Metarepresentation

Nicholas Allott

for Routledge Handbook of Pragmatics

Keywords:
communicative intention
direct quotation
higher-order representation
indirect quotation
informative intention
interpretive use
irony
metalinguistic negation
metarepresentation
mindreading
mixed quotation
pure mention
quotation
representation
speaker meaning

1. Introduction

Humans and other thinking beings can represent the world. My cat knows which cupboard her food is in: she has a mental representation of that state of affairs. Given that I believe that she knows where her food is, I have a mental representation of her mental representation. I also know where it is, and I know that I know that: I have a mental representation of my own mental representation about the food's location. What is more, using language I can describe all of these facts about representations.

The capacity to think and talk about representations, that is, to represent representations or *metarepresent*, is a species characteristic of human beings. All developmentally normal human beings, across all cultures, metarepresent, and with considerable facility; and metarepresentation is a central property of both human language use and human thought.

Quotation, both direct and indirect, as in (1 a and b) respectively, is metarepresentational: one representation is embedded in another.

1) a) John said "Mary's arrived!"

b) John said that Mary had arrived.

Some metarepresentational thoughts are described in (2). In each case the embedded sentence (the 'that'-clause) represents the world as being a certain way, and is embedded under an 'attitude' verb (these include *think*, *believe*, *doubt*, *desire*, *intend*, *wish*) which describes the relation that the referent of the subject of the sentence (John, in these cases) has to the lower-level representation.

2) a) John believes that Mary has arrived.

b) John doubts that Mary has brought wine.

Human communication is essentially metarepresentational according to the prevailing, broadly Gricean view. When a speaker produces an utterance the hearer's task is to work out what the speaker intended to convey. Thus the audience of the utterance starts with a mental metarepresentation like the one in (3a) and on that basis arrives at a mental metarepresentation like the one in (3b).

3) a) John said "It's snowing here."

b) John means (i.e. intends me to think that he intends me to think) that it is snowing in Oslo now.

In the sense that we are concerned with here, a representation is defined as a state that is *about* or *of* something. This property of 'aboutness' – called *intentionality* in philosophy – is generally agreed to be possessed by some or all mental states. Above, a metarepresentation has been defined as a representation of a representation. Another definition in the literature is that a metarepresentation is a higher-order representation with a lower-order representation embedded within it (Wilson, 2000: 411).

1.1 Higher-order metarepresentation

The metarepresentations in (1), (2) and (3a) are *first-order*: they consist of one level of metarepresentation of a representation. However, metarepresentations can themselves be embedded in other representations to form higher-order metarepresentations, as in (3b), and in (4), where numbered parentheses have been used to mark the different levels.

4) a) [$_2$Mary wrote [$_1$that John said [$_0$that she had arrived]]].

b) [$_3$Mary said [$_2$that she once wrote [$_1$that John said [$_0$that she had arrived]]]].

c) [$_2$Mary knows that [$_1$John knows [$_0$that she has arrived]]].

d) [$_3$John suspects [$_2$that Mary knows [$_1$that he doubts [$_0$that she has brought wine]]]].

Language supports metarepresentations of arbitrarily many orders, although processing higher-order metarepresentations is subject to capacity limitations on memory and attention: "Human limits on embedding are not impressive: only about four steps make our species uncomfortable." (Premack & Woodruff, 1978: 515–516)

1.2 Outline of the chapter

This chapter is concerned with metarepresentation as it relates to communication and language use. As section 2 explains, the leading pragmatic theories inherit from the work of the philosopher Paul Grice an account of communication as inference about what the speaker intended to convey, based on what the speaker uttered. Such accounts presuppose that both speaker and hearer have considerable facility with higher-order metarepresentations like (3b). This picture of communication, most thoroughly explored in Sperber and Wilson's relevance theory, construes it as a kind of metapsychology: an ability to infer others' mental states. The ability to infer others' mental states (beliefs, desires, intentions etc.) based on their ordinary, non-communicative behaviour is known as 'mindreading' or 'theory of mind' and has been extensively studied in recent decades. The relation between mindreading and utterance interpretation is discussed in this section of the chapter.

Section 3 provides a brief survey of the extensive cross-disciplinary literatures on quotation and metalinguistic negation, the clearest cases of language used to metarepresent. It also describes the now widely accepted claim made by Sperber and Wilson that verbal irony is a tacitly metarepresentational use of language. Finally it sets out the attempt made in relevance theory to provide a unified account covering interrogative sentences, in addition to quotation, metalinguistic negation and verbal irony, in terms of the *interpretive* use of language: that is, use of language to represent an utterance or a thought other than the speaker's own.

2. Metarepresentation and communication

It is a fundamental assumption of modern pragmatics that on the basis of an utterance the hearer attributes to the speaker an intention to convey a certain meaning. The hearer starts from a metarepresentation of what the speaker said and ends with a metarepresentation of the communicator's meaning, as in (3) above. This model of

communication derives primarily from Grice. The role of metarepresentation in his work on meaning and conversation is explained in sections 2.1 and 2.2.

