HAL
open science

# Semantic transparency, semantic opacity, states of affairs, mental states and speech acts

Anne Reboul

## To cite this version:

Anne Reboul. Semantic transparency, semantic opacity, states of affairs, mental states and speech acts. Say not to say. New perspectives in miscommunication, IOS Press, pp.46-71, 2001. halshs-00003826

**HAL Id: halshs-00003826**

**https://shs.hal.science/halshs-00003826**

Submitted on 4 Feb 2005

**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.

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

# 4 Semantic transparency, semantic opacity, states of affairs, mental states and speech acts

Anne REBOUL

IAGO Men should be what they seem,
Or those that be not, would they might seem none.
OTHELLO Certainly, men should be what they seem.
*Shakespeare,* Othello

If things were always what they seemed, how impoverished would be the imagination of man.
*L. Pursewarden*, *quoted by L Durrell,* Balthazar.

**Abstract:** There are two well-known views of linguistic communication: the code model and its counterpart, the hypothesis of the semantic transparency. If both of these views were correct, then there would be only one possible type of mishap in communication, that due to noise in the communication channel. However, none of these views is correct.

I will sketch a quick history of pragmatics relative both to the code model and to the hypothesis of semantic transparency. As we will see, the most recent pragmatic theories have discarded them. This enables them to account for the whole range of communication mishaps, from mere misunderstandings to sophisticated deception strategies. Incidentally it also enables them to propose a solution to Moore's paradox.

**Contents**

4.1 Introduction……2
4.2 The code model of communication and the birth of pragmatics……3
4.2.1 Two versions of the hypothesis of the semantic transparency of linguistic communication……...4
4. 2. 2 Austin and the notion of speech act……...5
4.2.3. Grice and the notion of non-natural meaning……6
4.2.4 Searle and the conventionalist trend in speech acts theory……7
4.2 Conclusion……8
4.3 Going beyond the code model: relevance theory……9
4.3.1 Misunderstandings without a noise in the communication channel……10
4.3.2 Relevance Theory……11

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

4.3.3 An extension of Relevance Theory to discourse interpretation……...14
4.3.4. Pragmatics and metarepresentation……..17
4.4 Lies……...19
4.4.1 Three interpretation strategies……..19
44.2 Lies and perlocutionary effects……20
4.4.3 Deception on Iago's mind……21
4.5 Moore's paradox……..23
4.5.1. A logical solution to Moore's paradox?……...24
4.5.2. A pragmatic solution to Moore's paradox……25
4.5.3 Moore's paradox and lies……27
4.6 Conclusion……...28
4.7 References……28

## 4.1 Introduction

It is often considered that language is — and has evolved as — a tool for communication and that it is semantically transparent relative to what is communicated by a given sentence. In other words, there would be nothing more to linguistic communication than encoding and decoding. In the scientific formulation of that hypothesis, which strongly relies on Shannon and Weaver's theory of communication [1], the only possible ground for misunderstanding is the noise which may interfere with the transmission of the signal.

There are two sides to the hypothesis of the semantic transparency of utterances: the first one, according to which utterances are semantically transparent relative to the states of affairs they represent — which was called by Austin [2] the *descriptive fallacy* —, has been largely discredited after the advent of speech acts theory; the other, according to which utterances are semantically transparent relative to the mental states of the speaker is still very active and speech acts theory has been instrumental in making it so. Whatever its version, semantic transparency is wrong not so much because language cannot represent reality, whether by *reality* one designates public states of affairs or private mental states, but because language is semantically underdetermined as has been shown by, notably, Grice [3]. The semantic underdeterminacy of language entails a corresponding semantic opacity of utterances and this, in its turn, can be a rich ground for miscommunication, be it mere (involuntary) misunderstanding or willful miscommunication.

The semantic opacity of utterances has a main consequence: the interpretation of utterances cannot be a simple matter of decoding and, more generally, the process of linguistic communication is not only a matter of encoding and decoding either. Rather the whole process rests on a complicated "system" of communicative and informative intentions on the part of the speaker and the (more or less correct) recovery of these intentions on the part of the hearer. Going from the purely linguistic interpretation of the utterance to its "complete" interpretation entails passing through a pragmatic process of "free enrichment", relying on a piecemeal context.

The present paper will be squarely in the theoretical framework of Relevance Theory [4] and will, among other things, assume its distinction between informative and communicative intentions. From that vantage point, I will try to give a general panorama of misunderstandings and various kinds of miscommunication, going from the case where the speaker is misleading his hearer unwillingly, because he himself is mistaken about the

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

state of affairs he is describing, to the case where the speaker is willfully miscommunicating, leading his hearer to believe that he (the speaker) is lying and not telling the truth, when, in fact, he is telling the truth but merely exploiting the mistrust of his hearer and leading him to the (false) belief that the propositional content of his utterance is false when it is in fact true. I will try to elucidate how those kinds of more and more Machiavellian and complicated miscommunication rely on the human ability to metarepresent one's own and others' representations or metarepresentations. I shall show that this is the basis of what Dennett [5] has called the interpretive stance and that it exploits the so-called principle of charity which is generally operative in human communication. Finally I shall draw the consequences of that for the Gricean and neo-Gricean accounts of communication and for the strong version of speech acts theory (Vanderveken's [6]) where some types of speech acts are supposed to *analytically* imply some types of mental states on the speaker's part.

## 4.2 The code model of communication and the birth of pragmatics

Suppose that you have two communicating devices, for instance human beings. These communicating devices are able to entertain internal representations, which we will call *messages*. They also are able to send *signals*, which, according to Sperber and Wilson [4], can be defined as "… a modification of the external environment which can be produced by one device and recognised by the other" (p. 4). In other words, the message which is an internal representation of communicating device CD1 is encoded as a signal, transmitted through a channel to communication device CD2, which then decodes that signal to uncover the message. Communication is successful iff, at the end of the process, CD1 and CD2 both entertain the message which, initially, was an internal representation of only CD1.

Under the code model of communication, the whole process, following a diagram by Shannon and Weaver [1], can be representated as follows (from Sperber and Wilson [4], p 5):

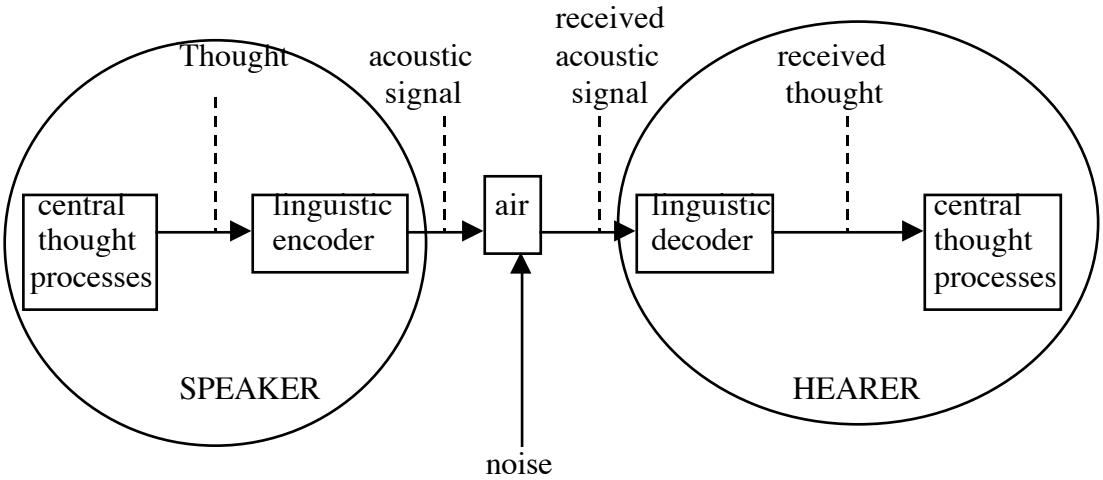


**Figure 4.1** The code model of communication

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

According to this view, there are two important hypotheses which must be taken into account in any analysis of communication:

The only possibility for miscommunication is a misunderstanding due to noise in the communication channel;

Communication is successful in as much as the message recovered is the same as the message sent, i.e. if both communicative devices share the same internal representation as a consequence of the communication process.

As we shall see, both of these hypotheses, which are still widely accepted in some linguistic circles, are mistaken.

### *4.2.1 Two versions of the hypothesis of the semantic transparency of linguistic communication*

The hypothesis of the semantic transparency of linguistic communication is obviously dependent on the code model of linguistic communication. Indeed, it is hard to see how semantic transparency could be secured outside of the code model. However, depending on the nature of the message which is communicated, it can be said that the hypothesis of the semantic transparency of linguistic communication has two sides: the first one is relative to the case in which the message being communicated is the representation of a state of affairs in the world; the second is relative to the case in which the message being communicated is not the representation of a state of affairs but is itself a state of mind of the communicative device which entertains it. A state of mind, in this instance, can be a belief, a desire, etc.

The hypothesis of the semantic transparency of linguistic communication in the case in which the message is a representation of a state of affairs in the world claims that a given signal can be decoded in such a way that it will yield a full proposition, i.e. a representation which can be evaluated in terms of its truth-value. The hypothesis of the semantic transparency of linguistic communication in the case in which the message is itself a state of mind claims that a given sentence is, so to speak, a window to the speaker's mind.

For a long time, philosophers considered that the aim of linguistic communication, and indeed the function of language was the representation of reality. They adopted the hypothesis of semantic transparency. A sentence was thus considered as a means of conveying the description of a state of affairs and its interpretation was considered as yielding a proposition, in the truth-conditional sense. In keeping with the code model, communication would be seen as successful in as much as the proposition recovered by the hearer is identical with the proposition that was communicated. Thus, until the fifties, the code model of linguistic communication as well as the semantic transparency hypothesis were concentrated on linguistic descriptions of states of affairs in the world.

