Local pragmatics : reply to Mandy Simons

François Recanati
Institut Jean-Nicod
(ENS/EHESS/CNRS, PSL Research University)

1

In the 1980s, several theorists interested in the foundations of pragmatics, including myself, launched an attack on some aspects of Grice's theory of conversational implicatures.[1] One of the claims was that, among the phenomena that are, indeed, pragmatic and 'Gricean' in a suitably broad sense, some do not fit Grice's analysis of conversational implicatures as resulting from an inference taking as input the fact that the speaker has said what she has said (in Grice's favoured sense of 'what is said').

Grice's analysis is often, and rightly, referred to as a 'two-stage' analysis : the interpreter first computes the proposition expressed by the utterance (semantic stage), and then infers what the speaker actually means by expressing that proposition in that context (pragmatic stage). Crucially, the pragmatic inference which makes up the second stage relies on two premises : (i) that the speaker has said what she has said (the conclusion of the first stage), and (ii) that the speaker is observing the Cooperative Principle and the attendant maxims. In this framework, conversational implicatures are 'post-propositional' in the sense that their calculation presupposes the prior identification of what is said (the proposition expressed). Grice's critics argued, inter alia, that there are pragmatic effects that are 'pre-propositional' and cannot be accounted for along the lines of Grice's two-stage analysis.

Such effects are still pragmatic and Gricean in a *broad* sense. As Mandy Simons puts it,

> any account of pragmatic inference which posits that an interpreter reasons about what the speaker meant and that this reasoning is guided by presumptions of rationality of the speaker is a Gricean account.

But we must distinguish accounts that are Gricean in the broad sense, because they appeal to intention-recognition and give a role to the Cooperative Principle, and accounts that are Gricean in the narrow sense of the two-stage model. Grice's critics assumed a Gricean account in the broad sense, but rejected the applicability of the two-stage model to some of the phenomena amenable to such an account. In effect, they claimed that we must distinguish between different varieties of pragmatic inference, only some of which give rise to 'post-propositional' effects.

As Grice's critics were the first to point out, and as Mandy Simons reminds us in her paper, the assumption that the Cooperative Principle is being observed plays a role at *all* levels of interpretation, not merely at the post-propositional stage. Prior to semantic composition, disambiguation has to take place : whenever an expression is ambiguous the interpreter has to select one of the possible meanings of the expression in order to feed it to the compositional machinery. It is (or should be) uncontroversial that disambiguation is pragmatic through and through : it can only proceed under the presumption that the speaker is rational and obeys the Cooperative Principle. Once disambiguation has taken place, some

[1] See in particular Wilson and Sperber 1981, Travis 1985, Sperber and Wilson 1986, Carston 1988, and Recanati 1987 and 1989.

pragmatic work is still needed to determine what is said, for the proposition expressed depends upon the context whenever the sentence includes an indexical expression or a free variable-like element. It is sometimes said that, in contrast to disambiguation, the contextual assignment of value to indexicals is governed by linguistic *rules* (e.g. the rule that 'I' refers to the speaker), so that no appeal to pragmatics is necessary at that stage. But demonstratives and free variables don't work that way : which value they are assigned in context depends upon the speaker's likely intentions, the recognition of which is made possible by the hearer's presumption that the Cooperative Principle is being observed. Here again, we find a pragmatic inference of the Gricean sort (in the broad sense), but no one is tempted to say that indexical values result from a two-stage process like that which Grice appeals to for the derivation of conversational implicatures. Even though the contextual assignment of values to demonstratives and free variables is sensitive to rationality considerations of the Gricean sort, it contributes to what is said (hence counts as a pre-propositional pragmatic process). In contrast, conversational implicatures are post-propositional and remain external to what is said. As Grice puts it, 'the truth of a conversational implicatum is not required by the truth of what is said (what is said may be true — what is implicated may be false)' (Grice 1975, in Harnish ed. 1994 : 72).

