento en la producción verbal (conceptualización, formulación y articulación) pueden ser inspeccionados y controlados de forma mutuamente independiente. Por otra parte, se ha observado que las pautas de interrupción y la forma de las correcciones están regidas por principios de buena formación gramatical que relacionan la corrección con el enunciado original, al objeto de preservar la fluidez del habla y la coherencia del discurso (Levelt, 1983²). Si estos principios de buena formación, de cuya ejecución se ocupa el formulador, se aplican de modo similar en distintas clases de errores, tendríamos una prueba del carácter modular del proceso de formulación. En concreto, si se demuestra que las consecuencias conceptuales de los errores no ejercen influjo sobre las pautas de corrección, entonces las correcciones ejecutadas por el formulador deberían considerarse procesos independientes de la conceptualización.

A continuación, voy a resumir los resultados de un análisis de 913 correcciones espontáneas de errores subléxicos y 806 correcciones de errores léxicos recogidas por escrito en el corpus de errores del castellano, que pueden resultar reveladores de cara a la independencia de los procesos de inspección y corrección con respecto a las propiedades conceptuales de los mensajes que contienen errores del habla. En primer lugar, se evaluaron las pautas de interrupción (en número de sílabas y/o fronteras sintácticas cruzadas) y al *locus* de retroceso al efectuar la corrección; en segundo lugar,

² Los criterios de buena formación gramatical de las correcciones que se han seguido en este análisis aparecen expuestos en Levelt (1983). En esencia, establecen que una corrección se halla bien formada siempre que el fragmento «error» y el fragmento «corrección» puedan acoplarse a una estructura coordinada sin violar restricciones estructurales.

se compararon las pautas de interrupción y reinicio en errores léxicos semánticamente congruentes y semánticamente anómalos. El propósito de este análisis era determinar si las propiedades conceptuales de los mensajes se reflejan en las pautas de corrección de los errores, que supuestamente tienen lugar bajo control de los procesos de formulación.

Por lo que a la pautas de corrección se refiere, se da una relación directamente proporcional entre el tamaño de la unidad lingüística objeto de error y la distancia del error al *locus* de interrupción, de suerte que los errores subléxicos tienden a ser interrumpidos con mayor prontitud que los errores léxicos. Por lo que atañe a la buena formación gramatical de las correcciones, hay que señalar que el 92 por ciento de las correcciones examinadas satisfacían los criterios señalados por Levelt.

En un análisis subsiguiente se dicotomizaron los errores léxicos provistos de corrección en dos categorías: una (N=279) constituida por errores conceptualmente anómalos («He visto un *lavabo* [mosquito] en el *mosquito* [lavabo]») y otra (N=527) integrada por errores conceptualmente apropiados («el pato [perro] parecía un perro [pato]»). Si las características conceptuales de los errores influyera en la forma de efectuar correcciones, cabría esperar que se dieran interrupciones más rápidas en los errores conceptualmente anómalos, bajo el supuesto de que las anomalías semánticas son detectadas de forma inmediata por el conceptualizador. Asimismo, habría que esperar mayores distancias de retroceso en la primera categoría de errores, en aras de facilitar una mejor comprensión de la corrección al interlocutor. No obstante, contrariamente a lo que cabía esperar, los resultados de este análisis indicaron que los errores semánticamente anómalos tendían a ser interrumpidos más tarde que los semánticamente coherentes, si bien los *loci* de retroceso de ambas subclases de errores no presentaban diferencia alguna. Una posible explicación del efecto de interrupción tardía de los errores semánticamente anómalos es que los procesos de inspección y control del habla no abarcan de forma simultánea los diferentes niveles de representación en que puede ser analizada la señal, pudiendo recibir el nivel semántico un análisis demorado con respecto a otros niveles de orden inferior.

A tenor de los datos examinados en este apartado, parece razonable concluir que al menos una parte de los procesos de formulación que están al servicio de la planificación de oraciones se hallan preservados del influjo de los procesos responsables de la generación de mensajes. La preservación de la buena forma lingüística de las correcciones, tanto en mensajes semánticamente incongruentes como congruentes (aunque ambos tipos fueran naturalmente inadecuados desde la perspectiva de la intención comunicativa del hablante), viene a indicar que las propiedades del *input* que suministra el conceptualizador al formulador tienen consecuencias limitadas para las operaciones que realiza este último, especialmente en lo relativo a la forma lingüística de los enunciados verbales. En consonancia con esto, se ha observado que la formulación de oraciones juega un papel muy limitado en el proceso de generación de mensajes, lo que sugiere la autonomía del conceptualizador frente al formulador. Según muestran los datos sobre semejanza fonológica, las propiedades formales de las palabras pueden ejercer un influjo en procesos de orden superior dentro del formulador, e. g. la recuperación léxica, pero este efecto de retroinformación no alcanza el nivel en el que se elaboran los mensajes preverbales.

3. La modularidad en los procesos de formulación: la relación entre la recuperación léxica y la planificación sintáctica

Prácticamente todos los modelos de producción del lenguaje dan por sentada la distinción entre niveles de representación correspondientes, de un lado, a las palabras y, de otro, a la estructura de la oración. Los modelos seriales o autónomos mantienen además que esta distinción afecta a los procesos o cómputos que el sistema realiza durante la formulación lingüística de los mensajes. Afirman que los procesos de recuperación de elementos léxicos y los procesos de construcción del armazón sintáctico operan independientemente unos de otros. Por el contrario, los modelos interactivos consideran que no existe autonomía de procesamiento entre el léxico y la sintaxis, e insisten en que la activación de representaciones léxicas influye directamente en la ordenación de los constituyentes en la estructura de superficie y, por tanto, en la construcción de la estructura sintáctica del enunciado.

Por otro parte, nos consta que tanto los procesos léxicos como los sintácticos son conjuntos relativamente heterogéneos de fenómenos que se hallan distribuidos en distintos niveles de procesamiento, llevando a cabo tareas bastante específicas en cada uno de ellos. Algunos modelos, por ejemplo, descomponen el proceso de lexicalización en dos subcomponentes autónomos que operan de forma serial: un «léxico semántico», del que se seleccionan los elementos léxicos a partir de sus propiedades conceptuales, y un «léxico fonológico», donde se recupera la forma de las palabras (Garrett, 1975; Kempen y Huijbers, 1983; Schriefers *et al.*, 1990). En lo que a la sintaxis se refiere, se suele establecer una diferenciación entre dos clases de procesos: de una parte, los implicados en la elaboración de la estructura subyacente o profunda del enunciado, encargados de acoplar los papeles temáticos a funciones sintácticas y de establecer dependencias sintácticas entre los constituyentes del enunciado; y de otra, los relacionados con la construcción de la estructura superficial, que se ocupan de fijar el orden de los constituyentes en la oración, recuperar los morfemas gramaticales que sirven de marcadores sintagmáticos y mantener la concordancia morfosintáctica entre los constituyentes. Estos dos niveles de procesamiento sintáctico corresponden a los niveles «funcional» y «posicional» del modelo de producción de oraciones de Garret (1975, 1976).

Este esquema estructural es un punto de partida idóneo para contrastar hipótesis sobre las relaciones funcionales entre los componentes de procesamiento internos al formulador, y en consecuencia, determinar su posible encapsulamiento informativo. El experimento que voy a exponer en este apartado se centra en las relaciones entre la recuperación léxica y la planificación sintáctica, aunque se ocupa tanto de las influencias que se dan dentro de cada nivel (funcional o posicional) como de las que se puedan producir entre los dos niveles, ya sea en sentido «ascendente» (bottom-up) o «descendente» (top-down). El problema tratado en este experimento es el de la integración de información léxica en el marco sintáctico oracional durante la planificación, problema conocido también como «de la coordinación» (Bock, 1987a) o del «orden serial» (Lashley, 1951). Las implicaciones de este problema para la modularidad de los procesos de formulación son evidentes: si el proceso de construcción de la estructura sintagmática resulta afectado de algún modo por el proceso de recuperación de los

elementos léxicos, entonces la planificación sintáctica no podrá contemplarse como un proceso informativamente encapsulado; si, por el contrario, se observa que la selección de elementos léxicos no altera la planificación de la estructura sintáctica, habrá que mantener la separación funcional de estos dos componentes de procesamiento.

El problema de la coordinación se halla sujeto a constricciones de dos tipos: primero, constricciones extralingüísticas que vienen dadas por las características perceptivas y conceptuales de los referentes y contenidos del mensaje, y por exigencias pragmáticas de la conversación. Segundo, constricciones lingüísticas impuestas fundamentalmente por el estado de activación de la información sintáctica, léxica o fonológica en el sistema en el momento de iniciar la formulación de un plan lingüístico. Por lo que se refiere a las primeras, cabe citar el llamado «principio del orden natural», que establece que los contenidos han de ser expresados según la sucesión natural de los acontecimientos en el mundo (e.g., orden cronológico, proximidad espacial, etc.), el «principio de carga mínima», que exige que las transiciones del discurso o la conversación impongan el mínimo esfuerzo de memoria posible para el interlocutor, o el «principio de viveza», en virtud del cual las entidades con mayor intensidad perceptiva o afectiva tienden a preceder en la oración a las comparativamente más neutras (Osgood y Bock, 1977; Bock, 1982; Levelt, 1989).

En lo concerniente al segundo tipo de constricciones, hay pruebas bastante persuasivas de que la activación o disponibilidad de representaciones lingüísticas puede determinar el orden serial de los constituyentes de la oración. Algunos trabajos (Levelt y Maassen, 1981; Bock, 1986, 1987b) han puesto de manifiesto que la disponibilidad léxica puede inducir cambios en la formulación sintáctica de un modo similar a como el análisis sintáctico en la comprensión puede ser objeto de reformulación cuando el oyente o lector es inducido a efectuar un primer análisis erróneo del *input* lingüístico.

El procedimiento empleado en algunos de estos estudios consistía en una tarea de descripción verbal de dibujos combinada con una técnica de «priming» auditivo con material lingüístico. Inmediatamente antes de la presentación de cada dibujo para su descripción verbal, el sujeto repetía en voz alta una palabra o una frase presentada auditivamente. El material auditivo servía como «prime», con el cometido de activar en la memoria de trabajo del sujeto cierta información de hipotéticamente podía inducir cambios en la estructura sintáctica de la descripción. En los estudios en que se examinaba el influjo de la disponibilidad léxica sobre la planificación sintáctica, se empleaban como «primes» palabras que podían guardar una relación de significado o de forma con entidades presentes en el dibujo (y, por tanto, en la representación conceptual del mensaje). Así, en una escena que representa un caballo dando una coz a una vaca, la relación entre cada «prime» y su correspondiente «target» podía ser semántica (pony-caballo; leche-vaca) o fonológica (cobaya-caballo; laca-vaca). Asimismo, el «priming» podía afectar a cualquiera de las entidades temáticas del dibujo; al agente (pony/cobaya/ «caballo»), o al paciente (leche/laca/ «vaca»). Si la activación de una representación léxica influye en el orden de mención de las entidades presentes en el dibujo, se dará una mayor probabilidad de descripciones con el elemento «primado» en posición inicial de la oración. Cuando dicho elemento sea el agente del dibujo, habrá mayor proporción de oraciones activas (*El caballo está coceando a la vaca*); cuando sea el paciente, aumentará la proporción de pasivas (*La vaca es coceada por el caballo*).

Es obvio que este patrón de relaciones entre la activación léxica y la construcción sintáctica presupone que el procedimiento de «priming» induce un efecto invariablemente «excitatorio» sobre la información léxica asociada a los «primes». Aunque este supuesto no presenta problema alguno en el caso del «priming» semántico, se ha observado que tanto en percepción y comprensión (Peterson *et al.*, 1989) como en producción del lenguaje (O'Seaghdha *et al.*, en prensa) la activación de información fonológica origina una inhibición de las representaciones (léxicas) que comparten elementos idénticos a los activados, lo que se traduce en mayores latencias de reconocimiento léxico, mayor proporción de errores de articulación o la aparición más tardía de las palabras objeto de «priming» fonológico. El efecto inhibitorio del «priming» fonológico en la producción de oraciones ha sido constatado en un reciente estudio de Bock (1987b) en el que se empleó el procedimiento de priming léxico antes descrito.

La hipótesis de la interacción entre la recuperación léxica y la planificación sintáctica, con su doble efecto de facilitación semántica e inhibición fonológica (Bock, 1986, 1987b), fue sometida a prueba en un experimento de «priming» léxico con una tarea de descripción de dibujos. El objetivo del experimento era determinar hasta qué punto la activación de representaciones léxicas es capaz de inducir cambios en la estructura sintáctica del enunciado durante la producción del lenguaje, en comparación con una variable conceptual no controlada léxicamente, como es la animación de las entidades incluidas en la representación del mensaje. Por otra parte, se trataba de examinar la posible disociación de efectos excitatorios e inhibitorios en función del tipo de «priming» empleado. Como en anteriores experimentos de esta naturaleza, el «priming» podía ser semántico o fonológico y podía afectar a los papeles temáticos del «agente» y «paciente», lo que supone que todas las escenas experimentalmente relevantes constaban de un agente y un paciente relacionados por medio de acciones transitivas. A causa de la manipulación de la animación de las entidades que aparecían en los dibujos, el experimento fue dividido en cuatro partes. Así, una cuarta parte de escenas tenían un agente y un paciente animados; otra cuarta parte, un agente animado y un paciente inanimado; otra, un agente inanimado y un paciente animado; y la última, un agente y un paciente inanimados.