However, some obvious problems were pointed out, notably regarding the use of deixis. Indeed, if John says sentence (1) in London in 1897, the proposition he expresses is quite different from the proposition that Giovani expresses when Giovani says the same sentence in Roma in 2001:

(1) I am here now.

This encouraged philosophers to distinguish between the *type* (or *sentence*) and the *token* (or *utterance*). The sentence or type is an abstract entity, which is realised on different occasions as different tokens or utterances. A given utterance of a sentence is interpreted relative to a set of parameters (i.e. speaker, addressee, time and place of utterance). There is no need to go beyond the code model. Sentences such as (1) can still be considered as expressing propositions when they are uttered, though the proposition expressed by a given utterance of a sentence can differ with that expressed by another utterance of the same sentence, if deictics are involved. Thus, the propositions expressed by John and by Giovani are respectively represented (informally) as (2) and (3):

(2) John is in London in 1897.
(3) Giovani is in Roma in 2001.

*4. 2. 2 Austin and the notion of speech act*

This was the situation in 1955 when John Austin, an Oxford philosopher, gave the *William James Lectures* at Harvard [2], roughly a year before a symposium at MIT which involved Allen Newell, Herbert Simon, Noam Chomsky and George Miller and which is considered as marking the birth of cognitive sciences. Austin's *William James Lectures* are the true origin of a discipline, which at that time did not see itself as a cognitive science, i.e. pragmatics. Austin described the position adopted by previous analytic philosophers as the *descriptive fallacy*. He attacked the notion of language as a tool for the description of reality, by pointing out that, even if only affirmative sentences are considered, some — if not most — of them cannot properly be analysed in terms of truth-conditions and as describing reality. Hence, evaluating them for truth-value was just ineffective, most of them not yielding to an analysis according to which they were true or false. Indeed, properly considered, without any *a priori* prejudice, affirmative sentences could be ranged in two classes depending on whether they actually were, or were not, describing reality: the first class is the class of *constative* utterances, while the second is the class of *performative* utterances, so called because such non-descriptive affirmative utterances were used to act on the world rather than to describe it. Below, examples (4)-(5) and (6)-(7) are respectively examples of constative and of performative utterances:

(4) The cat is on the mat.
(5) It's raining.
(6) I promise that I will come tomorrow.
(7) I baptise you in the name of the Father, the Son and the Holy Spirit.

What is more, constative and performative utterances differed not only in their functions but also in the way they could be evaluated. Constative utterances express propositions which can be evaluated for truth-conditions, while performative utterances are evaluated relative to their success, i.e. they are not true or false but felicitous or non-felicitous.

This was Austin's position in the first few lectures he gave in the *William James Lectures* cycle. However, in the following lectures, his position changed in as much as he realised that constative and performative utterances could not be contrasted relative either to their function or to the mode of their evaluation. Indeed, constative utterances are used to

modify the hearer's states of mind, i.e. to act and can be evaluated for felicity. And performative utterances do express propositions — for instance, (6) express the proposition that the speaker will come tomorrow — which can be evaluated for truth-value and their felicity may depend on the truth or falsity of the proposition expressed. This led Austin to a modification of his view, according to which any grammatical sentence uttered in "normal" conditions was *ipso facto* the accomplishment of a speech act of a particular kind, i.e. an *illocutionary act*.

Indeed, this new version of speech act theory introduced a triple distinction between kinds of speech acts:

*The locutionary act*: the act **of** saying something;
*The illocutionary act:* the act accomplished **in** saying something;
*The perlocutionary act*: the act accomplished **by** saying something.

In other words, locutionary acts are the very acts of uttering something; illocutionary acts are, for instance, promises, orders, assertions; perlocutionary acts are, for instance, persuasion. Utterances which in the first lectures would have been classified as constative were classified as illocutionary acts of *assertion* in the following lectures.

Austin's *William James Lectures* were the foundation of a new discipline in language studies, i.e. pragmatics. However, as we will see below, the *Lectures* were posthumously published and Austin's work was pursued by John Searle, an American philosopher.

### *4.2.3. Grice and the notion of non-natural meaning*

Before we turn to Searle, it is interesting to note that another milestone for pragmatics occurred when another philosopher, Paul Grice, published a seminal paper in 1957, soberly entitled *Meaning* [7]. Ten years later, he also gave the *William James Lectures* in Harvard [3], thus securing as important a place for himself in the history of pragmatics as that of Austin.

Grice's definition of *meaning* was formulated as follows: "'[the speaker] meant something by *x*' is (roughly) equivalent to '[the speaker] intended the utterance of *x* to produce some effect in an audience by means of the recognition of this intention'" ([7], p 377). This definition, in itself, does not indicate that the recognition of the speaker's intention should be done through a decoding process, neither does it indicate that any linguistic convention is involved. Thus, the Gricean definition of meaning can be considered as a first step out of the hypothesis of semantic transparency and the code model of linguistic communication.

This was reinforced when Grice pronounced the *William James Lectures* [3], describing the communication processes involved in conversation and linguistic communication. He made it obvious that, indeed, there was more to communication than mere encoding and decoding. Grice introduced a distinction between what is said and what is communicated. What is said may indeed be recovered through an encoding-decoding process and thus may be subject to the hypothesis of semantic transparency. However, what is communicated is a different matter. While what is said is a proposition, what is communicated is both what is said (i.e. the proposition expressed) and other things which can be recovered from what is said but which are not part of the explicit content of the utterance. These are called by Grice *implicatures* and there is no question that the recovery of implicatures quite simply is not done through an encoding-decoding process. Grice's hypotheses on the processes through which implicatures are recovered are well known:

according to him, conversation is a cooperative process in which the speaker and the hearer obey the *cooperative principle*. The cooperative principle is articulated in nine *conversational maxims*, which play a role in the communication process in as much as they give rise to implicatures. Let us take the first *maxim of quantity*: *Make your contribution as informative as is required*. Suppose that you are told:

(8) Anne has four children.

On a strictly semantic understanding, (8) can be interpreted as *Anne has four children and maybe more*. This, however, is clearly not the meaning which the speaker intended to convey. That meaning is: *Anne has not more than four children*. It is recovered through an inferential process in which the hearer, relying on the first maxim of quantity, makes the hypothesis that if the speaker had intended to communicate that Anne has at least four children, he or she would have said it explicitly and that his having said (8) is a good indication that he intended to communicate that Anne has no more than four children.

Thus, Grice's account of linguistic communication goes beyond the code model in that what is said does not exhaust what is meant or communicated. Thus, though the code model applies to the recovery of what is said, it does not apply to the recovery of implicatures. Hence, the Gricean model of communication is mixed: it is codic concerning the semantic interpretation of what is said, and inferential concerning the recovery of implicatures. What is more, the communicative intentions of the speaker, or hypotheses on these, play a role in this inferential process. In this sense, Grice clearly anticipated Dennett's notion of *intentional stance* [5], which can be informally defined as follows:

> The *intentional stance* is adopted whenever one interprets another's behavior on the basis of two general principles:
> 1. Other individuals are rational agents.
> 2. They have beliefs, desires and other mental states.

Though, as we will see, Grice's position is not entirely satisfactory, the intentional stance is very much to the fore of contemporary pragmatics.

### *4.2.4 Searle and the conventionalist trend in speech acts theory*

Austin died prematurely in 1960 and his *William James Lectures* had a posthumous publication in 1962. This was not the end of speech acts theory as an American philosopher, John Searle, carried on Austin's work and gave what became for the next fifteen to twenty years the foundation of pragmatics, i.e. a basic theory of speech acts [8].

Searle's main contribution was in articulating precisely much of what was left implicit or insufficiently developped in Austin's work. His first task was to make explicit the fact that speech act theory is based on two notions: *intention* and *convention*. In other words, a speech act implies an intention of the speaker which is recovered by the hearer through linguistic convention. In order to do this, Searle concentrated on only one of the three categories of speech acts distinguished by Austin, illocutionary acts. He then borrowed Grice's definition of *meaning*, though he gave it a twist. First, he redefined the recognition of the speaker's intention in such a way that it would depend exclusively on the linguistic conventions governing the sentence uttered, i.e. on its semantics. Second, given that

illocutionary acts are supposed to be linguistically encoded, he replaced the notion of the effect that the speaker intended to produce on the hearer through his utterance — which, in Grice's definition, could be interpreted as either perlocutionary (i.e. non semantic) or illocutionary (i.e. semantic) — by the notion of the **illocutionary** effect which the speaker intended to produce in his hearer through his utterance. Hence, Searle's main goal in his book seems to be to make pragmatics — given that speech acts theory pertains to pragmatics — a part of semantics.

Another contribution of Searle's which will be of interest here is that he tried to make explicit the conditions under which a given illocutionary act is felicitous. Predictably, felicity conditions, though they partly differ from one type of illocutionary act to the next, include conditions relative to absence of noise in the communication channel, community of linguistic code and a sincerity condition which I will now detail. The sincerity condition is to be found in felicity conditions regardless of the type of speech acts considered. Basically, what it says is that the illocutionary act in question can be felicitous iff the speaker is in the required state of mind: for instance, if the illocutionary act is an assertion, the speaker must believe in the truth of the proposition expressed, while if it is a promise, the speaker must have the intention to fulfill his promise, etc. The sincerity condition is the source of a variety of difficulties for speech acts theory as I pointed out in previous works ([9], [10], [11], [12], [13]). I will have more to say about it below.

Finally, Searle introduced a distinction between the *marker of illocutionary force* and the *marker of propositional content*. In a utterance such as (9), the marker of illocutionary force is *I promise that* and the marker of propositional content is *I will come tomorrow*:

(9) I promise that I will come tomorrow.