Relevance theory has developed the most comprehensive picture of the role of metarepresentation in communication, putting it in the context of work in psychology on "theory of mind" or "mindreading", the human ability to attribute thoughts to others on the basis of their behaviour. Given that utterance interpretation is the attribution of thoughts to the speaker on the basis of her communicative action, it is a form of mindreading. Sections 2.3 and 2.4 set out the role of metarepresentation in communication according to relevance theory, section 2.5 discusses the relation between mindreading and communication.

2.1 Speaker meaning

Two strands of Grice's work are concerned with communication: his work on speaker meaning and his theory of conversation. Both make essential use of metarepresentations.

Grice defined speaker meaning in terms of three intentions of the utterer, U:

> "By uttering x, U meant something" is true iff for some audience A, U uttered x intending:
>
> i) A to produce some particular response r.
>
> ii) A to recognize that U intends (i), and
>
> iii) A's recognition that U intends (i) to function, in part, as a reason for (i). (Grice, 1969: 151. See also Strawson (1964) on Grice's (1957) original proposal.)

Consider the first intention, (i), to get A to produce a certain response, *r.* Utterances of indicative sentences such as (5a) are aimed at getting the audience to entertain a belief, while utterances of imperatives such as (5b) aim at getting the hearer to intend to perform a certain action (Grice, 1989: ch. 6; Neale, 1992: 546, fn 53). So it seems that the intended response *r* is (or at least includes) a mental state (belief or intention) and thus intention (i) is always a metarepresentation.

5) a) Bertrand's book is on the table.

b) Shut the window!

The second and third intentions are also metarepresentational, if 'recognising' that a person has intention X entails mentally representing her as intending X. On that assumption, (ii) entails that the speaker intends the hearer to represent the speaker as

intending (i). That is, the utterer intends the audience to think that she (U) intends him (A) to have response *r* (i.e. to entertain a certain belief or intention).

Why should an utterer not aim at getting the audience to believe a certain proposition *directly*, without all of this metarepresentation? Grice envisaged such cases and argued that they were not examples of speaker meaning, but of a kind of manipulation. Suppose that the audience, A, is a detective investigating a crime and U knows the identity of the culprit. Wanting to keep herself out of it, U might contrive a false clue to the real criminal's identity, dropping his handkerchief at the crime scene for A to find. In such a case, U intends A to come to believe *p*, but does not intend that A come to believe that she, U, intends him to believe *p*.

Grice's point was that such examples lack something that is essential to speaker meaning: namely that the utterer openly intends the audience to come to a certain conclusion. That is why he defines speaker meaning in terms of an intention that the speaker intends the hearer to recognise.

2.2 Implicatures

Grice's theory of conversation attempts to explain how it is possible for speakers to mean something more than, or different from, what the words that they utter literally mean, and for hearers to understand them when they do this, as in John's indirect reply in (6a) and the attested example of verbal irony in (6b):

6) a) Mary: Would you like a bit of this mutton stew?

John: You're forgetting that I'm a vegetarian.

b) Germany also decided to shut down its nuclear power plants after the Fukushima crisis, due to the imminent risk of tsunamis in Bavaria. (From an article by George Monbiot in *The Guardian*, 5th February 2013.)

In cases like these, Grice proposed that the speaker intends to convey a conversational implicature, i.e. something i) other than what she literally says, which ii) she intentionally implies in (or by) saying what she does. Grice's introduction of conversational implicatures includes an account of how the hearer could derive them, on the assumption that the speaker is rational and cooperative, and thus is conforming with certain expectations about rational cooperative speakers: that they be relevant; do not say things they believe to be false etc. (Grice, 1967/1989)

Grice's 'working-out schema' is given below, where *p* is the proposition expressed by the sentence uttered, and *q* is the proposition implicated. Obviously, it is thoroughly metarepresentational:

> A general pattern for the working out of a conversational implicature might be given as follows: 'He has said that *p*; there is no reason to suppose that he is not observing the maxims, or at least the CP [Cooperative Principle]; he could not be doing this unless he thought that *q*; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that q *is* required; he has done nothing to stop me thinking that q; he intends me to think, or is at least willing to allow me to think, that q; and so he has implicated that q.' (Grice, 1967/1989: 31)

The psychological reality of the working out schema has often been challenged. Is it plausible that hearers reason like this, or that speakers expect them to? In making and understanding utterances we are not aware of having to engage in explicit, reflective reasoning or to entertain such complex metarepresentational thoughts, although we may do so in some cases. Such criticisms miss their mark, because Grice's concern was not that speakers and hearers explicitly reason like this, but that we could. That is, it must be possible that the hearer could work out that an implicature is present (Grice, 1967/1989: 31).

Elsewhere Grice argues that in reasoning we can either go 'the hard way' or 'the quick way', where the quick way, "made possible by habituation and intention", is to intuitively grasp the conclusion, skipping intermediate steps that would be needed to spell out the inference fully (Grice, 2001: 17). He did not claim that hearers have to consciously or explicitly entertain such complex metarepresentational thoughts as e.g. "he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that q is required".