La inclusión de una variable conceptual como la animación tenía por objeto contrastar el influjo de variables lingüísticas (internas al formulador) y extralingüísticas (procedentes del conceptualizador) sobre la planificación sintáctica (orden de mención de entidades en la oración), con el propósito de examinar el encapsulamiento o, en su caso, la permeabilidad informativa de los procesos de formulación con respecto a estas dos posibles fuentes de interferencia. Hay que advertir, además, que si bien el origen de las influencias es distinto (conceptual, en la animación de las entidades; léxico-semántico y léxico-fonológico, en el «priming»), el *locus* de éstas es siempre el mismo, dado que las posibles variaciones producidas en la estructura sintáctica implicarán por definición un cambio en la estructura funcional del enunciado (estructura profunda, en términos de teoría lingüística). Para analizar estos mismos efectos en la estructura posicional, habría que emplear materiales que indujeran variaciones lingüísticas más locales y que no comprometieran la estructura completa del enunciado (para un análisis más preciso de los fenómenos estructurales atribuibles a las representaciones funcional (profunda) y posicional (superficial), véase Sells, 1985).

Los resultados del experimento se presentan en cuatro figuras independientes, una por cada condición de animación. En la parte izquierda de cada figura se recoge la proporción de descripciones en que el agente figuraba en posición inicial del enunciado, obtenidas bajo tres condiciones de «priming» léxico al agente (semántico, fonológico y control, esto es, ausencia de «priming»); en la parte derecha de cada figura se da la proporción de respuestas con el paciente de la escena en posición inicial, bajo tres condiciones de «priming» léxico al paciente. Este desglose de los datos permite evaluar por separado la magnitud del efecto del «priming» léxico sobre cada entidad del dibujo.

Subexperimento 1 (agente y paciente animados): se obtuvo un efecto significativo del priming semántico con respecto a la condición de control, mientras que el priming fonológico apenas ejerció influencia. Esto implica que la activación de representaciones léxico-semánticas asociadas a entidades conceptuales provoca la anteposición de la entidad que alcanza mayor nivel de activación por efecto del priming. En cambio, la activación de representaciones léxico-fonológicas no resulta eficaz a este respecto. Hay que señalar, no obstante, que el efecto excitatorio del priming semántico es sensiblemente mayor en la condición de «priming al paciente», lo que indica la existencia de un efecto techo para el priming semántico al agente (Figura 5). Así pues, cuando las entidades que forman la representación de un mensaje transitivo son animadas, se tiende a emplear el orden «agente-acción-paciente» en su descripción. Esta tendencia podría interpretarse según un hipotético «principio de ‘agentividad’», que impondría un sesgo hacia la mención inicial del agente como condición no marcada en la conceptualización de mensajes.

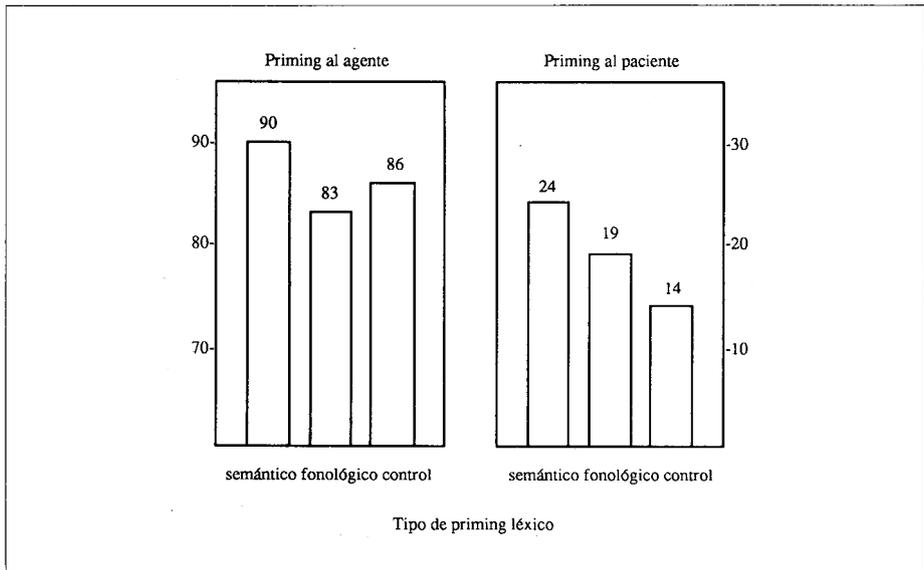


Figura 5. Proporción de respuestas (1) con el agente en posición inicial del enunciado (oraciones activas) [izquierda] y (2) con el paciente como constituyente tematizado al inicio de la oración (pasivas y dislocaciones a la izquierda) [derecha], bajo diversas condiciones de priming: (1) al agente y (2) al paciente, con el agente y el paciente animados (subexperimento 1).

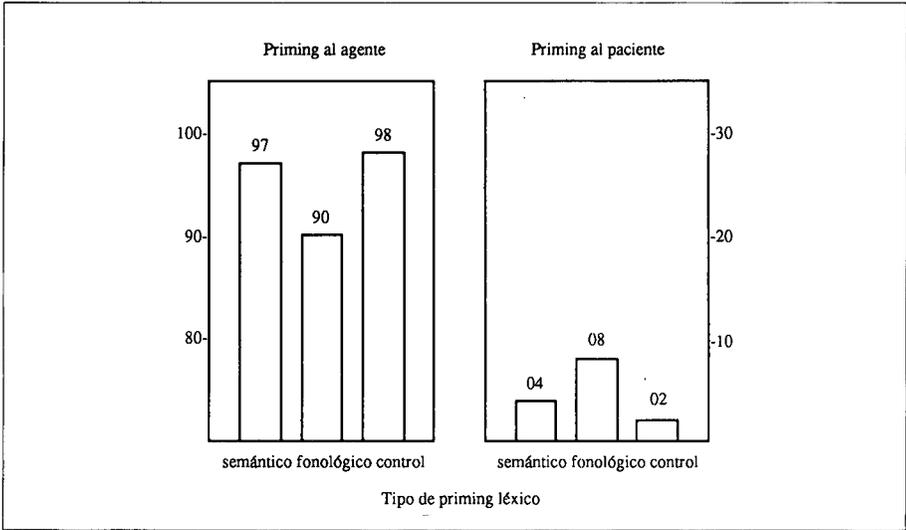


Figura 6. Proporción de respuestas (1) con el agente en posición inicial del enunciado (oraciones activas) [izquierda] y (2) con el paciente como constituyente tematizado al inicio de la oración (pasivas y dislocaciones a la izquierda) [derecha], bajo diversas condiciones de priming: (1) al agente (2) al paciente, con el agente animado y el paciente inanimado (subexperimento 2).

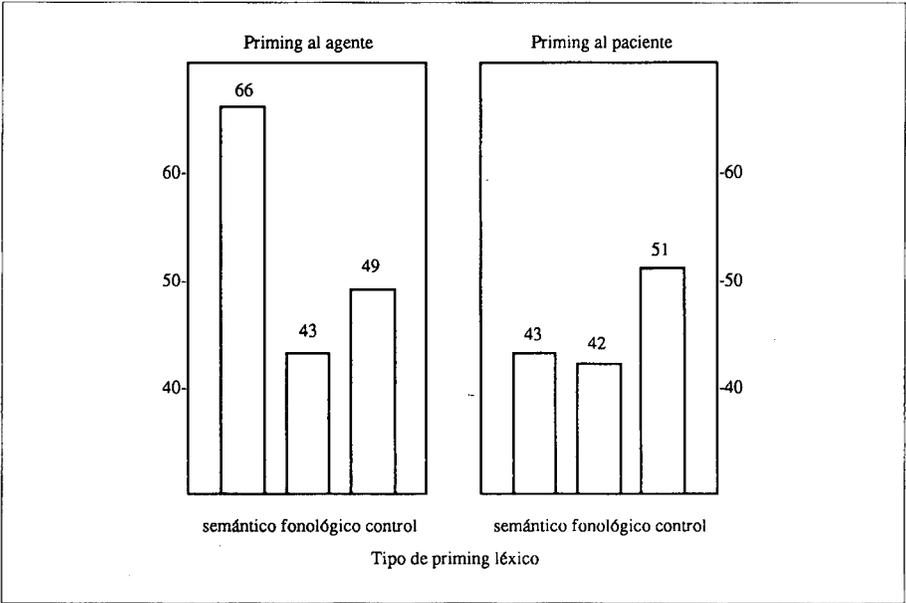


Figura 7. Proporción de respuestas (1) con el agente en posición del enunciado (oraciones activas) [izquierda] y (2) con el paciente como constituyente tematizado al inicio de la oración (pasivas y dislocaciones a la izquierda) [derecha], bajo diversas condiciones de priming: (1) al agente y (2) al paciente, con el agente inanimado y el paciente animado (subexperimento 3).

Subexperimento 2 (agente animado y paciente inanimado): no se apreciaron efectos de priming léxico en ninguna de sus modalidades sobre el orden de mención de las entidades conceptuales (Figura 6). En este caso, al principio de «agentividad» se le podría añadir un principio de «animación», que, en casos de asimetría como éste, impondría la anteposición de la entidad animada a la inanimada. Esta preferencia por la mención inicial de entidades animadas ha sido constatada en la literatura sobre producción y recuerdo de oraciones (Harris, 1978; Dewart, 1979).

Subexperimento 3 (agente inanimado y paciente animado): en este caso, como en el primer subexperimento, se da un efecto significativo del priming semántico, si bien este efecto es mucho mayor en el priming al paciente que en el priming al agente (Figura 7). Esto parece indicar que en condiciones de priming semántico existe un efecto techo en la descripciones con el paciente (animado) en primer lugar, efecto que puede atribuirse al predominio del principio de agentividad (anteposición del agente) sobre el de «animación» (anteposición de la entidad animada). En este subexperimento se aprecia asimismo una tendencia inhibitoria del priming fonológico, aunque ésta no alcanza significación estadística.

Subexperimento 4 (agente y paciente inanimados): se aprecian efectos significativos de priming léxico (tanto semántico como fonológico) sobre las respuestas de los sujetos, aunque sólo en la condición de priming al paciente (véase Figura 8). Estos efectos fueron excitatorios en ambos casos, es decir, la activación por priming de

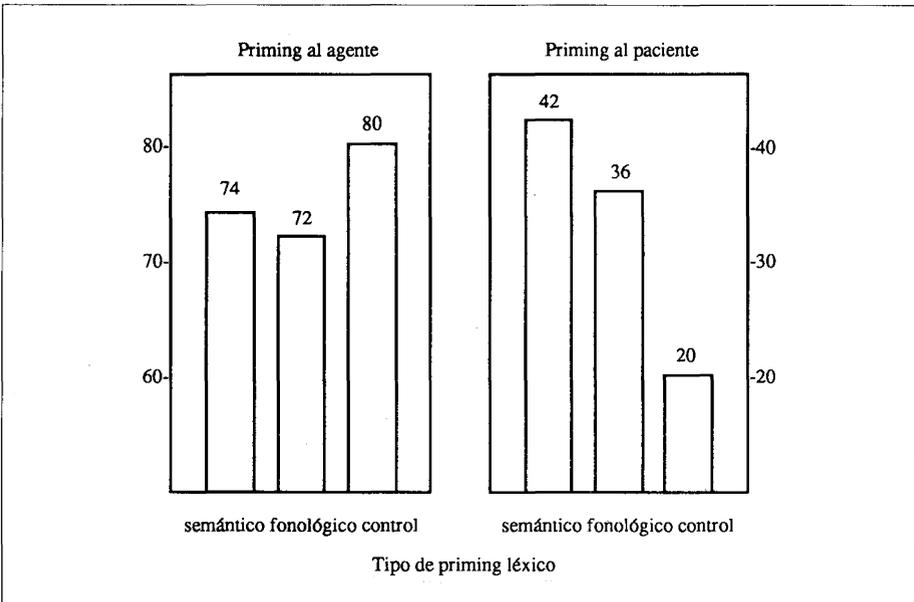


Figura 8. Proporción de respuestas (1) con el agente en posición inicial del enunciado (oraciones activas) [izquierda] y (2) con el paciente como constituyente tematizado al inicio de la oración (pasivas y dislocaciones a la izquierda) [derecha], bajo diversas condiciones de priming: (1) al agente y (2) al paciente, con el agente y el paciente inanimados (subexperimento 4).

representaciones léxicas (semánticas y fonológicas) originaba una tendencia estable a anteponer el paciente de la escena en la descripción.

Hay varios aspectos a destacar en estos resultados. En primer lugar, existe una convergencia entre los datos de los subexperimentos 1 y 4 en lo relativo al efecto excitatorio del priming semántico, aunque éste se ve sujeto a un efecto techo que limita la magnitud de la activación semántica en la condición de priming al agente; por otra parte, se observa en ambos una ausencia de efectos inhibitorios del priming fonológico. Esta convergencia se deja explicar apelando a factores conceptuales. Los subexperimentos 1 y 4 coinciden en que las entidades temáticas (agente y paciente) comparten la misma condición de animación (animados en 1 e inanimados en 4). Ello supone que cuando las entidades incluidas en la representación conceptual de un evento presentan el mismo valor de animación, el proceso de elaboración de la estructura funcional del enunciado que describe dicho evento puede verse influido por procesos de recuperación del significado léxico. Con todo, esta influencia se halla subordinada a las ciertas características conceptuales del mensaje.

En segundo lugar, la presencia de un efecto fonológico excitatorio en el cuarto subexperimento sugiere que las contricciones conceptuales (animación y agentividad) sobre la planificación sintáctica parecen ser menos rígidas cuando las entidades temáticas (y en particular el agente) son inanimadas; en tal caso, la activación de representaciones fonológicas de palabras, ubicadas en un nivel jerárquicamente inferior (posicional), puede influir también en la construcción de la estructura funcional. No obstante, el carácter excepcional de este efecto fonológico no permite aventurar conclusiones muy claras acerca del papel de la activación de formas léxicas en la construcción de la estructura sintáctica.

En suma, la disponibilidad de representaciones léxicas durante la planificación de la oración ha demostrado ejercer efectos relevantes en la elaboración del marco sintáctico. Estos efectos son más claros cuando tales representaciones léxicas contienen información semántica. Asimismo parece haber, como mínimo, un ligero efecto de representaciones fonológicas de palabras sobre el proceso de construcción de la estructura funcional, lo que revela un flujo de retroinformación desde los procesos inferiores de codificación fonológica hacia procesos de formulación de nivel superior. Contrariamente a la evidencia disponible, y por razones aún desconocidas, esta retroinformación fonológica tiene un carácter excitatorio, es decir, provoca la mención inicial del elemento fonológicamente asociado. Con todo, todas estas influencias léxicas se hallan controladas «desde arriba» por factores conceptuales, tales como la animación de las entidades que aparecen en el mensaje que se ha de formular o la condición de agente en dicho mensaje.