Thus, both illocutionary force and propositional content are linguistically marked and both are subject to the code model and are considered as semantically transparent.

### *4.2.5 Conclusion*

What then of the code model in the first decades of pragmatics? As we have seen, though what pragmatics called the "descriptive fallacy" was discredited by Austin in his *William James Lectures*, nevertheless the code model was preserved in the recovery of the proposition expressed by the utterance, as well, in the later Searlian theory of speech acts, as the recovery of the illocutionary force. It is interesting to note that on Searle's view, interpreting a utterance is, just as it is for Grice, recovering the intention of the speaker. However, and this is where Gricean pragmatics and Searlian pragmatics part, the Searlian model is just the code model applied both to the recovery of propositions and to the recovery of illocutionary force, exhausting the intention of the speaker, while the Gricean model limits the code model to the recovery of what is said and defends an inferential model for the recovery of further implicatures. This inferential model heavily relies on the notion of speaker's intentions as well as on the cooperative principle and conversational maxims. By contrast, there is nothing of the kind in the Searlian model. Thus, though both models rely on a double-barreled intentional definition, where the speaker has first the intention to communicate a given content through his or her utterance and the further intention that the first intention be recognised as such, that further intention plays a major role in the interpretation process according to Grice and plays no role in the interpretation

process according to Searle. Hence, Grice's pragmatics can be said to be authentically a model of double-intentionality whereas Searle's pragmatics cannot be described in such a way. And, as we will see, double-intentionality in this authentic form is exactly what is needed to go beyond the code model of linguistic communication.

### 4.3 Going beyond the code model: relevance theory

Before going beyond the code model, let us just go back to the semantic transparency hypothesis. According to Searle, utterance are semantically transparent — in the sense that they can be interpreted along lines suggested by the code model — both relative to their propositional content and relative to their illocutionary force. According to Grice, they are semantically transparent — in the same sense — relative to what is said. Nevertheless, the code model does not suffice to determine the implicatures of a given utterance and utterances are thus not semantically transparent relative to what is being communicated in excess of what is said. This, however, might be a simple matter of a terminological disagreement, where Searle and Grice would have two different uses (and understandings) of the phrase *what is communicated*. In other words, it might be that they agree on the definition of what is said as being the sum of illocutionary force and propositional content, but disagree in as much as Searle would see what is communicated as equivalent to what is said while Grice would see what is communicated as the sum of what is said and the implicatures of the utterance. Thus, the question would be, not whether what is said is semantically transparent, but whether Grice is right to include implicatures in what is communicated, i.e. to make them a part of what the speaker intended to communicate. After all, it might be argued that the recovery of what is said is quite enough to recognize the speaker's intention and that the recovery of anything further is not necessary to it.

Again, this would mean that all of what is necessary for the analysis of linguistic communication is given by the code model with its two corollaries:

> The only possibility for miscommunication is misunderstanding due to noise in the communication channel;
>
> Communication is successful in as much as the message recovered is the same as the message sent.

According to what is said above, a possible definition for miscommunication would be that it occurs whenever communication is not successful, i.e. when there is a discrepancy between the message recovered and the message sent. This would certainly be true according to the code model. However, it would be interesting to have a more intuitive clue that miscommunication has happened. I propose that miscommunication can be said to occur when the participants in the interaction seem to be talking at cross-purposes, as, for instance, when each participant is saying the same thing over and over again. This, it should be clear, can happen when there is noise in the communication channel. People would repeat themselves in the hope of being heard or read by their audience. Whether it only happens in such a case is however a good test of the validity of both the code model and Searle's view. If, indeed, there is nothing more to what is communicated than what is said, given that what is said is supposed to be semantically transparent, then without noise in the communication channel, no miscommunication should occur.

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

### *4.3.1 Misunderstandings without a noise in the communication channel*

However, regrettably for the code model of linguistic communication, such miscommunication can occur, and in fact frequently does. Some time ago, the transcription of a radio conversation at sea between Americans and Canadians was widely circulated on the web because of its comic impact:

(10) AMERICANS: Please reroute your ship by 15 degrees North to avoid a collision. Over.
CANADIANS: Please reroute YOUR own ship by 15 degrees South to avoid a collision. Over.
AMERICANS: Here the captain of a US navy ship is speaking. I repeat: please reroute YOUR ship. Over.
CANADIANS: No, please reroute YOUR ship.
AMERICANS: THIS IS THE USS AIRCRAFT CARRIER LINCOLN, THE SECOND SHIP IN THE US NAVY. WE ARE ACCOMPANIED BY THREE DESTROYERS AND A GREAT NUMBER OF OTHER SHIPS. I ASK YOU TO REROUTE YOUR SHIP BY 15 DEGREES NORTH OR RESTRICTING MEASURES WILL BE TAKEN TO ENSURE THE SECURITY OF OUR SHIP. OVER.
CANADIANS: This is a lighthouse.

Though it is very far from certain that example (10) actually is the transcription of an authentic radio exchange between the American and Canadian navies, it is nevertheless interesting in as much as it is a good example of a misunderstanding — most of it represents Americans and Canadians as speaking at cross-puposes — and there does not seem to be anything like noise involved.

In fact, what is happening in example (10) is something over and above the predictions of the code model. Every sentence seems to be perfectly well understood by its addressee but no common ground for communication is found. The conversation, indeed, consists mainly in each speaker in his or her turn repeating his or her previous utterance, adding, in the case of the last American intervention, a clear menace. The final Canadian intervention clears the misunderstanding and gives a clue as to its origin.

In (10), the miscommunication arises through a discrepancy between the informations accessible to the Americans and those accessible to the Canadians. The Americans know the spatial position of the Canadians and the Canadians know the spatial position of the Americans. Both the Canadians and the Americans know that their respective addressee belongs to the opposite navy. However, the Canadians have an information that the Americans do not. The Canadians know that they are not on a ship (i.e. a mobile object) but in a lighthouse (i.e. a non-mobile object). The American ignorance of this fact explains the insistence on their part that the Canadians reroute. The Canadian knowledge of this fact equally explains their own insistence on the Americans rerouting. It is only when the Canadians, in their last intervention, finally say that they are in a lighthouse that the misunderstanding is cleared.

Thus in this specific case, there is a clear miscommunication without any noise in the channel of communication. What is more, it is not only the code model of communication that is weakened by such an example. It is also the Searlian position as outlined above, that is, the equivalence between what is said and what is communicated. Clearly, what is communicated, notably in the last Canadian intervention is more than what is said. What is said, on a Searlian analysis, is a propositional content *This is a lighthouse* with an illocutionary force of assertion. What is communicated is however much more than that.

What is communicated is a refusal to obey the previous American order as well as an explanation of this refusal. The interpretation of the last Canadian intervention appeals not only to Gricean maxims of conversation but also to encyclopaedic knowledge about what lighthouses are.

This raises a few questions: though the Gricean interpretation mechanism is inferential and, hence, part of the interpretation process is inferential as well, it is not clear that conversational maxims are sufficient to account for these inferential interpretation processes. For instance, how does linguistic interpretation interact with encyclopaedic knowledge in the above instance? What is more, it is very far from certain that Grice would have endorsed a contextualist pragmatics (see, e.g., Recanati [14]) and it is hard to see how encyclopedic knowledge can interfere in the inferential interpretation process on an anti-contextualist view of pragmatic interpretation. This is where yet another pragmatic theory, which might be called *post-Gricean* rather than *neo-Gricean*, comes into play.

### *4.3.2 Relevance Theory*

Relevance Theory [4] can be said to be post-Gricean, rather than merely neo-Gricean, in that, though it adopts some of the major tenets of the Gricean theory of communication, it does not adopt all of them. For instance, as Grice does, Sperber and Wilson see the communication process as a mixture of both encoding-decoding and inference. However, the frontier between what belongs to the code domain and what belongs to the inference domain does not run in the same place in Grice's theory and in Relevance Theory. This, relative to the question of semantic transparency, may be Relevance most important departure from Gricean theory.

What, in fact, Sperber and Wilson say is that semantic underdetermination is more widespread than what was previously believed and that it infects not only what is communicated over and above what is said but also what is said. They point out that it is not so much that utterances are incomplete but that language does not semantically determinate the proposition expressed by a given utterance, even if a parametric approach (see § 4.2.1) is adopted. They point out the existence of frequent lexical or syntactic ambiguities as well as the semantic underdetermination of referring expressions and the even more frequent use of loose talk. In other words, though encoding-decoding processes play a part in linguistic communication, they are far from the whole story even if the proposition expressed is concerned. What is more, and this is one point of agreement with Grice, there is much more to what is communicated than what is said. However, there is no need to postulate a cooperative principle between communicators, nor to deal with nine maxims. A single cognitive principle can do the whole job.