2.3 Infinite regresses of metarepresentations?

Another point on which Grice's views have been thought psychologically unrealistic is the implication that speaker and hearer have to entertain infinite series of increasingly higher-order metarepresentations of speaker's intentions or of each other's knowledge. Grice's definition of speaker meaning tries to capture the idea that communication is overt or transparent in a certain sense. As discussed above, this requires not only i) that the speaker intend the hearer to entertain *p*, but also ii) that the speaker intends the hearer to recognise this intention. But it may be that the speaker also has to iii) intend the hearer to recognise the intention in (ii), and iv) to intend the hearer to recognise the intention in (iii) and so on, with no limit (Strawson, 1964; Grice, 1969: 156–157; Schiffer, 1972: ch. 2; Neale, 1992: 549–550). The argument for this requirement turns on certain complex counterexamples (from Schiffer and Strawson) to Grice's definition of meaning, which are aimed at showing that there are cases where intentions (i)–(iii) are present but which should not count as cases of meaning.

One response to these counterexamples was Schiffer's (1972) suggestion that intentions (i)–(iii) must be mutually known to hearer and speaker. That is, the hearer and speaker

must both know them, each know that the other knows them, and so on, without limit: an infinite series of metarepresentations.

Background knowledge might also need to be mutually known. Clark and Marshall (1981) argued that mutual knowledge is required for speaker and hearer to correctly coordinate on a referent. If the speaker does not know that the hearer knows a certain film is showing, she can't be sure that when she says "The film that is showing" she will be correctly understood. But the speaker's knowing that the hearer does know which film is showing is not enough. If the hearer does not know that the speaker knows that he (the hearer) knows which film is showing, he can't be sure that she is referring to *Monkey Business* rather than some other film which she wrongly thinks that he thinks is showing. By similar arguments it can be shown that there is no principled limit to the order of the metarepresentations required.

In early linguistic pragmatics, considerable effort was expended on a psychologically realistic solution to the problem of mutual knowledge (Clark & Carlson, 1981; Smith, 1982; Sperber & Wilson, 1986; Sperber & Wilson, 1990; Perner & Garnham, 1988). Interest in these problems has since faded, although Sperber and Wilson's solution (1986: 38–50), that communication depends not on mutual knowledge, but on a weaker condition, mutual manifestness, which does not require speaker or hearer to entertain an infinite series of metarepresentations, became one of the foundations of relevance theory.

2.4 The psychology of communication: relevance theory and mindreading

Relevance theory treats metarepresentation as central to communication in two ways (Noh, 2000: 4; Wilson, 2000: 424). First, it claims that utterance interpretation is inference from a metarepresented utterance to a metarepresentational intention attributed to the speaker, and it has made proposals about the order of metarepresentations involved in communication, and about the relation of this capacity to general mindreading abilities. These are the topics of the remainder of section 2. Second, relevance theory tries to give a unified account of various types of use of language as interpretive: i.e., representing thoughts or utterances attributed to others. This is dealt with in section 3.

Like most current theories of pragmatics, relevance theory differs from Grice in proposing a much greater role in utterance interpretation for inference, extending to disambiguation, reference resolution and other processes involved in inferring the proposition expressed. Thus, in contrast to Grice's theory of conversation, where a metarepresented proposition – what the speaker asserted – is the starting point for inference, relevance theory proposes that utterance interpretation starts from a metarepresentation of the speaker's *utterance*, as in (3a).

According to relevance theory, the communicative faculty is a module specialised for

inferring from an utterance the speaker's informative and communicative intentions, where these are defined as follows:

The informative intention:

The intention to inform an audience of something.

The communicative intention:

The intention to inform the audience of one's informative intention. (Wilson & Sperber, 2004: 611.)

Suppose that a speaker utters 'It's snowing here,' with the following informative intention, which is a first-order metarepresentation: an attitude to a belief (cf. Sperber, 1994):

7) John should believe [that it is snowing in Oslo.]

Therefore the hearer, John, if he grasps the informative intention, entertains a second-order metarepresentation:

8) She intends [me to believe [that it is snowing in Oslo]].

Note that if the hearer grasps the informative intention then communication has succeeded, since he has arrived at what it was that the speaker wanted to convey to him. This, then, is what speakers generally aim at. So, in general, a speaker has a third-order intention that her hearer recognize her informative intention. This is the communicative intention. In the example, this is the metarepresentation in (9):

9) John should believe [that [I intend [him to believe [that it is snowing in Oslo]]].

If the hearer becomes aware of the speaker's communicative intention, he forms a fourth-order metarepresentation – in this case, the one in (10):

10) She intends [me to believe [that she intends [me to believe [that it is snowing in Oslo]]]].

2.5 Theory of mind and communication

Do [we] have a meta-representational ability? Do birds fly? Do fish swim? Humans can

no more refrain from attributing intentions than they can from batting their eyelids.
(Sperber, 1994: 187)

Theory of mind or 'mindreading' (Baron-Cohen, 1995) is the ability possessed by all developmentally normal adult human beings to attribute to others mental states such as beliefs, desires and intentions – that is, to form metarepresentations of others' mental states – on the basis of their behaviour. We do it automatically and mostly without explicit reasoning. Suppose you see a man walk towards a door, pulling a key from his pocket. He puts it into the keyhole and applies turning force. It does not turn. He pulls it out of the keyhole and looks at it with a puzzled expression. We cannot help but metarepresent his thoughts: he wanted to open the door and wrongly thought that the key he had was the right one.

