Can We Say What We Mean?
Expressibility and Background

Jesús Navarro-Reyes
University of Seville

Abstract: The aim of this paper is to discuss a basic assumption tacitly shared by many philosophers of mind and language: that *whatever can be meant, can be said*. It specifically targets John Searle's account of this idea, focussing on his Principle of Expressibility (PE henceforth). In the first part of the paper, PE is exposed underlining its analyticity (1) and its relevance for the philosophy of language (2), mind (3), society and action (4). In the critical part, the notion of Background is taken into account in order to re-evaluate two basic distinctions: the one between sentence and utterance meanings (5), and the one between native and type speakers (6). PE is reconsidered in the light of the previous arguments as a methodological strategy that does not prevent uses of language from eventual semantic excesses and deficits (7), and a complementary Principle of Expression Fallibility is finally proposed (8).

Keywords: Expressibility, background, utterance/sentence meaning, individual/collective intentionality, native speaker, John R. Searle.

1. The Principle of Expressibility as an analytic judgment.

Back in his classic essay on *Speech Acts*, John Searle established what he called the Principle of Expressibility (PE), according to which *whatever can be meant, can be said* (Searle 1969: 19ff). Despite its explicit formulation, PE has received an astonishing little amount of criticism, probably due to the fact that it is considered unproblematic not only by Searle's followers, but also by many of his critics. In fact, the idea is generally present wherever propositional attitudes are considered to be descriptions of real mental events, and not merely unessential strategies for behavioural prediction.

PE is a key concept in Searle's account of speech acts, because it allows him to analyse language from what he considers to be *the simplest cases of meaning*: "those in which the speaker utters a sentence and means exactly and literally what he says" (Searle 1975: 30). In his opinion, any other case of language use can be reduced to those, since

for any meaning *X* and any speaker *S* whenever *S* means (intends to convey, wishes to communicate in an utterance, etc.) *X* then it is possible that there is some expression *E* such that *E* is an exact expression of or formulation of *X* (Searle 1969: 20).

It is important to emphasise that PE is not necessarily true *de facto*, but only *de iure*—a point that makes it less interesting in practice, but also less vulnerable to conceptual criticism. Searle does not claim that, whatever the speaker intends to say, there actually *is* an expression in his own language that she could use to say it. There might be no expression available indeed, due to speaker's ignorance or to lexical limitations, but Searle's point is that those boundaries can always be overstepped: the speaker could acquire new idiomatic abilities, or language could always be extended, in order to satisfy the expression necessities of its speakers. If there are limits for expressibility, that is "a contingent fact and not a necessary truth" (Searle 1969: 20).

There is no point in considering specific cases in order to prove or refute PE, since it is formulated as an analytic judgment, and analytic judgments are supposed to be immune to particular experiences.[1] E.g., it would be futile to look for a married bachelor in the Civil Registry in order to falsify the analytic judgment "All bachelors are unmarried", for all we could finally prove is that the Civil Registry was wrong. By the same token, no particular experiment could ever falsify PE, namely because such an experiment would be hardly conceivable: if I claimed that there is a particular content that could neither be expressed in fact by a particular speaker, nor in principle by anyone else, it would be a precondition for my experiment to be accepted as such, to indicate *what* is that content, the speaker is supposed not to be able to say. Would I not have to express it, in order to be my hypothesis considered by others? In that case, if that particular content is expressible *de facto*, at least for me, why shouldn't it also be expressible in principle for that speaker? PE seems to be proved by *reductio ad absurdum*, since there is apparently no way to formulate the alternative possibility.[2]

However, if PE were analytic—had we got to believe Searle—that would only mean that it is true *by definition*, but would not say much about the correctness or incorrectness of our definitions. On that point, we can only make a judgment considering the fertility of our theoretical system as a whole (Quine 1950: 1-8). Being analytic does not preserve a thesis from being reconsidered, questioned, or complemented, as I will try to do in the second part of this paper. Nonetheless, we should be very conscious of the effects that a reconsideration of PE would eventually

have because, as I will show in the first part of this paper, PE plays a much more crucial role in Searle's philosophy in general—and not only his speech act theory—than could be expected at first sight.

2. Expressibility and language: utterance and sentence meanings.

Searle's conception of meaning is constructed on Grice's 1957 proposal:

> To say that a speaker *S* meant something by *X* is to say that *S* intended the utterance of *X* to produce some effect in a hearer *H* by means of the recognition of this intention. (Searle 1969: 43).

According to Searle, although Grice's proposal is valuable, it neglects the crucial fact that linguistic expressions, as social institutions, have literal meanings, which are independent from the communicational purposes of the particular speaker. This is a conventional feature of the speech act that the speaker must assume when she performs it: it is certainly up to her to use *X* in order to say something different from what *X* means—hoping that the hearer will guess by the context that she is not speaking literally—but, if such an indirect act is possible, it is because both speaker and hearer do previously share the knowledge of *X*'s meaning. Namely, they share a conventional language.

Searle's analyse is in a sense the reverse of Grice's one. The latter tries to move from the particular meaning of an utterance—not yet conventionalised—to what he calls *timeless meaning* of the sentence uttered; the former, on the contrary, begins his account by conventionalised meaning and, from that starting point, tries to understand the working of non conventionalised expressions. Searle's point is that, without a previous knowledge of linguistic conventions, no pragmatic deduction of non-literal meanings would be possible. Metaphors, ironies and indirect speech acts are in general, in this sense, *parasitic* on literal, direct speech acts, as they are logically dependent on them: "In each of these cases, what the speaker means is not identical with what the sentence means, and yet what he means is in various ways dependent on what the sentence means" (Searle 1975: 77).[3]

We should consequently distinguish, each time the speaker made an utterance, what she meant to say from what she actually said; i.e., the meaning of her utterance from the meaning of the sentence she uttered. Any non literal feature will correspond to utterance meaning—for only utterances can be metaphoric or ironic, not sentences themselves. In fact, according to Searle (1979a: 118), "literal sentence" is a pleonasm, since sentences as such can only have literal meanings.

What PE would allow us to say is that, whenever the speaker utters a sentence intending to communicate something different from what the very sentence means, there does exist (or *can* exist) another sentence which means—of course, literally—what the speaker is intending to say. In fact, what makes the hearer's contextual, pragmatic deduction possible—i.e., her deduction from utterance to sentence meaning—is that there is (or *could* be) another sentence whose meaning would literally express the one of the utterance. What else could be the aim of her inquiry? How else could she guess what the speaker *means*, behind what she actually *says*, if it would not be eventually possible to *say* what she *means*? We cannot even speak about utterance meanings unless we identify them, in an implicit or explicit way, with particular sentence meanings, since otherwise they would not only be inexpressible, but probably unthinkable for us (I will return to this idea in the next section).[4]

By the title of his work, one could be led to mistake Searle's *Speech Acts* for a study of what Ferdinand de Saussure called *parole*: the individual act, by which a speaker *uses* a language, in contrast to the *langue*, as the abstract set of signs available to the community. It would not be an abstract theory about semantics or syntax, but a study of the use of language in particular situations, considered in their varying pragmatic context. Nevertheless, Searle offers his own theory as a study of the *langue*, not of the *parole*: his object is the socially shared aspect of language that the individual cannot change, but passively accept as a pre-condition for the speech acts that she intends to perform. Searle is not interested in the particular application of the rule, but in the rule itself, which, as such, does not belong to the specific situation of a particular speech act. It is PE what allows him to take that step, in that it assures the possibility of establishing a link between utterance and sentence meanings: for each *parole* act we can perform, there is (or can be) a *langue* structure whose rules we can describe. This is what has been called the *semantization* of Searle's speech acts theory (Dascal 1983, 31),

namely the idea that we can identify the rules of speech acts with the meanings of the sentences used in those acts:

> just as it is part of our notion of the meaning of a sentence that a literal utterance of that sentence with that meaning in a certain context would be the performance of a particular speech act, so it is part of our notion of a speech act that there is a possible sentence (or sentences) the utterance of which in a certain context would in virtue of its (or their) meaning constitute a performance of that speech act (Searle 1969: 18).