Let me briefly outline the theory. The first thing to note is that it may well be the first pragmatic theory to explicitly claim a place for itself among cognitive sciences. What is more, it is clearly chomskyan in its view of both what goes on in linguistics and what goes on in pragmatics, placing pragmatics squarely out of the linguistic domain. In its first guise, that is, on the first publication of [4], in 1986, Relevance Theory adopted the modular view of mind advocated by Fodor [15]. According to Fodor, the mind is organised in a hierarchical way, with vertical modules specialised in the interpretation of stimuli in a given modality and a horizontal, non specialised central system. Examples of the vertical modules would be, for instance, a visual module, an olfactory module, an auditory module, etc. Modules would output a first interpretation of a stimulus that would

then serve as input to the central system. Incidentally, on this view, the binding problem — i.e. the problem of how different stimuli all produced by the same source can be bound together to yield a recognition of this source — would be solved by the central system where inputs from the relevant modules could be considered simultaneously. Translated in the domain of linguistic communication, which was the main interest of Sperber and Wilson, the modular view of mind amounts to the following: there is a linguistic module specialised in the interpretation of linguistic stimuli, i.e. utterances. This linguistic module deals with the problems traditionally considered as pertaining to phonology, syntax and semantics. In keeping with Sperber and Wilson's view about the underdetermination of language, the output of the linguistic system is not a full propositional form, which can be evaluated for truth-value, but a logical form in which some elements (reference assignation for instance) are not fully determined. This logical form consists in a structured sequence of concepts. It is the input to the pragmatic processes which take place at the level of the central system. Pragmatic interpretation, as is the case in Grice's theory, is largely inferential and non-demonstrative. It operates not from the logical form of the utterance and from a set of Gricean maxims but on the logical form of the utterance and from a set of premises, i.e. a *context*. This, apart from the pragmatic intrusion in the determination of what is said, is the second major departure of Relevance Theory from Gricean pragmatics. Relevance Theory is and Gricean pragmatics is not contextualist. According to Relevance Theory, utterances are interpreted relative to a context, and the context is not given once and for all: it is rebuilt for each new utterance.

As said above, the context is a set of premises, i.e. of propositions which the hearer entertains as true or probably true. These propositions come from different sources:

the interpretation of previous utterances in the same conversation or discourse;
the situation in which the communication occurs, i.e. from the physical environment;
encyclopaedic knowledge, accessed through the concepts in the logical form.

However, it seems clear that these sources yield an enormous amount of potential premises and that a selection process has to apply in order to yield a manageable context. Similarly, applying deductive rules — which Sperber and Wilson describe as operating both on the premises in the context and on the logical form of the utterance — could be a non-finite process yielding ever more conclusions. Again some sort of restraining influence must be brought to bear on the interpretation processes if they are not to run away.

Relevance solution to these problems goes through a single and general cognitive process: the Principle of Relevance, to which the whole theory owes its name. The name of the principle is borrowed from Grice, but it has two main differences from its Gricean counterpart: it relies on a definition of Relevance; it subsumes all other Gricean maxims, hence its singularity. Relevance is defined (roughly) as follows:

**Relevance**:
a) The more effects a given stimulus has, the more relevant it is.
b) The less efforts a given stimulus demands for its processing, the more relevant it is.

Hence, relevance is a matter of equilibrium, of balance, between processing efforts and effects. In the last version of Relevance Theory ([4], 1995), Sperber and Wilson distinguish between two principles of Relevance, a general cognitive principle which

applies throughout cognitive processes to the effect that human cognition is geared to achieve maximum relevance and a communicative principle of relevance, limited to stimuli produced in the course of a communicative exchange. It is the second principle that we will be interested in here. It states that "Every act of ostensive communication communicates a presumption of its own optimal relevance" (p. 260). Before I articulate more precisely how the communicative Principle of Relevance should be interpreted, let me just say that it operates both on the selection of the context from the different sources indicated above and on the limitation of the inference process of interpretation.

I would like to begin the exegesis of the communicative Principle of Relevance by explaining what *ostensive communication* is. As indicated above, Relevance Theory, just as Gricean pragmatics, is an authentic double-intentionality theory and this manifests itself in that acts of communication are described as implying both an informative intention and a communicative intention. The *informative intention* is defined as the intention of "[making] manifest or more manifest to the addressee a [given] set of assumptions" (p. 58). The *communicative intention* is defined as the intention "to make mutually manifest to the communicator and the addressee that the communicator has this informative intention" (p. 61). "A fact is *manifest* to an individual at a given time iff he is capable at that time of representing it mentally and accepting its representation as true or probably true" (p. 39). The *cognitive environment* of an individual at a given time is the set of facts which are manifest to him at that time. Two individuals can share a mutual cognitive environment i.e. a cognitive environment in which it is manifest which people share it. In a mutual cognitive environment, every manifest assumption is *mutually manifest*. In other words, an act of communication is ostensive in as much as the communicator's informative intention is mutually manifest both to him and to his addressee.

As ostensive acts, communicative acts have a special claim to relevance: given that the communicator is claiming a non-negligible share of her addressee's attention and cognitive efforts, her act of communication must yield enough effects to be worth the addressee's attention. What this means is that communicative acts are interpreted in such a way that the interpretation obtained is consistent with the communicative Principle of Relevance, i.e. in such a way that the contextual effects are enough to offset the processing efforts.

Very briefly, contextual effects are cognitive effects, i.e. they exist in as much as they modify the representation which an individual entertains, at a given time, of the world. They can be of three sorts:

a new information is added to the existing representation of the world;

a pre-existing information becomes either more or less plausible, i.e. the degree of confidence with which it is entertained is modified either positively or negatively;

a pre-existing information is contradicted and either it or the new contradictory information drawn from the interpretation of the stimulus is erased, depending on the degree of confidence with which each is entertained.

Thus, *ceteris paribus*, the (right) interpretation will be the most accessible and, more generally, both the full propositional form and the implicatures of a given utterance will be considered in the order of their accessibility.

Finally, let us turn to speech acts, which were the basis of pragmatics at its birth. Sperber and Wilson very clearly part ways there with previous pragmatic accounts. According to them, two types of illocutionary acts should be clearly distinguished: those that are

socially defined (baptism, most administrative declarations, legislative acts, etc.) and which rely on non-linguistic conventions; those that rely on linguistic conventions and that are not socially determined. Though the first class may have a sociological interest, it does not pertain to pragmatics. The second class does pertain to pragmatics but, concerning it, Sperber and Wilson question one of the main assumptions of Searlian speech acts theory, that is, the hypothesis that each utterance should be classified as performing one or another illocutionary act, such as menace or prediction. This, according to Sperber and Wilson, is not the case, and the same sentence (e.g. *I will come tomorrow*) can be interpreted, depending on the context, as a promise, a menace, a prevision, etc. in defiance of the Searlian conventionalist view. Thus, it is more realistic to distinguish three classes of non-social illocutionary acts: acts of *saying that*, acts of *telling to* and acts of *asking whether*. The main thing to say is that success of one of these acts does not depend on the satisfaction of a sincerity condition as was the case in Searlian pragmatics. The central speech act for the present paper is the act of *saying that*, that Sperber and Wilson define as follows: "Let us define *saying that P*, where *P* is the propositional form of the utterance, as communicating that the thought interpreted by *P* is entertained as a description of an actual state of affairs" (p. 237). As we shall see below, this definition does not meet with the difficulties that plague accounts with sincerity conditions.

### *4.3.3 An extension of Relevance Theory to discourse interpretation*

Relevance Theory is, properly speaking, a theory of utterance interpretation, given that utterances are considered as a special case of ostensive communication. It can, however, also deal with all manners of ostensive communication from non-verbal communication to verbal communication extended over the limits of an utterance. In other words, it can account not only for utterance interpretation, but also for discourse interpretation (for a detailed defense of this claim, see [16], [17], [18], [19] and [20]). The first thing any theory of discourse should presumably do is to offer a definition of it. My own would run as follows: *A discourse is a non-arbitrary sequence of utterances.* This much should not be controversial. Controversy begins, however, with the seemingly inoffensive expression *non-arbitrary*, which can be interpreted in two quite different ways: as implying that the non-arbitrariness of discourse is due to the fact that there are linguistic rules governing the production and interpretation of discourse; as implying that discourse is non-arbitrary in a pragmatic sense, i.e. there are pragmatic rules governing the production and interpretation of discourse. These two different interpretations have widely different implications: in the first case, *discourse* can be — and has often been — seen as another linguistic unit, to be added to *phoneme*, *morpheme* and *sentence*; in the second, discourse can be seen as the product of pragmatic rules which would be specific to it or as the product of the same rules that govern the production and interpretation of utterances and it should be noted that the choice between the two depends on which version of pragmatic theory one favors. If one adopts a conventionalist view of pragmatics, then one will probably choose the first interpretation; if one adopts, as I do, a post-gricean theoretical framework, such as Relevance Theory, one will probably defend the second interpretation. The choice between those two interpretations is not trivial: choosing the first one would presumably lead to considering discourse as a pragmatic unit, whereas choosing the second would not (discourse would merely be considered as just another act of ostensive-inferential communication). The preceding distinction between a linguistic or conventional pragmatic

approach and a radical pragmatic approach has a corollary: if there are rules — linguistic or conventionally pragmatic — specific to discourse, those rules would govern the **form** of the discourse, not its **content**, just as syntactic rules govern the form of the sentence, not its content. Non conventional pragmatic rules would not only not be specific to discourse; they would just relate to its content, not its form. This is the view briefly outlined below.

In a linguistic theory of discourse, there are linguistic rules relative to discourse; in a conventionalist pragmatic theory of discourse, there are pragmatic rules relative to discourse. In both cases, the non-arbitrariness of discourse can be accounted through the existence of a set of rules which would be specific to it and whose application would guarantee its *coherence*, just as the application of syntactic rules guarantees the grammaticality of a sentence. In a pragmatic framework which does not hypothesize the existence of such a set of rules, how can one account for the non-arbitrariness of discourse, and — even more crucially according to some currents in discourse analysis — account for its coherence or incoherence?