Sperber and Wilson have pointed out that on a broadly Gricean view of communication, utterance interpretation is a type of mindreading: the hearer attributes a communicative intention to the speaker on the basis of her utterance. Support for this view is provided by evidence that mindreading and utterance interpretation pattern together to some extent both in childhood development and in the way that they break down due to pathology (Sperber & Wilson, 2002: 7–8). In autistic spectrum disorders, there are deficits in both general mindreading and non-verbal communication (Perner, Frith, Leslie, & Leekam, 1989; Langdon, Davies, & Coltheart, 2002). (Although Chevallier, Wilson, Happé, & Noveck, 2010; Chevallier, Noveck, Happé, & Wilson, 2011 show that some communicative impairment in ASD is not due to ToM deficit.) Mindreading ability correlates with communicative abilities including word learning (Bloom, 2000; Happé & Loth, 2002) and reference resolution (Mitchell, Robinson, & Thompson, 1999).

Standard tests for mindreading assess the ability to attribute false beliefs, since this reveals the ability to attribute a different belief from one's own. For some years, it was thought that children below about 4 years old could not mindread since they failed the standard (verbal) false belief task. That seemed problematic for the view that utterance interpretation is a type of mindreading, given that younger children can make and understand utterances. However, it has been demonstrated by more sophisticated experiments that children attribute beliefs well before they pass the verbal task (Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007).

There is some evidence that utterance interpretation uses a different mechanism from general mindreading (Wilson, 2000). Adults have difficulty processing metarepresentations of more than around four orders, but we are quite capable of talking and thinking about embedded speaker meanings, as in (11). If each layer of speaker meaning is processed as a fourth order metarepresentation then the sentence in (11) should be unprocessable, since it would require entertaining eighth- and twelfth-order metarepresentations.

11) I didn't mean that you *said* that the weather in Oslo was always bad; what I meant was that you strongly implied it, or that what you meant was that that was what Mary meant.

It also seems introspectively implausible that all these layers of metarepresentation are actively entertained every time we understand an utterance. Grice's definition of speaker meaning that implies that (e.g.) (12b) is (part of) the correct analysis of (12a) was a considerable theoretical achievement, not a simple consultation of intuitions.

12)a. John means that it is snowing in Oslo now.

b. John intends me to think that he intends me to think that it is snowing in Oslo now.

All of this suggests that we normally understand speakers without unpacking the notion of meaning. That is, the mental representation of a speaker meaning is more like (12a) than (12b).

This in turn suggests that hearers do not work out for each utterance that it is an attempt to get them to think that the speaker intends them to think something. It seems more likely that we have a specialised procedure that works out speaker meaning and makes representations like (12a) available to the rest of cognition.

3. Metarepresentational use of language

The previous section has explored the idea that the intentions underlying communication are metarepresentational. The content of speaker's meaning, i.e., what the speaker intends to communicate, can also be metarepresentational. This section looks at forms of language and types of language use that have been analysed as metarepresentational, from the obvious cases of quotation and metalinguistic negation, to verbal irony, which is now generally seen as essentially allusive, and concludes with relevance theory's attempt to provide a unified account of these phenomena, together with interrogatives, as instances of interpretive use.

3.1 Quotation

In both direct and indirect quotation, as in (13) a and b respectively, one representation is embedded in another and the embedded sentence is a representation of something uttered or thought: thus the whole is a metarepresentation. Direct quotation highlights the form of what is quoted, indirect, the semantic content.

13) a) John said to me, "You'll surely come up with something to save the company."

b) John said that I would think of a plan.

It is now usual to distinguish at least two further types of quotation: free indirect discourse (McHale, 1978) and 'mixed' quotation, as in (14) a and b respectively. There is also mention or 'pure' quotation, in which an abstract type is represented, as in (14c).

14) a) John's irrationally optimistic, as usual. I'm surely going to come up with a plan to save the company!

b) According to John, I'll "surely come up with" a plan. (Cf Wilson, 2000: 413)

c) "come up with" is a phrasal verb which means *devise* or *produce* an idea, plan or similar and has the main stress on the second word. (Cf Wilson, 2000: 413; Noh, 2000: 7, 11)

'Pure' and 'mixed' quotation are explicitly metarepresentational. Free indirect discourse is not. In (14a) the hearer has to infer that the speaker does not herself endorse what her second sentence seems to say, and that she is tacitly attributing a thought or utterance to John. Another type of tacitly metarepresentational quotation, free direct speech, is defined as direct quotation where the verb of saying is not present (Noh, 2000: 15). This is also known as 'zero quotation' (Mathis & Yule, 1994: 63; see also Buchstaller, this volume).

There are several overlapping literatures on quotation. It has attracted attention in philosophy (e.g. Davidson, 1979/1984), linguistic formal semantics (e.g. Partee, 1973), psycholinguistics (e.g. Wade & Clark, 1993), literary studies (e.g. Leech & Short, 1981: ch. 10) and corpus linguistics and stylistics (Semino & Short, 2004) as well as pragmatics (Coulmas, 1986; Clark & Gerrig, 1990; Noh, 2000: ch. 1; Wilson, 2000: 424–437). Cappelen & Lepore, 2012 is a recent survey from the point of view of philosophy, while Noh, 2000: ch 1, although slightly out of date, is the best general overview. Noh (2000) and Wilson (2000) are the best guides to the connections between quotation and metarepresentation.

