1 Introduction: pragmatics

Dan Sperber and Deirdre Wilson

Pragmatics is often described as the study of language use, as opposed to language structure. In this broad sense, it covers a variety of loosely related research programmes ranging from formal studies of deictic expressions to sociological studies of ethnic verbal stereotypes. In a more focused sense – the one we will use here – pragmatics contrasts with semantics, the study of linguistic meaning, and is the study of how contextual factors interact with linguistic meaning in the interpretation of utterances. Here we will briefly highlight a range of closely related, fairly central pragmatic issues and approaches that have been of interest to linguists and philosophers of language in the past thirty years or so. Pragmatics, as we will describe it, is an empirical science, but one with philosophical origins and philosophical import.

References to pragmatics are found in philosophy since the work of Charles Morris (1938), who defined it as the study of the relations between signs and their interpreters. However, it was the philosopher Paul Grice's William James Lectures at Harvard in 1967 that led to the real development of the field. Grice introduced new conceptual tools – in particular the notion of implicature – in an attempt to reconcile the concerns of the two then dominant approaches to the philosophy of language, Ideal Language Philosophy and Ordinary Language Philosophy (on the philosophical origins of pragmatics, see Recanati 1987, 1998, 2004a, 2004b). Ideal language philosophers in the tradition of Frege, Russell, Carnap and Tarski were studying language as a formal system. Ordinary language philosophers in the tradition of the later Wittgenstein, Austin and Strawson were studying actual linguistic usage, highlighting in descriptive terms the complexity and subtlety of meanings and the variety of forms of verbal communication. For ordinary language philosophers, there was an unbridgeable gap between the semantics of formal and natural languages. Grice showed that the gap could at least be reduced by drawing a sharp distinction between sentence meaning and speaker's meaning, and explaining how relatively simple and schematic linguistic meanings could be used in context to convey richer and fuzzier speaker's meanings, consisting not only of what was said, but also of what was implicated. This became the foundation for most of modern pragmatics.

1

Grice (1967/1989: 47) proposed a rather vague general principle (Modified Occam's Razor) for deciding whether some aspect of interpretation is semantic or pragmatic: *Senses are not to be multiplied beyond necessity.* However, judgements about what is necessary have too often been affected by disciplinary parochialism and opportunistic considerations. When the work of Montague and Davidson suggested that natural language semantics could be directly studied as a formal system, Gricean pragmatics offered a rationale for dismissing a variety of hard-to-handle intuitions as irrelevant to semantics. A good example is Nathan Salmon's claim that failure of substitutivity in belief contexts is only apparent, and can be explained away in terms of Gricean implicatures (Salmon 1986). However, when formal semanticists feel they have the tools to handle some specific regularity in interpretation, they tend to treat it as *ipso facto* semantic, and to see a pragmatic account as inferior and unnecessary. Thus, the treatment of natural language conditionals has proved a rich field for formal elaboration (e.g. Jackson 1991), while the Gricean pragmatic approach to conditionals has been neglected. By the same token, pragmaticists tend to assume that whatever they feel able to account for is automatically pragmatic, on the ground that pragmatic explanations are more general, albeit vaguer. A more principled and generally accepted division of labour between semantics and pragmatics will involve more collaborative work. The recent development of formal pragmatics (Stalnaker 1999; Kadmon 2001; Blutner and Zeevat 2003; Asher and Lascarides 2003) is to be welcomed in this context.

1.1 Three approaches to pragmatics

The approaches to pragmatics we will consider here all accept as foundational two ideas defended by Grice (1989: chapters 1–7; 14; 18) (for representative collections, see Davis 1991; Kasher 1998; Horn and Ward 2004). The first is that sentence meaning is a vehicle for conveying a speaker's meaning, and that a speaker's meaning is an overtly expressed intention which is fulfilled by being recognised.[1] In developing this idea, Grice opened the way to an inferential alternative to the classical code model of communication. According to the classical view, utterances are signals encoding the messages that speakers intend to convey, and comprehension is achieved by decoding the signals to obtain the associated messages. On the inferential view, utterances are not signals but pieces of evidence about the speaker's meaning, and comprehension is achieved by inferring this meaning from evidence provided not only by the utterance but also by the context. An utterance is, of course, a linguistically coded piece of evidence, so that comprehension involves an element of decoding. How far does linguistic decoding take the hearer towards an interpretation of the speaker's meaning? Implicitly for Grice and explicitly for John Searle (1969: 43), the output of decoding is normally a sense that is close to being fully

propositional, so that only reference assignment is needed to determine what is said, and the main role of inference in comprehension is to recover what is implicated. Following Recanati (2004a), we will call this a *literalist* approach to semantics. However, a major development in pragmatics over the past thirty years (which has gone much further than Grice envisaged) has been to show that the explicit content of an utterance, like the implicit content, is largely underdetermined by the linguistically encoded meaning, and its recovery involves a substantial element of pragmatic inference. Following Recanati (2004a), we will call this a *contextualist* approach.[2]

