



Semantic innocence and substitutivity

Paul Egré

► **To cite this version:**

| Paul Egré. Semantic innocence and substitutivity. 2003. <ijn_00000407>

HAL Id: ijn_00000407

https://jeannicod.ccsd.cnrs.fr/ijn_00000407

Submitted on 21 Oct 2003

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semantic Innocence and Substitutivity

Paul Égré*

IHPST- Université Paris 1

In “Opacity and the Attitudes” and in his book *Oratio Recta, Oratio Obliqua*, Recanati presents a semantic theory of belief reports that rests fundamentally on the principle of *semantic innocence*. The principle of semantic innocence, first stated by Davidson and promoted by Barwise & Perry, states that the semantic value of a referential expression ought to remain constant inside and outside the scope of a verb of attitude like “believe”. On the basis of this principle, Recanati is committed to a strong version of the principle of substitutivity, stated as a general principle of intensional replacement (Recanati 2000b: 41): two genuine singular terms (proper nouns, indexicals) having the same referential content, like two predicates expressing the same concept, or two sentences expressing the same proposition, can be substituted to each other *salva veritate* in any complex sentence. As Recanati admits, this principle cannot apply unrestrictedly, due to well-known examples of failures of substitutivity in belief contexts, but for Recanati these failures of application of the principle are to be accounted for at the level of pragmatics. A simplified Gricean account of this division of labor between semantics and pragmatics would be to say that contexts of attitude are “semantically transparent”, but “pragmatically opaque”. The picture suggested by Recanati is more subtle however, for Recanati thinks that the analysis of attitude reports is part of a program of “truth-conditional pragmatics”, where the pragmatic components affect the truth-conditions of sentences directly, and not secondarily. The object of this paper is to evaluate the merits and limits of this account of opacity phenomena, and to determine to what form of a division between semantics and pragmatics the principle of semantic innocence commits us. This requires a precise analysis of this notion of semantic innocence. Fundamentally, semantic innocence ought to mean *semantic constancy* (in a sense to be made precise), but under a certain conception of meaning, it means *semantic constancy + direct reference*. My claim is that the second construal is sometimes interpreted too narrowly, and that only the intuition of semantic constancy ought to be preserved. In the first section of the paper, I review some background on opacity and contrast Recanati’s pragmatic account with two other major approaches of substitutivity in belief contexts, roughly the “fine-grained meaning” approach and a Hintikka-type analysis. I argue that both of these are compatible with semantic innocence as semantic constancy, but moreover that this sense of the notion of semantic innocence is probably the only one we need to preserve Recanati’s intuitions about substitutivity, as discussed in the second section of the paper.

1. Semantic innocence and the opacity of belief contexts

Contexts of attitude are not only intensional contexts, namely contexts in which two expressions with the same extension cannot be substituted *salva veritate*, but they are also hyperintensional, to use an expression coined by Cresswell (1973): two expressions with the same intension cannot always be substituted *salva veritate* under the scope of a verb of

* Many thanks to F. Recanati, P. Schlenker and B. Spector for helpful comments.

attitude. On standard accounts, two singular terms (proper names or indexicals) have the same intension if and only if they denote the same individual in all possible circumstances; likewise, two predicates are cointensional if and only if they express the same property, that is if they denote the same set of individuals under all possible circumstances; and two sentences are cointensional if and only if they express the same proposition, if and only if they are necessarily true together or false together, namely if they are logically equivalent.

Rather than going into the formal definition of the semantic concepts involved in these definitions, it is simpler to consider examples.¹ The proper names “Cicero” and “Tully” both refer to the individual Cicero, and following the analysis of Kripke, it is assumed that they refer to the same individual in all the counterfactual worlds that we can think of, and even across these worlds (by the *rigidity* thesis). Likewise, in English the predicates “eye-doctor” and “ophthalmologist” pick the same extension in all the possible worlds (although not necessarily across the possible worlds). Finally, to use an example of Muskens², the two sentences “if door A is locked, then door B is not locked”, and its contrapositive “if door B is locked, door A is not locked”, are logically equivalent: on the assumption that the conditional is a material conditional in both cases, they are necessarily true together or false together. However we can find belief contexts in which these expressions cannot be substituted to each other *salva veritate*. It seems that we can consistently say:

- (1) Peter believes that Cicero was rich, but he doesn't believe that Tully was rich
- (2) Peter believes that John is an eye-doctor, but he doesn't believe that John is an ophthalmologist
- (3) Peter believes that if door A is locked, door B is not locked, but he doesn't believe that if door B is locked, door A is not locked

For (1), it suffices to imagine that Peter conceives of Cicero and Tully as two distinct individuals. For (2), we may imagine a situation where John takes ophthalmologists to be eye-doctors with miraculous powers, but doesn't think that all eye-doctors, including John, have this extraordinary power. (3), arguably, holds in a situation in which John has been told and accepts the instruction that if door A is locked, door B is open, but momentarily believes that A could be locked as he tries B first and finds it locked.