An answer to this question can be found when one considers that, indeed, discourse is a case of ostensive communication, and, as such, implies the same double intentionality as do utterances. What is more, just as the Principle of Relevance and non-demonstrative inference processes can account for the recovery of intentions behind utterance production, they can account for the recovery of the intentions behind discourse production. In effect, this means that the two types of intentions in utterance, i.e. informative intention and communicative intention, are also to be found in discourse. Thus, the speaker of a discourse has a global informative intention and a global communicative intention, concerning the whole of his discourse, just as she has a local informative intention and a local communicative intention for each utterance in her discourse. These global intentions are recovered through the interpretation of the utterances composing the discourse, though the process is not purely incremental (i.e. the product of the interpretation of a discourse is not a sequence — or a sum — of the products of the interpretation of the successive utterances in that discourse) and this is where non-arbitrariness creeps back in. Why do we need global intentions in addition to local intentions? Well, this of course hinges on the fact that local intentions in Relevance are taken to concern utterances, not sequences of utterances and though they are enough (together with the rest of Relevance) to account for the fact that utterances are interpreted relative to context (which includes much more than just the interpretation of previous utterances), they are not quite enough to explain why, in a discourse, a given utterance occurs when it does. This is, among other things, because speaker's intentions regarding discourse are not always straightforward: for instance, in literature, an author may willfully mislead her audience, that is, some utterances may lead the reader in a direction which is not consistent with the whole of the text, before surprising him with the conclusion (e.g. A. Christie in *The Murder of Roger Ackroyd*). In this specific case, the reader will eventually recover as the global informative intention the culpability of the doctor, whereas he might legitimately have recovered most local informative intentions as indicating the doctor's innocence. By the way, this is also why the interpretation of the discourse is not equivalent to the sum of the interpretation of the successive utterances, though each of these local interpretations is done relative to a context. And, again, this is a matter of contents and relations between them and not of form.

The non-arbitrariness of discourse is due to two main factors: the whole of the discourse falls under the global communicative intention (i.e. it **is** an act of **willful** communication)

and the local informative intention for each utterance is determined at least in part by the global informative intention. This is also what accounts for coherence and here is how it does it. *Coherence* has been traditionally regarded in linguistic discourse analysis as the discourse equivalent of grammaticality in syntax: it was thought that just as grammaticality is defined in syntax as the property of sentences, coherence can be defined as the property of discourses. The threat of circularity is to be avoided in both cases through the fact that (syntactic or discursive) rules are made explicit and are exhaustively described. One can criticize syntax for not being exhaustive; however, and that is much worse, one can criticize linguistic discourse analysis for not being explicit at all, i.e. linguistic discourse analysis does not seem to have produced any set of rules. Given that, it seems that coherence is at best a pre-theoretical notion which cannot account for anything (and certainly not for the non-arbitrariness of discourse), but which itself is in need of an explanation. I think that the framework proposed above can do precisely that without in any way presupposing that coherence is a theoretical notion and without using it theoretically.

In saying that coherence is a pre-theoretic notion, I mean that one should not try to give it a definition which gets beyond its vernacular use. There is no way to deny that people do have spontaneous judgment of coherence about discourses, just as they have spontaneous judgment of grammaticality about sentences. However, grammaticality can attain theoretical status through the production of an explicit set of rules aiming at exhaustivity (Generative Grammar comes to mind). As said above, it is doubtful that this is the case for coherence. Hence, coherence is still in need of an explanation. The explanation which can be offered in a relevance oriented pragmatics of discourse such as that proposed above is the following: a discourse will be judged as more or less coherent by its hearer depending on the ease with which the global informative intention can be recovered and depending on the richness and complexity of the information transmitted. That is, depending on the effort necessary for the interpretation of the discourse and on its effects, i.e. **depending on its overall relevance**. Note that all discourses will not provoke the same expectations of coherence (i.e. of relevance): the coherence expected of a scientific text is presumably much higher than the coherence expected of a discussion between friends in a restaurant. But that does not detract for the ability of a relevance oriented pragmatics of discourse to both account for the non-arbitrariness of discourse and explain coherence judgements.

Here is a short, authentic and unabridged example:

> (11) Dare I tell the story which was confided to me while taking the shade under the shadow of a graveyard wall in a delightfully green square of alfafa? Why not? I am already dishonored as saying truths which shock the fashion of 1838:
>
> The curate was not old; the maid was pretty; people were talking which did not discourage a young man from a near village to court the maid. One day, he hides the kitchen tongs in the maid's bed. When he came back a week later, the maid said:
>
> "Tell me where you put the tongs which I have looked for everywhere since your departure. This is a very bad joke"
>
> The lover kissed her, with tears in his eyes, and went away.
>
> Stendhal, *Voyage dans le midi*, My translation.

I will take it for granted that no one would find this text incoherent and that all judgments of coherence would be favorable. The question is *why is it (judged to be) coherent?*

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

The first thing to note is that there are no connectives in this text and very few pronouns. In fact, almost the only expressions which could related to the judgment of coherence are temporal (e.g. *a week later*), though they are certainly not sufficient to account for it. I think that the best way to answer the question is in terms of the hypotheses which Stendhal sows in the reader's mind as to what he intends to tell (i.e. these hypotheses concern not so much the local informative intention of each successive utterance than the global informative intention of the whole text) and this begins with the first utterance of the text. He then exploits the reader's encyclopaedic knowledge with the first utterances of the story itself, i.e. curates are male and maids female and young people of both sexes can fall in love. The following utterances of the text mainly confirm these hypotheses.

Why is the extension of Relevance Theory to discourse important for our present concern? As indicated in the introduction, I will be interested in what follows both in falsity masquerading as truth and in truth masquerading in falsity, i.e. in lies. It is a commonplace to say that the main difficulty involved in lying is not so much on the inventive side (i.e. in the content of the lie), but in keeping that content consistent with the rest of one's discourse, i.e. the main difficulty occurs not at the level of the utterance, but at the level of discourse. Before turning to lies of both kinds, I, however, want to outline the relations between pragmatics (specifically the interpretation of utterances) and metarepresentation.

### *4.3.4. Pragmatics and metarepresentation*

Before turning to lies of all hues, I would like to outline the potentially complex relations between pragmatic interpretation and metarepresentation. First of all, let me say what I mean by *metarepresentation*, relying on the very sophisticated discussion given in Perner's book [21]. Perner squarely belongs to the representational tradition in cognitive sciences adopted here. He distinguishes three kinds of representations:

- *primary representations*, representations of an object, i.e. a photograph of myself is a primary representation of myself;
- *secondary representations,* voluntary *misrepresentations* of an object, i.e. representations in which a property is falsely attributed to an object, as, for instance, in hypothetical reasoning;
- *metarepresentations*, "represent[ing] that something (another organism) is representing something" (p. 7), and how it is representing it.

This may be clearer from a figure (see figure 4.2).

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

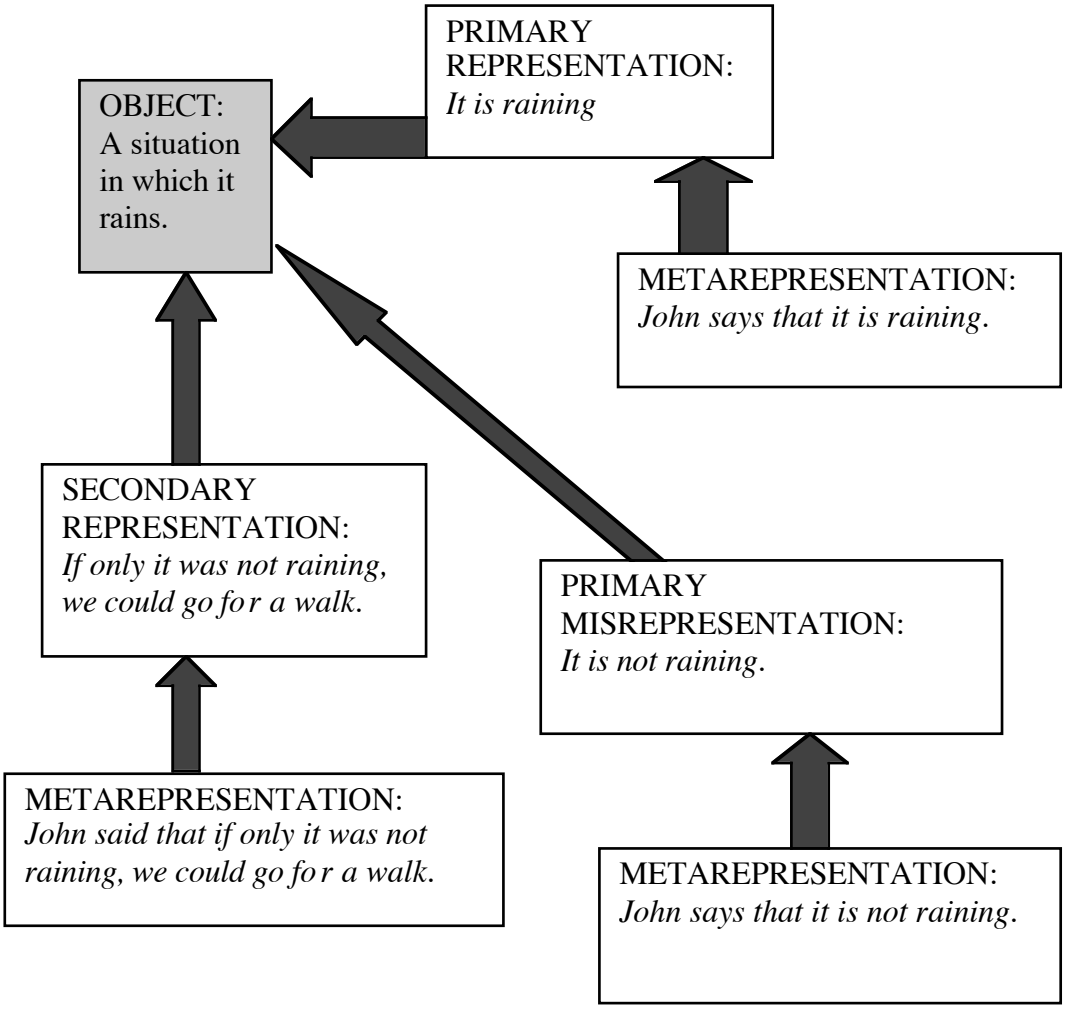


**Figure 4.2** The different sorts of representation

Let us forget secondary representations and misrepresentations for the moment and concentrate just on representation and metarepresentation. On a representational view of cognition, mental states are, or at least involve, representations. This is the case, for instance, for intentions. Thus the intentions which are recovered in utterance interpretation are representations. However, given that what is recovered is an attributed intention, the form of utterance interpretation will be a representation (on the hearer's part) of the intention of the speaker, hence a metarepresentation. This much is the common ground behind all pragmatic accounts, whether conventionalist or not, and whether contextualist or not. Can we go farther than that? On the Searlian conventionalist account, as said above (see § 4.2.5), it does not seem to be the case. What, however, can be said of the Gricean and post-Gricean accounts?