The second foundational idea defended by Grice is that, in inferring the speaker's meaning, the hearer is guided by the expectation that utterances should meet some specific standards. The standards Grice proposed were based on the assumption that conversation is a rational, cooperative activity. In formulating their utterances, speakers are expected to follow a Cooperative Principle, backed by maxims of Quantity (informativeness), Quality (truthfulness), Relation (relevance) and Manner (clarity) which are such that 'in paradigmatic cases, their observance promotes and their violation dispromotes conversational rationality' (Grice 1989: 370):

Cooperative Principle (Grice 1967/1989: 26–27)
Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Quantity maxims
1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.

Quality maxims
Supermaxim: Try to make your contribution one that is true.
1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Maxim of Relation
   Be relevant.

Manner maxims
Supermaxim: Be perspicuous.
1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).[3]
4. Be orderly.

When an utterance has several linguistically possible interpretations, the best hypothesis for the hearer to choose is the one that best satisfies the Cooperative Principle and maxims. Sometimes, in order to explain why a maxim has been (genuinely or apparently) violated, the hearer has to assume that the speaker believes, and was trying to communicate, more than was explicitly said. Such

implicitly communicated propositions, or implicatures, are widely seen – along with presuppositions and illocutionary force – as the main subject matter of pragmatics.[4]

Most current pragmatic theories share Grice's view that inferential comprehension is governed by expectations about the behaviour of speakers, but differ as to what these expectations are. Neo-Griceans such as Gazdar (1979), Levinson (1983, 1987, 2000), Horn (1984, 1989, 1992, 2000, 2004, 2006) and Atlas (2005) stay relatively close to Grice's maxims. For instance, Levinson (2000) proposes the following principles, based on Grice's Quantity and Manner maxims (and given here in abridged form):

Q-Principle (Levinson 2000: 76)
Do not provide a statement that is informationally weaker than your knowledge of the world allows.

I-Principle (Levinson 2000: 114)
Produce the minimal linguistic information sufficient to achieve your communicational ends.

M-Principle (Levinson 2000: 136)
Indicate an abnormal, nonstereotypical situation by using marked expressions that contrast with those you would use to describe the corresponding normal, stereotypical situations.

Each principle has a corollary for the audience (e.g. 'Take it that the speaker made the strongest statement consistent with what he knows') which provides a heuristic for hearers to use in identifying the speaker's meaning.

For many philosophers and linguists, an attraction of the neo-Gricean programme is its attempt to combine an inferential account of communication with a view of language strongly influenced by formal semantics and generative grammar. The aim is to solve specifically linguistic problems by modelling pragmatics as closely as possible on formal semantics, assigning interpretations to sentence–context pairs without worrying too much about the psychological mechanisms involved. The following comment by Gerald Gazdar gives a flavour of this approach:

The tactic adopted here is to examine some of the data that would, or should, be covered by Grice's quantity maxim and then propose a relatively simple formal solution to the problem of describing the behaviour of that data. This solution may be seen as a special case of Grice's quantity maxim, or as an alternative to it, or as merely a conventional rule for assigning one class of conversational meanings to one class of utterances. (Gazdar 1979: 49)

Accordingly, neo-Griceans have tended to focus on *generalised* conversational implicatures, which are 'normally (in the absence of special circumstances)' carried by use of a certain form of words (Grice 1967/89: 37), and are therefore codifiable to some degree. For example, the utterance in (1a) would normally convey a generalised implicature of the form in (1b):[5]

(1)   a. Some of my friends are philosophers.
         b. Not all of my friends are philosophers.

Levinson (2000) treats generalised implicatures as assigned by default to all
utterances of this type, and contextually cancelled only in special circumstan-
ces. *Particularised* implicatures, by contrast, depend on 'special features of the
context' (Grice 1967/1989: 37), and cannot be assigned by default. For exam-
ple, the speaker of (2a) would not normally implicate (2b), but this implicature
might be conveyed if (2a) were uttered (in England) in response to the question
'Are the pubs open?':

(2)   a. It's midnight.
         b. The pubs are closed.

Neo-Griceans, and formal pragmaticists in general, have little to say about
particularised implicatures.[6] The result is a significant narrowing in the domain
of pragmatic research, which has yielded valuable descriptions of data from this
domain, but is driven largely by methodological considerations.