Many attempts have been made to handle these failures of substitutivity within belief contexts. Here I shall restrict myself to the consideration of three main options, in order to contrast their relation to the notion of semantic innocence and to locate precisely the conception defended by Recanati. I should point out that these options are not exclusive of each other: they do correspond to different strategies at the conceptual level, but they happen to share some representatives.

Let us start with an important preliminary. The notion of intension, as defined above, has two sides. On one side it is tied to the notion of *referential content* of an expression: two expressions have the same intension if and only if they have the same reference in all the possible worlds.³ On the other side it is intended to capture the meaning, semantic value, or

¹ See Gallin (1975) for a classic exposition of intensional logic.

² Cf. Muskens (1991), who adapts it from R.C. Moore.

³ By the *referential content* of an expression, I really mean, here and elsewhere in the text, its *objective* referential content, as opposed to the subjective reference a speaker might mistakenly associate with it. Identity

truth-conditional contribution of an expression. What the above examples suggest is that expressions with the same intension or referential content may not have the same truth-conditional contribution within belief contexts. It seems therefore that the meaning of an expression cannot be reduced to its intension, as traditionally understood.⁴

(i) One option is therefore to refine the notion of intension, to find a level of granularity of meaning that would be sufficient to give “Tully” and “Cicero”, or likewise “eye-doctor” and “ophthalmologist”, or “if door A is locked, door B is not locked” and its contrapositive, distinct semantic values, even outside of belief contexts.

(ii) A second option is to maintain the usual notion of meaning as referential content, but to suppose that the semantic value of an expression shifts under the scope of an attitude verb like ‘believe’.

(iii) A third option, finally, is to hold on to the notion of meaning as referential content, but this time to contest that the semantic value of an expression is shifted under the scope of belief operators. Instead, their introduction under the scope of a belief verb triggers specific pragmatic effects.

The first option is the more radical. On this conception, it is simply misguided to assume, for instance, that the semantic contribution of two coreferential proper names, even outside of attitude contexts, is simply their referential content. The two sentences “Cicero was rich” and “Tully was rich” simply don't have the same meaning, and it should come as a direct prediction that they cannot be substituted to each other *salva veritate* in belief contexts. On this view, the morphology of expressions is part and parcel of their customary meaning. Disregarding details and important theoretical differences, I would include among the main representatives of this view the advocates of syntactic meaning (Quine 1960), structured meanings (Carnap 1947, Lewis 1970, Cresswell 1985), ‘hyperfine-grained’ meanings (Muskens 1991, Church 1989), but also of subjective or informational meaning (Stalnaker 1987, Haas-Spohn 1994).

The second option differs from the first mainly on the idea that in contexts free of attitude operators, the semantic value of an expression is its objective referential content. On this view, the simple sentences “Tully was rich” and “Cicero was rich” do have the same meaning, and the semantic contribution of the proper nouns “Tully” and “Cicero” is their common reference in the actual world. But belief contexts are special, in them the semantic value of “Tully” and “Cicero” is not necessarily their ordinary referential content. In Hintikka-type epistemic logic, for instance, the referential content of “Tully” and “Cicero” is the same in the actual world and in the counterfactual worlds that are relative to the actual world, but it may vary with respect to the epistemic worlds of the agent of the ascribed belief.

On the third option, the semantic contribution of the proper nouns “Tully” and “Cicero” is simply their reference in the actual world, and the sentences “Tully was rich” and “Cicero was rich” necessarily have the same meaning. But this time, the semantic value of “Tully” and “Cicero” remains their ordinary reference even when the sentences are embedded under an attitude verb. What happens in cases of substitutivity failure is a pragmatic phenomenon of the nature of a conversational implicature. This corresponds to what Recanati (1993) calls the

of referential content can sometimes simply mean identity of extension (actual reference), but it is primarily captured by the notion of identity of intension (coreference in all the possible worlds).

⁴ At least, as Lewis (1970) writes: “intensions are part of the way to meaning”, cf. Gallin (1975: 9).

Implicature Theory of belief reports. Recanati considers himself a representative of this conception, although with some caveats.