The speaker may accomplish her speech act by uttering words whose literal meaning are not identical—or are even opposed—to the very act she's doing (e.g., she could be ironical); but in that case, if PE is sound, it would always be possible to translate that particular utterance meaning into the sentence meaning of a different expression. Utterance meaning, if it does not coincide with the meaning of the sentence that is being uttered, is nothing but the meaning of a different sentence. Cases where the speaker is not speaking in a direct, literal way can be left aside for the sake of theory, since it is always possible to translate what the speaker is intending to say into words that actually say it. Perlocutive or rhetorical effects would be different (see Searle 1979a: chapter IV), what Frege disdainfully called the *coloration* would be lost, as well as ambiguity and vagueness (which might be part of what the speaker intended to communicate). However, thanks to PE, if the speaker was not intending to say q when she uttered q, but something different like p, then she could have just said p. Then, from a theoretical point of view, we can limit our scope to the general rules of use of expression p—*langue* rules that do not only govern that particular situation, but any situation of that type, in which p is literally uttered.

There can be no science of particulars, *magister dixit*: Searle does not want to construct his solid theory on the unstable, changeable and unpredictable particularity of specific use, but on the general and socially shared structure, which supports linguistic practices. From this point of view, performance of speech acts can be analysed as the act of uttering sentences literally, in the correct situation, following socially established rules. The particularity of abnormal uses where these rules are left aside will be explained in form of corollaries (Searle 1979a), but should not be mistaken for a constitutive aspect of the theory itself.

3. Expressibility and the mind: urbanising the cognitive space.

Previous section was focused on meaning and language and, under that light, sentence meaning had priority over utterance meaning. However, this section will focus on philosophy of mind and, from this perspective, it is the utterance or, more specifically, the speaker, who takes priority, as that is the only place where we can find what Searle calls *intrinsic* intentionality.

Intentionality is primarily a feature of some mental states such as perceptions, beliefs, or desires, which are addressed to something else. Although language is also intentional, since words generally refer to something, they do so in quite a different way. If we consider language itself, we will only find ink on paper, or sounds in the air: just ordinary physical phenomena that are not intentional on their own. They are simply there, like sunsets or volcanic eruptions, not referring to anything else. According to Searle's view, intentionality can only be attributed to linguistic events if they are considered as the effect of a mind—the one of a speaker or a writer. Linguistic intentionality is *derived* from mental intentionality, but the latter does not derive from anything else: it is just the effect of the brain, which acquired through our evolutionary history the ability to be in intrinsically intentional states (Searle 1992: 78-82). Unlike Putnam (1981) or Dennett (1987), this is considered by Searle as a brute fact that would require no further philosophical reduction.

Besides the different character of mental and linguistic kinds of intentionality, Searle considers them to be articulated by one same logic, which he tries to make explicit (Searle 1983). Understanding the mind of a person is attributing intentional states to her, in the form of propositional attitudes such as *believe that…*, *desire that…*, or *having the intention to…*, all of them attitudes towards contents, which can be expressed in propositions. The intentionality of the mind is accessible to philosophical study given that it is essentially expressible in linguistic propositional conventions. This common articulation is indispensable if we want to understand the phenomenon in question: if the intentionality of the mind were not translatable into the intentionality of language, hardly could we explain how is it possible to say in words what we are thinking, or how can we understand what the other person is thinking when we hear what she is saying. Just as we previously established that any utterance meaning could be expressed in a sentence meaning, we will have to accept now that any mental

intentional state can be expressed linguistically. In this way, Searle seems to presuppose what Frank (1984: 399), following Derrida, called a "pre-established harmony" between intentions and expressions, which guarantees the possibility of their eventual identification.

> In general, a speech act with a propositional content is an expression of the corresponding intentional state, and the propositional content of the speech act is identical with the representative content of the intentional state (Searle 1979b: 192).

Even if Searle does not explicitly appeal to PE in this particular point, it seems to be an unavoidable condition of possibility of the kind of philosophy of mind that he considers significant. The *a priori* establishment of this access to the mind is a must for Searle, whose position is epistemically controversial in a hostile environment, dominated by functionalism and its derivatives. By its side, functionalism attempts to solve the puzzle of mental inexpressibility before it arises: if language is able to express what happens in the external physical world, i.e. behavioural patterns, then it is able to express what happens in the inner mental world, since inner mental assessments are, according to functionalism, essentially linked to behaviour (functional computational states, even if unobservable in themselves, are inferred as casual explanations of observable behaviour). But the distinctiveness of Searle's position—and the key of the appeal that it holds—is that he tries to grasp the very essence of mental states, considering their first person perspective as an essential feature, and not linking them by definition to any kind of external behaviour whatsoever (Searle 1992: 77-94). The point seems to be reasonable: we are perfectly able to entertain particular beliefs, desires or intentions, but not show any behavioural effect at all, even not in principle. Mental states are not essentially linked to behaviour: if functionalists do assume this link as a necessary condition, it is not because of the very nature of the phenomenon they are studying, but because of some unjustified empiricist prejudices they have inherited from behaviourism. Nevertheless, Searle's view leads to a problem that does not even arise in the functionalist approach: the difficulty to articulate mental and behavioural statements, i.e., descriptions of subjective and objective events.

The claim that conscious mental states are not essentially linked to behaviour is shared by Searle with authors such as Galen Strawson or Colin McGinn, but he does not want to be commited, as they are, to strong scepticism or 'mysterianism' about

subjectivity. Therefore, he needs a guarantee that our common language is able to express what we inwardly intend to communicate, once it has been acknowledged that what we believe or desire *in the inside* is not tied anymore by definition to what we do *on the outside*. And, of course, that guarantee is PE, which states that, even if descriptions of mental states do not derive from behavioural statements, they are *essentially* expressible by linguistic conventions.

By establishing PE, Searle seems to be *urbanising* the cognitive space, allowing sentence meanings to go through it. In the wild land of thought, PE assures the possibility of building roads and streets that will eventually allow the transit of social meanings; meanings that do not only belong to the individual, but to the inter-subjective community of speakers of a common language. If the mind is accessible for linguistic expression, it is not thanks to the possibility of a behavioural translation, but to the ability of the thinker, as a native speaker, to recognise her own mental states and verbalise them, i.e., to say what she believes, desires or intends, and to mean what she says. PE is the guarantee (at least a *de iure* one) that the mind's state can be expressed through linguistic manifestations; namely, that the intrinsic intentionality of the mind can assume the shape of the derived intentionality of language.[5]

In fact, PE helps to solve a double problem in philosophy of mind: in the first place, thoughts probably do not have to be linguistic in order to be intentional, and there must be a way to express the pre-linguistic intentional states in linguistic propositional utterances.[6] And in the second place, even when we do use language to entertain certain thoughts, the identity of thinker's and speaker's meaning should not be taken for granted. As Dascal has shown, uses of language by a solitary thinker "are not even 'potentially' communicative, for, *ex hypothesi*, there is no communicative intention whatsoever in them." (Dascal 1983, 45). Both problems would be hard to solve without PE, since sentence meanings would not be able to express those pre-linguistic or privately linguistic kind of thoughts. Access to what is intrinsically mental would be lost, and the conceptual problem of other minds would be intractable. Philosophy of mind could not even begin to explain what Searle considers to be the very nature of its object.