On a Gricean account, it seems obvious that metarepresentation has a role to play in the interpretative process itself, notably when conversational maxims are flouted. Then, the hearer must go through a complex and rather cumbersome reasoning process as to what the intentions of the speaker could be, etc. On a post-Gricean account, however, things are

more simple: no such reasoning process is necessary though metarepresentation is the reason why the communicative Principle of Relevance comes into play. However, it does not necessarily play a role in the communication process, as long as everything goes well. The story, however, may become different when communication is not straightforward, i.e. in case of miscommunication, or in case of the detection of lies.

## 4.4 Lies

As said above, I will be concerned here with two kinds of lies which I will designate, rather ornately, as *falsity masquerading as truth* and as *truth masquerading as falsity*. Both have a common aim: leading the hearer to a false belief or, more precisely, leading the hearer to believe something which the speaker believes to be false. I will have more to say about lies below, but before I turn to them, I would like to borrow a few reflections on metarepresentation and utterance interpretation proposed by Dan Sperber [22].

### *4.4.1 Three interpretation strategies*

Not surprisingly, Sperber bases his paper on the assumption that language is semantically underdetermined. The consequence of this is that utterances are semantically opaque, that is, no amount of decoding is going to yield the interpretation of a given utterance. Hence, the linguistic decoding interpretation process has to be supplemented by a pragmatic inferential interpretation process. This implies a modicum of metarepresentation. However how much metarepresentation is involved in the pragmatic interpretation process is going to depend on the degree of sophistication of the interpretation strategy adopted by the hearer. On lines reminiscent of the three little pigs of the children tale, Sperber describes three strategies ranging from the less sophisticated to the more sophisticated. Interestingly, the less sophisticated here means not only the less complex as far as processes go, but also the more likely to lead the hearer in the clutches of a less than benevolent speaker, i.e. to lead the hearer to believe any lie which the speaker might intend him to believe. By contrast, the more sophisticated the strategy, the less risks there is that the hearer will be led to false beliefs.

As Sperber points out, given Relevance Theory, there are two ways in which the communication process can flounder, i.e. not recover the speaker's intentions:

the speaker may be incompetent, thus leading the interpretation process astray by not producing an utterance which can be interpreted so as to yield an interpretation consistent both with the communicative Principle of Relevance and with the speaker's informative intention;

the speaker may be less than benevolent, i.e. she can want to deceive the hearer.

Given those two ways of producing miscommunication, the three strategies are distinguished by whether they take them into account or not. The first strategy is *naive optimism.* In naive optimism, the hearer (let us call him John, following Sperber [22]) considers the speaker (let us call her Carol, still following Sperber) as both competent and benevolent. The only thing he has to do to interpret Carol's utterance is to follow the communicative Principle of Relevance and accept the first interpretation available to him which is consistent with it. The second strategy is *cautious optimism.* In cautious optimism, Carol is presumed to be benevolent, but not necessarily competent. In this

strategy, John will again follow the principle of least effort, but he will not adopt the first interpretation consistent with the communicative Principle of Relevance. Rather, he will adapt to what he believes true of Carol, and, notably, what he believes that Carol believes about him. In other words, he will adopt not the first interpretation consistent with the communicative Principle of Relevance but the first interpretation that Carol might have believed to be relevant to him, John. This entails a good amount of metarepresentation on John's part. The third strategy is *sophisticated understanding*. Under sophisticated understanding, the speaker is believed to be neither competent nor benevolent. She is assumed to **seem** competent and benevolent.

The three strategies differ not only in the amount of metarepresentation involved, but also in the level of metarepresentation involved. Let us look at the following examples where different degrees of metarepresentation are involved:

(12) **Primary representation:** *It is raining.*
(13) **First-order metarepresentations:** *John says/believes/intends to communicate [that it is raining]*
(14) **Second-order metarepresentation:** *John intends [me to believe [that it is raining]].*
(15) **Third-order metarepresentation:** *John intends [me to believe [that he believes/intends to communicate [that it is raining]]].*
(16) etc.

There are, in principle, no logical limits to the orders of metarepresentation, though cognitive limits do exist. The naive optimism strategy implies the capacity in the speaker and the hearer of forming first-order metarepresentations. The cautious optimism strategy implies a capacity to form second-order metarepresentations, while the sophisticated understanding strategy implies the capacity of forming at least third-order metarepresentations.

It should also be noted that each strategy is more costly in terms of processing efforts than is the previous one. Thus, *ceteris paribus*, in most circumstances, we will tend to adopt the naive optimism strategy. It is only in specific cases, for instance when we know that the speaker is likely to be incompetent (if, for instance, it is a small child), or when the stakes are particularly important to us that we resort either to cautious optimism or to sophisticated understanding. This leaves a loophole in the communication process, which unscrupulous speakers can and frequently do exploit.

### *44.2 Lies and perlocutionary effects*

Lies are one of the stumbling blocks of Searlian speech acts theory. The difficulties it raises for it mainly depend on the redefinition of meaning by Searle and on the place of the sincerity condition in the felicity rules for assertion. In order to see why this should be so, let me begin by a definition of a lie. It is generally thought that lying is saying something false. However, as was pointed out by Davidson [23], this is incorrect: "It is sometimes said that telling a lie entails what is false; but this is wrong. Telling a lie requires not that what you say be false but that you think it false" (p. 258). I will however not define lying through the nature of what is said: whether we take lies to entail saying what is false or saying what we believe to be false, this does not make lies different from most metaphors.

Hence it is hardly a defining characteristic of lies. I will rather define a lie through a general characteristic of the speaker's intention. The speaker of a lie has the intention to produce a specific effect: that his hearer should believe something which he, the speaker, takes, rightly or wrongly, to be false. What is more, he intends to obtain this effect through his utterance(s). This, it should be clear, is not an illocutionary effect. It is a perlocutionary effect. As such, it does not tally with Searle's redefinition of meaning along (what Searle takes to be) Gricean lines. In that definition, Searle insist that the effect intended cannot be anything but illocutionary.

I want to defend the idea that there are two ways of achieving the intended effect of a lie: you can say something which you believe to be false and get your hearer to think that you have said something which you believe to be true on good grounds; you can say something which you believe to be true and get your hearer to think that you have said something that you believe to be false on good grounds, leading him to adopt a contradictory belief (which you think false). There are thus two varieties of lies though the first one is the one generally considered. To avoid misunderstandings, I will reserve the word *lie* to the first variety, which I will characterize as falsity masquerading as truth. I will call the other variety, truth masquerading as falsity, *truthful deception*. There is a third kind of deception, apart from lies and truthful deceptions: it occurs when you lead your hearer to derive a false conclusion. I will call it *Machiavellian deception*.

Why, apart from the fact that deception in general aims at a perlocutionary rather than illocutionary effect, do lies make any difficulty for Searle's theory of speech acts? It seems impossible to describe lies in terms of speech acts and yet it seems rather weird to deny that lying is accomplishing a speech act. Let us suppose that lies are indeed a type of speech act. It seems clear, given that deception is involved, that lies cannot be illocutionary acts. This, however, does not mean that they cannot be another type of speech act and that they cannot have felicity conditions. Let us suppose that this is the case. Lies are perlocutionary acts and their felicity condition is that the intention of the liar should be satisfied, i.e. the hearer believes that the propositional content of the lie is true. Under speech acts theory, any utterance must correspond to the accomplishment of an illocutionary act. I will here adopt the simplifying assumption that all lies correspond to the performance of an illocutionary act of assertion. However, this cannot be the case, given that a defining condition of assertion is the sincerity condition, which is obviously not satisfied in lies. However, it does seem that if a lie is to be successful, the corresponding assertion also should be successful. But this cannot be the case if the sincerity condition is not satisfied. Thus, it does not seem to be the case that lies can be given a satisfying description in Searlian speech acts theory.

It should be noted, however, that, on Sperber and Wilson's definition (see § 4.3.3) of the act of *saying that*, lies are just a variety of saying that. Indeed, *saying that P* is merely described "as communicating that the thought interpreted by *P* is entertained as a description of an actual state of affairs", which does not imply any kind of sincerity condition.

### *4.4.3 Deception on Iago's mind*

Iago is well known as the epitome example of a deceitful man. *Othello* will be my main illustration in the remaining part of this paper. Its scenario is well known, but I will briefly recapitulate it. Iago has been bypassed by Othello, a Moor employed by Venice in its fight

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

against the Turks, who has chosen Cassio, of whom he is fond, as his officer, leaving to Iago the less prestigious role of ensign. Iago, to revenge himself, decides to make Othello believe that his wife, Desdemona, is Cassio's mistress. To this aim, he gives Othello a few hints and tops his work by a "proof", an handkerchief of Desdemona, given to her by Othello, which he steals and plants in Cassio's room. Othello, maddened by jealousy, kills Desdemona, understands his error and commits suicide on his wife's body. The turning point in the play occurs in the third scene of the third act. It begins when Cassio has been disgraced through a plot of Iago and has asked Desdemona to intercede with Othello on his behalf. This is when Iago begins to lead Othello slowly but surely to the belief that his wife is unfaithful. Scene III, 3 begins with Othello secure in his marital bliss and ends with Othello sure of his cuckold status. The question is how does Iago manage to bring about this change?