In his writings, Recanati tends to distinguish two options instead of three, however. According to him, we ought to distinguish accounts of belief sentences that rest on a principle of “semantic deviance”, and accounts that preserve semantic innocence. Option (ii) clearly exemplifies what Recanati understands by the notion of semantic deviance. Thus Recanati writes (Recanati 2000a: 395):

According to Hintikka (1962: 138-141), failures of substitutivity in belief contexts show that two co-referential singular terms, though they pick out the same individual in the actual world, may refer to different objects in the ascriber's belief world. That option is ruled out in the present framework; for we want the ontology to be that of the ascriber all along: we want the singular terms to refer to the same objects, whether we are talking of the actual world, or about the ascriber's belief world. That is the price to pay for semantic innocence.

The analyses of belief sentences put forward by Frege and Quine are also mentioned by Recanati as instances of semantic deviance (Recanati 2000b, c. 2). As an extensionalist, Quine takes the semantic value of a proper name to be its reference, and the value of a declarative sentence to be a truth-value. But for him belief contexts behave like quotation contexts: under the scope of “believe”, the value of a sentence is no longer its ordinary value. Likewise, Frege held that two distinct proper names, in belief contexts, denote their senses, and no longer their reference. To that extent, following the typology I presented, Frege is a representative of option (ii) rather than option (i). However, what are we to say of accounts of type (i) above? Should we say that they also depart from semantic innocence?

Strictly speaking, the answer to this question should be negative. If we refer to the original quotation where Davidson introduced the notion of semantic innocence, a semantic theory departs from semantic innocence, according to Davidson (1968: 830), if in this theory

the words ‘the earth moves’, uttered after the words ‘Galileo said that’, mean anything different, or refer to anything else, than is their wont when they come in different environments

Consequently, a theory that would predict that the simple sentences “Tully is bald” and “Cicero is bald” have distinct meaning in ordinary contexts, and that meaning of “believe” applies straightforwardly to these, would fully deserve the label of “semantic innocence”. Option (i), strictly speaking, *is* compatible with semantic innocence. This conclusion, I believe, may nevertheless surprise the reader, because the principle of semantic innocence tends to be identified with a version of the theory of direct reference, according to which the meaning of singular terms, even within belief contexts, is exclusively their actual reference. The above quotation from Recanati illustrates such a view.

Before going deeper into this issue, the question I would like to address in the next section is that of the plausibility of option (iii), with respect to the view of substitutivity that Recanati defends. In the following section, finally, I will examine to what extent option (ii) is necessarily incompatible with the principle of semantic innocence.

2. Recanati's account of substitutivity failures

2.1. The principles of intensional replacement

As explained in the introduction of this paper, Recanati defends a version of the principle of substitutivity in the form of three principles of intensional replacement (Recanati 2000b: 41):

- (a) A genuine singular term t (e.g. a name or an indexical) can be replaced *salva veritate* by some other term t' provided t and t' have the same extension (hence the same content).
- (b) A predicative expression F can be replaced *salva veritate* by another predicative expression G if F and G express the same concept.
- (c) A sentence S can be replaced *salva veritate* by a sentence S' in any complex sentence in which S occurs if S and S' express the same proposition.

What are the predictions associated with these principles? Are these principles necessarily too strong?

Strictly speaking, only principle (a) directly allows to make predictions that are problematic. Indeed, principle (a) predicts that the inference from (4) to (5) is semantically valid:

(4) Peter believes that Cicero was rich

(5) Peter believes that Tully was rich

In a situation where it would be appropriate to utter (1) above, this yields an outright contradiction. As for principles (b) and (c), however, Recanati does not specify the sense in which two predicates 'express the same concept', not the sense in which two sentences 'express the same proposition'. On a fine-grained view of the nature of concepts, propositions, and of the notion of expression, as described in option (i) above, these two principles may be perfectly acceptable. By contrast, on a classical intensional conception of referential content, as in standard possible world semantics, principles (b) and (c) directly conflict with the consistency of (2) and (3). Elsewhere, Recanati generally adopts a Russellian conception of the nature of propositions, in the spirit of Barwise & Perry (1983). Although this corresponds to a more fine-grained conception of the notion of referential content in general, I don't think it should make different predictions with regard to (2) and (3). To illustrate principle (b), Recanati takes precisely the pair "eye-doctor" and "ophtalmologist". For full sentences, the situation is more delicate. It seems doubtful, for instance, that principle (c) should validate the substitution of any pair of mathematically true (resp. false) sentences in belief contexts. Intuitively, we wouldn't want to say that "two plus two equals four" and "there are infinitely many primes" have the same referential content, even in the sense in which we are ready to say that "eye-doctor" and "ophtalmologist" have the same referential content. A Russellian theory of propositions can account for the difference, but things are less clear with example (3). I think reasonable to assume that principle (c), on Recanati's view, would licence the substitution of two logically equivalent sentences like "if door A is locked, door B is not", and "if door B is locked, door A is not", if only because these two sentences can be uttered to express the same thing (more plausibly, at least, than the foregoing arithmetical sentences).