A la vista de los resultados obtenidos, se puede avanzar un modelo esquemático del proceso de formulación de enunciados en los siguientes términos: el conceptualizador construye una representación del mensaje a partir de la representación visual del dibujo. Esta representación es suministrada al formulador, que inicia en paralelo dos procesos: la recuperación de elementos léxicos en función de sus propiedades semánticas y la construcción de la estructura funcional del enunciado. En este último se efectúa una asignación de papeles temáticos (agente, paciente, etc.) a funciones sintácticas (sujeto,

objeto directo, etc.). El orden en que tiene lugar este proceso de asignación viene determinado por factores tanto conceptuales (i.e., «agentividad» o «animación») como de formulación, como es el caso de la activación de representaciones de significados léxicos. Así pues, la construcción de la representación funcional del enunciado es un proceso penetrado tanto por influencias conceptuales como por efectos de disponibilidad léxica.

4. Conclusiones

¿Qué interpretación se puede ofrecer de los resultados expuestos en este artículo a la luz de la hipótesis de la modularidad de los procesos de formulación lingüística? La evidencia relativa al análisis de las pautas de error y de corrección parece compatible con la idea de que los procesos de planificación de oraciones *considerados en su conjunto* exhiben por lo menos algunas propiedades típicas de los sistemas modulares. Primero, estos procesos se hallan informativamente encapsulados en la medida en que se muestran insensibles a información no lingüística (i.e. conceptual) mientras efectúan sus operaciones; en segundo lugar, los sistemas centrales (i.e. el conceptualizador) tienen acceso limitado a las representaciones que computan; y tercero, parecen actuar con bastante rapidez. No obstante, la investigación sobre errores y correcciones espontáneas no parece suficiente para someter a prueba otras posibles propiedades modulares del formulador. Por ejemplo, no está claro que los procesos referidos sean obligatorios, ya que si bien se ha observado que las transgresiones a la buena formación lingüística son una excepción, no es menos cierto que éstas ocurren en ocasiones. Por otro lado, la cuestión de la especificidad de dominio no parece fácil de justificar con la sola evidencia de los errores, dado que no existe una correspondencia directa entre clases de error y tipos de información lingüística.

En lo que concierne a la modularidad de los procesos internos al formulador de planes lingüísticos, los datos del experimento priming léxico expuesto en el apartado 3 de este trabajo indican que los procesos de planificación sintáctica del nivel funcional no se hallan preservados del influjo de otros componentes de procesamiento. En particular, se ha comprobado que tales procesos se encuentran informativamente penetrados por propiedades conceptuales del mensaje, por una parte, y por el estado de activación de las propiedades semánticas de los elementos léxicos que forman parte del enunciado, por otra. Así pues, la elaboración de la estructura funcional o subyacente de la oración parece más una zona de «interfaz» entre procesos centrales de conceptualización y procesos modulares de formulación de planes lingüísticos. Según esto, la construcción de este nivel de representación corresponde a la categoría de procesos de Jackendoff (1987) denomina «integradores», en tanto en cuanto cumple el cometido de integrar información conceptual, y también léxica, a tenor de los datos presentados aquí, en una estructura sintáctica unificada. Junto a este tipo de procesos se encuentran los llamados «procesos de traducción», cuyo papel consiste en transformar de forma automática y obligatoria información codificada en un nivel de representación a otro nivel diferente. Una posibilidad no examinada en este trabajo es que el nivel «posicional» de representación sintáctica responda a la caracterización de proceso de traducción, en cuyo caso deberá exhibir las propiedades de encapsulamiento informativo y obligatoriedad de las que carece el nivel funcional de representación.

Las dificultades que plantea la demostración de la modularidad de los procesos sintácticos no sólo se hacen patentes en la producción del lenguaje, sino también en el análisis sintáctico en la comprensión. Ello obedece sin duda a la propia complejidad de los procesos sintácticos. Es sabido que el componente sintáctico comprende un conjunto heterogéneo de fenómenos que entroncan, por un lado, con procesos de conceptualización e inferencia y, por otro, con procesos más superficiales» de codificación y categorización fonológica. Ello hace suponer que el procesamiento sintáctico abarca operaciones de naturaleza muy diversa, que van desde la elaboración de representaciones abstractas de carácter global (i.e., la estructura funcional) hasta la recuperación de estructuras precomputadas bastante locales (i.e., fragmentos de sintagma o elementos de concordancia). Por eso parece excesivamente simplista, como pretende Fodor en su propuesta, considerar el análisis o la planificación sintáctica como procesos modulares en su conjunto.

Un inconveniente que presenta la noción fodoriana de módulo radica, a mi juicio, en que no logra resolver el problema del interfaz entre los módulos y los sistemas centrales. La distinción entre procesadores de integración y procesadores de traducción podría contribuir a arrojar luz sobre este espinoso problema. En lo que atañe a la producción del lenguaje, no parece conveniente considerar los procesos de recuperación de representaciones léxico-semánticas y de construcción de la estructura sintáctica funcional como procesos isotrópicos, pese a que, según todos los indicios, tampoco se trata de procesos modulares. Un modo de reducir la ambigüedad de estos procesos sería incluirlos en una nueva categoría de operaciones que sirviera de puente entre los módulos y los sistemas genuinamente centrales (e.g., fijación de creencias, realización de inferencias no demostrativas). Hasta qué punto ello alteraría la propuesta original de Fodor, acarreando con ello más problemas de los que pudiera resolver, es, sin embargo, un peligro a tener en cuenta.

Bibliografía

- BOCK, J.K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89, 1-47.
- BOCK, J.K. (1986). Meaning, sound and syntax: Lexical priming in sentence production. *Journal of experimental Psychology: Learning, Memory and Cognition*, 12, 575-586.
- BOCK, J.K. (1987a). Co-ordinating words and syntax in speech plans: En A.W. Ellis (Ed.), *Progress in the Psychology of Language*, Vol. 3. (pp. 337-390). London: LEA.
- BOCK, J.K. (1987b). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language*, 26, 119-137.
- BUTTERWORTH, B. (1982). Speech errors: Old data in search of new theories. En A. Cutler (Ed.), *Slips of the Tongue*. (pp. 73-108). Amsterdam: Mouton.
- CUTLER, A. (1982). The reliability of speech error data. En A. Cutler (Ed.), *Slips of the tongue*. Amsterdam: Mouton.
- DELL, G.S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, 27, 124-142.
- DELL, G.S. y REICH, D. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611-629.
- DEWART, M.H. (1979). Role of animate and inanimate nouns in determining sentence voice. *British Journal of Psychology*, 70, 135-141.

- FERREIRA, F. y CLIFTON, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.
- FODOR, J.A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press. *La modularidad de la mente*. Madrid: Morata, 1986.]
- FODOR, J.A. (1987). Modules, frames, fridgeons, sleeping dogs and the music of the spheres. En J.L. Garfield (Ed.), *Modularity in Knowledge representation and Natural Language Processing*. (pp. 25-36). Cambridge, MA: MIT Press.
- FORSTER, K.I. (1976). Accessing the mental lexicon. En R.J. Wales y E.C.T. Walker (Eds.), *New Approaches to Language Mechanisms*. (pp. 257-287). Amsterdam: North Holland. [Versión castellana en F. Valle y otros (Eds.), *Lecturas de Psicolingüística*, Vol. 1. Madrid: Alianza.]
- FRAZIER, L. (1987). Sentence processing: A tutorial review. En M. Coltheart (Ed.), *Attention and Performance*, Vol. XII: *The Psychology of Reading*. (pp. 559-585). Hillsdale, N.J. : LEA.
- GARRETT, M.F. (1975). The analysis of sentence production. En G. Bower (Ed.), *The Psychology of Learning and Motivation*. Vol. 9 (pp. 133-177). New York: Academic Press.
- GARRETT, M.F. (1976). Syntactic processes in sentence production. En R.J. Wales y E.C.T. Walker (Eds.), *New Approaches to Language Mechanisms*. (pp. 231-256). Amsterdam: North Holland.
- HARLEY, T.A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, 8, 191-219.
- HARRIS, M.(1978). Noun animacy and the passive voice: A developmental approach. *Quarterly Journal of Experimental Psychology*, 30, 495-501.
- JACKENDOFF, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.
- KEMPEN, G. y HUIJBERS, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, 14, 185-209.
- ŁASHLEY, K. (1951). The problem of serial order in behavior. En L.A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior*. New York: Wiley.
- LEVELT, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41-104.
- LEVELT, W.J.M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- LEVELT, W.J.M. y MAASSEN, B. (1981). Lexical search and order of mention in sentence production. En W Klein y W.J.M. Levelt (Eds.), *Crossing the Boundaries in Linguistics*. (pp. 221-252). Dordrecht: Reidel.
- O'SEAGHDHA P.G., DELL, G.S. PETERSON, R.R. y JULIANO, C. (en prensa). Modelling form-related priming effects in comprehension and production. En R. Reilly y N.E. Sharkey (Eds.), *Connectionist Approaches to Language Processing* (Vol. 1)).
- OSGOOD, C.E. y BOCK, J.K. (1977). Saliency and sentencin: Some production principles. En S. Rosenberg (Ed.), *Sentence Production: Developments in Research and Theory*. (pp. 89-140). Hillsdale, N.J.: LEA.
- PETERSON, R.R., O'SEAGHDHA, P.G. y DELL, G.S. (1989). Phonological competition in form-related priming. Póster presentado en la 13ª reunión anual de la Psychonomic Society, Atlanta, Georgia.
- SCHRIEFERS, H, MEYER, A.S. y LEVELT, W.J.M. (1990). Exploring the time course of lexical access in language production: Picture-word interference tasks. *Journal of Memory and Language*, 29, 86-102.
- SEIDENBERG, M.S. y TANENHAUS, M.K. (1986). Modularity and lexical access. En I. Gopnik y M. Gopnik (Eds.), *From Models to Modules: Studies in Cognitive Science from the McGill Workshops*. (pp. 135-157). Norwood, N.J.: Ablex.
- SELLS, P. (1985). *Lectures on Contemporary Syntactic Theories*. Stanford, CA: Center for the Study of Language and Information.

- SHATTUCK-HUFNAGEL, S.R. (1979). Speech errors as evidence for a serial order mechanism in sentence production. En W.E. Cooper y E.C.T. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies presented to Merrill Garrett*. pp 295-342. Hillsdale, N.J.: LEA.
- STEMBERGER, J.P. (1985). An interactive activation model of language production. En A.W. Ellis (Ed.), *Progress in the Psychology of Language* Vol. 1. (pp. 143-186). London: LEA. [Versión castellana en F. Valle y otros (Eds.), *Lecturas de Psicolingüística*, Vol. 1, Madrid: Alianza.]
- SWINNEY, D. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 14, 645-660. [Versión castellana en F. Valle y otros (Eds.), *Lecturas de Psicolingüística*, Vol. 1. Madrid: Alianza.]
- VISO, S. del (1990). Errores espontáneos del habla y producción del lenguaje. Tesis doctoral no publicada. U. Complutense de Madrid.
- VISO, S. del, IGOA, J.M. y GARCÍA-ALBEA, J.E. (1987). Corpus de Errores Espontáneos del Castellano. Informe de investigación no publicado. Universidades de Oviedo y Complutense de Madrid.
- VISO, S. del, IGOA, J.M. y GARCÍA-ALBEA, J.E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research*. 20, 161-185.

Taula (UIB) núm. 15, 1992

Confirmación isotrópica-quineana reconsiderada

J. Padilla-Gálvez

Universidad de León

Abstract

The points raised by J. A. Fodor about carrying out scientific confirmation have special interest with respect to reasoning in scientific methodology. His investigation proposed that the connection between confirmation and basic interpretation is theoretical, it is about how the structure of scientific thought is exposed by means of our language. It is clear then, that such confirmation has to be established after a careful examination of the data relevant to our thoughts. On the basis of determined data he develops an interpretation that comes from the Quinean thesis with which the isotropic Quinean platform is specified. Following Fodor, the fixation of scientific thought by undemonstrative means present the aforementioned properties. It is well worth underlining the point such that it emphasizes the modularity hypothesis made on the grounds of semantic proposals.

But what is isotropism and Quineanism? One of the characteristic meanings of isotropism is that the data that one uses in the confirmation of scientific hypothesis can be encountered anywhere in the universe of empirical truths. This isn't a problem of interest for us. However, one cannot characterize Quineanism positively, without it being concerned with the criticism made by Quine of the role played by a priori (dis)confirmation. Following Fodor, to reject a priori experimental observations, is a cut easily made which appears linked to the conventionalist arguments. The aforementioned cut is strongly criticized on the basis of the difficulties of the analytic/synthetic distinction that affect irreparably the distinction between a priori and a posteriori truths.

The problems mentioned have already been broached in the framework of individual thought fixation, it is well analysing it under this light. If the existence of a priori truths don't have the slightest scientific value, then, they are not methodologically justifiable, science cannot precede the raising of hypothesis a priori. Despite this, hypothesis are corrected by our experiences and our empirical data. An example that reinforces this nuance is concerned with singular prediction as well as with inductive generalization. The first question -a key for our data- is concerned with the establishment of degrees of confirmation based on available empirical data and of the singular hypothesis that an individual fact even through not observed has a certain property. This is precisely to develop an inductive method suitable, that is to say a function of appropriate confirmation that one can not provide with mere empirical data.