There are two competing explanations for the difference in status between the pragmatic inference at work in indexical resolution and in conversational implicature. The most widespread one is the 'minimalist' explanation, which appeals to a general distinction between two types of pragmatic inference, depending on how the inference is *triggered*. Indexicals call for 'saturation', in virtue of their linguistic properties. Whatever inference takes place in indexical resolution is triggered, if not guided, by linguistic meaning. Inferences that are semantically triggered in this way are mandatory : nothing is said unless a value is assigned to the expression in need of saturation. By contrast, inferences that are pragmatically triggered only take place if the context demands it. Being based on such inferences, which may or may not take place, conversational implicatures are cancellable. They do not belong to semantic content (which is fixed by the rules of the language, modulo ambiguity and indexicality) but come in addition to it.

This explanation has several shortcomings, but one particularly obvious problem has to do with the principle it appeals to : that, because pragmatic inferences that are semantically triggered are mandatory, nothing is said unless a value is assigned to the expression in need of saturation. That does not seem to be right. Grice distinguished conventional implicatures from conversational implicatures on the grounds that conventional implicatures are triggered conventionally in virtue of the meaning of some expression (e.g. 'but'). Contrary to what many theorists seem to believe, conventional implicatures are not, or not always, directly associated by the conventions of the language with the expressions that carry them. In the case of 'but' the conventional implicature, though semantically triggered, must be worked out, and that involves a pragmatic inference. Suppose I say : 'It's a beautiful day (P), but I am tired (Q)'. To understand this, according to Ducrot's analysis of 'but', one must contextually identify some proposition R such that the speaker takes P to be a reason for holding R, and Q to be a stronger reason for holding not-R. (In this example, R might be 'we should go for a walk'.) Identifying R is like assigning a value to an indexical[2] : it depends upon the context, and the reasoning which leads the hearer to the right value appeals to the presumption that the speaker obeys the Cooperative Principle. So a pragmatic inference is involved in conventional implicatures, and it is semantically rather than pragmatically triggered. Still, in contrast to the case of indexicals, conventional implicatures do *not* affect what is said. They are not part of

[2] Jakobson noted the similarity between 'but' and indexicals (see his 1957, §1.5).

the 'at-issue' content of the utterance, where at-issue content corresponds to the Gricean notion of what is said.

The 'contextualist' explanation of the difference between indexical resolution and conversational implicature — the fact that the former, but not the latter, affects truth-conditions — appeals to a specific property of conversational implicatures (Recanati 1989, 1993, 1995, 2004). Conversational implicatures don't affect truth-conditions because, as per the two-stage analysis, their calculation depends upon, hence presupposes, a prior identification of what is said. If this is the correct explanation, then we should not rule out from the outset the possibility of a pragmatic inference that is both pragmatically triggered and pre-propositional. Conversational implicatures don't affect truth-conditions because what is said (or the speaker's saying of it) is a premise serving as input to the inference. The speaker's actually saying that *p* implies that *q* (given the presumption that the CP is obeyed, etc.), where '*q*' is the implicature. In the case of indexicals and free pronouns, the hearer does not know what is said until the value of the context-sensitive expression is computed, so the fact that the speaker has said that *p* cannot be one of the premises used in carrying out the inference to the intended value. Rather, the relevant premise is that e.g. the speaker has used the pronoun 'he' referringly, hence must have some male person in mind of whom she wants to predicate the property denoted by the VP. The semantic rule governing the pronoun triggers the search for an appropriate value, constrained by the presumption that the speaker who intends that value obeys the Conversational Principle. But there is no reason why a *pragmatically triggered* inference might not similary affect truth-conditions if the premise it uses as input is not the fact that the speaker has said that *p*, but some other fact, e.g. the fact that the speaker has used a certain expression, the literal meaning of which generates an interpretation for the utterance which clashes with the presumption that the speaker obeys the CP.

Examples are not difficult to come by :

(1) The *ham sandwich* is getting restless
(2) There is a *lion* in the middle of the piazza
(3) She likes to wear/eat *rabbit*

The first sentence can be used to say that the ham sandwich *orderer* is getting restless. (2) is likely to mean that there is a *statue of* a lion in the middle of the piazza. In (3), 'rabbit' is interpreted as rabbit *fur* or as rabbit *meat* depending on whether it is the object of 'eat' or 'wear'. For such cases of 'pragmatic modulation', the intuitive truth-conditions of the utterance *are* affected by the pragmatic inference (Recanati 2004). Moreover (and relatedly), the pragmatic effect that results from the inference is *local* and can embed, thus giving rise to what Mandy Simons calls 'embedded pragmatic effects' :