Iago's deceptions cover the three strategies outlined above: lies, truthful deception and Machiavellian deception. Let us, however, recognise that Othello is the ideal victim, only stirring out of naive optimism when Iago leads him to it.

Let us try and describe the whole process through the categories defined above. What then is Iago's global informative intention? His global informative intention is to make it manifest to Othello that Desdemona is Cassio's lover. He begins by subtly drawing Othello's attention to the fact that Cassio, who was speaking with Desdemona and asking her help, is leaving subrepticely. To do this, Iago mutters: "Ha! I like not that". His aim is to be heard, but not understood by Othello and to induce in Othello the belief that his utterance was an involuntary exclamation. His local informative intention is to make Othello believe that he, Iago, uttered an exclamation of dismay. Questioned by Othello on what he says, he reinforces Othello's initial interpretation by saying: "Nothing my lord. Or if, I know not what". In this exchange, Iago adopts the Machiavellian deception strategy. It is not the content of his utterances which Othello should believe: it is the inference that Othello draws from his interpretation of Iago's attitude which leads him to a dim kind of disquiet. The next episode where Iago again is active consists in Iago asking Othello whether Cassio was involved in Othello's wooing of Desdemona. This leads Othello to question him on his opinion of Cassio's honesty. This is where Iago resorts to truthful deception. His answers affirm Cassio's honesty, and he does believe Cassio to be honest. The whole exchange between Othello and Iago at this point deserves to be quoted in full:

(17) IAGO For Michal Cassio,
I dare be sworn I think that he is honest.
OTHELLO I think so too.
IAGO Men should be what they seem,
Or those that be not, would they might seem none.
OTHELLO Certain, men should be what they seem.
IAGO Why then, I think Cassio's an honest man.
OTHELLO Nay, yet there's more in this.
I prithee speak to me as to thy thinkings,
As thou dost ruminate, and give thy worst thoughts
The worst of words.

The situation in this exchange is the following. Iago knows Cassio to be honest. He says so. But he says it in such a way that Othello is led to the belief that Iago thinks Cassio to be dishonest. In terms of metarepresentation, this could be formulated as follows:

(18) Othello believes [that Iago thinks [that Cassio is dishonest]]

Iago's intention is the following:

(19) Iago intends [that Othello believes [that Iago thinks [that Cassio is dishonest]]]

His strategy can be articulated as:

(20) a. Iago intends [that Othello believes [that Iago says [that Cassio is honest]]]
b. Iago intends [that Othello believes [that Iago does not believe [that Cassio is honest]]]
c. To achieve b., Iago intends that [Othello believes [that Iago lies [in saying that Cassio is honest]]]
d. To achieve c, Iago intend that [Othello believes [that Iago does not want [Othello to worry about Cassio's and Desdemona's relationship]]]

This, in effect, is exactly what happens. Let us now turn to a lie.

One occurs later in the scene, after Iago has adjured Othello not to fall into suspicion regarding Cassio and Desdemona. Othello assures him, falsely, that he has not. Iago then says:

(21) IAGO Should you do so, my lord,
My speech should fall into such vile success
Which my thoughts aimed not.

Here, what Iago says is that Othello succumbing to jealousy was not his (Iago's) intention. It is of course a blatant lie. What is Iago's intention?

(22) Iago intends [that Othello believes [that Iago did not intend [Othello to be jealous]]]

All of this should now be assessed relative to the notions of semantic transparency and semantic opacity. Linguistic underdetermination has a corollary: utterances are semantically transparent neither relative to their propositional content nor relative to the speaker's state of mind. This is why Machiavellian deception is possible: it relies on exactly the same inferential processes as utterances where no Machiavellian deception applies, as for instance, in the interpretation of example (11). This is also why truthful deception is possible: utterances are not always taken to be sincere or truthful and they are not considered as semantically transparent relative to the speaker's state of mind. So far so good, but if this is the case, how are lies possible? Why does the sophisticated understanding strategy not apply across the board? To answer this question, I propose to quickly examine a well known philosophical puzzle, Moore's paradox.

## 4.5 Moore's paradox

Moore's paradox was proposed by the Cambridge philosopher George Moore and can be enunciated in the following way:

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

(23) It rains but I don't believe it rains.

The first thing to note is that, strictly speaking, Moore's paradox is not a contradiction — it does not contravene the law of excluded middle —, neither is it a paradox in the sense in which, for instance, the liar paradox is:

(24) I'm lying.

Thus the paradox in Moore's paradox resides in the fact that we consider such a utterance as (23) as weird in some mysterious sense, though it does not seem to have any of the anomalies listed in catalogues of semantic teratology. So why is Moore's paradox intuitively bizarre and what is its relationship with lies?

The present section will be devoted to an answer to the first part of this question and this answer will allow us to deal with its second part.

*4.5.1. A logical solution to Moore's paradox?*

A possible solution to Moore's paradox would be to show that, despite appearances, it in fact is a contradiction. This seems to be the position adopted, notably, by Vanderveken [6] and Levinson [24]. In other words, the solution would consist in considering the utterance in (23) as semantically equivalent with either (25) or (26):

(25) I believe it is raining and I don't believe it is raining.
(26) It is raining and it is not raining.

The first thing to note is that there is no problem in replacing *but* in (23) by *and* in (25)-(26), given that *but* is generally considered as logically equivalent to *and*. However for (25) to be equivalent to (23), it should be considered that saying *It's raining* is logically equivalent to saying *I believe it's raining*. Similarly, for (26) to be equivalent to (23), it should be considered that saying *I don't believe it is raining* is logically equivalent to saying *It is not raining*. How likely is that?

Saying that two linguistic expressions are semantically or logically equivalent is saying that they make the same contribution to the truth-conditions of any sentence where they might occur. In other words, in a given sentence, replacing one of two semantically equivalent expressions by the other will not change the truth-value of the whole sentence. When the utterance being considered is a compound of two sentences, each of which expressing a proposition, saying that one of the sentence in the compound is semantically equivalent with another sentence is just saying that replacing the sentence in the compound with the semantically equivalent sentence will not change the truth value of the compound. It is on this reasoning that advocates of the logical solution to Moore's paradox base their argumentation. As, for instance, *It is raining* is semantically equivalent with *I believe that it is raining*, (25) is semantically equivalent with (23) and thus both are necessarily false because they are contradictions.

This is good reasoning in itself apart from the fact that it begs the main question which is that of the logical equivalence of either *it is raining* and *I believe it is raining* or *I don't believe it is raining* and *It is not raining.* What ground is there for thinking that it is so?

Logical equivalence between sentences means quite simply that the two sentences have the same truth-values. One way of ensuring that is to make sure that they have the same truth-conditions, i.e. that the truth of either of them implies the truth of the other. In other words, for *It is raining* and *I believe it is raining* to be equivalent, both (27) and (28) — where the ⇒ sign is the logical sign for material implication — would have to be true:

(27) It is raining ⇒ I believe it is raining
(28) I believe it is raining ⇒ It is raining

For (27) to be true, semantic transparency relative to the speaker's state of mind would have to be true and, as we have seen before, this is just not plausible. For (28) to be true, an even less likely condition should be met, i.e. that the speaker is infallible.

Let us now turn to the second possible equivalence. Again, for *I don't believe that it is raining* to be semantically equivalent to *It is not raining*, both (29) and (30) should be true:

(29) I don't believe that it is raining ⇒ it is not raining
(30) It is not raining ⇒ I don't believe it is raining

(30) corresponds to an operation called *Negation raising* and is seen as legitimate. However, no such legitimacy is attached to (29), which would depend on two operations. The first one would be to consider that *I don't believe that it is raining* is semantically equivalent to *I believe that it is not raining*. The second one would consist in an elimination of the preface *I believe.* None of these two operations are legitimate: I can not believe that it is raining without simultaneously believing that it is not raining. For instance, I may just not know whether it is or it is not raining. And, as pointed out before, unless I am infallible, my beliefs do not necessarily come true and, hence, it may well be the case both that *I believe that P* and *not-P*. This, incidentally, is why Moore's paradox is not a contradiction.

Thus, it does not seem to be the case that there is a logical solution to Moore's paradox and neither can it be seen as contradiction. As a last argument against the logical solution, let me borrow (with modifications) two examples from Tsohatzidis [25]:

(31) I don't believe that he feels sick but as he says he does and as his words would be accepted as the best evidence of his feeling sick, **I assert** that he feels sick.
(32) *I don't believe that he feels sick but as he says he does and as his words would be accepted as the best evidence of his feeling sick, **I believe** that he feels sick.

Examples (31) and (32) are identical except for the replacement of *I assert* in (31) by *I believe* in (32). (32) is definitely contradictory while (31) is perfectly acceptable. If semantic transparency relative to the speaker's state of mind was a sound hypothesis, assertion — as in (31) — and belief — as in (32) — would be equivalent. The comparison between (31) and (32) shows that this is not the case.

### *4.5.2. A pragmatic solution to Moore's paradox*

Thus, Moore's paradox is not a contradiction. What should be done, then, is explaining why (23) is, nonetheless, weird. The solution outlined below is a pragmatic solution,

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

where *pragmatic* implies cognitive considerations. The first thing I would like to insist upon here is the indexicality of Moore's paradox. Let us compare (23) and (33) and (34):

(23) It is raining but I don't believe it is raining.
(33) Mary says it is raining, but I don't believe it is raining.
(34) It is raining in Caracas, but I don't believe it is raining in Lyons.