Introducción

El análisis del génesis, desarrollo y desmoronamiento de nuestras creencias es un tema que afecta a todos los procesos cognitivos en la ciencia por igual. Así pues, el foco de atención acerca de dichas creencias como disciplina de investigación trata de la estructura de nuestro razonamiento ordinario acerca de los requerimientos. Siendo así, supuestamente, su planteamiento ha de considerar las siguientes cuestiones: (i) cómo revelar y clarificar los criterios de razonamiento válidos en relación a dichos problemas; (ii) cómo iluminar la estructura de captación de información; y, (iii) cómo nos proveemos de entendimiento acerca de la estructura lógica del lenguaje en que están inmersas nuestras creencias. Las tres cuestiones que abordamos dirigen su atención al aspecto lógico y representativo de cómo se fijan nuestras creencias en los actos mentales. Consecuentemente, una investigación de nuestras creencias que se propone como «confirmación»¹ es, junto con su interpretación primaria, una teoría acerca de la estructura de nuestras creencias expuesta mediante nuestro lenguaje. Es claro, entonces, que dicha «confirmación» debe constituirse después de un examen cuidadoso de datos pertenecientes a nuestras creencias.

Muchos sistemas de creencias han sido propuestos. Algunos de ellos han sido ampliamente investigados con respecto a sus propiedades lógicas y con respecto a su suficiencia filosófica. Este último tipo de investigación ha generado una serie de críticas respecto los problemas que el propio J. A. Fodor toca en sus análisis, a saber, problemas que tienen que ver con la confirmación, testabilidad o la reducibilidad, por nombrar algunos. Todos estos problemas están influenciados por las denominadas paradojas de la confirmación.

No quiere ser este trabajo un espacio donde se planteen dichas paradojas, si bien quiere contribuir a aclarar algo que a primera vista es plausible - como el propio J. A. Fodor nos lo quiere hacer ver - esto es, que la confirmación científica es «isotrópica» y «Quineana». Esta tesis es una de las premisas fundamentales en determinados argumentos de la *modularidad de la mente* y cuyo objeto es mostrar que las propiedades en el proceso de fijación de las creencias de nuestras estructuras mentales hacen referencia a las propuestas epistemológicas de W. v. O. Quine así como a los principales hitos de la orientación diferencial. Este punto de vista es central en el tratamiento de varios problemas de filosofía de la mente que, a mi juicio, me parecen fundamentales.² Mi

¹ Véase al respecto: J. A. Fodor, *The Modularity of Mind. An Essay on Faculty Psychology*. The MIT Press. Cambridge Mass. 1983, p. 105. Trad. al castellano de J. E. García Albea, Ediciones Morata, Madrid, 1986, p. 148. Una discusión interesante, pero que sale de nuestro ámbito de trabajo, ha sido suscitada en la discusión entre G. Rey y J. A. Fodor. Para ello véase: G. Rey, 'Quinity, isotropy, and Wagnerian rapture'. *The Behavioral and Brain Sciences*, 8:1 (1985), pp. 27-28. La contestación se encuentra en J. A. Fodor, 'Reply module'. *The Behavioral and Brain Sciences*, 8:1 (1985), pp. 33-42, sobre todo p. 38. Esta réplica contiene una amplia bibliografía.

² No entraré en este trabajo en el estudio de la significación de este punto de vista, ni en los pormenores de la formulación de los argumentos desarrollados en la filosofía de la mente ya que desbordaría el problema que queremos tratar.

aportación será demostrar las consecuencias que se deben de extraer de dicha confirmación y el porqué han de ser o bien transformadas puntualmente o rechazadas.

I. Fijación de creencias

Hemos planteado algunas cuestiones de entrada en la introducción y un modo de ordenar dichas preguntas se puede efectuar mediante los «sistemas» cuya función es la de repartir y ordenar los procesos mentales. Así pues dicha ordenación se lleva a cabo mediante determinados parámetros, a saber: el *sistema de entrada* aglutina, por un lado, la difícil tarea de examinar simultáneamente las representaciones que suministran los distintos sistemas, la información almacenada, el empleo del lenguaje así como la conversión que se efectúa sobre la percepción y la necesidad misma del sujeto. Por otro lado, el *sistema central* examina los datos suministrados por el sistema de entrada y lo sobrepesa con la información almacenada en la memoria con el fin de fijar las creencias por medio de *inferencias no demostrativas*.³ En dicho examen se constriñe la elaboración de la «hipótesis más probable». Es sobre dicha hipótesis sobre la que vamos a tratar en este trabajo.

Para empezar, ha de llevarse a cabo una restricción en el ámbito de la fijación de creencias científicas ya que J. A. Fodor está interesado en analizar ámbitos de inferencia no demostrativos.⁴ Al menos, eso se desprende de la lectura cuando afirma: «La fijación de ... creencias científicas por medios no demostrativos presentan dos propiedades ampliamente reconocidas» (p. 148). Dichas propiedades, que caracterizan la confirmación científica, son denominadas «*isotropía*» y «*quineanismo*».⁵ En nuestro trabajo prestaremos menor atención a la «isotropía» y nos centraremos en el «quineanismo» pero antes vamos a decir algunas cosas sobre el carácter de los no-demostrativos.

Prima facie se entiende por fijación, sin adentrarnos en las propuestas de Fodor, cuando se comparte un aparato referencial doxástico de modo que la referencia de los términos es determinable por la indicación que efectuamos.⁶ Las restricciones que

³ Probablemente no es casual el hecho de que tengamos que vérmolas con ámbitos no-demostrativos. Como es bien sabido se ha llevado a cabo una discusión entre la tesis de la irreducibilidad de los ámbitos demostrativos (es decir ahí donde aparece la indexicalidad) y la tesis según la cual, en principio se puede reducir el significado de los ámbitos demostrativos a los no-demostrativos. El marco de dicha discusión me induce a pensar que en la elección arriba expuesta, J. A. Fodor toma partido por la segunda alternativa. Véase: C. McGinn, *The Subjective View. Secondary Qualities and Indexical Thoughts*. Clarendon Press, Oxford, 1983, p. 67.

⁴ Al parecer las caracterizaciones elaboradas por J. A. Fodor son de otro cariz para los ámbitos donde se constituyen nuestras creencias de un modo demostrativo. Téngase en cuenta - sólo como dato auxiliar - que el modo demostrativo es ambigüo y es uno de los factores que originan la inescrutabilidad de la referencia de los términos observacionales.

⁵ La propiedad con el título «quineanismo» hace alusión al lógico y filósofo americano W. v. O. Quine y en concreto a la crítica que efectúa éste a los dogmas del empirismo de las propuestas elaboradas por R. Carnap. Véase: W. v. O. Quine, 'Two Dogmas of Empiricism'. En: *From a Logical Point of View*. Harvard University Press. Cambridge, Mass. 1953, pp. 21-46.

⁶ Esto tiene vigencia a la inversa en tanto que si el universo de un hablante nos es conocido, se podría ir determinando inductivamente los elementos de los que consta su lenguaje y que cumplen las funciones gramaticales de nuestra lengua.

se llevan por este lado, y las propuestas por otro, permite así, analizar la estructura de nuestras creencias científicas.⁷ Una estrategia similar nos pide Fodor que sigamos al reconocer que los procesos que se llevan a cabo en los sistemas centrales son inconscientes. Con el fin de inferir algo sobre los procesos centrales nos exige que analicemos aquellos procesos que son explícitos para la inferencia. Para asegurar esto firmemente, consideremos el siguiente dato:⁸

(1)Un individuo describe el siguiente hecho singular que se observa al lavar la vajilla, a saber: al sacar un vaso del agua caliente que contiene jabón, y colocarlo invertido sobre la superficie del fregadero, se forma al borde de dicho vaso una pompa de jabón que primero se dilata y luego se contrae.⁹

(2)Armados de dicho dato postulamos la creencia de dicho individuo del siguiente modo:

X cree que ... si, y sólo si X confirma '...',
en donde '...' ha de entenderse como una hipótesis científica que, según J. A. Fodor puede ser tomada de cualquier rea del universo que denomina «*verdades empíricas*»¹⁰

(3)Conclusión: La fijación de las creencias del individuo X, relevante de cara a la confirmación, no pueden ser *a priori*.¹¹

La pregunta pertinente de los datos expuestos, es saber lo que tiene que ver con la filosofía de la mente. La contestación a dicha cuestión viene de la mano de la siguiente explicación: al encontrarnos ante procesos inconscientes, como ya hemos dicho, el único recurso que nos queda para analizar la estructura que los apoya es, plantear datos análogos que nos permitan analizar procesos explícitos. El mismo Fodor afirma al

⁷ Desde nuestro punto de vista, las propuestas así construidas son típicamente débiles. Lo que se necesita es, en cambio, sistemas amplios que puedan esclarecer el lenguaje de nuestras creencias, comenzando por aquellas que el autor restringe y que tienen que ver con el modo demostrativo, siguiendo los aspectos quasi-demostrativos que tienen que ver con la referencia al 'yo' y desembocando en aquellas creencias científicas cuyo ámbito no es demostrativo. No me voy a parar a tratar aspectos de la demostración, quasi-demostración, para ello pueden ver los análisis llevados a cabo por H.-N. Castañeda, 'Indicators and Quasi-Indicators', *American Philosophical Quarterly*, 4 (1967), pp. 85-100; H.-N. Castañeda, *Sprache und Erfahrung. Texte zu einer neuen Ontologie*. Suhrkamp. Frankfurt a.M., 1982.

⁸ La elección del dato, que seguidamente exponemos, sigue al pie de la letra los casos que propone J. A. Fodor, a saber:

(i) La inferencia no demostrativa en la que se desea determinar el grado de verdad de un mecanismo causal;

(ii) Proponer una solución de problemas mediante la fijación individual de creencias.

Véase al respecto: J. A. Fodor, *The Modularity of Mind. An Essay on Faculty Psychology*. The MIT Press. Cambridge Mass. 1983, p. 105. Trad. al castellano de J. E. García Albea, Ediciones Morata, Madrid, 1986, p. 149.

⁹ Este dato enunciado en la premisa está sacado de C. G. Hempel, *Aspects of Scientific Explanation. En: Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. Collier Mcmillan Publ., London, 1965, pp. 331-496, pp. 335s. que al mismo tiempo lo asume de J. Dewey, *How We Think*. D. C. Heath & Co., Boston, 1910, Cap. VI.

¹⁰ Véase: J. A. Fodor, *The Modularity of Mind. An Essay on Faculty Psychology*. The MIT Press. Cambridge Mass. 1983, p. 105. Trad. al castellano de J. E. García Albea, Ediciones Morata, Madrid, 1986, p. 178.

¹¹ La conclusión expuesta es una consecuencia de la propuesta quineana.

respecto que para inferirse algo sobre el funcionamiento de los procesos mentales, a partir de los procesos explícitos de inferencia no demostrativa nos hemos de basar en lo que sabemos acerca de la inferencia empírica en la ciencia.

Los datos descritos arriba van a ser planteados en base a lo que se podría denominar la solución de un problema, en donde la inferencia no demostrativa juega un papel crucial, de modo que se puede plantear: (i) en base al grado de verdad de un mecanismo causal; o, (ii) en base a la fijación individual de creencias. Consecuentemente, deberás haber una noción de «semántica para los datos». Los datos que aparecen son analizados en términos de una estructura en la que aparece la actitud proposicional. Así pues, la estructura de las creencias expuesta en (2) presupone una descripción de la estructura que puede ser abstraída del dato expuesto en (1). De este modo podemos tomar la estructura semántica, generada por el dato (1) y expuesto mediante la creencia (2), como una estructura asignada mediante las expresiones que la contienen. La semntica del dato expuesto en (1) será pues, la semántica «óptima de la forma que genere dicho dato. En (3) llegamos a una conclusión global sobre el proceso llevado a cabo en (1) y (2).

Según (3) la fijación de creencias no puede ser *a priori* ya que con ello bloquearíamos irremediamente aquellas creencias que pueden ser corregidas o alteradas por la experiencia. La confirmación es pues - según Fodor - isotrópica en tanto que los datos de la hipótesis son tomados del área del universo *deverdades empíricas*. Sobre dicha conclusión no nos vamos a centrar másya que sería un tema excepcional para otro trabajo.

En conclusión, el ejemplo (1) nos proporciona la batería de pruebas de adecuación con respecto el retrato de la estructura mental que caracteriza el sistema central de Fodor y nos permite saber si las fijaciones de creencias, mediante las propiedades isotrópicas, son apropiadas o, por el contrario, han de hacerse determinadas observaciones y reservaciones al respecto. Los datos cruciales de nuestra investigación se pueden resumir de la siguiente manera: Los sistemas de creencias vienen caracterizados, junto con su interpretación básica, mediante una estructura mental. Siendo así, a menos de que encontremos una razón de peso para proceder de otra forma, debemos considerar las propiedades, en el proceso de fijación de creencias, como claves para distinciones importantes en la estructura mental.¹² Empecemos, pues, teniendo cuidado en emplear la mejor metodología, con un examen de la estructura mental del ejemplo anterior.

II. Dualidad estructural básica

Consideremos el dato arriba descrito, puesto que es el más complejo de todos y puede, por lo tanto, revelar aspectos más amplios de la estructura en la formación de

¹² Dicha intuición está de acuerdo con el dato general que se postula en la semántica filosófica al teoretizar sobre actitudes psicológicas ya que éstas son actitudes que instancian propiedades cuyo contenido son proposiciones. Véase al respecto: J. Padilla Gálvez, (¿Pueden aplicarse los argumentos informales contra el punto de vista formalista? En: *Structures in Mathematical Theories* (Ed. A. Diaz, et. al.) Universidad del País Vasco, San Sebastián, 1990, pp. 461-467.

las creencias. Sobre el dato (1) se puede sostener que la descripción de una observación singular puede ser considerada relevante de cara a la confirmación de una teoría de la dilatación de los cuerpos. Los miembros de los que consta el dato (1) tienen una estructura peculiar. En (2) se presenta latente la ambigüedad de los enunciados que informan sobre determinadas actitudes. Ya que a nosotros no nos interesa dicha ambigüedad, proponemos una desambiguación, en tanto que vinculamos directamente la creencia, el estado mental de creencia, con el asentimiento de un enunciado - como es por caso '...' - , o, en la terminología de J. A. Fodor, la «confirmación» de dicho enunciado. El propósito inmediato de dicha vinculación no es otro que el de mejorar nuestro conocimiento acerca de las propiedades funcionales de lo que denomina «sistema central», a saber, la propiedad isotrópica-quineana que se da en la fijación de creencias.