> *Embedded Pragmatic Effects*: Cases where the propositional content which falls under the scope of a linguistic operator (at some stage of interpretation) includes content which is the output of pragmatic inference. (Simons, this issue, p. 4)

To see that modulation embeds, let us modify the noun 'lion' whose meaning is modulated in (2) by adding an adjective, e.g. 'old' :

(2') There is an old lion in the middle of the piazza

We can distinguish two readings for (2'): it can be used to say either that there is a statue of an old lion, or that there is an old statue of a lion, in the middle of the piazza. The latter

reading involves an embedded pragmatic effect : the meaning of 'old' applies to the modulated meaning of 'lion'. Just as exhaustification in the Fox *et al.* theory of scalar implicatures, modulation is a local process that can affect any constituent, at any level.

The foregoing suggests that the two distinctions (semantically *vs* pragmatically triggered inferences, and inferences which do *vs* inferences which do not contribute to truth-conditions) are actually orthogonal to each other. There are four types of case :

Table 1 : Pragmatic inferences

	Affect truth-conditions	*Do not affect truth-conditions*
Semantically triggered	indexical resolution	conventional implicatures
Pragmatically triggered	pragmatic modulation	conversational implicatures

Along these lines, I offered an analysis of pragmatic modulation which reverses the order prescribed by the two-stage model : instead of a global inference process taking place on top of the compositionally determined content, as for conversational implicatures, I argued for a local inference process which contributes to the determination of content by modulating the meaning of the parts undergoing composition.[3] Ivan Sag, inspired by Nunberg's work, had argued for such a reversal in an early paper :

> What is the role of context in examples such as (1)? Is this a case of an absurd literal meaning (an attribution of restlessness to a culinary object) rescued from pragmatic absurdity by the Cooperative Principle augmented by some ancillary principle which guides Gricean inferencing? Or is the shift from ham sandwich to ham sandwich orderer somehow more directly involved in the semantics of such utterances? Perhaps the shift from ham sandwich to individual who is in some relation to a ham sandwich (possibly different from context to context) is like the shift in denotation that accompanies indexical expressions as they are uttered in various contexts. (...) This approach, rather than one of the first kind, where all examples like (1) are pushed off to pragmatic theory and are abstracted away from in semantic analysis, is intuitive on the grounds that these transfers seem very different in kind from the kind of inferential operations that lead one from *It's hot in here* to the sense of "Please open the window," which clearly deserve treatment of the first type. (Sag 1981:275-6)

[3] 'We do not have a "global" transfer from the absurd proposition that the sandwich itself is getting restless to the more plausible proposition that the ham sandwich orderer is getting restless, as Bach and many others claim, but a local transfer from the literal interpretation of the description to a non-literal interpretation, only the latter going into the interpretation of the utterance in that context and undergoing the composition process which yields the semantic value of the whole on the basis of the semantic values of its parts. On the picture I am advocating, the composition process takes place "after" various primary pragmatic processes, including transfer, have applied locally. The latter do not presuppose the prior computation of some basic propositional value for the utterance; on the contrary, it is the process of propositional composition which presupposes the prior operation of primary pragmatic processes, since they provide the (relevant) semantic values of the parts on which the composition process operates to yield the semantic value of the whole.' (Recanati 1993 : 265-66 ; see also Recanati 1995, 2004, 2010, 2012)

However, an advocate of the two-stage model has an obvious response to make. She can argue that pragmatic modulation is much more like conversational implicatures than it is like indexical resolution. Not only is the pragmatic inference at work in modulation pragmatically triggered, in contrast to indexical resolution, which is semantically triggered. In the case of indexicals there is, prior to the assignment of contextual value, no proposition which is expressed and could serve as a premise in the inference. But in the case of modulation, just as in the case of conversational implicatures, there is. The proposition literally expressed by (2) arguably is the (obviously false) proposition that there is a (real) lion in the middle of the piazza. According to Grice and his followers, it is the obvious falsity of that proposition that triggers a conversational implicature. So the two-stage model applies to pragmatic modulation after all. The first stage is the computation of what is said (i.e. what is literally said). The second stage is a pragmatic inference triggered by the realization that the proposition literally expressed cannot be what the speaker means. The second stage involves a local operation, namely substituting the modulated meaning of the problematic constituent for its literal meaning, but what triggers the search for the nonliteral meaning intended by the speaker is the realization, at the global, utterance level, that the speaker's saying what she literally says conflicts with the CP. That is, basically, the position Mandy Simons takes in her paper with respect to embedded pragmatic effects.