The weirdness of (23) disappears in all of (33) and (34). In other words, it is essential to Moore's paradox that the person who says that it is raining should be the same as the person who says that he/she does not believe that it is raining, and it is just as essential that the location where it rains should be the same in the first and the second clauses of the sentence. This should, on the face of it, be mysterious: there is no location explicitly designated in (23).

An answer to that problem has been suggested by American philosopher John Perry from example (35):

(35) It is raining.

As Perry remarks, the truth-value of this utterance can only be evaluated relative to a precise place, though no place is mentioned in (35). He introduces the notion of *unarticulated constituent* (see Perry [26]). An unarticulated constituent is a constituent of truth-conditions which is solved directly on the situation in which the communication takes place, but which is not mentioned explicitly in the utterance. Thus unarticulated constituents share with indexicals their deictic aspect. What is more, Perry introduced another notion, the notion that some indexicals are *essential*. Presumably, the best way of explaining Perry's ideas on the matter is to quote his example ([27], p. 33):

(36) "I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch.
I believed at the outset that the shopper with a torn sack was making a mess. And I was right. But I did not believe that I was making a mess. That seems to be something I came to believe. And when I came to believe that, I stopped following the trail around the counter and rearranged the torn sack in my cart. My change in beliefs seems to explain my change in behavior."

The notion of *essential indexical* applies to those indexicals which are indispensable to explain the behavior or the changes in behavior. In the above example, replace the indexical by a co-referring expression (e.g. *John Perry*) and you will loose the explanatory weight of *I am the shopper with a torn sack*. In this case, indexicality is essential in that it is impossible to eliminate it. Not all indexicals are essential however. Only indexicals which both are interpreted relative to the situation in which the communication takes place and which refer to components of that situation are essential in the above sense. This means that essential indexicals are *I*, *here*, *now* and the unarticulated constituents. Unarticulated constituents intervene to fix the location where it is raining though the

speaker does not believe it in (23). However, they are not indispensable to the weirdness of Moore's paradox: identity of location, however it is ensured, is.

Let us extend the notion of unarticulated constituent out of truth-conditions and include in it a non-vericonditional preface any utterance: in other words, let us suppose that each utterance is accompanied by the unarticulated constituent *<I speak>*. Thus, (35) — which is identical with the first part of (23) — should be understood (non-vericonditionally) as:

(37) <I speak> it rains

This allows the following (not semantically equivalent to (35)) deduction:

(38) (<I speak> it rains) ⇒ I say that it rains

Let us now resort to the intentional stance (see § 4.2.3). According to it, the interpretation of behavior relies on the attribution to other people of states of mind, through the hypothesis that people are rational. Suppose that we combine the intentional stance with a naive optimism strategy regarding utterance interpretation. Then from the fact represented in (38), we will deduce that indeed the speaker of (35) believes that it is raining. However, when we apply the intentional stance and the naive optimism strategy to Moore's paradox, we find ourselves faced by a contradiction between the state of mind we attribute to the speaker from the interpretation of the first sentence of the compound and the state of mind he describes himself as being in in the second sentence of the compound. There is a contradiction, but that contradiction is not between the two sentences of the compound. It is between a deduction which the first sentence seems to allow and the propositional content of the second sentence. That, in itself would not be sufficient to explain the weirdness because, after all, we only have to change our strategy and abandon naive optimism in favor of sophisticated understanding. This would leave us with a difficulty, however, in that the intentional stance would still apply but its second requisite would not be complied with. Indeed, it seems that there is no way to attribute to the speaker of (23) one or several states of mind consistent with the rule that the speaker is considered as a rational individual.

Let me explain shortly why this is so. Let us suppose that the speaker of (23) is sincere: thus, when he says *It is raining* he believes that it is raining; similarly when he says *I don't believe it is raining* he does not believe that it is raining; however, it is not possible to believe both that it is raining and that it is not raining of the same place; hence, the speaker of (23) cannot be rational. Let us now suppose that the speaker of (23) is not sincere, i.e. he is lying: when he says *It is raining*, he intends his hearer to believe that he (the speaker) believes that it is raining; similarly, when he says *I don't believe that it is raining* he wants to make his hearer believe that he (the speaker) does not believe that it is raining; it is impossible to have and to satisfy simultaneously two contradictory intentions; hence, the speaker of (23) is not rational.

In other words, Moore's paradox draws its weirdness from the fact that it is impossible to attribute to its hypothetical speaker a state of mind consistent with the rationality requirement of the intentional stance.

### *4.5.3 Moore's paradox and lies*

We are free to adopt any interpretative strategy we want, as long as it yields results consistent with the intentional stance. As we saw before, though the most reliable strategy presumably is sophisticated understanding, it also is, unsurprisingly, the most costly strategy. This is why we generally adopt naive optimism unless we have good reasons to believe that the speaker is neither competent nor benevolent. This also is why we tend to believe the lies we are told, as long as the liar is competent enough not to get himself into obvious incoherence. This is also why we tend to believe that it is raining when in fact it is not and why there is a tendency to believe that Moore's paradox is a contradiction. However, it is not.

## 4.6 Conclusion

Semantic transparency is just a myth. The code model of linguistic communication is far from satisfying. However, there are models that can do justice to the intrinsic complexity of human communication. Relevance Theory is such a model as I hope to have shown above. What is more, a plausible model of communication should cover all the avatars of communication, from perfect functioning to gross imperfection. This, I think, is just what the code model and the hypothesis of semantic transparency cannot do. Relevance Theory can and, hence, it is, now, the most promising model for communication.

## 4.7 References

[1] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949.
[2] J. Austin, *How to do things with words*. Oxford: Clarendon, 1962.
[3] P. Grice, *Studies in the way of words*.Cambridge, MA: Harvard University Press.
[4] D. Sperber and D. Wilson, *Relevance: Communication and Cognition*. Oxford: Basil Blackwell, 1996 (2nd edition).
[5] D. Dennett, *The intentional stance*. Cambridge, MA: The MIT Press, 1987.
[6] D. Vanderveken, *Les actes de discours*. Bruxelles: Mardaga, 1988.
[7] P. Grice, Meaning, *Philosophical Review* **66** (1957) 377-388.
[8] J.R. Searle, *Speech Acts*. Cambridge: Cambridge University Press, 1969.
[9] A. Reboul,: The logical status of fictional discourse: what Searle's speaker can't say to his hearer, in: *Speech Acts, Meaning and Intentions: critical Approaches to the philosophy of John R. Searle*, A. Burkhardt, Ed. Berlin/New York,:de Gruyter, 1990, pp. 336-363.
[10] A.Reboul, The description of lies in speech acts theory, in *Pretending to communicate*, H. Parret, Ed. Berlin: de Gruyter, 1993, pp.292-298.
[11] A. Reboul, La fiction et le mensonge : les « parasites » dans la théorie des actes de langage, in *L'assertion en débat. La description du monde dans la conversation*, Ch. Brassac, Ed. *Psychologie de l'Interaction* **5-6**, (1997) 87-125.
[12] A. Reboul, Le paradoxe du mensonge dans la théorie des actes de langage, *Cahiers de linguistique française* **13** (1992) 125-147.
[13] A.Reboul, Represented speech and thought and auctorial irony: ambiguity and metarepresentation in literature, in *Hommages à Ronald Landheer*, J. Rooryck, P. Smith and P. Bogaarts, Eds. Amsterdam: Rodopi, to appear.
[14] F. Récanati, Contextualism and anti-contextualism in the philosophy of language, in *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, S. Tsohatzidis, Ed. London: Routledge, 1994, pp. 156-166.
[15] J. Fodor, *The modularity of mind. An essay on faculty psychology*. Cambridge: MIT Press, 1983.
[16] A. Reboul and J. Moeschler, Le dialogue n'est pas une catégorie naturelle scientifiquement pertinente, *Cahiers de linguistique française* **17** (1995) 229-248.

(2001), in Anolli, L. & Ceciri, R. (eds), *Say not to say. New perspectives on miscommunication*, IOS Press, Amsterdam, 46-71

[17] A. Reboul and J. Moeschler, Faut-il continuer à faire de l'analyse de discours?, *Hermes, Journal of Linguistics* **16** (1996) 61-92.
[18] A. Reboul and J. Moeschler, Reduction and contextualization in pragmatics and discourse analysis, in *Pragmatik,* B. Eckard Ed. *Linguistische Berichte, Sonderheft* **8**, (1997) 283-295.
[19] A. Reboul and J. Moeschler, *La pragmatique aujourd'hui. Une nouvelle science de la communication*. Paris: Le Seuil, 1998.
[20] A. Reboul and J. Moeschler, *Pragmatique du discours. De l'interprétation de l'énoncé à l'interprétation du discours*. Paris: Armand Colin, 1998.
[21] J. Perner, *Understanding the representational mind.* Cambridge, MA: The MIT Press, 1991.
[22] D. Sperber, Understanding verbal understanding, in *What is intelligence?*, J. Khalfa, Ed. Cambridge: Cambridge University Press, 1994, 179-198.
[23] D. Davidson, *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, 1984.
[24] S.C. Levinson, *Presumptive Meanings: The Theory of Generalized Conversational Implicatures*. Cambridge, MA: The MIT Press, 2000.
[25] S. L. Tsohatzidis, Speaker meaning, sentence meaning and metaphor, in *Foundations of speech act theory. Philosophical and linguistic perspectives*, S. L. Tsohatzidis, Ed. London: Routledge, 1994, 365-373.
[26] J. Perry, Indexicals, Contexts and Unarticulated Constituents, in *Proceedings of the 1995 CSLI-Amsterdam Logic, Language and Computation Conference*. Stanford: CSLI Publications, 1998.
[27] J. Perry, The problem of the essential indexical, in *The problem of the essential indexical and other essays*. Oxford: Oxford University Press, 1993, 32-52.