La vinculación mencionada parte de la observación de que hay algo importante que es común a nuestras estructuras de la creencia que aceptan, en un determinado momento y lugar, la descripción expresada en (1). Y no es muy difícil averiguar lo que es: la tendencia de un vaso sacado del agua caliente y que contenía jabón a desarrollar una pompa de jabón. Por supuesto, no se trata de una generalización universal determinista sino de una propensión que están sujeta a una *ley probabilística*.¹³

El argumento de J. A. Fodor puede ser reconstruido de la siguiente manera: los procesos centrales responsables de la fijación de creencias son típicamente procesos de inferencia racional no demostrativa que contienen las propiedades isotrópicas quineanas. Uno de los significados que caracteriza la isotropía es que los datos para la confirmación de una hipótesis científica pueden ser encontrados en cualquier reza del universo de verdades empíricas.¹⁴

III. Aprioricidad y confirmación científica

El Quineanismo tiene que ver con una crítica al papel de la (des)confirmación *a priori*. Según Fodor descartar *a priori* observaciones experimentales es una coartada fácil que aparece por lo general vinculada al convencionalismo. Como uno de los motivos centrales es la fijación individual de creencias, bien vale la pena analizar dicho aspecto desde su vertiente metodológica.

¹³ La generalización universal determinista vendría expresada mediante el siguiente enunciado:

(1') Todos los vasos lavados en agua caliente que contengan jabón en la circunstancia S y bajo los presupuestos p desarrollan una pompa de jabón. Sin embargo, damos por aceptada una restricción y entendemos (1') como un enunciado sujeto a una ley probabilística por lo que el caso (1') no es relevante para nuestro análisis.

¹⁴ Dicho de paso, dicha caracterización se podría llamar también holista, en tanto que la totalidad de los datos empíricos parece jugar un papel preponderante en el análisis de la aceptación de una hipótesis científica determinada. De aquí resulta que una hipótesis particular pueda estar en una relación determinada con el resto de las hipótesis aún sin que éstas pertenezcan al mismo ámbito de investigación. Véase al respecto el trabajo de C. Glymour, Fodor's holism. En: *The Behavioral and Brain Sciences*, 8:1 (1985), pp. 15-16 y la respuesta y puntualizaciones de J. A. Fodor, Reply module. En: *The Behavioral and Brain Sciences*, 8:1 (1985), pp. 33-39, sobre todo p. 36 s.

La exégesis del ejemplo expuesto en (1)- (3) y la lectura que hemos efectuado de la propuesta acerca de la fijación de creencias, nos ha llevado a unas conclusiones determinadas. Una propuesta que quiera hacer honor a su nombre, que posea una cierta riqueza, ha de superar, pues, cualquier contraargumento. Ahora bien, si lo que queremos es indagar en la comprensión de lo que acontece en la mente de un hablante, es importante saber lo que ocurre en la lógica propuesta por dicho hablante cuando se refiere a sus (propios) procesos mentales. Dicha lógica, como todos sabemos, trata *grosso modo* sobre las categorías nocionales que pueden, en principio, recibir la expresión lingüística que usamos. Con ello, centramos a la vez el problema que envuelve el «quineanismo».

El origen del problema que suscita el quineanismo radica, en mi opinión, en las dudas acerca de la existencia de un concepto presistemático de *analiticidad* lo bastante verosímil que pueda hacer de *explicandum*. Ya que la analiticidad es considerada por los opositores de W. v. O. Quine como una categoría clave, bien vale la pena repasar su caracterización de modo genérico. Según éstos, el valor de verdad de un enunciado depende, tanto del sentido de las expresiones usadas, como de los hechos a los que se refiere. Si el valor de verdad depende de ambos factores, denominamos dicho enunciado un *enunciado sintético*. Si depende sólo y exclusivamente del sentido de las expresiones usadas entonces hablamos de un *enunciado analítico* (si es verdadero, y contradictorio si es falso). De este modo, si se aceptan ambos criterios, las verdades sintéticas son enunciados sobre el mundo mientras que las verdades analíticas son productos subsidiarios del lenguaje. De este modo, podemos caracterizar la *tesis analítica* (abreviadamente (TA)), contra la que argumenta Quine, del siguiente modo:

(TA) Un enunciado es analíticamente verdadero si, y sólo si su verdad puede ser establecida con la sólo ayuda del análisis del sentido de las expresiones.

Está claro, que (TA) no recurre a la experiencia por lo que la investigación de los hechos mentales pierde así una riqueza de sumo valor.

Uno de los intentos más radicales se debe a la sugerencia de R. Carnap que propuso ofrecer conceptos semánticos en calidad de *explicata* de conceptos pragmáticos.¹⁵ Con dicha propuesta varía sensiblemente la polémica en torno a la dicotomía analítico-sintético, llevada a cabo entre R. Carnap y W. v. O. Quine. Este último pretende demostrar que no existe diferencia entre la dicotomía analítico-sintético, ya que la dificultad radica en que falta un criterio o procedimiento capaz de indicar puntualmente si un enunciado determinado de un lenguaje es analítico o sintético.

Como es sabido, W. v. O. Quine diferencia dos nociones de analiticidad. La primera noción atiende al siguiente argumento: un enunciado es analítico si es obtenido

¹⁵ Véase al respecto: R. Carnap, *Meaning and Synonymy in Natural Language*. En: *Philosophical Foundations of Physics*. Basic Books, Inc., New York, 1966, Cap. 27. Una dura crítica, desde la posición de Fodor, es llevada a cabo al afirmar: «Conventionalisms as our account of theoretical constructs; sharp distinction between the observational and inferential language of theories; uncritical reliance upon the analytic-synthetic distinction» J. A. Fodor, *Psychological Explanation: An Introduction in the philosophy of Psychology*. MIT. Randon House. New York, 1968, p. xiv. La relación entre *explicata* y *explicandum* ha sido tratada en: J. Padilla-Gálvez, *La verdad y descripción de estados*. En: *Meeting on Logic and Philosophy of science. Rudolf Carnap & Hans Reichenbach in Memoriam*, Madrid, 1991, pp. 415-422.

de una verdad lógica en la que se sustituyen sinónimos por sinónimos. Según el segundo argumento - y eje de este apartado - un enunciado es analíticamente verdadero si puede ser confirmado. Al solaparse esta última argumentación con el concepto tradicional de aprioricidad, H. Putnam llevó a cabo una lectura propia de la afirmación quineana y dedujo que no existe distinción entre verdades analíticas y sintéticas, lo que equivale a decir *summa sumorum* que tampoco hay una diferencia entre verdades *a priori* y verdades *a posteriori*.¹⁶ Este resultado intuitivo está de acuerdo con la propuesta vigente en la filosofía de la ciencia de que no hay cuestiones irrefutables por principio.¹⁷

A pesar del consenso al que se ha llegado, optaré por la siguiente propuesta: si bien es correcto lo que se ha afirmado, los argumentos arriba esbozados han de ser matizados. Es decir si bien, como se ha probado, no hay razón para aceptar verdades *a priori*, ya que no se pueden reconocer la existencia de enunciados absolutamente irrevisables, sin embargo esto no supone que no sea conveniente recurrir en la ciencia a *hipótesis a priori*. Dichas hipótesis *a priori* no son, pues, irrefutables, sino que sólo actúan estructuralmente de tal modo. Se podría decir que actúan con marco con respecto a los datos que se elaboran en el marco central. De este modo cabe incluir la pregunta por la caracterización de las hipótesis *a priori* en el sistema central. Si caracterizamos en términos sustantivos una hipótesis *a priori* entonces nos encontramos ante una función que elabora los datos del sistema central. Dicha caracterización puede definirse descriptivamente cuando se indica que para examinar los datos suministrados por el sistema de entrada y elaborarlos con la información almacenada en la memoria nos hacemos eco de un marco informativo que por lo general no se pone en cuestión. Dicho marco actúa como hipótesis *a priori* aunque bien pudiera ponerse en cuestión. La hipótesis *a priori* contribuye a fijar las creencias en el sistema central. Dicho marco actúa, por así decirlo, de modo inconsciente con respecto los datos de entrada. El término «hipótesis *a priori*» es, pues, usado con una acepción especial, a saber, como punto de referencia y marco que elabora los datos de entrada.¹⁸

Indudablemente, estoy proponiendo que la aceptación de verdades *a priori* y la aceptación de *hipótesis a priori* son siempre distintas, de modo que la premisa de Fodor, entendida literalmente, es válida, pero no le sirve en su argumento en contra del papel que juegan las hipótesis *a priori* en la fijación de las creencias relevantes de cara a la confirmación. Me da la impresión que Fodor sabe esto ya que en su ensayo defiende premisas en base a una batería de datos que tienen que ver con el rechazo a la aceptación de verdades *a priori*.

De este modo me propongo defender mi tesis principal, concediéndole a Fodor su tesis local. Me explico: Fodor tiene razón al afirmar que la aceptación de verdades

¹⁶ Véase H. Putnam, 'Two Dogmas' Revisited. En: H. Putnam, *Realism and Reason. Philosophical Papers*. Vol. 3. Cambridge University Press. Cambridge, 1983, pp. 87-97.

¹⁷ Con dicha intuición tiene vigencia el principio de invalidez de la inducción de D. Hume. De este modo no se aceptan enunciados absolutamente irrevisables.

¹⁸ Véase de un modo más detallado algunos argumentos similares al que esbozamos aquí en: A. Rivadulla, 'Analiticidad, aprioricidad e inferencia inductiva'. En: J.J. Acero y T. Calvo (Eds.) *Symposium Quine*. Universidad de Granada. Granada, 1987, pp. 178- 205.

a priori puede ser refutada por el quineanismo que postula, pero esta premisa no favorece en nada a los argumentos en contra de la aceptación de hipótesis *a priori*. Las refutaciones desarrolladas por W.r.O. Quine sólo afectan a la irreversibilidad absoluta de los enunciados y no a las hipótesis. Por lo tanto, sigo convencido de que la imposibilidad de elementos apriorísticos que posteriormente puedan ser refutados continúa indemostrada por el quineanismo.

IV. Fijación de creencias, quineanismo y predicción singular

Armados de dichos argumentos y en base a los datos elaborados, podemos exponer la tesis de este trabajo, a saber: mi tesis central es que para la fijación de creencias, como es en el caso de una predicción singular, el quineanismo que postula J. A. Fodor ha de puntualizarse ya que utiliza en sus términos una sutileza y una ambigüedad, que, a la vista de los argumentos que hemos esbozado, ha de superarse mediante puntualizaciones más nítidas. Por una parte, se puede dudar - como hemos esbozado arriba - el que el sentido de la propiedad denominada «quineanismo» sirva realmente para su propósito, es decir, agote el campo más vago, pero quizás más intuitivo de analiticidad, entendida como crítica a la insuficiencia de verdades *a priori*. Por otra parte, se puede encontrar insatisfactorio el modo en que se llegó a este sentido más amplio y, por tanto, querer reemplazarlo por un sentido enteramente diferente. Este último punto es el que hemos postulado en nuestro trabajo. Consecuentemente, la crítica de Fodor es relevante de cara a la refutación de verdades *a priori*, pero esto no supone que desarrolle una conjetura adecuada contra las hipótesis *a priori* que podemos usar en la predicción singular en la fijación de creencias ya que los procesos de fijación no son excluibles de la conceptualización.¹⁹

A la vista de los datos elaborados, y siguiendo las conclusiones a las que hemos llegado mediante, el caso de la predicción singular, es necesario llevar a cabo unos ajustes mínimos. Primeramente un colorario importante de nuestro análisis afirma que los grados de confirmación dependen en gran medida del lenguaje.

El caso más indicado para ello es el que tiene que ver con la *predicción singular*. Los casos de predicción singular como los manifestados en la creencia expuesta en (2) conciernen, como ya hemos dicho, el establecimiento del grado de confirmación, en base a la evidencia empírica que disponemos, de la hipótesis singular que predice que un hecho individual (que puede que aún no haya sido observado) posea una propiedad

¹⁹ Vuelvo a resumir el planteamiento general ya que en la discusión posterior a la presentación de esta ponencia se confundieron, desde mi punto de vista, niveles argumentativos importantes. Primero, no acepto verdades *a priori*, porque no creo que halla cuestiones irrefutables por principio. Segundo, han de diferenciarse, verdades *a priori* y hipótesis *a priori*. Tercero, las hipótesis *a priori* actúan como marco con respecto a los datos que se elaboran en el sistema central, examinando y evaluando los datos. La refutación de las hipótesis *a priori* es más complicada que las conjeturas que se puedan hacer de la descripción de los datos, porque no permiten conjeturas ingenuas. Esto no significa que no sean refutables. Las hipótesis *a priori* son punto de referencia y marco de evaluación. La dificultad de la refutación de las hipótesis *a priori* tienen que ver con el contenido, y de ahí su dificultad.

determinada. Dicha propiedad puede ser descrita por la formación de la pompa de jabón en unas determinadas condiciones. Para ello, nos valemos de un método inductivo adecuado, por lo que se desarrolla una función de confirmación apropiada. La función de confirmación de (2) depende:

- (i) del *factor empírico*,
- (ii) la *frecuencia relativa* observada en la muestra, y
- (iii) un *factor lógico* que asume valores de cero a infinito.

Para aquellos casos en que el factor lógico descansa en los valores cero o infinito, los métodos inductivos correspondientes descansan en la suposición *a priori* de que el grado de uniformidad del dominio considerado es respectivamente máximo y mínimo. Cuando el factor lógico es un número finito positivo, el valor otorgado corresponde igualmente a una suposición *a priori* de un cierto grado de uniformidad.²⁰ Si dicha conceptualización es aceptada, y hasta la fecha no parece ser que halla algún argumento que invalide nuestra pretensión, se demuestra que la aceptación o rechazo de un enunciado analítico tiene más bien que ver con el sistema conceptual que escogemos. La dicotomía *a priori* - *a posteriori* vuelve así a tener sentido aunque este sea de un modo muy restrictivo.