Philosophy and formal semantics have focussed on direct quotation and pure mention, plus indirect quotation seen as a propositional attitude, while pragmatics and literary studies have been interested in all types except pure mention. Recently there has been some dialogue between these camps, with debate over whether a 'semantic' theory of quotation is possible, that is, one that treats the various forms of quotation as contributing compositionally to truth-conditions.

Quotation (other than pure quotation) is often called reported speech, but these terms are not synonymous, since not all quoted utterances are reports. That is, quotation is always metarepresentational, but the attitude to what is metarepresented can vary. Consider the examples in (15):

15) a) John: Are you a student?

Mary: Am I a student? (Noh, 2000: 16)

b) John: Leave me alone.

Mary: Leave you alone! (Noh, 2000: 16)

Mary's utterances are not reports but echoes of John's utterances. In (15a) she might be wondering aloud, reposing John's question to herself; or she might be echoing John's utterance in order to ridicule it. The examples in (15) are free indirect speech, but indirect speech and mixed quotation can also be used echoically (Noh, 2000: 16–17) as in (16) a and b respectively:

16) Peter: I'm the best damn cook in England.

a) Mary: [You say/You think] you're the best cook in England. (Cf. Noh, 2000: 17)

b) Mary: You say you're "the best damn cook" in the country.

This links quotation to verbal irony, understood as echoic allusion (see below).

3.2 Metalinguistic negation

Metalinguistic negation involves the representation of an utterance or another representation such as a thought and its rejection on grounds other than its falsity, as in the examples in (17). The phenomenon was noted by Fillmore (1971) and Ducrot (1974) and brought to wider attention by Horn (1985; 1989: ch 6).

In uttering (17a), the speaker objects to an utterance or thought that represents Mary as *sometimes* late; in (17b), the negation targets the speech sounds used; in (17c) the objection is to representing Poul Anderson as merely 'one of the best'. As this example demonstrates, notwithstanding its name, metalinguistic negation does not require the use of 'not' nor indeed any morphologically overt negation (Noh, 2000: 117–118; Carston, 2002: 300-301).

17) a) Mary isn't *sometimes* late – she is *always* late. (Noh, 2000: 109)

b) These aren't [təmɑ:təuz]; they're [təmeɪɾouz]. (Cf. Carston, 2002: 300)

c) Everybody knows that Poul Anderson is one of the best: they're wrong – he's better! (Jerry Pournelle, quoted on the back cover of a 1983 edition of Poul Anderson's *Mirkheim:* Langford, 2010).

Several diagnostics have been suggested to distinguish metalinguistic negation from

ordinary negation. The most clear-cut is lexical incorporation (e.g *not happy →
unhappy*). Ordinary negation can lexically incorporate, as in (18a), but metalinguistic
negation cannot, as (18b) demonstrates. Another hallmark of metalinguistic negation is
that there is often contrastive stress on the locus of the objection and correction, as
indicated by the italics in (17a).

18) a) She's [not happy/unhappy].

b) She's [not happy/*unhappy] – she's ecstatic.

Metalinguistic negation has been the focus of several connected debates. Is negation
linguistically ambiguous between truth-conditional and metalinguistic senses or is there
just one negation operator used for both 'ordinary' and metalinguistic negation? Is
metalinguistic negation truth-functional? Can it operate only on actual prior utterances;
or possible utterances; or utterances and thoughts, as assumed in the definition given
above (see Carston, 2002: 296–297)?

The most developed account of metalinguistic negation as a form of
metarepresentational use of language is found in a series of publications by Carston and
Noh (Carston, 1996; Carston, 1998; Carston, 1999; Carston, 2002: 291–302; Carston &
Noh, 1996; Noh, 2000: ch. 3). Their main claim is that it involves echoic use: part of a
potential or actual utterance or thought is metarepresented and the result is what is
negated. For example, the linguistic material "these aren't [təmɑːtəuz]" in (17b) is
pragmatically enriched to be understood as something like *these aren't what is properly
called [təma:təuz]*. This is propositional, so the ordinary, truth-functional negation
operator can be applied to it. With a similar enrichment of "they're [təmeɪɾouz]" we then
have (19) as what is expressed by a (typical) use of (17b).

19) It is not the case that (these are what is properly called [təmɑːtəuz]); they are what
is properly called [təmeɪɾouz].

This account of metalinguistic negation has no need to assume that 'not' is ambiguous
between a sense that operates on propositions and a sense that operates on other types
of material. This means that the account requires no extra stipulations to give an
analysis of cases like (17c) where the word 'not' is not present.

A further advantage is the theoretical unification of metalinguistic negation with other
cases of metarepresentational use that do not involve negation but which are otherwise
similar. An example is the use of the sentence in (20) to mean *the person who has been
described as/thought of as 'the intelligent bloke' certainly is stupid*:

(20) The intelligent bloke certainly is stupid. (cf. Carston, 2002: 301)

3.3 Irony

Sperber and Wilson's account of verbal irony (Sperber & Wilson, 1981; Wilson & Sperber, 1992; Wilson, 2006; Wilson, 2009; Wilson & Sperber, 2012) and recent alternative theories analyse irony as metarepresentational use of language. It is not possible to define verbal irony in a theory-neutral way, but there is considerable agreement about paradigm examples, such as (6b) above and insincere utterances of the sentences in (21):

21) a) What lovely weather! [said in a hailstorm]

b) I love your tie.