Principles (a)-(c), therefore, all make predictions that conflict with the consistency of examples (1)-(3). According to Recanati, however, these cases of substitutivity failure are to be accounted for in pragmatic terms, they do not threaten the validity of principles (a)-(c) *qua* semantic principles. The proper way to understand this division of labor between semantics and pragmatics is to go deeper into Recanati's analysis of belief reports.

2.2. The *what* and *how* of beliefs and belief attributions

For Recanati, belief sentences are a particular instance of what he calls metarepresentational statements. Metarepresentational statements express representations about representations. For instance, if John utters (4) in a given context, John expresses a representation about a representation of Peter's. If John is sufficiently competent, whenever he uses the noun "Cicero" in uttering the plain sentence "Cicero was rich", it is to refer to Cicero. Semantic innocence corresponds to the idea that the word "Cicero", whenever used by John, ought to keep the value it has in the language of John in all contexts, including metarepresentational contexts. On this conception, the principle of substitution of cointensional expressions in contexts of attitude ascriptions is a principle of transparency *relative to the ascriber*. In particular, the substitution of cointensional, yet morphologically distinct expressions under a verb of attitude need not imply a change in the mode of presentation of the content ascribed to the agent. In *Direct Reference*, using a distinction of McGinn, Recanati motivates this idea in the following way (Recanati 1993: 333):

There are two aspects in a belief content: 'what is represented, and how it is represented' (McGinn, 1982, p. 214) – the truth-conditional of the belief content and the narrow content, in the terminology of chapter 4. A belief ascription is more informative if it specifies not only the former, but also the latter. (...). To be sure, there are contexts in which the *how* is irrelevant – contexts in which the only thing that matters is the truth-conditional content of the belief.

Importantly, Recanati distinguishes two levels here: the level of belief, and the level of belief sentences. Adapting his later terminology, I will call the first the representational level, and the second the metarepresentational level. Also, I shall rather talk of *referential content* instead of *truth-conditional content*, for the reasons mentioned above.⁵ To each level, there corresponds further two aspects, the "what" and the "how". At the representational level, the "what" corresponds to the referential content of the belief (what makes the belief true or false), and the "how" to a subjective mode of presentation. At the metarepresentational level, the "what" now corresponds to the referential content of the embedded sentence used to ascribe the belief. But the specification of the "how", for Recanati, is optional, often it can simply be implicated. An example will help to clarify these distinctions.

If Peter points to a statue of Cicero and sincerely utters "this man must have been rich", I may say: "Peter believes that Cicero was rich". The individual Cicero is part of the referential content of Peter's belief, so is the property of richness, and so is the fact that their combination makes his belief a true belief. But now it seems reasonable to suppose that the psychological mode of presentation of this belief corresponds to the way Cicero appears to Peter in this case, for instance represented as a statue. At the metarepresentational level, the "what" of the belief corresponds to the referential content of the sentence "Cicero was rich": following the identification suggested by Recanati, it coincides with the referential content of the ascribed

⁵ In Recanati 2000b: 396, Recanati talks of "(narrow) contribution to the truth-conditions": by this he has in mind exactly what I offered to call the *referential content* of an expression, in order to contrast it from its *truth-conditional contribution*.

belief. However nothing in the embedded sentence specifies the subjective mode of presentation of this content relative to Peter. Peter may have never heard of someone called “Cicero”. I might as well have said, for this reason: “Peter believes that Tully was rich”.

What happens in attitude contexts where substitutivity fails, according to Recanati (2000a, 2000b), is precisely the opposite. Suppose I am in the situation where it is appropriate to utter (1):

(1) Peter believes that Cicero was rich, but he doesn't believe that Tully was rich

In this case, according to Recanati, the “how” of the belief is implicated by the use of two distinct proper names, and the failure to substitute can be accounted for by the presence of an unarticulated constituent in the verb, of the form:

(6) Peter so-believes that Cicero was rich

where the modified verb “so-believes” is meant to reflect the contribution of the very form of the embedded sentence, or of some salient elements of it, to the truth-conditions of the attribution. The expressions themselves then serve as an *index* of the psychological mode of presentation underlying the ascribed belief.⁶ What (6) ought to mean is something like:

(7) Peter believes that Cicero was rich, under a mode of presentation $m = f(\text{Cicero was rich})$

Here f is a speaker function taking syntactic expressions as arguments and ranging over modes of presentations. A natural constraint on f is to suppose it one-to-one, so that in a given context, it relates distinct syntactic strings to distinct modes of presentation. In this case it would be wrong to substitute “Tully” to “Cicero”, because this would yield:

(8) Peter so-believes that Tully was rich

namely:

(9) Peter believes that Tully was rich, under a mode of presentation $m' = f(\text{Tully was rich})$

Because f is one-to-one, m and m' are distinct modes of presentation, and the substitution is thereby blocked.