Como conclusión inmediata se nos presenta aquí el hecho de que la reconstrucción lógica del procedimiento inductivo por el que percibimos la experiencia, muestra que las hipótesis *a priori* juegan un papel destacado si bien son susceptibles a ser corregidas posteriormente por la experiencia. Vale pues la pena saber si dicha conclusión repercute en los mecanismos de la fijación de nuestras creencias.²¹

Conclusión

Estoy, pues, de acuerdo con Fodor que aceptar enunciados analíticos y estos a la vez ser aceptados como enunciados verdaderos supone un alto grado de riesgo para la filosofía de la mente. Como hemos podido comprobar, encierra insuficiencias irreducibles y altamente peligrosas. Sin embargo para determinados ámbitos, como pueden ser la confirmación, una modalidad especial de analiticidad, no es sólo posible sino aconsejable.

²⁰ W. K. Essler ha propuesto que no se asuma dicho valor como verdadero *a priori* sino como una hipótesis sintética acerca del mundo, pudiendo ser revisable por el hecho de que la experiencia puede corregir nuestra suposición del grado de uniformidad del dominio a medida que aumentan las observaciones. Véase: W. K. Essler, *Induktive Logik. Grundlagen und Voraussetzung*. Karl Alber. Freiburg, 1970, pp. 185 ss.

²¹ El problema ahí enunciado es tema para otro estudio que tendría que resolver, ante todo, el movimiento inverso.

Taula (UIB) núm. 15, 1992
Recensions

ANTON PACHECO, José Antonio, *Un libro sobre Swedenborg*, Publicaciones de la Universidad de Sevilla, Sevilla 1991 (95 pp.)

Swedenborg (1688-1772) no es ciertamente una de las primeras figuras de la historia de la filosofía, y con razón. Con todo resulta ser un intelectual suficientemente complejo como para que permita un estudio sobre su filosofía. Ciertamente su producción literaria es enorme y abarca prácticamente todos los saberes de su tiempo. El autor se propone una “desmitologización del visionario escandinavo” (p. 6) (“nunca fundó una iglesia positiva”, “nunca perteneció a ninguna logia masónica”, “criticó explícitamente la fe en los milagros y todo tipo de adivinación” p. 13), presentando un perfil unitario y armonioso de Swedenborg, que une la formación ilustrada de científico y el impulso renovador cristiano; rechaza que exista “un primer Swedenborg científico y un segundo visionario” (p. 12); sin negar la crisis religiosa de 1741, el autor comprende su trayectoria intelectual como una “maduración progresiva de una creciente inquietud filosófico-religiosa” (p. 12), que va desde una temprana vocación humanista, pasando por un intenso interés y dedicación a la ciencia, inscribiéndose dentro del mecanicismo cartesiano, y que culmina con la consumación de la crisis existencial, dedicándose posteriormente a su obra teológica, tan inmensa como la científica anterior.

Dos capítulos (el II y el III) parecen ser los que más se adentran dentro de la filosofía de Swedenborg, exponen las líneas básicas de su metafísica. El segundo: “Unidad y determinación en Swedenborg” da cuenta de su concepción de la realidad, que en contra del dualismo cartesiano y en un intento explícito por superarlo, se fundamenta en la afirmación de la unidad

de la realidad. Esta unidad, sin embargo, no se resuelve en un unitarismo abstracto y absolutizador que borre las diferencias, y de ahí la otra característica de “determinación”, que tiene el sentido de concreción en particulares diferentes, algo así como los modos de la sustancia una de Spinoza, según el principio de “omnis determinatio est negatio”. De este modo quiere dar cuenta de la unidad y variedad de la realidad. Pero con una diferencia capital respecto a Spinoza, puesto que la determinación es entendida fundamentalmente como personalización (p. 36); de ahí “la prioridad que el sujeto adquiere para la visión totalizadora de Swedenborg” (p. 35); a “El papel del sujeto en Swedenborg” está dedicado todo el III capítulo.

Los demás capítulos giran fundamentalmente en torno a dos temas: la hermenéutica y la recepción de Swedenborg en España. El componente “visionario” es considerado como una hermenéutica que enlaza con el gnosticismo en el intento de unir filosofía religiosa y hermenéutica espiritual; pero frente a otras interpretaciones el autor resalta el carácter de exégesis de este gnosticismo, por su referencia explícita al texto de la biblia (p. 49ss.). Este elemento -aunque con diferencias peculiares- le situa dentro de una amplia tradición que va desde el gnosticismo antiguo, el misticismo, el neoplatonismo y la teosofía. Respecto al influjo de Swedenborg en la literatura hispánica, el autor estudia las relaciones de Unamuno con Swedenborg (pp. 73ss.), y la presencia de Swedenborg en Eugenio D’Ors (pp. 81ss.), además de una comparación de la religiosidad de Jorge Luis Borges con Swedenborg (pp. 89ss.).

Gabriel Amengual

CORDUA, Carla, *El mundo ético. Ensayos sobre la esfera del hombre en la filosofía de Hegel*. Ed. Anthropos, Barcelona 1989 (223 pp.)

Se trata de un conjunto de estudios sobre la Filosofía del Derecho de Hegel, articulados en torno a algunos temas básicos, fundamentalmente tres: la comprensión de 'mundo ético', la voluntad, la modernidad. Con ser una recopilación de artículos, están muy bien trabados y unidos temáticamente. En definitiva es un estudio sobre la teoría del espíritu objetivo, una "exploración del orden y de la variedad interna del mundo 'creado' por la voluntad racional para sí misma" (p. 13).

El libro se abre con un estudio sobre "La realidad humana entendida como 'mundo ético en la historia'", donde se explica la concepción hegeliana del mundo ético, que la autora -con toda razón- entiende como el propiamente humano, como aquel que une individualidad y comunitariedad, subjetividad y objetividad, tal como puede y debe concebirse en la modernidad. (Respecto a la exclusión del espíritu subjetivo de la cualificación propiamente humana, en cierto modo es comprensible que los ámbitos de la antropología y de la fenomenología sean considerados que tratan de la humanidad aún como una especie animal [p.17], pero en cambio difícilmente podría afirmarse esto mismo de la psicología; en cambio sí podría decirse que todas estas esferas tratan más bien de la génesis de lo humano). A partir de este enfoque se traza el desarrollo de la Filosofía del Derecho. El segundo capítulo "El conocimiento del mundo ético y su rango en la filosofía de Hegel" sigue la línea de presentación general de la filosofía práctica hegeliana, destacando que la mayor aportación de Hegel -y quizás también uno de los grandes escollos para su comprensión- consiste en el intento de

unir lo lógico y lo empírico, situándose la autora en la auténtica perspectiva especulativa que supera las interpretaciones basadas en la disyuntiva entre una comprensión lógico-absolutista o bien antropológica.

Siguen tres capítulos unidos por un tema básico en la filosofía práctica hegeliana (y en cualquier otra): la voluntad racional. Primero se analiza el concepto de voluntad a base de los párrafos de la "Introducción" de *Fundamentos de la Filosofía del Derecho*, fijándose en el instrumental conceptual (universalidad y negatividad) y con qué trasfondo o marco conceptual: mediante la idea y mediante un recurso directo a la conciencia de sí (p. 49), o la génesis del yo y de la conciencia (p. 52). Situada en el contexto sistemático y lógico, se pasa a presentar el desarrollo constitutivo de la voluntad ("breve fenomenología de la voluntad" p. 69), desarrollo que se lleva a cabo mediante el recorrido de cinco figuras. Quizás podría discutirse no sólo si puede hablarse de cinco figuras (la misma autora reconoce que la reflexividad se encuentra "tanto en su origen en el yo como también en las tres figuras que hemos presentado ya" p. 88), sino también si se puede hablar de figuras; la autora cuida de argumentarlo (p. 75, 93-94) tomando partido por una de las corrientes de interpretación (la de Ilting, aunque sin nombrarlo). El capítulo V "La libertad en el mundo ético" justifica la importancia dada al estudio de la voluntad, siendo ésta el principio para entender la libertad y su realización en el mundo ético.

La voluntad como principio es una característica de la modernidad. A la modernidad se le dedica el capítulo más

largo “El mundo moderno y la filosofía”, y todo un ejemplo de matización, distinguiendo muy bien los diversos aspectos o ámbitos de la modernidad (religiosa, política y filosófica) y su conexión, pero fijándose en la propiamente filosófica. Concluye el libro un capítulo sobre “Las condiciones político-culturales de los derechos de los ciudadanos”, cuya tesis es que “derechos individuales, en un sentido imprecisamente comparable al del liberalismo de la segunda mitad del siglo pasado y primera del nuestro, hay en la teoría de Hegel sólo dentro del Estado” (p. 176), aunque por otra parte la pertenencia del individuo al Estado (que es un deber moral) o sea “la existencia política del ciudadano está fundada en su carácter de sujeto libre, capaz de autodeterminarse” (ibid.).

En su conjunto creo que podrían destacarse dos características de este estudio. La primera y en referencia a su contenido creo que el primer capítulo es significativo (la interpretación del mundo ético como el mundo humano), designa todo un enfoque que, después del predominio de lo político y a diferencia de él, se podría calificar de antropológico. Ciertamente no antropológico en sentido de opuesto a lógico o absoluto, tal como por ejemplo siguieron las interpretaciones desde Kojève hasta Garaudy e incluyendo

al mismo Bloch, cuyo antropologismo significaba un reduccionismo de las dimensiones absolutas y lógicas equiparando espíritu (también absoluto) con espíritu humano, sino por resaltar otras dimensiones (la voluntad, sobre todo, pero también el contexto histórico-cultural, la modernidad), y que por tanto no conlleva ningún reduccionismo, sino todo lo contrario, una ampliación del punto de mira y del ámbito de consideración. La segunda característica, más formal, viene dada por un cierto aire de leer directamente a Hegel, presentando un Hegel pensado por sí mismo, desde sus propios textos (apenas si se cita alguna obra sobre Hegel a lo largo de toda la obra, por más que la obra incluya una buena bibliografía, mientras que continuamente se cita a Hegel mismo), y cierta pretensión de “desfacer entuertos” hechos por la literatura secundaria o las interpretaciones corrientes; una cierta acribia, en suma, y rigor. Por todo ello creo que es una obra que enriquece enormemente el panorama de la literatura castellana sobre Hegel, constituyendo uno de los claros signos de cierta revitalización de la producción filosófica propia a la altura de la mejor europea o americana.

Gabriel Amengual

HEGEL, G.W.F., *Vorlesungen über die Geschichte der Philosophie*. Teil 4: Philosophie des Mittelalters und der neueren Zeit. Herausgegeben von Pierre Garniron und Walter Jaeschke. Felix Meiner Verlag, Hamburg 1986 (G.W.F. HEGEL, Vorlesungen. Ausgewählte Nachschriften und Manuskripte, Bd. 9) (437 pp.)

HEGEL, G.W.F., *Vorlesungen über die Geschichte der Philosophie*. Teil 2: Griechische Philosophie I. Thales bis Kyniker. Hrsg. v. P. Garniron u. W. Jaeschke. Felix Meiner Verlag, Hamburg 1989 (G.W.F. HEGEL, Vorlesungen. Ausgewählte Nachschriften und Manuskripte, Bd. 7) (442 pp.)

Dentro del esfuerzo de revisión textual de las obras de Hegel que lleva a cabo el Hegel-Archiv le ha tocado el turno también a las *Lecciones* berlinesas; unas obras de segunda categoría, por cuanto no son propiamente de Hegel, ni pensadas para su publicación, sino simples apuntes que alumnos tomaron de los cursos de Hegel; por ello se lleva a cabo no propiamente una edición crítica, sino una “Studienausgabe” más conforme con la base textual de que hoy se dispone. Con ser de segunda categoría, estos escritos han sido de gran influencia, por ser a veces los únicos testimonios de Hegel sobre algunas cuestiones (Estética, Filosofía de la Historia, Filosofía de la Religión y evidentemente Historia de la Filosofía) y por haber contribuido poderosamente a la creación de tópicos sobre la filosofía de Hegel, por tratarse de unos textos más ligeros y fluídos y por tanto de más fácil acceso que las obras que Hegel mismo escribió y publicó. Justamente entre otros efectos que tiene esta nueva edición de las *Lecciones* es el de contraponer a la imagen convencional de un Hegel filósofo sistemático del estado prusiano que todo lo tiene ya encasillado dentro de su férrea lógica, un Hegel más bien en búsqueda del concepto y del hilo sistemático, en intentos sucesivos, como si en cada curso estuviera experimentando en búsqueda de la exposición más adecuada del tema.

Ciertamente las *Lecciones sobre la Historia de la Filosofía* no son el lugar más idóneo donde pueda observarse esta búsqueda del esquema sistemático adecuado, dado que éste viene dado por el orden cronológico. Las diferencias entre los diversos cursos son en este caso menores, afectando solamente al menor o mayor desarrollo de alguna cuestión o al acento que en ellas se pone. Por eso los editores se han decidido por editar un solo curso: las lecciones del semestre de invierno de 1825-26, que son además las mejores por su contenido, las mejor documentadas, y por coincidir en sus líneas generales con las de otros semestres. Seguramente por estas mismas razones ya el primer editor de dichas lecciones de historia de la filosofía, Michelet, las tomó como base de su edición, aunque ampliándolas con añadidos de otros cursos, añadidos a los que esta nueva edición ha renunciado en favor de la autenticidad del texto, liberado de manipulaciones. Este hecho representa una gran diferencia respecto a las *Lecciones sobre filosofía de la religión* en las que se editaron las lecciones de todos los cursos por separado, dado que cada una representa una unidad por ella misma, con esquema y desarrollo propios, cosa que no se da en estas de historia de la filosofía.