According to Searle, the next step in philosophy of mind will be the search for neurological correlates for mental states, for we will not find an explanation of the mind until we do not fully understand how those states are produced in the brain. On the one hand, neural states are increasingly accessible through new techniques as Magnetic

Resonance Imaging (MRI), which offer inter-subjective, observable data about them; and, on the other hand, mental events are accessible from the first person perspective, and translatable into inter-subjective language, thanks to the abilities of native speakers to perform speech acts. The articulation between both spheres, even if it is still quite far from reachable, is at least conceivable, given that PE assures a shared access to the content of the mind. Without that principle, our search for the neural correlate would be misguided, for we cannot make a rope taut just by holding one of its sides.

4. Social philosophy and theory of action.

Besides the distinction between intrinsic and derived intentionality, the one between individual and collective intentionality is also very relevant here. It is introduced by Searle (1995) in his search for a philosophical foundation for social institutions such as money, marriage, civil laws, honours or political offices, which do only exist as far as we consider them to be real, and act accordingly to that consideration. What makes them effective is the intentionality of our minds and, as socially shared realities, they stem from a kind of intentionality, which is not just individually hold by each one of us, but socially shared in the form of collective intentionality.

This aspect of his theory is particularly controversial since, according to Searle, collective intentionality is not reducible to any compound form of individual intentionality. In his opinion, individualistic accounts of social reality are ill-conceived, because it is not enough to describe our common beliefs as a result of putting together several individual intentional states (I believe *p*, you believe *p*, I believe you believe *p*, you believe I believe *p*, and so on); on the contrary, we need to introduce into our logic states which are collectively intentional (*we* believe *p*).[7] Furthermore, most social realities cannot even exist without the intervention of language, for our ability to mention abstract entities in a common language is required in order to participate in the kind of complex collaborative actions, which are the basis of our societies.

Once again, we find PE as the tacit condition of possibility for the theory to advance: if the meaning of a particular utterance, performed by an individual speaker, were not be translatable into the intentionality of the sentence she uttered—which belongs to language, as a socially shared institution—, the transit from individual to social intentionality would not be conceivable. What is more, we would not be able to

know if collective intentionality has even taken place, since the only way to check if we are thinking the same, each one by his side, is the possibility of making it explicit. How could we say what *we* think, if PE were not there, not to assure success itself, but at least its possibility?[8]

From an individualistic point of view, collective intentionality could be rejected as a chimera, as in the last analysis particular individuals seem to be the only existing entities. Nevertheless, strictly considering the logic of intentional attributions, what could be rejected is the priority of the individual over the collective, since collective states seem to precede individual ones: we learn to read reading together, learn to speak speaking together, and even learn to think thinking together. We are eventually able to define the contents of many of our own individual mental states thanks to the fact that we previously acquired the ability to mention them in a collective language, using semantic conventional sentence meanings whose origins are in the social world. From the point of view of the theory, considering its conceptual architecture, collective intentionality is not more controversial than individual intentionality (maybe quite the opposite). In this sense, PE is not only a condition of possibility of communication, but even of thought, at least in its more evolved and specifically human forms.

Finally, PE is not only essential to conceive what we think or say, but also what we *do*, as far as our acts follow a model of rationality built on the idea of a *reason to act* (Searle 2001: 97ff). When we perform acts as rational agents, according to Searle, we choose the reasons we act on, reasons that can be individualised and separately considered by us as propositional intentions-in-action. E.g., if I have to decide whether to accept or not a job offer, the rationality of my decision is built on the possibility of considering one by one the reasons that address me one way or the other (different salary, interest of the job, mobility requirements…): an individualisation of reasons that allows me to weight them and choose sensibly. That specific consideration is possible because we can express the intentional states that support them, such as believes, fears, ambitions, and any other kind of mental attitude. Furthermore, we generally do not consider an agent as a fully rational one, until she is not able to *give* reasons for her actions, and reconsider them critically. Being a rational agent implies a disposition to assume a collective consideration of reasons, and all of this is based on PE, which, from this perspective, is not just a condition of possibility of linguistic expression, but even of rational action itself.

5. Meaning and Background.

Establishment of PE as an analytic *a priori* judgement has turned out to be an essential thesis for the development of Searle's theoretical system, from *parole* to *langue*, from language to mind, from individual to society, and from reasons to actions. This is giving PE much more relevance than what Searle himself would probably expect, who has rarely made any reference to it beyond *Speech Acts*. Is it sensible to challenge it, or even to point out the possibility of its rejection? In the rest of the paper I will try to show that this dangerous step is crucial, if we really want to grasp the scope and intention of the whole theory. It will be my aim in what follows to draw a kind of internal criticism to Searle's approach, since my intention is to fully accept his system in order to question its own coherence from the inside. I will therefore not consider here some external criticism that could be levelled at Searle from inferentialist accounts of mind and language (like Brandom 1998), no matter how promising they might seem even to myself. Restricted to the limits of Searle's own account, I will begin by examining one of its fundamental concepts: the one of *Background*.

According to Searle, philosophy of language is part of the theory of action, for language is just a rule-governed activity. Wittgenstein's influence can clearly be found here, an influence that forces Searle to face the kind of problems that Wittgenstein had to confront himself, as to the notion of rule and the difficulty to determine whether someone is following a particular one or not (e.g. Wittgenstein 1953: §143). The origin of this problem can be found in the confusion between two different kinds of knowledge: knowing-that and knowing-how. We could try to give an account of knowledge of word meanings from both perspectives. On the one hand, the semantic definition of words seems to be analysable in terms of knowing-that: we can tell if someone knows a word by her ability to give a definition. When we ask someone if she knows the meaning of a word or sentence, that is usually the kind of intra-linguistic answer we expect from her: *X* means *Y*. But, on the other hand, if we want to know how that piece of language is effectively related to the environment, we need to know what that person is able to do with them, and that is a matter of knowing-how.

It is well known that Knowing-how cannot easily be translated into knowing-that, since that translation would require an infinite series of explanations and qualifications that, in the last analysis, would require practical, and not merely linguistic

demonstrations.[9] We cannot learn to swim just by reading a book. In the same way, if we want to learn to speak, we have to be trained in the practical ability of linguistic use, a kind of knowledge that cannot be incorporated into dictionaries, nor exhaustively described in the form of explicit conventions. To paraphrase Aristotle once again, *what we have to learn to do, we learn by doing*.