In a debate with Kent Bach over these same issues, I had pointed out that, due to the local character of the pragmatic inference at stake, the (absurd) proposition literally expressed by (1) does not have to be actually computed in the course of interpreting the utterance (Recanati 1993 : 263-66). This is in contrast to conversational implicatures : they are subject to a 'working out' requirement (Grice 1975, in Harnish ed. 1994 : 65) from which it follows that the proposition which serves as input to the inference must be 'available' to the interpreter (Recanati 1989, 2002, 2004). Still, one can insist (as Mandy Simons does, and as Kent Bach himself did) that a fully explicit *rational reconstruction* of the hearer's implicit reasoning in cases of pragmatic modulation must start from the premise that what the speaker has literally said cannot be what he means.[4] Even if in actual processing the literal proposition does not have to be computed, the theorist can still acknowledge its role in rational reconstruction. Let's assume that this is right. Then, can't we describe the inference as a *global* inference that starts from the fact that the speaker has said something obviously false, and tries to identify the speaker's communicative intention ? That is Mandy Simons' position, and that is also how Grice himself describes cases of metaphor or irony. In such cases, Grice argues, what the speaker says is blatantly false, so the hearer starts searching for a 'related' interpretation that makes the utterance compatible with the assumption that the CP is being observed. In her paper, Mandy Simons claims that embedded pragmatic effects can be accounted for in this manner too :

> Local pragmatic enrichment is driven by the need to make sense of the utterance as a whole, that is, to understand the speech act as a whole as being in accord with the requirements of Cooperativity. In these cases, the process of pragmatic enrichment appears to be triggered by the interpreter's observation of apparent global violations of Cooperativity. In the relevant examples, however, the global problem has a local solution. (Simons, this issue, p. 8)

[4] 'To give a Gricean account of some pragmatic phenomenon, we must provide an explicit rational reconstruction of the reasoning that an interpreter might engage in to calculate the speaker's meaning on the basis of the compositionally derived content of the speaker's utterance' (Simons, this issue, p. 3).

The solution is local because, as I said already, it consists in substituting the modulated meaning for the literal meaning of the relevant constituent (in the scope of whatever operator happens to operate on the meaning of that constituent). The inference that drives the substitution is global, however, even though the effect it induces is local.

Before proceeding, we should note that, among the critics addressed to Grice's two-stage model in the eighties, some specifically targeted his treatment of cases like metaphor and irony, which is the model for Simons' treatment of embedded pragmatic effects. Thus Sperber and Wilson wrote :

> A speaker may sometimes intend to convey something other than one of the literal senses of his utterance. When he wants to convey something IN ADDITION TO one of the literal senses, the notion of conversational implicature is relevant… However, in the case of figurative language, the speaker normally intends to convey something INSTEAD OF one of the literal senses of his utterance ; the implicature has to be seen as SUBSTITUTING FOR the literal sense. The idea that an implicature could actually contradict the literal sense of an utterance — as it would in the case of irony — does not square with Grice's central claim that implicatures act as premises in an argument designed to establish that the speaker has observed the maxims of conversation in saying what he said. It follows that the interpretation of ironical utterances cannot be reduced to the search for conversational implicatures without grossly distorting the notion of implicature itself. (Sperber and Wilson 1981 : 299 ; see also Holdcroft 1978 : 123 and Recanati 1987 : 228-32)

In the case of irony, subsequent work has shown that Grice's critics were right. Most theorists nowaday believe that irony involves some kind of *staging*, a phenomenon different from implicatures.[5] Be that as it may, when it comes to pragmatic modulation, Grice's critics insist that the type of inference at stake is different from that which underlies genuine conversational implicatures. In favour of this view, I have myself offered various arguments, one of which appeals to the fact that the effects brought about by pragmatic modulation embed, as shown by the 'old lion' example. That they embed shows that they are local and pre-propositional, and that is enough to establish that the two-stage model, which yields *post-propositional* effects, does not apply to them. In her paper, however, Mandy Simons criticizes my argument as resting on a confusion. Even though I do not agree with her on this point, I think her discussion focusses on the right set of issues, and I take her criticism as a welcome request for clarification.