The upshot of Recanati's theory is the following: in a belief report, the expressions of the embedded sentence keep the semantic value they have in the language of the ascriber. But the ascriber can implicate the existence of certain modes of presentation by resorting to distinct expressions in his attributions. In *Direct Reference*, Recanati conceived of the “that”-clause of a belief report as a “dependent expression”, susceptible to name distinct propositions depending on the expressions occurring in it. In *Oratio Obliqua*, by contrast, the semantic value of the embedded sentence remains constant, and it is the prefix, “believe that”, which is conceived as a “dependent expression”. Under the former analysis, an attitude verb like

⁶ See Recanati (2000a: 397): the notion of *index* is from Nunberg (1993). It is needed because the psychological mode of presentation need not coincide with the embedded sentence itself, or with the relevant part thereof: this allows to bypass some classic objections against quotationalism.

“believe” expresses a binary relation between an agent and a variable proposition. Under the latter, “believe” expresses a relation of variable arity, depending on the context: either a binary relation between an agent and a proposition (when the “how” is irrelevant); or a ternary relation between an agent, a proposition, and a mode of presentation (when the “how” becomes relevant). For Recanati, only the second analysis is semantically innocent, because whether an attitude report is read transparently or opaquely, the proposition expressed by the embedded sentence remains constant.

2.3. Ascriber's meaning vs ascribee's meaning

An argument in favor of Recanati’s analysis of substitutivity is the observation that opacity phenomena may come from an ambiguity in the embedding verb, rather than from a shift of semantic value of the embedded sentence. In a situation like that of example (3), we specified the context by saying that John hears and accepts the instruction that if door A is locked, door B is not locked. There is a strong implicature, in this case, that what John *hears*, *accepts*, and therefore *believes*, is a particular sentence. His belief therefore has a metalinguistic component, and it is very sensible to analyze the failure of substitution as due to a special sensitivity of the verb in this context. But there are indeed many contexts where the substitution of two logically equivalent sentences under the scope of the verb “believe” is correct, namely situations where “believe” does not have this metalinguistic component. Suppose it is appropriate to say (to use a classic example from Bigelow⁷):

(10) Mary believes that Robin will win

In a context where only the “what” of the belief is relevant, it is then perfectly fine to infer and utter:

(11) Mary believes that anyone who does not compete or loses will do something that Robin won't do

The only restrictions on substitution that may apply in this case have to do with the ascriber’s mastery of logical equivalence, and with conversational maxims: uttering (11) instead of (10) is most likely a violation of Grice's maxim of manner, but it is clear that limitation on the range of available substitutions is then entirely ascriber-oriented, and has nothing to do with the putative mode of presentation of the ascribee's belief. This insight casts light, incidentally, on the scope of the so-called problem of logical omniscience: if Recanati is right, a principle of substitution of logically equivalent sentences under a belief verb need not ascribe exorbitant logical capacities to the belief subject. It is only if belief sentences are taken to encode the “how” of beliefs systematically that it must be so.

This intuition that the ascriber’s point of view prevails on the ascribee’s point of view in attitude reports is an important aspect of the principle of semantic innocence, and it can be further clarified. For each of the examples (1)-(3) given above, a common inference can be made that the ascribee misidentifies the meaning of certain expressions. For instance, on reading (1):

(1) Peter believes that Cicero was rich, but he doesn't believe that Tully was rich

⁷ Quoted by Heim & Kratzer (1998: 310), from Cresswell (1985).

we commonly infer: “For Peter, Tully and Cicero are two distinct individuals”. From example (2), we infer that: “For Peter, being an eye-doctor is not the same thing as being an ophtalmologist”; and similarly in (3), with the conditional sentence and its contrapositive. In all these cases, two expressions that have the same meaning in the language of the ascriber are inferred to have distinct meanings for the ascribee. But we can imagine the opposite situation, where the ascribee conflates the meaning of two expressions that have distinct meanings in the language of the ascriber. Let us imagine, following an example of Burge, that (12) and (13) both hold in a certain context:

(12) Alfred believes that a fortnight is a period of ten days

(13) Alfred believes that Bertrand will be gone precisely ten days

Would we infer (14) in this case?