This is what Searle calls *Background*: the set of practical abilities that, not being intentional themselves, allow the connection between language and world. In his opinion, it is a purely naturalistic concept, and there is nothing mysterious or perplexing about it: it is just the result of a set of faculties, which are in human brains and bodies as an effect of our evolutionary history (Searle 1992: chapter 8). It is not possible to understand intentionality, linguistic meaning in particular, neglecting the role of Background; not even literal meaning, since there is no way to determine conditions of satisfaction without the Background (e.g., there's no way to know what is supposed to be true when someone assesses something, or what would satisfy the need of the speaker when she asks something). In this way, the concept of Background prevents Searle from assuming what Recanati (2003: 189) has called *the determination view*, i.e., the idea that meaning completely determines truth conditions (non indexicals directly, and indexicals with respect to a particular context). If we only had in mind the semantic definitions of language in terms of knowing-that, we would not know how to relate words to the world: we have to know how to apply them, and that practical knowledge is the most essential part of what we learn when we learn a language.

Let us consider, e.g., the verb "to cut", one of whose literal meanings is: "to divide into parts with an edged tool".[10] This same meaning, according to Searle, is shared by expressions like "cut the grass" and "cut the cake": they are not two different senses, and none of them is metaphorical. On the contrary, they are both plain cases of the same literal meaning. But, if we asked someone to cut the cake and she answered by mowing it, we would have to consider whether she did not understand what we were *literally* asking her to do, or whether she played us a really bad joke. Knowing what a sentence means, even at the most literal level, is knowing-how to apply it in contexts of use: it is not enough to know the abstract explicit definition. Background itself cannot become a part of the definition, as that would make it just boundless: we would have to delimit and specify each and every possible misunderstanding, and it would be unavoidable to reach a level of not purely linguistic explanations—like deixis, pictures, ostension, or incitation to a certain imitative behaviour. In this sense, what really makes

us grasp the meaning of a word is the huge and increasing set of particular legitimate applications or "source situations" in which the word can be used (see Recanati 2003, Acero-Fernández 2006, and Bejarano-Fernández 2008: 334).

The problem now is how to make two essential aims of Searle's approach compatible: on the one hand, he does not want to be committed to *the determination view*, as Background is for him a *sine qua non* condition to determine linguistic content; but, on the other hand, his intention is to develop speech acts theory as a semantic study of the *langue*. As we have previously seen, it is PE what allows him to carry out this *semantization* of speech acts rules, since it claims that the gap between literal sentence meaning and speaker's utterance meaning can always be closed. But is this a really unproblematic step, if the Background is involved in it?

Searle does think so for, in his opinion, even if it is not possible to make every Background assumption explicit at the same time, nothing prevents us from making each one of them explicit separately. E.g.: to cut a cake you have to use something similar to a knife, not a lawnmower; the knife should face the cake by the cutting edge; pieces resulting should not exceed a certain measurement, etc. No matter how long our succession of qualifications is, there will always be room for new unexpected misunderstandings: what does the similarity with a knife consist in? How are we supposed to calculate that measurement? Can we cut the pieces horizontally, instead of vertically? Can we cut pieces in circles? Can we just cut the surface of the cake?... Obviously, each one of them can be expressed in the form of propositions—and we can *learn that* that is not the case—but what we cannot do is making all of them finally explicit, for that succession would simply be infinite, unpredictable and, which is more important, we would be making a serious conceptual mistake, since Background, in principle, is not propositionally articulated: it is something different from a set of presuppositions, and does not work that way.[11] This misunderstanding of the nature of Background is, according to Searle, in the origin of the puzzlement in some of the most recurrent and well-worn philosophical problems, like the existence of the outer world, or of other minds.[12]

Background can just not be gathered in propositional descriptions, since it cannot be made fully explicit without bringing more Background assumptions. Global expressibility is therefore out of our reach, and there will always be room for an essential 'implicitness' in the use of language: the kind of *chronic insufficiency* Dascal has pointed out (1983: 89). However, this fact does not have to be incompatible with a

certain kind of local expressibility, as has been shown by Recanati (2003: 199): each and every element of the Background can be locally expressed, but they cannot all be expressed at the same time, as they are globally inexpressible.

6. Native, type and token speakers.

Background does not only change the *coloration* of what is being said but even its content, by affecting its conditions of satisfaction. If, as Searle claims, this is not only true with regard to utterance meaning but also to sentence meaning, it seems that we have a problem, since this forces us to question the kind identity PE postulates between sentence and utterance meanings. PE just cannot indentify their respective conditions of satisfaction, due to the fact that sentence is unable to determine any condition of satisfaction whatsoever, *qua* sentence. Recanati makes a similar claim when he says that

> Even if 'sentence meaning' is understood as the meaning of the sentence with respect to contextual assignments of values to indexicals, it is still much more indeterminate, much more susceptible to background phenomena, than the content of the speech act or the content of the expressed psychological state." (Recanati 2003: 202)[13]

Unless we postulate some kind of Background for sentence meaning, the equation PE tries to establish between sentence and utterance meaning will simply be inconsistent, as their elements will not be at the same logical level. In a way, we would be making a sort of categorial mistake, identifying the part—the sentence—with the whole—the utterance.

We need therefore a Background, in order to transform sentence meaning into something that could be identified with utterance meaning. The question is: where is that Background? Or rather: *whose* Background is it? Because Background cannot be in the sentence itself, which is just a piece of code, but in its speaker (just as intrinsic intentionality cannot be in the sentence either, but only in the mind of the person that utters it). And what kind of speaker can we attribute to sentence meaning, prior to its being uttered by anyone?

Searle's intention is to solve this problem by placing heavy methodological reliance on the notion of *native speaker*, as the source of literal meaning intuitions:

The 'justification' I have for my linguistic intuitions as expressed in my linguistic characterizations is simply that I am a native speaker of a certain dialect of English and consequently have mastered the rules of that dialect, which mastery is both partially described by and manifested in my linguistic characterizations of elements of that dialect (Searle, 1969: 13-15).[14]

The concept of a native speaker is interweaved with the one of *fungible* use—both of them extremely interesting concepts, but lacking of rigorous definitions. In Searle's words, "The notion of the literal meaning of a sentence is in a sense the notion of conventional and hence fungible intentionality" (Searle 1979a: 131). According to this, literal meanings of a native speaker's utterances are just transparent, immediate, and fungible to her (i.e., they are consumed in their use, and do not remain as reified instruments she should consciously make use of).

This idea of a fungible use allows Searle to tacitly obviate the difference between speaker and hearer for, as far as they are native speakers of a same language, the only kind of ignorance that they have of each other is about the contents of their future expressions, but not about the conventional, linguistic forms those contents could assume. In a certain way, speaker and hearer—in case of being both native speakers of one *same* language—are identified in Searle's theory as the *same* speaker as, from the point of view of the theory, they would just be *the* native speaker: someone who would *feel like home* in her own literal meanings, and would have fully interiorised her own language, up to the point of being just equivalent to any other speaker. From this perspective, native speaker is almost a kind of *type* speaker. Each and every native speaker of a language would be somehow *the* native speaker.

Searle does not explicitly assure this, but this sort of identification between native and type speakers is inconspicuously assumed in his view. Nevertheless, it is manifestly inconsistent, namely because native speaker is something real, in the flesh, whereas type speaker is just a theoretical abstraction. Native speakers are tokens, not types. Otherwise we would not be able to say platitudes like 'I am a native Spanish speaker'. Just as I can plainly be a Spanish citizen, whereas I just cannot be *the average Spaniard*—even if each and every one of my personal features were exactly in the average—, I can be a *native* Spanish speaker, but could never be a *type* speaker of the

Spanish language. One more time, we would be facing a categorial mistake, or at least a categorial ambiguity.