Para esta edición se ha tomado como base el manuscrito del capitán von

HEGEL, G.W.F., *Vorlesungen über die Geschichte der Philosophie*. Teil 4: Philosophie des Mittelalters und der neueren Zeit. Herausgegeben von Pierre Garniron und Walter Jaeschke. Felix Meiner Verlag, Hamburg 1986 (G.W.F. HEGEL, Vorlesungen. Ausgewählte Nachschriften und Manuskripte, Bd. 9) (437 pp.)

HEGEL, G.W.F., *Vorlesungen über die Geschichte der Philosophie*. Teil 2: Griechische Philosophie 1. Thales bis Kyniker. Hrsg. v. P. Garniron u. W. Jaeschke. Felix Meiner Verlag, Hamburg 1989 (G.W.F. HEGEL, Vorlesungen. Ausgewählte Nachschriften und Manuskripte, Bd. 7) (442 pp.)

Dentro del esfuerzo de revisión textual de las obras de Hegel que lleva a cabo el Hegel-Archiv le ha tocado el turno también a las *Lecciones* berlinesas; unas obras de segunda categoría, por cuanto no son propiamente de Hegel, ni pensadas para su publicación, sino simples apuntes que alumnos tomaron de los cursos de Hegel; por ello se lleva a cabo no propiamente una edición crítica, sino una “Studienausgabe” más conforme con la base textual de que hoy se dispone. Con ser de segunda categoría, estos escritos han sido de gran influencia, por ser a veces los únicos testimonios de Hegel sobre algunas cuestiones (Estética, Filosofía de la Historia, Filosofía de la Religión y evidentemente Historia de la Filosofía) y por haber contribuido poderosamente a la creación de tópicos sobre la filosofía de Hegel, por tratarse de unos textos más ligeros y fluídos y por tanto de más fácil acceso que las obras que Hegel mismo escribió y publicó. Justamente entre otros efectos que tiene esta nueva edición de las *Lecciones* es el de contraponer a la imagen convencional de un Hegel filósofo sistemático del estado prusiano que todo lo tiene ya encasillado dentro de su férrea lógica, un Hegel más bien en búsqueda del concepto y del hilo sistemático, en intentos sucesivos, como si en cada curso estuviera experimentando en búsqueda de la exposición más adecuada del tema.

Ciertamente las *Lecciones sobre la Historia de la Filosofía* no son el lugar más idóneo donde pueda observarse esta búsqueda del esquema sistemático adecuado, dado que éste viene dado por el orden cronológico. Las diferencias entre los diversos cursos son en este caso menores, afectando solamente al menor o mayor desarrollo de alguna cuestión o al acento que en ellas se pone. Por eso los editores se han decidido por editar un solo curso: las lecciones del semestre de invierno de 1825-26, que son además las mejores por su contenido, las mejor documentadas, y por coincidir en sus líneas generales con las de otros semestres. Seguramente por estas mismas razones ya el primer editor de dichas lecciones de historia de la filosofía, Michelet, las tomó como base de su edición, aunque ampliándolas con añadidos de otros cursos, añadidos a los que esta nueva edición ha renunciado en favor de la autenticidad del texto, liberado de manipulaciones. Este hecho representa una gran diferencia respecto a las *Lecciones sobre filosofía de la religión* en las que se editaron las lecciones de todos los cursos por separado, dado que cada una representa una unidad por ella misma, con esquema y desarrollo propios, cosa que no se da en estas de historia de la filosofía.

Para esta edición se ha tomado como base el manuscrito del capitán von

Griesheim, sin duda el mejor en todos los sentidos, en razón de la gran riqueza de contenido, aunque alguna vez vaya completado por los manuscritos de Hc (anónimo) y Pinder. Los títulos y subdivisiones, que no son de Hegel, han sido puestos por los editores, siguiendo los tópicos comunes. Otra peculiaridad de estas lecciones es que no cuenta con un texto de Hegel. Sin embargo, estas lecciones están tan bien documentadas que, comparando los diversos apuntes, uno casi podría establecer dicho texto. Respecto al establecimiento del texto los editores informan tanto en la introducción como más detalladamente al final del texto antes de las notas; al principio del primer volumen (aún no publicado) darán cuenta de los criterios que guían toda la edición. En cambio algo que comparten estas lecciones con otras de esta nueva edición es la enorme riqueza de notas explicativas -que van al final, a diferencia de las de crítica textual que van a pie de página-, aduciendo los textos a que hace referencia Hegel en el texto, dando citas, fuentes y toda clase de datos; baste considerar que más de la mitad de las páginas de cada volumen están dedicadas a las notas, siendo además estas páginas impresas en un tipo de letra por lo menos doblemente más apretado que el del texto.

Pasando ya a la consideración del contenido, hay algo que llama enormemente la atención: de los 4 volúmenes que comprenden estas lecciones de historia de la filosofía dos estarán dedicados a la filosofía clásica griega; el primero estará dedicado a la introducción y a la filosofía oriental; el cuarto comprende desde la filosofía patrística y escolástica hasta Schelling; la filosofía griega clásica se lleva dos terceras partes del total. De todos modos ello no

representa ninguna novedad, puesto que la misma proporción daban ya las ediciones anteriores. Que Hegel dedicara tanto espacio al período griego no parece atribuible a que al principio del semestre demorara en los temas y que, al ver que el semestre tocaba a su fin, acelerara el ritmo pasando como por sobre ascuas por encima de los temas más modernos y casi contemporáneos, sino que todos los indicios apuntan a que lo hacía intencionadamente. Ello se explica por el aire de familiaridad que el europeo experimenta -según comenta el mismo Hegel- al pasar de la consideración del arte, la religión o la filosofía de los pueblos orientales a la de los griegos. Esta familiaridad no responde puramente al filohelenismo o entusiasmo por la Grecia clásica, de moda en su tiempo, sino al hecho de que -como afirma Hegel- “toda ciencia y arte, todo lo que adorna la vida espiritual y la hace digna... ha llegado a nosotros saliendo en parte directamente de los griegos y en parte por el rodeo de los romanos”. Esta simpatía por la filosofía griega no le impide afirmar la superioridad de la moderna, aunque también dicha superioridad es -según él- herencia griega.

La filosofía moderna está dividida en cuatro apartados: 1. Bacon y Böhme, 2. Descartes y Spinoza (y Malebranche), 3. Locke y Leibniz (además de Grocio, Hobbes, Wolff, “filosofía metafísica y popular” -donde se evoca a Rousseau, Voltaire, d’Holbach- y David Hume, cuyo escepticismo constituye el paso inmediato a la filosofía kantiana [p. 146]). La ausencia más notable quizás sea Berkeley, además de algunos ilustrados franceses: La Matrie, Helvetius, etc. El cuarto apartado está dedicado a “Kant, Fichte y Schelling”, pero con la peculiaridad de que no se exponen sucesivamente, sino juntándolos por temas; se empieza

por la estética y analítica trascendentales kantianas (pp. 149-156) y la parte teórica de la *Doctrina de la ciencia* de Fichte (pp. 156-162), después viene la dialéctica trascendental (pp. 162-164), Jacobi (pp. 165-169), a lo que sigue la filosofía práctica kantiana con referencias a Fichte (pp. 167-169), y finalmente la crítica al juicio teleológico y a los postulados de la *Crítica de la razón práctica* o la prueba moral de la existencia de Dios de la *Crítica del Juicio*, de nuevo en conexión con el concepto de fe de Jacobi y el concepto de saber inmediato (pp. 169-179). Este entretrejimiento de los autores modernos -los que ahora podríamos llamar del idealismo alemán- o su

tratamiento de modo unitario y sistemático parece ser una constante de dichas lecciones de historia de la filosofía.

No es claro el lugar que Hegel asignaría dentro de su sistema a estas *Lecciones sobre la historia de la filosofía*, si hay que considerarlas parte del sistema o incluso su coronación conclusiva; pero en todo caso está fuera de duda la importancia que les concedió Hegel, recogiendo en parte una tendencia propia de su tiempo, pero dándole además una impronta muy propia: el estudio de la historia de la filosofía es estudio de la filosofía misma.

Gabriel Amengual

LAUTH, Reinhard, *Transzendente Entwicklungslinien von Descartes bis zu Marx und Dostojewski*, Meiner, Hamburg 1989 (486 pp.)

El autor es sin duda uno de los mejores conocedores de Fichte, director de la nueva edición crítica de la obra de dicho autor se está haciendo bajo el patrocinio de la Academia Bávara de Ciencias, además de autor de numerosas investigaciones sobre su pensamiento. Pero no es un investigador aséptico, sino entusiasta seguidor de Fichte, que le entiende como la culminación de la filosofía, en contra de toda posible superación de su filosofía trascendental que no puede ser más que una exageración y por tanto degeneración, como es el caso del idealismo absoluto de Schelling y Hegel. Como afirma en el prólogo, todos sus estudios, incluso los aparentemente más historiográficos y filológicos, están dominados por la intención fundamental de separar inequívocamente la filosofía como ciencia o trascendental de la pura especulación (p. VII).

El libro es una recopilación de estudios, la mayoría de los cuales ya publicados por separado, algunos aún inéditos, casi todos con el mismo origen: conferencias o ponencias en congresos. Sin ser un estudio historiográfico llevado a cabo de modo sistemático, el contenido responde exactamente al título. Hay estudios sobre Descartes, Rousseau, Kant, Fichte, Reinhold, Jacobi, Hegel, Schelling, Dostoievski, siendo casi todos ellos estudiados en sus relaciones: visión que Fichte tiene de Spinoza, crítica de Fichte a Kant, relación entre Kant y Fichte o entre Fichte y Jacobi, interpretación jacobiniana de Descartes, etc. Con ello se echa ciertamente mucha luz sobre un período (el que podríamos llamar clásico de la filosofía moderna) y sobre una filosofía: la trascendental. Al final de cada estudio se indican

otros estudios del autor sobre el mismo tema; estas referencias dan cuenta del documentado conocimiento del tema, así como de la amplitud de la investigación del autor.

Como se ha insinuado, estos estudios, aún siendo historiográficos, están lejos de ser puramente historicistas. Seguramente se les podría aplicar lo que según el autor respondería Descartes a los historicistas: «no me interesa si y por qué lo dijeron, sino si es verdadero lo que dijeron» (p. 16). Y ello por fidelidad al propio planteamiento de la filosofía trascendental. La filosofía trascendental es entendida como aquella filosofía que fue inaugurada por Descartes, desarrollada en la crítica de la razón kantiana y en la doctrina de la ciencia fichteana y que se basa en el pensamiento fundamental de que la filosofía, como ciencia del conjunto de principios de la realidad, todo lo que afirma ha de mostrarlo como sabido en verdad (p. 360). La filosofía trascendental, por tanto, no es entendida puramente como filosofía de la conciencia, en contraposición a una filosofía del ser o del objeto (ni como filosofía moderna en contraposición con la antigua), sino como síntesis de idealismo y realismo, precisamente como la superación del subjetivismo/escepticismo y del objetivismo/dogmatismo; puesto que la filosofía de la conciencia y del pensar remite al ser que lo posibilita, el cual aparece en el pensar; pero a la vez el pensar es tal si es del ser o de la verdad y no al revés. Esta concepción, que subyace en todos los estudios, está claramente expuesta en el que versa sobre la filosofía de Descartes como proyecto de la filosofía moderna (pp. 20ss.).

La filosofía trascendental es además esencialmente práctica, es una filosofía de la libertad, por ello no llega a su plenitud cuando se desarrolla como ciencia, sino cuando se realiza en una forma de vida. De aquí que el autor integre dentro de estas líneas de evolución de la filosofía trascendental no sólo a los autores más tópicos como Descartes, Kant, Fichte, Reinhold, etc. (además de Schelling y Hegel que son estudiados por contraste, cfr. especialmente pp. 360ss.), sino también Rousseau, Marx y Dostoievski. Sobre todo que éste último sea integrado dentro de esta historia puede extrañar. Este hecho puede

ser considerado una muestra de la amplitud de intereses y saberes del autor; de hecho es un buen conocedor de Dostoievski, a quien cita en el original ruso (y en general de la literatura rusa) y del que más de una vez ha ofrecido lecturas filosóficas o lo ha integrado dentro de la filosofía; a veces como contrapunto a la filosofía marxista -hasta ahora imperante en aquellas repúblicas- como si quisiera mantener viva la memoria del alma rusa, que por mucho que se haya volcado en la defensa del sufriente no por ello ha perdido el corazón ni la fe.

Gabriel Amengual

MAYOS SOLSONA, Gonçal, *Entre lògica i empiria. Claus de la filosofia hegeliana de la història*. Prefaci de Ramon Valls Plana. Barcelona 1989, 259 pp.

La filosofia de la història ha estat considerada freqüentment com la gran aportació del pensament de Hegel i en moments recents de revisió d'aquest pensament havia estat postergada, per immadura, per haver-la deixada Hegel quan la tenia encara en fase d'elaboració, i per la magre i barrejada base textual. L'autor la mostra com una peça en la que s'hi reflecteix tota la filosofia de Hegel, sense però confondre-la en cap moment amb el tot, sinó distingint-la ben be tant pel seu objecte com pel seu mètode.

El present treball, publicació d'una part de l'extensa tesi doctoral dirigida per Ramon Valls Plana, té dues parts ben diferenciades. La primera tracta uns «cossos de doctrina» que «serveixen a Hegel com a base conceptual i clau hermenèutica a partir de la qual interpretar i donar sentit a la massa ingent de dades i esdeveniments de la història universal» (p. 19). Aquests cossos de doctrina, exposats en els 5 primers capítols (que van precedits d'una introducció en la que es situa el tema tant en el conjunt del pensament modern com en el de la filosofia hegeliana), són: «La llibertat, la meta de la història», «L'estat i les constitucions polítiques», «El paper de la religió en la història», «Esperit versus natura». El darrer, dedicat a «L'astúcia de la raó, o com la lògica s'imposa en i per mitjà de l'empíria», ja no és un cos qualsevol, sinó «el mecanisme que els unirà [els altres cossos anteriors] i dotarà de dinamicitat el conjunt» (p. 20). És un capítol clau pel paper que aquest tema juga en la filosofia de la història de Hegel i per a la comprensió de la temàtica de la tesi de Mayos, la unió de lògica i empiria, i també per lo ben tractat

que està.