    Relevance theory (Sperber and Wilson 1995; Carston 2002a; Wilson and
Sperber 2002, 2004), while still based on Grice's two foundational ideas, departs
from his framework in two important respects. First, while Grice was mainly
concerned with the role of pragmatic inference in implicit communication,
relevance theorists have consistently argued that the explicit side of communica-
tion is just as inferential and worthy of pragmatic attention as the implicit side
(Wilson and Sperber 1981). This has implications not only for the nature of
explicit communication but also for semantics. As noted above, Grice and others
such as Searle and Lewis who have contributed to the development of an
inferential approach to communication have tended to minimise the gap between
sentence meaning and speaker's meaning. They treat sentences as encoding
something as close as possible to full propositions, and explicit communication
as governed by a maxim or convention of truthfulness, so that the inference from
sentence meaning to speaker's meaning is simply a matter of assigning referents
to referring expressions, and perhaps of deriving implicatures. Relevance theo-
rists have argued that relevance-oriented inferential processes are efficient enough
to allow for a much greater slack between sentence meaning and speaker's
meaning, with sentence meaning typically being quite fragmentary and incom-
plete, and speaker's explicit meaning going well beyond the minimal proposition
arrived at by disambiguation and reference assignment.

    Relevance theory also departs substantially from Grice's account of the expect-
ations that guide the comprehension process. For Griceans and neo-Griceans,
these expectations derive from principles and maxims: that is, rules of behaviour
that speakers are expected to obey but may, on occasion, violate (for instance,
because of a clash of maxims, or in order to trigger an implicature, as in Grice's

account of tropes). For relevance theorists, the very act of communicating raises precise and predictable expectations of relevance, which are enough on their own to guide the hearer towards the speaker's meaning. Speakers may fail to be relevant, but they can not, if they are genuinely communicating (as opposed, say, to rehearsing a speech), produce utterances that do not convey a presumption of their own relevance.

Relevance theory starts from a detailed account of relevance and its role in cognition. Relevance is defined as a property of inputs to cognitive processes (whether external stimuli, which can be perceived and attended to, or internal representations, which can be stored, recalled, or used as premises in inference). An input is *relevant* to an individual when it connects with available contextual assumptions to yield *positive cognitive effects*: for example, true contextual implications, or warranted strengthenings or revisions of existing assumptions. Everything else being equal, the greater the positive cognitive effects achieved, and the smaller the mental effort required (to represent the input, access a context and derive these cognitive effects), the greater the relevance of the input to the individual at that time.

Relevance theory is based on two general claims about the role of relevance in cognition and communication:

Cognitive Principle of Relevance (Sperber and Wilson 1995: 260–66)
Human cognition tends to be geared to the maximisation of relevance.

Communicative Principle of Relevance (pp. 266–72)
Every act of overt communication conveys a presumption of its own optimal relevance.

As noted above, these principles are descriptive rather than normative. The first, or Cognitive, Principle of Relevance yields a variety of predictions about human cognitive processes. It predicts that human perceptual mechanisms tend spontaneously to pick out potentially relevant stimuli, human retrieval mechanisms tend spontaneously to activate potentially relevant assumptions, and human inferential mechanisms tend spontaneously to process them in the most productive way. This principle has essential implications for human communication. In order to communicate, the communicator needs her audience's attention. If attention tends automatically to go to what is most relevant at the time, then the success of communication depends on the audience taking the utterance to be relevant enough to be worthy of attention. Wanting her communication to succeed, the communicator, by the very act of communicating, indicates that she wants the audience to see her utterance as relevant, and this is what the Communicative Principle of Relevance states.

According to relevance theory, the presumption of optimal relevance conveyed by every utterance is precise enough to ground a specific comprehension heuristic which hearers may use in interpreting the speaker's meaning:

Presumption of optimal relevance (Sperber and Wilson 1995: 266–78)
(a) The utterance is relevant enough to be worth processing.
(b) It is the most relevant one compatible with the communicator's abilities and preferences.

Relevance-guided comprehension heuristic (Sperber, Cara and Girotto 1995: 51)
(a) Follow a path of least effort in constructing an interpretation of the utterance (and in particular in resolving ambiguities and referential indeterminacies, in going beyond linguistic meaning, in supplying contextual assumptions, computing implicatures, etc.).
(b) Stop when your expectations of relevance are satisfied.

A hearer using the relevance-theoretic comprehension heuristic during online comprehension should proceed in the following way. The aim is to find an interpretation of the speaker's meaning that satisfies the presumption of optimal relevance. To achieve this aim, the hearer must enrich the decoded sentence meaning at the explicit level, and complement it at the implicit level by supplying contextual assumptions which will combine with it to yield enough conclusions (or other positive cognitive effects) to make the utterance relevant in the expected way. What route should he follow in disambiguating, assigning reference, constructing a context, deriving conclusions, and so on? According to the relevance-theoretic comprehension heuristic, he should follow a path of least effort, and stop at the first overall interpretation that satisfies his expectations of relevance. This is the key to relevance-theoretic pragmatics.