Once native and type speakers are clearly distinguished for their different logical status, it is easier to find out which of them is the best candidate to get conditions of satisfaction out of sentence meaning, not turning it into mere utterance meaning. If it were a native speaker, that is a token speaker, the difference between sentence and utterance meaning would be blurred, and we would probably be relying too much on the abilities of a particular individual, as I will try to show in the next section. However, if we consider type speaker as the one whose Background is at stake when we are trying to understand the conditions of satisfaction that PE requires from sentence meaning, its logical status would preserve the sentence from just becoming an utterance. Type speaker is simply as abstract as the sentence itself is: it just defines the class of all token speakers that would attribute the same conditions of satisfaction to the same sentences under similar situations—i.e, the class of individuals who share a language.

The problem now is that Searle's uncritical identification of all speakers of a language in the notion of *the* native speaker calls for further justifications. Whenever we face effective speech act situations, we always have to acknowledge that both speaker and listener are different token individuals, and we cannot simply assume anymore that they fall under the same type *just because*. Justifications for that statement will prove to be rather elusive.

By appealing to the intuition of the native speaker, what Searle vindicates is the importance of first person perspective in order to explain meaning and understanding. It is ultimately there where all the system bears out: in the fact that I know what I mean when I say what I say. But first person authority has certain limitations, and one of them is particularly fundamental here: first person authority does not extend to the formulation of explicit rules. Being a native speaker of English is something very different from knowing by heart the definitions of an English dictionary, or the rules of the English grammar. There is nothing contradictory in the idea of a competent native speaker unable to give any explicit definition of the words she uses, nor of the grammatical rules she follows. Now the point is that the fungible character of native speaker's utterances stems from the knowing-how of her habits, not from the knowing-that of explicit linguistic conventions. As Searle himself pointed out in a brilliant passage (1995, 140ff), speakers do not have to know explicitly the rules they follow in

order to follow them. What a native speaker is quite sure to know is just *the convenience of each particular application*. Rules can be inferred from practice, but first person authority arrives to this inference notoriously weakened.

Even if explicit rules are usually not the cause of linguistic behaviour, their formulation plays an essential role in the establishment of shared meanings. We can check if our definitions are common by formulating them explicitly, expecting the acceptance of others. The problem is that we cannot check in that way whether Background is being shared: Recanati's local PE would allow us to check specific parts of Background, maybe each one of them, but not at the same time. And that is not a method to check the commonality of Background: it would be like believing, as in Augustin's famous legend, that we are perfectly able to empty the sea due to the fact that, firstly, we can fill a shell with sea water and throw it out of the sea and, secondly, there is no part of water in the sea that could not fit into that shell. We can check if we share definitions, but there is no way to check if we share Backgrounds, which cannot be made explicit by any definite set of conventions. It is rather a kind of fundamental assumption, which makes the acceptance of any convention possible. In this respect, we can establish definitions by convention, but we cannot establish Background by convention, since Background is the basic assumption of any convention.

We cannot prove that Background is being shared: we can just expect it to be so. And that is something that cannot be verified but only—in Popper's terms—*falsified*, when we notice that two speakers, that seemed to be willing to accept a common definition of a sentence, are in fact conferring to it different conditions of satisfaction (see Dascal 1983: 90). Only the commonality of explicit conventions, but not of Backgrounds, can be simply verified.

7. Saying more (or less) than what we mean to say.

I will now use Searle's well-known example of the dependence of literal meaning on Background in order to criticise the idea that Background can simply be *shared*. The original example is as follows:

(a)   I am sitting in a restaurant, and order a hamburger. The waiter serves it to me encased in a block of concrete (or sends it to my apartment, or serves a 580 Kg. one…).

According to Searle, such a response would not satisfy the conditions of satisfaction of my petition, as Background assumptions would not have been observed. The waiter would not have served what I *literally* ordered, just like the joker who mowed the cake did not do what I literally asked him to do, when I asked him to *cut* it. The argument seems to be convincing, but let's look at the following counterexample:

(b)   I am sitting in a restaurant, and order a hamburger. The waiter serves a pizza to me, encased in a block of concrete (or sends a pizza to my apartment, or serves a 4 meters in diameter one…).

If Searle were right, and outlandish interpretations were simply excluded of the literal meaning of the sentence, (a) and (b) would simply be two cases of the same kind of misinterpretation. But it seems to be clear that they are not so: there is something in (a) which is not present in (b): the hamburger, i.e., *what I ordered*. In (a) the waiter served it to me in quite an unusual way, but he *did* serve it, whereas in (b) he just did not. The reason why (a) is a meaningful example—whereas (b) just makes no sense at all—is that it is logically possible to include what the waiter served in the literal meaning of the petition, even if it infringes the conditions of satisfaction expected by the speaker, and his specific Background assumptions.

Since it was Searle who offered the first disconcerting example, I beg the reader to let me add a new one: imagine we are sitting in an inter-planetary restaurant, where food must be served encased in concrete, in order to preserve it from long space trips. The very simple question now is: would that excess of imagination also force us to imagine a different language? Would the words "I would like to order a hamburger" *mean* something different? Would I be able to order a hamburger there in English, in the very same English language? I think the answer is quite plain: yes I would. I simply would be able to order that hamburger with the same sentence, in the same language and with the same literal meaning, but my particular conditions of satisfaction—if I knew this specific need of space trips—would be different, and would probably not include hamburgers served in ordinary dishes. In that imagined situation, (1) would

perfectly make sense, even under the supposition that the English language is exactly the same. On the contrary, in order to imagine a context for (2) that could transform it into a meaningful situation, we would be forced to change the very conventions of the English language.

The moral of the story is that each time a speaker makes an utterance, even if she is a *native* speaker, she can only consider the bounds and limits of her own Background expectations, but cannot foresee what the community of native speakers of her language would accept or not as a literal interpretation of her sentence. No native token speaker can know *a priori* the Background of the *type* speaker: a character that, as a kind of linguistic Leviathan, emerges from the global community of speakers, herself included. She cannot know what English native speakers would accept as a literal satisfaction of her speech act in a different situation, or even in her own particular situation, but considered from a different perspective.

By my lights, Searle is right in rejecting *the determination view*, and holding that there are no conditions of satisfaction in sentences themselves: they *do* require Backgrounds in order to be really meaningful, for otherwise their connexions with beliefs, perceptions, and other intentional states would be lost (Searle 1979a: 136). But it cannot be denied that the limits of literal meaning—meaning of sentences, not of utterances—cannot depend on the particular Background of a token speaker, no matter how native she is. Sentences can mean much more than what each speaker is able to foresee; the literality of their meaning cannot be as limited as her specific expectations.

At first glance, this idea seems to be demolishing for PE: it just seems that, each and every time the speaker makes an utterance, her sentence says more (or less) than what she means to say with it. Summarised, the argument is as follows: (1) the conditions of satisfaction of the sentence used cannot be determined by explicit semantic conventions, since Background assumptions are requested; (2) the Background that should be considered in order to determine the conditions of satisfaction of the sentence *qua* sentence is not the one of the token speaker, native or not, but the one of the type speaker; (3) we cannot know in principle the degree to which token speaker's and type speaker's Backgrounds are shared, since all we can do is trust their commonality, as far as it has not been falsified yet; (4) by using the sentence, the speaker is obliged to accept the conditions of satisfaction that would result from the type speaker's Background assumptions. In conclusion, (5) each time that a speaker

produces an utterance, the sentence she utters means more (or less) than what she means to say with it (*more* because, as a sentence, it is also satisfied by literal interpretations which are unexpected by the particular speaker; and *less* because, no matter how much she tries to limit those misunderstandings, conditions of satisfaction will always remain underdetermined, since Background is globally inexpressible).