The classical characterisation of verbal irony, still given by most dictionaries, is the use of words or a sentence to meant the opposite of what they normally, or conventionally, mean. Thus (21a) might be taken to express 'What horrible weather' and (21b) to express 'I hate your tie'. On this account, there is nothing essentially metarepresentational about irony. But the classical definition is problematic. Grice pointed out that one cannot freely mean the opposite of what one says. His example in (22) is an utterance which, intuitively, is not successful as irony:

22) Look, that car has all its windows intact. [said of a car with broken windows] (Grice, 1967/1989: 53)

Sperber and Wilson propose that in irony i) the words uttered are used to represent a thought or another utterance (actual or possible) and ii) to express a negative ("dissociative") attitude to the thought or utterance represented. An additional feature of fully ironic utterances is iii) that both (i) and (ii) are tacit, in the sense that they are not explicitly spelled out in the linguistic form of what is uttered.

According to this analysis, an ironic utterance of (21a), if echoing and mocking an expectation, could be roughly glossed as 'It was ridiculous to expect lovely weather,' or – if mocking a prior utterance – as 'It was ridiculous to say that there would be lovely weather'.

This account "treats ironical utterances as forming a natural class with other types of interpretive, attributive or echoic use." (Wilson, 2006: 1736) That fits well with irony's affinity for quotation, semi-quotation and allusion. On this understanding of irony, not only the utterances in (21) are ironic, but also the use of 'intelligent bloke' in (20) and the second sentence in (14a). The account correctly predicts that (22) will be felicitous in a context where it can be understood as echoing an expectation or utterance (Sperber & Wilson, 1986: 240–241; Wilson, 2006: 1728, 1732, 1735).

The echoic account implies that irony makes greater demands of the hearer's metarepresentational abilities than non-ironic utterances that are otherwise similar; and evidence has been found that the ability to comprehend irony correlates with fairly sophisticated metarepresentational ability (as measured by a second-order false belief task) and makes greater demands than the interpretation of metaphor (Happé, 1993). For a summary of the experimental work and discussion in the light of recent developments in mindreading research, see Wilson, 2009: 188–190.

Rivals to Sperber and Wilson's theory include the *pretence* account, a development of a suggestion made by Grice (Grice, 1967/1989: 54; Clark & Gerrig, 1984). The claim is that the speaker of an ironic utterance is only pretending to perform a speech-act – e.g., the speaker of (20b) is pretending to praise the hearer's tie without really doing so – and generally there is in addition a negative or mocking attitude towards the speech act that is simulated.

Recent pretence accounts of irony adopt from Sperber and Wilson the view that it is metarepresentational. These are *allusional pretence* accounts (Walton, 1990; Kumon-Nakamura, Glucksberg, & Brown, 1995; for discussion see Wilson, 2009: 205–210). Thus the consensus recognises that irony is metarepresentational use of language, since:

> Both [the echoic and allusional pretence accounts] recognise that irony involves the attribution of a thought (or perspective, or point of view) to a specific person or type of person, or to people in general, and the expression of a dissociative attitude to the attributed thought. (Wilson, 2006: 1736)

3.4 Interpretive and echoic use

Sperber and Wilson (1986: 224–254; Wilson & Sperber, 1988; Wilson, 2000) and Noh (2000) propose a unified account of metarepresentational uses of language. They claim that each utterance is relevant by representing another representation: every utterance resembles a thought of the speaker's. That resemblance may be relevant because the thought resembles a state of affairs (they call this 'descriptive use'), or because it resembles another thought or utterance ('interpretive use') (1986: 230–231). Echoic utterances, reported speech and irony are all cases of interpretive use.

A second claim is that the relation between the utterance and the representation it represents is one of resemblance (not, in general, identity). Resemblance comes in degrees. Sperber and Wilson define interpretive resemblance as greater the more implications are shared in context by the two representations (1986: 224–254). Identity is the limiting case in which all implications are shared.

Resemblances may be of different kinds. Wilson suggests that a difference between interpretive and metalinguistic resemblance underlies the different types of quotation:

> "[t]ypically, direct quotation [...] increases the salience of formal or linguistic properties, and indirect quotation [...] increases the salience of semantic or logical properties. We might call these resemblances *metalinguistic*, on the one hand, and *interpretive*, on the other [...] Mixed quotation [...] exploits both metalinguistic and interpretive resemblances, while reports of thought, and metarepresentations of thought in general, are typically interpretive." (Wilson, 2000: 426)

Previous sections have set out echoic-use theories of irony and metalinguistic negation. Noh also develops metarepresentational accounts of echo questions (1998; 2000: ch. 4) and metarepresentational uses in conditionals (2000: ch. 5), and Papafragou (1996) analyses metonymy as interpretive use.

Wilson and Sperber (1988) give an account of imperative and interrogative sentences based on the two claims above, with one further assumption. In descriptive use, a thought may describe either an actual state of affairs or a desirable one; equally, in interpretive use a thought can be an interpretation of either an *attributed* thought or utterance or a *desirable* one. They argue that the imperative sentence type encodes a descriptive relation to a desirable state of affairs, while the interrogative sentence type encodes an interpretive relation to desirable thoughts or utterances, that is, relevant answers.