2

The argument which Simons criticizes, to the effect that no genuine conversational implicature can embed, is the following (Recanati 1989, 1993 : 271-72, 2003 : 89-90):

> (a) Conversational implicatures are pragmatic consequences of an act of saying something.
> (b) An act of saying something can be performed only by means of a complete utterance, not by means of an unasserted clause such as a disjunct or the antecedent of a conditional.

[5] The word 'staging' comes from Herb Clark (see e.g. Clark 1996), but the general idea is common to a wide family of views of irony, among which I include both the 'pretense' theory (Clark and Gerrig 1984) and the 'mention' theory put forward by Sperber and Wilson 1981.

(c) Hence, no implicature can be generated at the sub-locutionary level, i.e. at the level of an unasserted clause such as a disjunct or the antecedent of a conditional.
(d) To say that an implicature falls within the scope of a logical operator is to say that it is generated at the sub-locutionary level, viz. at the level of the clause on which the logical operator operates.
(e) Hence, no implicature can fall within the scope of a logical operator.

Mandy Simons thinks step (d) is faulty. It conflates two distinct things, she says : the fact that a pragmatic effect is embedded, as when 'old' applies to the modulated meaning of 'lion', and the fact that the pragmatic inference responsible for the effect takes place locally, rather than globally at the level of the whole utterance. She argues that the inference is global — it is an attempt to rescue the presumption that the speaker, in saying what she says, conforms to the Cooperative Principle — even if its effect is local. So she denies that an embedded pragmatic effect has got to be 'generated at the sub-locutionary level, viz. at the level of the clause on which the logical operator operates'. Like other conversational implicatures, she claims, embedded pragmatic effects are post-propositional, in the sense that it is *the speaker's saying what she says* that clashes with the CP and triggers the search for an alternative, nonliteral interpretation of the utterance.[6] This is compatible with their being pre-propositional in the sense that the effect brought about by reconstrual occurs within the scope of an operator.

I welcome Mandy Simons' distinction between embedded/local pragmatic effects and embedded/local pragmatic inferences. I agree with her that a pragmatic inference to the intended meaning can be triggered globally by e.g. the absurdity of the literal meaning of the utterance undergoing interpretation, even though the effect resulting from the pragmatic inference is the local substitution of a nonliteral (e.g. pragmatically enriched) interpretation for the literal interpretation of some constituent in the sentence. But there is another distinction we need, which relates to that between Gricean accounts in the broad sense and Gricean accounts in the narrow sense. The distinction I have in mind is crucial to a proper understanding of my argument that embedded pragmatic effects are not, indeed cannnot be, genuine conversational implicatures. It is a distinction which Mandy Simons herself makes, when she presents what she calls the 'standard pattern' for calculating a relevance implicature :

(i) Speaker has said that p.
(ii) Speaker's saying that p is not prima facie cooperative. If the speaker said and meant (only) that p, she would be in violation of the requirement to be Relevant.
(iii) But there is no reason to think that the speaker does not intend to be cooperative.
(iv) Therefore the speaker means something somehow inferable from p which would render the conversational contribution Relevant.
(v) On the basis of further inference, I deduce that speaker means q instead of, or in addition to, p.

(Simons, this issue, p. 7)

Simons says we need to distinguish two sub-parts in this reasoning :

One part involves identification of an apparent violation of cooperativity, leading to the conclusion that the speaker means something other than her utterance meaning. This part of the reasoning is what I will call *Gricean reasoning*, and I will call the

[6] Borg (2012 : 65-67) uses 'post-propositional' in that sense.