Of course, in ordinary life, Background is so widely shared, or so it seems, that we usually get what we want in restaurants, and this intrinsic limitation of expressibility does not generally affect us. However, language theoreticians should not neglect it.

8. Expressibility reconsidered.

With the previous argument, I was not intending to prove what Searle repeatedly tries to deny: that PE can fail in particularly marginal or irregular occasions, as there are thoughts that are intrinsically inexpressible. It has not been my aim to prove that there is something in the outer limits of ordinary language that cannot be said, whether in art, religion or other kinds of mysticism. I was rather trying to emphasise, as Scharfstein points out, that "there is something unformulated and perhaps unformulable and finally mysterious in the prosaic, everyday successes and failures of words" (Scharfstein 1993: xvii). The reason why this might happen is that, in each and every use of language, there is certain incommensurability between sentence and utterance meanings. Not being susceptible of propositional expression, the intervention of Background seems to prevent the success of expressibility in general—which would not make PE just false, in particular occasions, but impossible as a rule. It would be completely impossible simply to identify what the speaker means with the meaning of the sentence she uses to say it.[15]

Nevertheless, as I argued at the beginning, I am not trying to falsify a principle that, being analytic, cannot be directly confronted to facts: PE is not a particular thesis that could simply be true or false, but a methodological strategy that makes Searle's approach possible, an *a priori* that provides the backbone for research in different philosophical fields. In this light, we would do well to rethink PE not as a kind of *discovery* about the working of language—as Searle seems to present it—but as a sort of methodological *invention*, and one of a kind that it would not be easy at all to go on without it, as I will try to show now.

Let's imagine I simply reject the validity of PE, and perform a speech act strictly considering that there is the kind of inadequacy between sentence meaning and utterance meaning I have just pointed out. In that situation, I utter the sentence:

(1)   The cat is on the mat.

My utterance does have a specific intentional meaning because I am able to determine its conditions of satisfaction. As a native speaker, I have interiorised its rules of use, which are *fungible* to me when I make them work with *my* Background. In that way, I know what *I* mean when I say "cat", "on" and "mat". However, if by uttering (1) I were just saying this:

(2)   What I mean by the word "cat" is in the kind of relationship I mean by the word "on" with regard to what I mean by the word "mat",

I would not be communicating anything at all. I would definitely not be communicating anything about the world, since the hearer would not know what aspects of reality I am talking about, but it would be a mistake to believe that I would be communicating something about my own mind, for the hearer would not know what I am talking about either—she would have no idea what my "beetle in the box" is (Wittgenstein 1953: §293). If she is able to identify what I mean by those words, that is because their meanings are not only mine, but ours; i.e., their content is able to assume the shape of collective intentionality, and therefore be named in our common language. But then, when I say (1) I seem to be saying something similar to this:

(3)   What a type speaker means by the word "cat" is in the kind of relationship that a type speaker means by the word "on" with regard to what a type speaker means by the word "mat".

That is to say: if I am effectively performing a communicative speech act, the words that I use do not only mean what I mean by them, but what *we* mean by them. But, by the same token, the meanings of *my* words are not determined by *my* Background, but by the one of the type speaker. Now, if I reject PE, and do not accept that it is always possible to identify native and type speaker's Backgrounds, I seem to

be lead to a dilemma: either I am encased in a private language, like in (2), or I will just never be able either to say what I mean nor to mean what I say, loosing control of my own utterances, as in (3).

It is obvious that a simple rejection of PE would be quite disturbing for social philosophy, since the commonality of content would be unconceivable; but it would be a mistake to believe that those devastating consequences would be restricted to the collective aspect of the problem. We would also have to pay an excessively high price in philosophy of mind if we just decided to give up PE, since that would prevent us from studying mental intentional states through propositional attitudes. If individual thoughts were not expressible in sentences, we would simply not be able to articulate mind from language, not even from the individual perspective. The problem would not be restrained to the issue of communication, because language does also accomplish an essential role in the development of our cognitive *inner* abilities. In other words: it is not only a matter of identifying utterance and sentence meanings, but also speaker's and thinker's ones. What do the words that I use *in my mind* mean, not when I utter them on the outside, but when I use them in *inner speech*, putting my thoughts in order for myself? If they only mean what I want them to mean, (2), how can I say what I think? And if they only mean what the type speaker means by them, (3), how do I know what I think? The attribution of propositional attitudes, not only to others but also to oneself, would be extremely problematic.

The possible identity of native and type speakers' Backgrounds seems to be unavoidable, as it is a necessary condition for me to be able to tell, even to myself, what I am thinking—and maybe even to think it. But we should be aware that this is only a presupposition, and the kind of presupposition that cannot be checked to be true, but only falsified in particular occasions. I do not think then that we should just reject PE, but I think we should acknowledge that the equation it tries to postulate will always remain suspicious, since it cannot avoid the possibility of being affected by semantic excesses or deficits.[16] In that sense, the opposite alternative should be acknowledged as equally possible: a sort of Principle of Expression Fallibility (PEF) that would be as follows:

For any Expression E and any speaker S whenever S utters E as an exact expression or formulation of meaning X, it is possible for E to mean something different from what S means (intends to transmit or communicate) by uttering it.

Once again, it would be a *de iure* principle, not a *de facto* one: if PE claims that it is always possible to say what you mean, PEF claims that this expression could always fail, and is essentially subject to further revisions. PE and PEF are not incompatible principles, since both are modal judgements of possibility that do not exclude each other.[17] On the contrary, they complement each other for, when we do not find a particular expression, PE assures that it *can* exist whereas, when we think we have found it, what PEF assures that we can be wrong. If we do not find it, it could exist; if we think we found it, it could not be.

Under this light, expressibility should be regarded as a process, instead of as a product: the kind of task that could not be definitely accomplished, but must essentially remain in progress, and always to be done. At this point it could be alleged—as Kannetzki (2001b)—that PE should just be abandoned, and substituted by a sort of "principle of explicability", but this step would be inadvisable at least for two reasons: in the first place, because without PE—as I have tried to show in the first part of the paper—the backbone that articulates speech acts, mental intentional states, social collective institutions and reasons for action would be broken, and that would certainly shake the foundations of the whole system. And, in the second place, because it is not clear if explicability would be understandable at all if we didn't keep expressibility as a kind of regulative principle—for the same reasons that Popper did not abandon the notion of truth, even if falsability became unpreventable.[18]

In conclusion: if PE allows a rational approach to language, mind and action, PEF limits the scope of this approach, preserving the elusiveness of human rationality, and advocating us to confront in the theory the unavoidable indefiniteness that constitutes, in fact, processes of communication, expression, translation, or even understanding, that are, and will always be, matters of degree. The acceptance of the tandem PE/PEF would help us to grasp the peculiar character of speech acts, and to be aware of the limits of systematisation in a scientific approach to the nature of thought and language.[19]

Notes

1. For this reason, I find Searle's defence of expressibility more challenging than Katz's one, as Katz seems to consider his "Principle of Effability" as an empirical truth, alleging observational evidences for it, and expecting for empirical corroboration (Katz 1972: 18-23).