(14) Alfred believes that Bertrand will be gone precisely a fortnight

Burge made a more precise description of the situation, in supposing that Alfred had sincerely uttered: “Bertrand will be gone precisely a fortnight”. He then made the following observation on this puzzle (Burge 1978: 132):

I have found nonphilosophical native speakers on both sides. But most lean negatively. An *affirmative* answer is typically based on a desire to maintain a close relation between sincere assertion and belief.

I tend to think that, if we set aside Burge's additional specification about Bertrand's assertion, the inference from (12)-(13) to (14) is clearly invalid. This suggests that the expression “fortnight” in (14), just like the expression “ten days” in (13), keeps its ordinary semantic value. In other words, there is no shift to the idiosyncratic language of Alfred when an ascription like this is made. The observation that most speakers “lean negatively” even in the situation where belief tends to be assimilated with assertion, seems just to confirm the priority of this semantic intuition. If this analysis is correct, this example provides further motivation for the idea that: *whether it is explicit or inferrable from a belief ascription that an agent misidentifies the meaning of certain expressions, these expressions nevertheless keep their ordinary semantic value in such contexts*. This thesis is not in opposition with the view that some of the linguistic idiosyncracies of the ascribee can be accomodated in the language of the ascriber. For instance, Recanati reports:

(15) My daughter believes that I am a *philtosopher*

The word “philtosopher” is just taken from childish language and integrated into the ascriber's language. What semantic innocence does rule out corresponds to situation of *unmarked context-shifting*, where an already existing expression is given a deviant meaning. But attitude verbs *per se* are not context-shifting operators: we can imagine a reading of (14) where the ascriber uses “fortnight” in the sense of “period of ten days”, but situations of this kind have to be marked by a special intonation, as happens in cases of irony, echoic uses, or pretense.⁸

⁸ See Recanati (2000b), IV-V, for a detailed analysis of such context-shifting phenomena.

3. Semantic innocence in a Hintikkean framework

From what we saw above, it should appear that the notion of semantic innocence, as Recanati uses it, is not of a single block. There are two main components in this notion. The first is the intuition that an expression should retain its ordinary semantic value inside and outside belief contexts: this corresponds to what I call *semantic constancy*. The second component is the identification of the ordinary semantic value of an expression with its referential content, a thesis which may be more problematic. On a fine-grained view of meaning (option (i)), opacity phenomena should be taken as primitive, and transparent readings of belief sentences may be viewed as the exception. On the view of meaning as referential content defended by Recanati (option (iii)), transparency is the default rule, and departures therefrom are of a pragmatic nature. As I argued in the first section, the latter option is not the only one to preserve “semantic innocence”, so does the first. The choice between option (i) and option (iii) is matter of theoretical economy, but it is also an empirical matter: if *semantic innocence* is not only *semantic constancy*, but encompasses a view of meaning as referential content, then maybe option (iii) is better founded. I shall not try to adjudicate between these two options. Instead, I would like to examine more closely the last of the three options we considered, namely option (ii), and to argue that, *pace* Recanati, it needn't conflict with his views on semantic innocence. If you remember the quotation given p. 3, Recanati contrasts his approach with that of Hintikka and writes:

we want the ontology to be that of the ascriber all along: we want the singular terms to refer to the same objects, whether we are talking of the actual world, or about the ascriber's belief world. That is the price to pay for semantic innocence.

My claim is that this commitment to the sole ontology of the ascriber is exorbitant, and is not forced upon by the attachment to the principle of semantic innocence. In other words, a more liberal ontology, in the style of Hintikka, is permitted, but even more than that, it can very well serve the variety of truth-conditional pragmatics that Recanati calls for. To show this, it is necessary to distinguish, more clearly than Recanati does, the principle of semantic innocence from certain commitments of the thesis of the direct reference of singular terms.

Let's consider again an opaque belief report like (1):

(1) Peter believes that Cicero was rich, but he doesn't believe that Tully was rich

According to Recanati, when an ascription of this sort is made by someone who knows that the proper names “Cicero” and “Tully” are coreferential, the reference of these two names ought to remain the individual Cicero. This comes from the fact that, in the ascriber's language, the proper names are supposed to be directly referential. Yet, as Recanati admits, (1) is consistent precisely because in this context the names “Cicero” and “Tully” become the *indices* of specific modes of presentation. As we saw earlier, the way we typically interpret the consistency of (1) is by inferring:

(16) For Peter, Cicero and Tully are distinct individuals

By this, we commonly understand:

(17) For Peter, the proper names ‘Cicero’ and ‘Tully’ refer to distinct individuals

This suggests that the modes of presentation are the names themselves. However (17) is semantically deviant as a paraphrase of (16): Peter may be a speaker of French, who hears of “Cicero” as “Cicéron”, and of “Tully” as “Marcus Tullius”. An innocent and adequate paraphrase of (1) should be:

(18) Peter believes of Cicero, under the name $f(\text{Cicero})$, that he was rich, but it is not the case that he believes of Tully, under the name $f(\text{Tully})$, that he was rich.