> conclusion that something further is meant (step iv. in the reasoning above) the *Gricean conclusion*. From the Gricean conclusion, the interpreter proceeds to an inference to the best explanation, trying to answer the question: What is the speaker most likely to have meant, given what she actually said? I'll call this second stage in the reasoning the *interpretative step* (recognizing that this "step" may consist of many sub-steps). (Simons, this issue, pp. 7-8)

The distinction between the Gricean reasoning and the interpretative step corresponds to that which I myself made between a *triggering* inference and a *generating* inference. A triggering inference appeals to the fact that the speaker's locutionary act is not prima facie compatible with the presumption of cooperativity, and concludes that what the speaker means must differ from what she says. I call that a triggering inference because its function is to *trigger another inference, the role of which is to identify the speaker's intended meaning*. That second inference (Simons' interpretative step) I call a generating inference because its function is to *generate* the correct interpretation for the utterance, over and above the interpretation generated by the compositional semantics (which the triggering inference shows to be pragmatically deficient).

In terms of that distinction, the point I want to make in response to Mandy Simons' criticism of my argument can be easily stated : my argument against the possibility of embedded conversational implicatures is exclusively concerned with the *generating* inference. When it comes to generating inferences, there are two very different types of case, according to me. In the case of (genuine) conversational implicatures, the generating inference is global because it starts from the premise that the speaker has said what she has said. In pragmatic modulation cases the generating mechanism which is needed to yield the correct interpretation is not (and cannot be) a global inference — it has no use for the premise that the speaker has said what she has said. That premise is used by the triggering inference, which thereby counts as global, but this is irrelevant to the issue at stake. Again, the issue is exclusively about the generating inference. My claim is that *the generating inference is strikingly different in the two cases*. That difference is also, I believe, the reason why Sperber and Wilson rejected Grice's extension of the notion of implicature to cover figurative speech. Mandy Simons' response – to the effect that in all cases (standard conversational implicatures, pragmatic modulation, figurative speech) the *triggering* inference is global — has no impact upon the issue.

Here is, in simplified form, how the generating inference looks like in cases of conversational implicature :

> <u>Generating inference for conversational implicatures :</u>
> (i) The speaker has said that *p*.
> (ii) The speaker's saying that *p* violates the Cooperative Principle, unless (the speaker believes that) *q*.
> (iii) There is no reason to think that the speaker is not observing the Cooperative Principle.
> (iv) Therefore, (the speaker believes that) *q*.
> (v) The speaker knows that the hearer is able to go through the (i)-(iv) reasoning and has done nothing to prevent the hearer from concluding that (the speaker believes that) *q*.
> (vi) Therefore, the speaker intends the hearer to think, or is at least willing to allow the hearer to think, that *q* ; and so she has implicated that *q*.

What is remarkable about this reasoning is that it provides the interpreter with a *recipe* for

caclulating conversational implicatures. The key principle is that 'to calculate a conversational implicature is to calculate what has to be supposed in order to preserve the supposition that the Cooperative Principle is being observed' (Grice 1975, in Harnish ed. p. 72). In a nutshell : the speaker's utterance implicates that *q* because (i) the speaker says that *p*, (ii) she observes the CP, and (iii) her saying that *p* violates the CP unless *q*.

I call that pattern of reasoning a generating inference because the conclusion of the inference *is* the conversational implicature which, together with what the speaker says, constitutes what the speaker means by her utterance. The generating inference is global because it uses the premise that the speaker has said that *p*. The conversational implicatures thus calculated are post-propositional because the generating inference through which they are calculated presupposes the prior identification of what is said (the proposition that *p*).

Within that pattern, we may distinguish a variety of cases, as Grice does. In some cases, the speaker intentionally produces an apparent violation of the CP in order to trigger the generating inference. In other cases (e.g. Grice's 'garage' example), the proposition that *q* is taken for granted by the speaker but the point of the utterance is not to convey that proposition to the hearer. One might be tempted to exclude the latter type of case from the category of implicatures, on the grounds that the relevant communicative intention is missing ; but one reason to follow Grice in including them is that there may be a continuum of cases, rather than a sharp division between the two categories. Be that as it may, there is another type of case, also categorized as conversational implicature by Grice, which is too different to be included in the same category. That is the 'substitution' type.