2. There is some resemblance between Searle's PE and a central idea of Wittgenstein's *Tractatus Logico-Philosophicus*: that "What can be said at all can be said clearly" (Preface). Nevertheless, even if Wittgenstein's idea of *clarity* were the same as Searle's idea of *literality*—which is not the case—, an important difference should be pointed out: Wittgenstein assures that what *can* be said, can be said clearly, whereas Searle's point is that whatever the speaker intends to say, can be said. Searle's claim is stronger than Wittgenstein's, since the former assumes that everything can be said, whereas the latter only claims that what can be said, can be said in a particular way. The author of the *Tractatus* would probably not accept PE, as it would abolish the distinction between *saying* [sagen] and *showing* [zeigen] (4.1212): for him, there is something we cannot say, something we can only show, the mystical, "Whereof one can not speak" (7). And obviously, in his opinion, it is possible to intend to communicate what cannot be said, since he recommends us not to do it.

3. Most griceans would probably not accept this dependence, since it would imply a kind of circularity in the analysis of meaning—unless we assume that analysis must not be reductive to be enlightening (see Avramides 1989). Furthermore, Grice does not restrict communication to the utterance of sentences.

4. Criticism of literal meaning is sometimes directed towards its empirical lack of evidence, as people in fact do not seem to go across literal meanings in order to grasp utterance meanings; see for instance the Dascal-Gibbs debate (Dascal 1987, 1989 and Gibbs 1989, 2002). In any case, what is at stake here is not the psychological role accomplished by literal meaning, but the logical necessity of literality in order to provide a philosophical account of speech acts. In fact, many of those who explicitly criticise literality from an empirical point of view, offer argumentations that are logically dependent on it. Once again, analytical principles seem to be immune to empirical approaches.

5. According to Searle, mental intentionality seems to be logically previous to its expression in a shared language, which leads to a puzzling dilemma (Kannetzky 2001a): how should we consider this priority, other than by the postulation of a private language? And how are we supposed to establish its connections and translations with public languages? If we deny the existence of private languages, the development and enlargement of language seems to be a mystery (where do new meanings come from?). But if we accept that they do exist, and we therefore consider the speaker's intention as something given before it is expressed in a public language, what we are unable to explain is mutual comprehension. Kannetzky is right when pointing out the relationship of this puzzle with PE, but he does not seem to notice that it is precisely Searle's intention to avoid the appearance of the dilemma *a priori*. Blinkey (1979) also

considers PE incoherent by its pretensions to compare expressed meanings of utterances with unexpressed meanings, which would not be identifiable on their own.

6. In Scharfstein's words (1993: 11), "we generate ideas from a substratum that is opaque to consciousness and resists transmutation into concepts. It is true that part of our more abstract thinking may be done in snatches of inner speech, but external speech is not ordinarily generated from preceding speech of any kind".

7. The refusal to introduce collective intentions in gricean analyses of meaning seems to be the reason why they cannot avoid to keep a kind of *vanishing point*, and always seems to leave us "waiting for the next ingenious counter-example" (Avramides 1999: 73).

8. I am not saying that only literal speech acts do lead to collective intentionality. On the contrary: metaphors, ironies, or even non-conventionalised ways of communication are perfectly able to fulfil that task. My point is that, whether in conventionalised or non-conventionalised situations, at least from the point of view of the analyst, literality must be acknowledged as a possibility, since that is the only final move which could make the collectivization of content explicit: the possibility of *literalizing* what is being said and, in that way, discarding any ambiguity in the utterance. In this sense, Grice's analysis of meaning (Grice 1957), which attempts to give an account of the rise of literality from non literal uses of language, might also presuppose PE, for it accepts that, whatever the speaker intends to communicate (literally or not), it can assume the shape of a particular intention that can be expressed by the first clause of the *analysans*—a clause expected to be *literally understood* by the community of scholars who read his paper. Even if literality is absent from the *analysandum*, and we can get rid of the code-model in order to explain some communicative acts, Grice has to accept it as a precondition for the use of the meta-language that paraphrases it. From this point of view, code-models and inference-models of communication are not so opposed as it was originally thought, a point that facilitates their articulation in complex theories of communication—as Sperber and Wilson (1986: chapter 1.5) or Dascal (2003: 507-520) have pointed out. The reason why they are not incommensurable might be that they do share a basic common assumption: that expressibility is, in the end, at least possible.

9. For a seminal view on this issue, see Ryle (1949, chapter 2). Brandom (1994) has taken this as a key point in his account of explicitness.

10. Meaning 3.b of the *Merriam-Webster Online Dictionary* (www.merriam-webster.com), last accessed January 30, 2009.

11. The idea of Background should be complemented with the one of Network: intentional states are not completely independent from each other; on the contrary, many of them require other intentional states in order to be effective. In this way, most beliefs and desires are articulated in a Network of other beliefs and desires (Searle 1983: 65). But I will not focus on this idea here because, unlike Background, Network

is susceptible of being exhaustively studied in a propositional way and, for that reason, is a much weaker challenge for PE.

12. Although Background is a crucial concept in Searle's account of intentionality, its lack of development in his work is manifest. Current studies on enactivism and non-propositional attitudes might through light on this issue (see Clark 1997, Hutto 2006 or Gallagher 2006), even if they are generally not part of representationalist programs, since they might help to understand in which way intentional propositional attitudes are rooted in bodily predispositions to interact with the environment.

13. Both in Recanati's argument and here, it is the inconsistent identity between sentence meaning, content of the speech act, and content of the intentional state, what is at stake. Nonetheless, I will move away from Recanati's argument from here on, since he makes no reference at all to a notion that will be crucial in what follows: the one of a *type speaker*.

14. The difference between Searle's and Austin's methods for the study of speech acts is quite suggestive; Austin does not rely so much in the intuition of a particular individual speaker, as in the possibility of an inter-subjective contrast between different linguistic intuitions (Austin 1956).

15. What makes this argument different from other famous arguments that also intend to undermine the solidity of meaning is its reliance in the notion of Background. E.g., Hillary Putnam's *twin earth* argument forces us to accept that meanings rely in the relationship between the community of speakers (and not only the particular speaker) with the physical world, since we delegate in specialists the meanings of the words we use (Putnam 1975). But nothing leads to inexpressibility there, since this delegation could result in a set of finite rules of use, all of them propositionally expressible. Besides, Quine's *gavagai* argument (Quine 1960) shows the difficulty of really grasping the content of an utterance performed in a foreign language with no connection to our own. However, in the last analysis Quine's argument relies in a behaviouristic scepticism about intensionality and the first person perspective. The argument here advanced, on the contrary, does accept this perspective but, as it has previously been shown, first person authority does not solve the puzzle of expressibility, since it is limited to the fungible and concrete application of the rule, and does reach neither the formulation of the rule itself, nor the scope of the type speaker's Background. Mainly because, it might not be useless to point it out, first person authority does not apply to the first person *in the plural*.

16. "Searle […] ne peut pas exclure (et cela dans sa théorie) la *possibilité structurelle* de ce que l'intention et l'expression ne se recoupent pas tout simplement, sans parler de la possibilité qu'elles coïncident. Il doit cependant insister sur leur recouvrement entier s'il ne veut pas renoncer à la possibilité de maîtriser les intentions par la convention et par conséquent, il doit renoncer au modèle du code du langage." (Frank 1984: 403).