4. Concluding remarks

Both thought and language have metarepresentational resources. There is consensus that human language and our metarepresentational abilities are closely linked, but no agreement about which is more fundamental. Debate has mostly focussed on evolutionary priority.

Sperber argues that mindreading ability was a prerequisite for the emergence of human language, since our use of language depends on mindreading. Human language profoundly differs from animal communication systems such as bee dances and vervet monkey calls in that it provides only partial clues to the speaker's meaning. If human linguistic communication was once purely a matter of coding and decoding, then "there is no reason to assume that our ancestors had the resources to become aware of the representational character of their signals, any more than bees or vervet monkeys do." (Sperber, 2000: 122) On the other hand, if human language always only provided partial clues to speaker's meaning, then interpretation has always depended on mindreading ability, so language cannot have preceded the ability to metarepresent.

However an argument can be made for the priority of language, if we assume that complex thought is enabled by language and that metarepresentation requires recursion, the ability which enables us to construct embedded representations. Chomsky and colleagues have argued that recursion is the essential property of language (Hauser, Chomsky, & Fitch, 2002) and that natural language is the language of thought (Chomsky, 2007: 22–26). It may be that there were no recursive thoughts before humans had language and *a fortiori*, no metarepresentation.

Further reading

Sperber, D. (1995). How do we communicate? In J. Brockman & K. Matson (Eds.), *How Things Are: A Science Toolkit to the Mind* (pp. 191-199). New York: W. Morrow.

A brief, accessible article in which Sperber argues that "human communication is a by-product of human meta-representational capacities".

Wilson, D. (2000). Metarepresentations in linguistic communication. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 411-448). Oxford: Oxford University Press.

A key paper which discusses both the role of mindreading in communication and the metarepresentational resources of natural language.

Papafragou, A. (2002). Mindreading and verbal communication. *Mind & Language, 17*, 55-67.

A technical, but still accessible, paper on the role of mindreading in communication, with a focus on children's abilities.

Carruthers, P. & Smith, P. K. (Eds.). (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

A useful collection of papers on mindreading.

Noh, E.-J. (2000). *Metarepresentation: A Relevance-Theory Approach*. Amsterdam: John Benjamins.

An ecumenical and interesting guide to the metalinguistic resources of natural language, including quotation and metalinguistic negation.

Chapter 4 of Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

An invaluable guide to the semantics and pragmatics of negation, including metalinguistic negation.

Abbott, B. (2003). Some notes on quotation. *Belgian Journal of Linguistics, 17*(1), 13-26. A short introduction to the recent state of the art on quotation in philosophy and formal semantics. The other papers in the same issue of the journal also concern quotation.

References

Baron-Cohen, S. (1995). *Mindblindness : An Essay on Autism and Theory of Mind*. Cambridge, Mass.: MIT Press.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, Mass: MIT Press.

Cappelen, H. & Lepore, E. (2012). Quotation. In N. Z. Edward (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2012 ed.).

Carston, R. (1996). Metalinguistic negation and echoic use. *Journal of Pragmatics*, *25*(3), 309-330.

Carston, R. (1998). Negation, 'presupposition' and the semantics/pragmatics distinction. *Journal of Linguistics*, *34*(2), 309-350.

Carston, R. (1999). Negation, 'presupposition' and metarepresentation: a response to Noel Burton-Roberts. *Journal of Linguistics*, *35*(2), 365-389.

Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Carston, R. & Noh, E.-J. (1996). A truth-functional account of metalinguistic negation, with evidence from Korean. *Language Sciences*, *18*(1–2), 485-504.

Chevallier, C., Noveck, I., Happé, F. & Wilson, D. (2011). What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia*, *49*(3), 507-517.

Chevallier, C., Wilson, D., Happé, F. & Noveck, I. (2010). Scalar inferences in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *40*(9), 1104-1117.

Chomsky, N. (2007). Of minds and language. *Biolinguistics*, *1*, 009-027.

Clark, H. H. & Carlson, T. B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 313-330). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, H. H. & Gerrig, R. (1984). On the pretense theory of irony. *Journal of Experimental Psychology*, *113*, 121-126.

Clark, H. H. & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber & I. A. Sag (Eds.), *Elements of Discourse Understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Clark, H. H. & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, *66*(4), 764-805.

Coulmas, F. (Ed.). (1986). *Direct and Indirect Speech*. Berlin: Mouton de Gruyter.

Davidson, D. (1979/1984). Quotation. In *Inquiries into Truth and Interpretation* (pp. 79-92). Oxford: Clarendon Press, Oxford University Press.

Ducrot, O. (1974). Les échelles argumentatives. In *La preuve et le dire: langage et logique*

(pp. 225-285). Paris: Mame.

Fillmore, C. (1971). Types of lexical information. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics; An Interdisciplinary Reader in Philosophy, Linguistics and Psychology* (pp. 370-392). Cambridge: Cambridge University Press.

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

Grice, H. P. (1967/1989). Logic and conversation: William James lectures. In *Studies in the Way of Words (1989)* (pp. 1-143). Cambridge, Mass: Harvard University Press.