In the standard cases, the implicature is sufficent to restore the utterance's conformity to the Conversational Principle. If the literal meaning of the utterance is insufficiently informative, adding a conversational implicature to the literal meaning completes the informational content of the utterance and reconciles it with the quantity maxim infringed at the level of what is said. If the literal meaning of the utterance is not relevant (i.e. if it does not provide a response to the question under discussion), adding the implicature makes the utterance relevant by providing an indirect answer to the question under discussion. But in cases like irony and metaphor, which Grice treats as apparent violations of the first maxim of quality, the utterance can be reconciled with the CP only by assuming that the speaker is not really 'saying' (asserting) what she makes as if to say. These are the cases in which the implicature has to be seen as *substituting for* the literal content instead of *complementing* it. I agree with Sperber and Wilson that including these cases among the conversational implicatures distorts the notion. Whenever substitution is involved, we lose the benefit of the Gricean recipe for calculating the implicature. The generating inference no longer works ; the only thing that is left is the triggering inference that tells us that the speaker is not really saying what she makes as if to say. Some alternative proposition meant by the speaker has to be identified, but to determine that alternative proposition we can no longer rely on the generating inference which works for standard conversational implicatures. Premise (ii) is no longer available : there is no proposition *q* such that, if *q*, then the speaker's saying that *p* can be reconciled with the CP. In substitution cases, the speaker's saying that *p cannot* be reconciled with the CP ; so, in order to preserve the supposition that the CP is being observed, the interpreter is led to give up premise (i) — that the speaker has said that *p* — in favour of a weaker premise : that the speaker has 'made as if to say' that *p*. This leads to the conclusion that what the speaker means is something distinct from the proposition literally expressed by the utterance. But the inference which leads to that conclusion is a triggering inference ; it is not a generating inference. Grice simply provides no recipe for calculating the 'implicature' in such cases, and that is why I think it is a mistake to see the standard notion of conversational implicature as covering this type of case, to which the Gricean recipe does not apply.

Cases of pragmatic modulation are themselves substitution cases. The metonymic

reading of 'ham sandwich' substitutes for its literal reading. Similarly, in (3), the enriched reading of 'rabbit' (*rabbit meat* or *rabbit fur*, as the case may be) substitutes for the literal reading *rabbit stuff*. In (2) the representational sense *statue of a lion* substitutes for the literal sense of 'lion'. The strategy here is one of *reconstrual*. As a result of the triggering inference, the interpreter knows that what the speaker means must be some proposition alternative to the proposition literally expressed. To identify that proposition, the strategy is to go through the process of semantic composition once again, while giving to some constituent in the sentence a modulated meaning distinct from its literal meaning. Possible candidates for the status of modulated meaning (determined on associative grounds) will have to compete on the basis of their ability to yield an interpretation that is pragmatically acceptable, i.e. compatible with the CP. But the role of the CP here is no different from the role it plays in the interpretation of free pronouns. Consider the following example, discussed in *Direct Reference*: 'John was arrested by a policeman yesterday. He had just stolen a wallet'. The pronoun 'he' is interpreted as referring to John, not as referring to the policeman. Why? Because the former interpretation is more plausible than the latter and fits better what we can expect the speaker to commununicate. In both the pronoun case and the pragmatic modulation case, the choice of the relevant candidate (for the status of referent, or for the status of modulated meaning) is constrained by the assumption that the speaker is rational and cooperative, but that constraint is a general constraint on pragmatic inference, a constraint which is *Gricean in the broad sense* and applies to *all* the pragmatic aspects of interpretation : disambiguation, indexical resolution, reference assignment, quantifier domain restriction, pragmatic modulation, figurative speech *and* conversational implicatures alike.

To sum up, I claim that conversational implicatures are generated by a global inference which uses as a premise the fact that the speaker has said that *p*. In cases of pragmatic modulation, only the triggering inference is global. What *generates* the correct interpretation is a process of reconstrual, which locally maps the literal meaning of a constituent to a modulated meaning and composes that meaning with that of the other constituents. That process is constrained by Gricean considerations (in the broad sense) but that is true of all pragmatic aspects of interpretation, whether pre-propositional or post-propositional. Just as indexical resolution, though pragmatic and constrained by Gricean considerations, does not fit the two-stage model through which Grice accounts for conversational implicatures, so pragmatic modulation can't be accounted for in terms of that model despite the fact that, like conversational implicatures and in contrast to indexical resolution, modulation is a pragmatic process which is pragmatically rather than semantically triggered.