17. In this sense, PEF is not a principle of ineffability at all, like the one Buyssens seemed to defend when he claimed that language, as a matter of principle, is unfit to express thought (see Dascal 1983: 54). I would like to thank an anonymous referee for suggesting me to reformulate this principle in a less misleading way.

18. Both reasons could of course be questioned, but this would be the aim of the kind of external criticism I briefly pointed out in section 6. As I said there, I do not deny that this might be an interesting way to follow, but I would say that the kind of criticism of PE I have tried to level—which is internal to Searle's position—would be deflected by accepting PEF.

19. An earlier version of this paper was presented at the XXVI AESLA International Conference, Almería, April 2008. I would like to thank Teresa Bejarano-Fernández, Pedro Chamizo-Domínguez, Federico Rodríguez Gómez, Margarita Planelles Almeida, Carlos Thiebaut, José Medina, two anonymous referees and Marcelo Dascal for helpful discussion and comments on its previous drafts. The idea to write this paper stemmed from my reading of the famous quarrel that took place between Searle and Derrida (see Derrida 1977, Searle 1977 and Searle 1994), even if none of them seemed to consider that neither PE nor the Background was at stake in their discussion. I focus on this episode in Navarro-Reyes (2009, forthcoming).

References.

Acero-Fernández, J.J. 2006. "Searle y el significado literal". *Revista de filosofía* 31:2, 9-30.

Austin, J.L. 1956. "A plea for excuses". In Austin, J.L. 1961. *Philosophical Papers*, Oxford: Oxford University Press, 175-204.

Avramides, A. 1989. *Meaning and Mind: an Examination of a Gricean Account of Language*. Cambridge, MA: The MIT Press.

Avramides, A. 1999. "Intention and convention". In Hale, B. and Wright, C. (editors): *A Companion to the Philosophy of Language*. Oxford: Blackwell Publishing, 60-86.

Bejarano-Fernández, T. 2008. "*The Construction of Social Reality*, de John Searle", *Thémata* 39, 331-334.

Binkley, T. 1979. "The principle of expressibility". *Philosophy and Phenomenological Research* 39.3, 307-325.

Brandom, R.B. 1994. *Making It Explicit. Reasoning, Representing, and Discursive Commitment*. Cambridge, MA/London: Harvard University Press.

Clark, A. 1997. *Being there: putting brain, body, and word together again*. Cambridge, MA: MIT Press.

Dascal, M. 1983. *Pragmatics and the Philosophy of Mind I. Thought in Language.* Amsterdam/Philadelphia: John Benjamins Publishing Co.

Dascal, M. 1987. "Defending literal meaning". *Cognitive Science* 11, 259-281.

Dascal, M. 1989. "On the roles of context and literal meaning in understanding". *Cognitive Science* 13, 253-257.

Dennett, D. 1987. "True Believers". In *The Intentional Stance*. Cambridge, MA: MIT Press.

Derrida, J. 1977. "Signature, event, context". *Glyph* I, 172-197.

Derrida, J. 1988. *Limited Inc*. Evanston, IL: Northwestern University Press.

Frank, M. 1984. "La loi du langage et l'anarchie du sens. À propos du débat Searle-Derrida". *Revue Internationale de Philosophie* 151:38, 396-421.

Gallagher, S. 2006. *How the body shapes the mind.* Oxford: Oxford University Press.

Gibbs, R. 1989. "Understanding and literal meaning". *Cognitive Science* 13, 243-251.

Gibbs, R. 2002. "Marcelo Dascal and the literal meaning debates". *Manuscrito* 25:2, 199-224.

Grice, H.P. 1957. "Meaning". *The Philosophical Review* 66, 377-388.

Hutto, D. D. 2006. "Unprincipled Engagements. Emotional Experience, Expression and Response", in Menary, R. (ed.): *Radical Enactivism: Intentionality, Phenomenology, and Narrative: Focus on the Philosophy of Daniel D. Hutto.* Amsterdam/Philadelphia: John Benjamins Publishing Co, 13-38.

Kannetzky, F. 2001a. "The Principle of Expressibility and Private Language". *Acta Philosophica Fennica* 69, 191-212.

Kannetzky, F. 2001b. "Expressibility, explicability, and taxonomy. Some remarks on the principle of expressibility". In: G. Grewendorf & G. Meggle (eds.): *Speech Acts, Mind, and Social Reality. Discussions with John R. Searle (Studies in Linguistics and Philosophy).* Dordrecht: Kluwer.

Katz, J. J. 1972. *Semantic Theory*. New York: Harper & Row.

Navarro-Reyes, J. 2009 (forthcoming). *Cómo hacer filosofía con palabras. A propósito del desencuentro entre Searle y Derrida*. Madrid: Fondo de Cultura Económica.

Putnam, H. 1975. "The Meaning of 'Meaning'. In *Philosophical Papers: Volume 2. Mind, Language and Reality*. Cambridge: Cambridge University Press, 215-271.

Putnam, H. 1981. "Brains in a Vat". in *Reason, Truth, and History*. Cambridge: Cambridge University Press.

Quine, W.v.O. 1950. *Methods of Logic*. New York: Holt.

Quine, W.v.O. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Recanati, F. 2003. "The limits of expressibility". En Smith, B. (ed.), *John Searle*. Cambridge: Cambridge University Press, 189-213.

Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson.

Sáez Rueda, L. 1996. "Segregación o domesticación de la experiencia prerreflexiva". *Volubilis* 4, 35-53.

Scharfstein, B.-A. 1993. *Ineffability. The Failure of Words in Philosophy and Religion*. New York: SUNY Press.

Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.

Searle, J. 1977. "Reiterating the differences: a reply to Derrida". *Glyph* I, 198-208.

Searle, J. 1979a. *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Searle, J. 1979b. "Intentionality and the use of language". In Margalit, A. (ed.): *Meaning and Use*, Jerusalem: The Magnes Press, 181-197.

Searle, J. 1980. "The Background of meaning". In *Speech Acts Theory and Pragmatics*, edited by Searle, Kiefer & Wierwisch. Dordrecht: B. D. Reidel.

Searle, J. 1983. *Intentionality. An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Searle, J. 1994. "Literary theory and its discontents". *New Literary History* 25:3, 637-667.

Searle, J. 1995. *The Construction of Social Reality*. New York: The Free Press.

Searle, J. 2001. *Rationality in Action*. Cambridge, MA: MIT Press.

Strawson, Peter F. 1950. "On referring". *Mind* 59:235, 320-344.

Wittgenstein, L. 1921. *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Basil Blackwell.

Author's address

Jesús Navarro-Reyes

Departamento de Metafísica y Corrientes Actuales de la Filosofía,

Universidad de Sevilla

C/Camilo José Cela, s/n

41018 Sevilla

Spain

Email: jnr@us.es

http://personal.us.es/jnr

About the author

Dr. Jesús Navarro-Reyes teaches Contemporary Philosophy at the University of Seville (Spain). He has been Post-Doctoral Fellow at the University of California-Berkeley and Visiting Scholar at Duke University. He is author of two books and various articles on Montaigne's scepticism, philosophy of subjectivity, mind and language. A forthcoming book is entitled *How to Do Philosophy with Words: Reflexions on the Searle-Derrida Debate*.