Grice, H. P. (1969). Utterer's meaning and intentions. *The Philosophical Review*, *78*(2), 147-177.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, Mass: Harvard University Press.

Grice, H. P. (2001). *Aspects of Reason*. Oxford: Clarendon Press.

Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101-119.

Happé, F. & Loth, E. (2002). 'Theory of mind' and tracking speakers_' intentions. *Mind & Language*, *17*(1&2), 24-36.

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569-1579.

Horn, L. R. (1985). Metalinguistic negation and pragmatic ambiguity. *Language*, 121-174.

Horn, L. R. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.

Kumon-Nakamura, S., Glucksberg, S. & Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, *124*(1), 3-21.

Langdon, R., Davies, M. & Coltheart, M. (2002). Understanding minds and understanding communicated meanings in schizophrenia. *Mind & Language*, *17*(1&2), 68-104.

Langford, D. (2010). Blurb masterclass. *Ansible*, *279*.

Leech, G. N. & Short, M. (1981). *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. London: Longman.

Mathis, T. & Yule, G. (1994). Zero quotatives. *Discourse Processes*, *18*(1), 63-76.

McHale, B. (1978). Free indirect discourse: a survey of recent accounts. *PTL: A Journal for Descriptive Poetics and Theory of Literature*, *3*, 249-287.

Mitchell, P., Robinson, E. J. & Thompson, D. E. (1999). Children's understanding that utterances emanate from minds: Using speaker belief to aid interpretation. *Cognition*, *72*(1), 45-66.

Neale, S. (1992). Paul Grice and the philosophy of language. *Linguistics and Philosophy*, *15*(5), 509-559.

Noh, E.-J. (1998). Echo questions: Metarepresentation and pragmatic enrichment. *Linguistics and Philosophy*, *21*(6), 603-628.

Noh, E.-J. (2000). *Metarepresentation: A Relevance-Theory Approach*. Amsterdam: John Benjamins.

Onishi, K. H. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255-258.

Papafragou, A. (1996). On metonymy. *Lingua*, *99*(4), 169-195.

Partee, B. (1973). The syntax and semantics of quotation. In S. R. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 410-418). San Francisco: Holt, Reinhart and Winston.

Perner, J., Frith, U., Leslie, A. M. & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, *60*(3), 689-700.

Perner, J. & Garnham, A. (1988). Conditions for mutuality. *Journal of Semantics*, *6*(1), 369-385.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515-526.

Semino, E. & Short, M. (2004). *Corpus Stylistics : Speech, Writing and Thought Presentation in a Corpus of English Writing*. London: Routledge.

Schiffer, S. R. (1972). *Meaning*. Oxford: Oxford University Press.

Smith, N. V. (Ed.). (1982). *Mutual Knowledge*. London: Academic Press.

Southgate, V., Senju, A. & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587-592.

Sperber, D. (1994). Understanding verbal understanding. In J. Khalfa (Ed.), *What is Intelligence?* (pp. 179-198). Cambridge: Cambridge University Press.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 117-137). New York: Oxford University Press.

Sperber, D. & Wilson, D. (1981). Irony and the use-mention distinction. In P. Cole (Ed.), *Radical Pragmatics* (pp. 295-318). New York: Academic Press.

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and Cognition* (2nd ed. 1995). Oxford: Blackwell.

Sperber, D. & Wilson, D. (1990). Spontaneous deduction and mutual knowledge. *Behavioral and Brain Sciences*, *13*(1), 179-184.

Sperber, D. & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*(1&2), 3-23.

Strawson, P. F. (1964). Intention and convention in speech acts. *The Philosophical Review*, *73*(4), 439-460.

Surian, L., Caldi, S. & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580-586.

Wade, E. & Clark, H. H. (1993). Reproduction and demonstration in quotations. *Journal of*

Memory and Language, *32*(6), 805-819.

Walton, K. L. (1990). *Mimesis as Make-Believe : On the Foundations of the Representational Arts*. Cambridge, Mass.: Harvard University Press.

Wilson, D. (2000). Metarepresentations in linguistic communication. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 411-448). Oxford: Oxford University Press.

Wilson, D. (2006). The pragmatics of verbal irony: Echo or pretence? *Lingua*, *116*(10), 1722-1743.

Wilson, D. (2009). Irony and metarepresentation. *UCL Working Papers in Linguistics*, *21*, 183-226.

Wilson, D. & Sperber, D. (1988). Mood and the analysis of non-declarative sentences. In J. Dancy, J. M. E. Moravcsik & C. C. W. Taylor (Eds.), *Human Agency: Language, Duty, and Value; Philosophical Essays in Honor of J. O. Urmson* (pp. 77-101). Stanford, Calif: Stanford University Press.

Wilson, D. & Sperber, D. (1992). On verbal irony. *Lingua*, *87*(1/2), 53-76.

Wilson, D. & Sperber, D. (2004). Relevance theory. In L. R. Horn & G. L. Ward (Eds.), *The Handbook of Pragmatics* (pp. 607-632). Malden, Mass: Blackwell.

Wilson, D. & Sperber, D. (2012). Explaining irony. In *Meaning and Relevance* (pp. 123-145). Cambridge: Cambridge University Press.