The reason why pragmatic modulation embeds, in this framework, is that it is a local mechanism that takes place as part of the process of reconstrual. The output of that process is the intuitive truth-conditional content of the utterance — what the speaker means. Because modulation is local, it determines a content for a *constituent*, a content that composes with the content of the rest of the sentence. If, in the course of composition, a semantic operation applies to the content of that constituent, it is the modulated content that is operated on. Whenever that happens, an embedded pragmatic effect is produced. Such effects cannot be accounted for in the two-stage model because they are local, while the generating inference appealed to by the two-stage model is a global inference from the fact that the speaker has said what she has said. This is wholly compatible with Mandy Simons' point that there is a global triggering inference in the background.[7]

[7] This paper has benefitted from support from the French Agence Nationale de la Recherche under grant agreement n° ANR-10-LABX-0087 IEC and grant agreement n° ANR-10-IDEX-0001-02 PSL. I wish to thank Mandy Simons for her challenging paper and the discussion we

References

Borg, E. (2012) *Pursuing Meaning*. Oxford : Oxford University Press.
Carston, R. (1988) Implicature, Explicature, and Truth-Theoretic Semantics. In R. Kempson (ed.) *Mental Representations : The Interface Between Language and Reality*, 155-81. Cambridge : Cambridge University Press.
Clark, H. (1996) *Using Language*. Cambridge : Cambridge University Press.
Clark, H. and R. Gerrig (1984) On the Pretense Theory of Irony. *Journal of Experimental Psychology : General* 113 : 121-26.
Ducrot, O. (1973) *La Preuve et le Dire*. Paris : Mame.
Ducrot, O. (1980) *Les Echelles argumentatives*. Paris : Minuit.
Grice, P. (1975) Logic and Conversation. *Syntax and Semantics* 3 : 41-58. Reprinted in R.M. Harnish (ed.) (1994), 57-73.
Harnish, R.M. (ed.) (1994) *Basic Topics in the Philosophy of Language*. Hemel Hempstead, Hertfortshire : Harverster Wheatsheaf.
Holdcroft, D. (1978) *Words and Deeds*. Oxford : Clarendon Press.
Jakobson, R. (1957) *Shifters, Verbal Categories, and the Russian Verb*. Russian Language Project, Department of Slavic Languages and Literatures. Harvard University.
Recanati, F. (1987) *Meaning and Force*. Cambridge : Cambridge University Press.
Recanati, F. (1989) The Pragmatics of What Is Said. *Mind and Language* 4 : 295-329.
Recanati, F. (1993) *Direct Reference : From Language to Thought*. Oxford : Blackwell.
Recanati, F. (1995) The Alleged Priority of Literal Interpretation. *Cognitive Science* 19 : 207-32.
Recanati, F. (2002) Does Communication Rest on Inference ? *Mind and Language* 17 : 105-26.
Recanati, F. (2003) Embedded Implicatures. *Philosophical Perspectives* 17 : 299-332.
Recanati, F. (2004) *Literal Meaning*. Cambridge : Cambridge University Press.
Recanati, F. (2010) *Truth-Conditional Pragmatics*. Oxford : Clarendon Press.
Recanati, F. (2012) Compositionality, Flexibility and Context-Dependence. In W. Hinzen, E. Machery et M. Werning (eds.) *Handbook of Compositionality*, 175-91. Oxford : Oxford University Press.
Sag, I. (1981) Formal Semantics and Extralinguistic Context. In P. Cole (ed.) *Radical Pragmatics*, 273-94. London : Academic Press.
Simons, M. (this issue) Local Pragmatics in a Gricean Framework.
Sperber, D. and D. Wilson (1981) Irony and the Use-Mention Distinction. In P. Cole (ed.) *Radical Pragmatics*, 295-318. London : Academic Press.
Sperber, D. and D. Wilson (1986) *Relevance : Communication and Cognition*. Oxford : Blackwell.
Travis, C. (1985) On What is Strictly Speaking True. *Canadian Journal of Philosophy* 15 : 187-229.
Wilson, D. and D. Sperber (1981) On Grice's Theory of Conversation. In P. Werth (ed.) *Conversation and Discourse*, 155-78. London : Croom Helm.

---

had in St Andrews when she presented an earlier version. Thanks also to Herman Cappelen for hosting the discussion, and to a group of linguists and psychologists involved in email discussion of these issues in the early 2000s.