The standard account of this situation in a Hintikkaean framework of modal predicate logic would be the following: let w be the world of utterance of (1). Relative to w , suppose Peter entertains two doxastic alternatives i and j . At i and j , let “Cicero” refer to individual **a**, and “Tully” to individual **b**, distinct from **a**. Let **a** fall within the extension of the predicate “rich” at i and j , and **b** out of this extension in at least one of these. We then get a model of Peter's doxastic situation that satisfies (1). Yet, in the actual world w , “Cicero” and “Tully” refer to one and the same individual. For that reason, Recanati concludes that this account is semantically deviant, because the proper names do not refer to the same individual across all these worlds.

But this is so only because all the worlds are put on a par. In a Hintikkaean framework, semantic innocence and rigidity can both be preserved, provided the relevant worlds are sorted out. Consider a construal of (1) such as (19), where “you” refers to the hearer, and “I” to the utterer. (19) and (1) are at least co-assertible in English:

(19) Peter believes that the individual you and I call “Cicero” was rich, but doesn't believe that the individual you and I call “Tully” was rich

I claim that (19) captures the very idea of (18). Clearly, someone who utters (19) is still using the words “Cicero” and “Tully” *as referring to the individual Cicero*. In the actual world w of the hearer and speaker, the descriptions “the individual you and I call ‘Cicero’” and “the individual you and I call ‘Tully’” denote the same individual, and likewise in any of the counterfactual worlds relative to w : these descriptions are co-rigid and play exactly the role of the corresponding proper names relative to the ascriber and hearer. But relative to Peter, they do not necessarily select one and the same individual. This paraphrase doesn't affect the rigidity of proper names, nor does it amount to a descriptivist view of their meaning: someone like Peter can misperceive the coreferentiality of proper names “Cicero” and “Tully”, or of their counterparts in his language, and yet interpret each of these rigidly relative to his belief worlds. As Maria Aloni writes (Aloni 2001: 44-45):

The failure of substitutivity of co-referential terms (in particular proper names) in belief contexts does not depend on the ways in which these terms actually refer to objects (so this thesis is not in opposition with Kripke (1972)'s analysis of proper names), it is simply due to the possibility that two terms that actually refer to one and the same individual are not believed by someone to do so.

If we follow (19), we can still say that Peter's belief is *about* the individual you and I call “Tully” and *about* “the individual you and I call Cicero”. To that extent semantic innocence is preserved.

This analysis applies to example (20) that Recanati (1993: 352-353) presents as particularly problematic for his understanding of semantic innocence. Consider the case of

Jean-Pierre Dupuy's son who "has hardly heard of President François Mitterrand", but who knows a certain ass under the name "Mitterrand". In this case, Dupuy would utter (20):

(20) My son believes that Mitterrand is an ass

Recanati notes:

If he (Dupuy) were to utter 'Mitterrand is an ass', his utterance 'Mitterrand is an ass' would express a singular proposition about François Mitterrand, but if he utters (20) instead, the same words 'Mitterrand is an ass' will no longer express that same proposition: rather they will express the proposition that *the individual called 'Mitterrand'* is a certain sort of animal, or more perspicuously, that the name 'Mitterrand' refers to a sort of animal; it is that metalinguistic proposition which is the content of the belief ascribed to Dupuy's son in (20) (italics mine)

I disagree with Recanati's interpretation of this example. The description "the individual called 'Mitterrand'" is indeed ambiguous, given that distinct individuals happen to be called "Mitterrand" in the actual world. When Dupuy utters (20), he normally means (21):

(21) My son believes that the individual you and I call "Mitterrand" is an ass

Contrary to Recanati's analysis, this suggests that (20), like (21), is still *about* François Mitterrand: if it were about the donkey Mitterrand, Dupuy had better say "know" instead of "believe" to make his ascription.⁹ For Dupuy's son, when people talk about Mitterrand, they talk about a certain animal. Yet when Dupuy utters (20), *he is still referring to François Mitterrand*, and saying that his son takes François Mitterrand to be an animal; he is clearly ascribing to his son an incorrect belief in this context. In this sense, once again, semantic innocence is preserved: the name "Mitterrand" retains its normal semantic value. But although the ascription of belief is *about* François Mitterrand, the belief of Dupuy's son does not refer to François Mitterrand directly, rather it involves an epistemic *counterpart* of the individual commonly called "Mitterrand" by the speaker and hearer. For that reason, I think the opposite of Recanati's conclusion is called for: semantic innocence does not rule out an ontology of epistemic worlds on top of the ontology of the ascriber (actual world and counterfactual worlds). Such an ontology is perfectly able to represent the dual contribution of referential expressions in opaque belief contexts, namely their ordinary reference relative to the speaker, and their subjective content relative to the ascriber.