La segona part (formada per l'amplíssim capítol 7) està dedicada a un aprofundiment del status de la filosofia hegeliana de la història: el seu lloc sistemàtic i la relació entre lògica i empiria, a més d'alguns interessants apartats sobre el progrés i la fi de la història. El lloc sistemàtic de la filosofia de la història no és pensat com a pura referència a les altres parts del sistema, sinó segons el model del tot en la part: totes les altres parts (expressament la filosofia de l'esperit, però indirectament també la filosofia de la natura i la Lògica) s'hi reflecteixen. La connexió amb la filosofia de l'esperit és formulada amb aquesta «tesi evident i d'àmplies conseqüències: la filosofia de la història universal conté la totalitat de la filosofia de l'esperit hegeliana» (p. 159). La diferència respecte a l'exposició sistemàtica de la filosofia de l'esperit rau en que mentres aquesta segueix una ordenació lògico-sincrònic-sistemàtica, en la filosofia de la història hi trobam una ordenació lògico-diacrònic-històrica (p. 160). Aquesta ordenació constitueix l'especificitat de la filosofia de la història que «rau en la realització empírica de l'esperit universal en un poble existent naturalment i exteriorment» (p. 162) i per tant constitueix la diferència respecte al sistema i a altres obres, com són la història de la filosofia i la *Fenomenologia de l'esperit*, la diferència entre les quals «rau bàsicament en la diversa referència a l'empíria» (p. 179). D'aquesta concepció de l'especificitat de la filosofia de la història en surt el tema clau: la relació entre lògica i empiria, ja que «en cap altre text com en aquest, Hegel no es compromet

tant a cercar aquesta lògica sota, en i per mitjà de l'existència empírica» (p. 171). La relació entre lògica i empiria és d'immanència: «En la filosofia de la història ens trobam sempre una lògica que viu i vitalitza immanentment la materialitat empírica de la història. [...] No hi ha, doncs, una lògica transcendent de l'empíria ni tampoc aquesta no és transcendent a la primera, sinó que totes dues s'impliquen i poden ser considerades com a moments de la idea» (p. 204).

Un poc problemàtica me pareix la relació establerta entre la filosofia de la història i l'esperit absolut, de tal manera que crec que delata una certa contradicció en la interpretació de Hegel que se proposa. Per una part s'afirma que l'esperit absolut és en certa manera representat per la religió (p. 77) i com a tal és tractat; la religió és entesa sobretot per la seva funció social: educadora i legitimadora (pp. 71ss.); així es ve a donar la primacia a allò polític, a l'esperit objectiu, ja que l'absolut és posat al seu servei; però per altra part s'assenyala la importància del saber, ja que la reconciliació en darrer terme es dona en l'autoconeixement, tal com indica sobretot l'esperit absolut que culmina en la filosofia, l'exposició de la qual és a més rematada amb la cita aristotèlica sobre el noesis noeseos. Segurament el text mateix de les *Lliçons sobre la filosofia de la història universal* afavoreixen una tal primacia d'allò polític, i per tant s'havia de confrontar amb una acurada consideració del lloc concret de la filosofia de la història en el sistema, com a conclusió de l'esperit

objectiu i pas a l'absolut, la qual sembla que hauria pogut donar raó millor del paper i de l'estatus també de l'esperit absolut. De totes maneres és una qüestió que només de manera indirecta afecta al tema tal com està enfocat de l'estudi, és una qüestió que hi ressona només com a música de fons, encara que alguna vegada passi al primer pla.

No hi ha dubte que s'afronta un tema bàsic de la filosofia de la història en general i de la hegeliana en particular, així com també un tema fonamental de la filosofia hegeliana en general. De fet, explicant aquesta relació entre lògica i empiria s'expliquen moltes peces de l'engrenatge del complex sistema hegelian, com per exemple la distinció entre enteniment i raó, realitat fàctica i realitat efectiva, captació racional de la realitat, la identificació entre real i racional, etc. Es també un bon moment -i l'autor l'aprofita amb intel·ligència- per aclarir prejudicis, imatges precipitades, falses o unilaterals, presentacions simplistes que no estan a l'altura ni poden donar compte de la complexitat d'aquesta filosofia, de les que sempre ha estat plegada la vulgarització de la filosofia hegeliana. El treball està dedicat a la filosofia de la història i per tant sobretot a les *Lliçons sobre filosofia de la història universal*, però, pel mateix enfocament del tema i la comprensió de la filosofia de la història hi ressona tot el sistema i l'autor mostra un bon coneixement ben aprofundit del conjunt.

Gabriel Amengual

MUGUERZA, Javier, et al., *El fundamento de los derechos humanos*, Ed. Debate, Madrid 1989.

L'Institut de Drets Humans de la Universitat Complutense de Madrid, responnent a un encàrreg de la Fundació Tanner, organitzà durant els dies 19 i 20 d'abril de 1988 unes conferències i seminaris a Madrid sota el lema «la fonamentació dels Drets Humans». L'encontre constava d'una conferència principal, la del professor J. Muguerza, amb un posterior col.loqui i debat, més la defensa per part de la resta dels participants de les corresponents comunicacions alternatives. Sense precedents al si de l'Estat Espanyol, es podria recordar en canvi el famós col.loqui de L'Aquila, celebrat el 1964 baix el patrocini de l'Associació Internacional de Filosofia, amb la participació de Bobbio, Perelman i altres («Le fondement des droits de l'homme», *La Nuova Italia*, Florència, 1966). Tot i que, com deim, un debat d'aquestes característiques, amb la concurrència de filòsofs i juristes, resultava ben novedós, sí que existia (i existeix) en el si de la cultura jurídica espanyola un tractament acadèmic i seriós de la qüestió del fonament dels drets, concretament dins l'àmbit de la Filosofia del Dret: és el cas de les obres de A.E. Pérez-Luño, Eusebio Fernández, Francisco J. Laporta, Luis Prieto, A. Fernández-Galiano, Manuel Atienza, o el propi G. Peces-Barba, organitzador de l'encontre. A l'esmentat debat s'incorporaren altres especialistes, en aquest cas filòsofs, començant pel propi J. Muguerza. El resultat d'això és un contrast de punts de vista que es reflexa en el contingut de les ponències, segons es provingui d'un camp o de l'altre (de la Filosofia o del Dret).

A més del treball que encapçala el llibre, del professor Muguerza, es poden trobar les col.laboracions dels professors (-es) I. Ara Pinilla, R. Asís Roig, M.

Atienza, J. Ballesteros, A. Calsamiglia, V. Camps, B. De Castro Cid, A. Cortina, J. Delgado Pinto, E. Díaz, E. Fernández, A. Fernández-Galiano, José A. Gimbernat, José M. González García, E. Guisán, G. Gutiérrez, F. Laporta, N. López Calera, J. de Lucas, A. Llamas Gascón, Manuel-Reyes Mate, A. Ollero Tassara, J. R. de Páramo, G. Peces-Barba, A. E. Pérez-Luño, F. Puy Muñoz, A. Rivero, G. Robles, A. Ruiz Miguel, J. Seoane, C. Thiebaut y R. J. Vernengo. El resultat és la contribució col.lectiva més important sobre el tema de la fonamentació dels DH dintre del panorama editorial en llengüa castellana.

Degut a les limitacions d'espai no ens ha resultat possible presentar ressenyes de cadescuna de les aportacions, i ens limitarem únicament a fer, al final d'aquesta introducció, un breu resum de la que encapçala el llibre, la del professor J. Muguerza (ben polèmica per altra banda, i a la qual fan freqüent referència els restants treballs).

Algunes de les contribucions recollides en aquest llibre havien estat publicades prèviament durant l'any 1988, concretament als «Anales de la Cátedra F. Suárez», nº 28. Això és una mostra de la vigència que ja tenia aquest debat sobre fonamentació dels DH dins l'àmbit de la Filosofia del Dret. El llibre que comentam, doncs, es presenta com una continuació d'aquest debat, però sobretot com l'intent de portar-lo a un terreny interdisciplinar, comú a (la Filosofia del) Dret i a la Filosofia (moral), fet que pensam no fa sinó incrementar l'interès del volum.

La reflexió entorn al contingut dels DH navega, efectivament, per aigües comunes al Dret i a la Ètica, o el Dret i la Filosofia en general, però de fet ha estat en

el terreny del Dret on s'ha cultivat amb més profusió la referida reflexió, via Filosofia del Dret. En canvi a les Facultats de Filosofia hom s'ha ocupat poc d'un tema que potser es veu massa allunyat de la «pura» reflexió o del concepte abstracte, per molt que realment no sigui així. En efecte: n'esteim segurs que allò «conceptual» o «filosòfic» s'ha de definir sempre per oposició a allò «pràctic»? Ben al contrari, dins l'àmbit de la racionalitat humana tot es troba relacionat amb tot. Ho veim també a Kant: les preguntes «què m'és donat conèixer?» i «què haig de fer?» es subsumeixen baix la pregunta (juntament amb «què me cal esperar?») única: «què és l'home?». Com deien els racionalistes, el saber és ú perquè la raó és una, i, podríem afegir, perquè la realitat també és una. Això pot ser una bona raó en favor dels enfocos multidisciplinars. Ningú no té la patent per poder parlar de «fonament» en un sentit excluint: la fonamentació jurídica no exclou la fonamentació filosòfica, i la fonamentació filosòfica no exclou la fonamentació jurídica.

Finalment, només recordar que del total de les aportacions devuit provenen del camp del Dret (proporció dos a un, aproximadament), set de la Filosofia Moral, i després hi trobam quatre aportacions de membres del CSIC (Centre superior d'Investigacions Científiques) i dues de l'IDH (Institut de Drets Humans), depenent de la Universitat Complutense de Madrid.

L'aportació de J. Muguerza es titula «La alternativa del disenso». Després d'unes consideracions inicials, Muguerza agafa com a punt de partida la definició de DH donada per Pérez-Luño. La interpreta en termes d'una teoria «dual» distingint entre «valors» i «normes jurídiques». Existeix, efectivament, un «substrat ètic»

constituït per les «exigències» de *dignitat, llibertat i igualtat* humanes, exigències que són de naturalesa ètica (i no «drets morals», recusant la coneguda qualificació). Aquestes exigències *han de ser* reconegudes positivament, pel seu propi caràcter d'exigències. Els ordrenaments jurídics són els encarregats de recollir-les i concretar-les, a nivell nacional i internacional, convertint-les, en sentit estricte, en *dret* (no podem parlar de «dret» a res sense la corresponent sanció jurídica). Muguerza s'oposa a aquella mena de *positivisme* que redueix la pregunta pel fonament a la sola facticitat del reconeixement d'uns drets o uns altres per l'ordenament jurídic corresponent. En certa manera, aquesta seria la postura de Bobbio, segons el qual la qüestió del fonament es troba «resolta» des del 10 de desembre de 1948.

Bobbio troba només *tres vies* per a la fonamentació dels DH: 1.- La que deriva els DH de la *naturalesa humana* (iusnaturalisme, entre altres); 2.- La que els deriva d'algun *principi d'evidència* (els valors serien evidents en ells mateixos); 3.- La que els deriva del *consens* (múltiples postures, entre les quals hi ha la del mateix Bobbio. Analitzant aquesta darrera via, Muguerza fa un recorregut pel contractualisme modern i el neo-contractualisme (Apel, Habermas) per concloure que totes aquestes no són sinó versions del consensualisme, inspirades en la màxima d'universalització de l'imperatiu categòric kantian. Muguerza rebutja aquesta via per insuficient (legítima suficientment la democràcia, però no abarca tot l'àmbit de la raó pràctica). Per a ell, cap acord col·lectiu de caire fàctic (ni tan sols un «consensus omnium gentium») pot contenir en ell el seu propi fonament racional, donat que la facticitat dels acords no és per ella mateixa garantia de la seva

racionalitat. Proposa la distinció entre condicions formals de l'argumentació o comunicació per una banda, i el contingut mateix de l'argumentació (necessitats, interessos reals, etc.) per una altra.

En aquest punt apareix la proposta de J. Muguerza en relació a la fonamentació dels DH, com a alternativa a la via de fonamentació exposada anteriorment. Muguerza proposa estalonar-se en la segona versió de l'imperatiu categòric kantian, i no tant en la primera. La segona versió es basa en la idea que l'home existeix com un fi en ell mateix i no ha de ser mai utilitzat per ningú (ni tan sols per Déu) únicament com un mitjà, sense ser al mateix temps un fi. Segons Kant, «el fi no s'ha de concebre com a fi a realitzar, sinó com un fi independent i per tant de manera purament negativa, és a dir, allò contra el que hom no s'ha d'obrar en cap cas». Aquesta idea portaria, segons Muguerza, a aplicar a la història social i política un esquema similar al de Kuhn en la història de la ciència (ciència normal enfront de ciència revolucionària): ordre establert enfront de moment revolucionari.

Després d'unes referències a Tugendhat, que treballa i desenvolupa el significat d'aquest segon imperatiu, passa Muguerza a la distinció entre *subjecte moral* i *subjecte empíric*. D'aquesta subjectivitat moral, no reduïble a la empírica, brollen la nostra autoconsciència i la nostra autodeterminació, essent alhora el fonament de la nostra *dignitat humana*; és a dir, allò que fa que siguem «subjectes» i no solament «objectes». El més bàsic de tots els drets seria el dret del «subjecte moral» a ser reconegut com a «subjecte de drets»; és a dir, el «dret a tenir drets». Qui concedeix aquest dret?. L'individu mateix, en autoafirmar-se, en desalienar-se, una volta es sap subjecte (negant la seva reducció de subjecte a objecte). Aquesta autoafirmació del subjecte es concreta, per acabar, en la figura del «disident», aquell que nega l'ordre establert per a, rebelant-se, afavorir l'establiment d'un ordre (sistema de dret) nou, més just.

Antoni Barber Orfila



Universitat de les Illes Balears
Departament de Filosofia