4. Conclusion

The principle of semantic innocence can be ambiguous. For Recanati, as originally for Davidson, semantic innocence means preservation of the ordinary meaning of expressions in attitude contexts, and in particular, preservation of their ordinary reference. For Recanati, however, semantic innocence is tied more specifically to the theory of direct reference. In this paper, I have suggested to separate out the two aspects, to look at semantic innocence in the first sense, as a principle of *semantic constancy*, namely as the thesis that within contexts of attitude, *the occurring expressions retain the meaning they have more generally in the language of the ascriber*. Semantic innocence rules out, in this sense, cases of what I called unmarked context-shifting. But although semantic innocence requires that referential expressions retain their meaning and reference in unshifted contexts, it can't rule out the fact that in opaque contexts of attitude, referential expressions contribute more than their actual

⁹ Assuming "ass" is taken literally.

reference. As Davidson granted in the same move in which he introduced the idea of “semantic innocence”, the “role” of words “in *oratio obliqua* is in some sense special” (1968: 830). For Recanati himself, “semantic innocence” is compatible with “pragmatic enrichment”. In this respect, a Hintikkaean semantics of belief sentences is perfectly compatible with an integrated program of “truth-conditional pragmatics” of the sort entertained by Recanati.

References

- Aloni M. (2001), *Quantification under Conceptual Covers*, Amsterdam, ILLC dissertation series.
- Barwise J. & Perry K. (1983), *Situations and Attitudes*, MIT Press, repr. CSLI, 1999.
- Burge T. (1978), “Belief and Synonymy”, *The Journal of Philosophy*, 75, 3: 119-138.
- Carnap R. (1947), *Meaning and Necessity*, Chicago: University of Chicago Press.
- Church A. (1989), “Intensionality and the Paradox of the Name Relation”, in J. Almog, J. Perry & H. Wettstein eds., *Themes from Kaplan*, c. 8: 151-165, Oxford UP.
- Cresswell M. (1973), “Hyperintensional Logic”, *Studia Logica* 34: 25-38.
- Cresswell M. (1985), *Structured Meanings*, MIT Press.
- Davidson D. (1968), “On Saying That”, *Synthese*, 19, quoted from P. Ludlow ed., *Readings in the Philosophy of Language*: 817-831, MIT Press, 1997.
- Gallin D. (1975), *Intensional and Higher-Order Modal Logic*, North-Holland.
- Heim I. & Kratzer A. (1998), *Semantics in Generative Grammar*, Blackwell.
- Haas-Spohn U. (1994, forthcoming), *Hidden Indexicality and Subjective Meaning*, English transl. by E. T. Zimmerman from Haas-Spohn 1994, *Versteckte Indexikalität und Subjektive Bedeutung*.
- Hintikka (1962), *Knowledge and Belief, an Introduction to the Logic of the Two Notions*, Cornell UP.
- Kripke S. (1972), *Naming and Necessity*, Blackwell 1980.
- Lewis D. (1970), “General Semantics”, repr. in D. Lewis, *Philosophical Papers I*, c. 12: 189-229, Oxford UP.
- McGinn C. (1982), “The Structure of Content”, in A. Woodfield ed., *Thought and Object*, Oxford, Clarendon Press: 207-258.
- Muskens R. (1991), “Hyperfine-Grained Meanings in Classical Logic”, *Logique & Analyse* 133-134: 159-176.
- Nunberg G. (1993), “Indexicality and Deixis”, *Linguistics & Philosophy* 16: 1-43.
- Quine W.V.O. (1960), *Word and Object*, MIT Press.
- Recanati F. (1993), *Direct Reference, From Language to Thought*, Blackwell.
- Recanati F. (2000a), “Opacity and the Attitudes”, in A. Orenstein & P. Kotatko ed., *Knowledge, Language and Logic*: 367-406, Kluwer Academic Publishers.
- Recanati F. (2000b), *Oratio Obliqua, Oratio Recta, an Essay on Metarepresentations*, MIT Press.
- Stalnaker R. (1987), “Semantics for Belief”, repr. in R. Stalnaker, *Context and Content*, c. 6: 117-129, Oxford UP, 1999.