The Gricean, neo-Gricean and relevance-theoretic approaches are not the only theoretical approaches to pragmatics (even in the restricted sense of the term that we are using here). Important contributors to pragmatic theorising with original points of view include Searle (1969, 1975b, 1979); Stalnaker (1974, 1999); Fauconnier (1975, 1985, 1997); Travis (1975, 2001); Harnish (1976, 1994); Kasher (1976, 1982, 1984, 1998); Clark (1977, 1993, 1996); Katz (1977); Bach and Harnish (1979); Hobbs (1979, 1985, 2004); Lewis (1979, 1983); Dascal (1981); van der Auwera (1981, 1985, 1997); Anscombre and Ducrot (1983); Ducrot (1984); Bach (1987, 1994a, 1999, 2001, 2004); Recanati (1987, 1995, 2002a, 2004a); Neale (1990, 1992, 2004, in press); Sweetser (1990); Vanderveken (1990–91); Hobbs, Stickel, Appelt and Martin (1993); Asher and Lascarides (1995, 1998, 2003); van Rooy (2003); Blutner and Zeevat (2003). However, the approaches outlined above are arguably the dominant ones.

In the rest of this chapter, we will briefly consider four main issues of current interest to linguists and philosophers of language: literalism versus contextualism in semantics (1.2), the nature of explicit truth-conditional content and the borderline between explicit and implicit communication (1.3), lexical pragmatics and the analysis of metaphor, approximation and narrowing (1.4), and the communication of illocutionary force and other non-truth-conditional aspects of meaning (1.5). We will end with some comments on the prospects for future collaboration between philosophy and pragmatics.

1.2   Literalism and contextualism in semantics

Grice's distinction between saying and implicating is a natural starting point for examining the semantics–pragmatics distinction.[7] One of Grice's aims was to show that his notion of speaker's meaning could be used to ground traditional semantic notions such as sentence meaning and word meaning (Grice 1967/89: chapter 6). In his framework, a speaker's meaning is composed of *what is said* and (optionally) *what is implicated*, and Grice sees sentence meaning as contributing to both. What a speaker *says* is determined by truth-conditional aspects of linguistic meaning, plus disambiguation, plus reference assignment. Thus, identifying what the speaker of (3) has said would involve decoding the truth-conditional meaning of the sentence uttered, disambiguating the ambiguous word 'pupil' and assigning reference to the indexicals 'I' and 'now':

(3)   *I* have two *pupils now.*

The resulting proposition is sometimes called the literal meaning of the utterance, or the proposition expressed. Grice saw the truth value of a declarative utterance like (3) as depending on whether this proposition is true or false. By contrast, the meanings of non-truth-conditional expressions such as 'but', 'moreover' or 'so' are seen as contributing to what is *conventionally implicated* rather than what is said; in Grice's terms, conventional implicatures involve the performance of 'higher-order' speech acts such as contrasting, adding and explaining, which are parasitic on the 'central, basic' speech act of saying (Grice 1989: 359–68).[8] For Grice, the semantics–pragmatics distinction therefore cross-cuts the saying–implicating distinction, with semantics contributing both to what is said and to what is implicated.

However, although he allows for semantic contributions to implicit content, and although his Quality maxims ('Do not say what you believe to be false', 'Have adequate evidence for what you say') are presented as applying at the level of what is said, Grice seems not to have noticed, or at least not to have pursued the idea, that pragmatic inference might contribute to explicit content apart (perhaps) from helping with disambiguation or reference assignment. It therefore seemed quite feasible to many (apparently including Grice himself) to combine a literalist approach to semantics with a Gricean approach to pragmatics.[9] The result was a division of labour in which pragmaticists concentrated on implicatures, semanticists concentrated on literal meaning, and neither paid sufficient attention to potential pragmatic contributions to the proposition expressed.

As noted above, literalist approaches to semantics treat sentences as encoding something close to full propositions. Extreme forms of literalism, found in early versions of formal semantics, were adopted by neo-Griceans such as Gazdar (1979), whose slogan *Pragmatics = meaning minus truth conditions* was very influential. On an extreme literalist approach, the sense and reference of (3) are

seen as determined by purely linguistic rules or conventions, whose output would generally coincide with the intended sense and reference, but might override them in the case of a clash. More moderate literalists see the output of semantics as a logical form with variables for indexicals and other referential expressions, needing only reference assignment to yield a fully propositional form.

On a contextualist approach to semantics, by contrast, sentence meaning is seen as typically quite fragmentary and incomplete, and as falling far short of determining a complete proposition even after disambiguation and reference assignment have taken place. A considerable body of work in semantics and pragmatics over the last thirty years suggests strongly that the gap between sentence meaning and proposition expressed is considerably wider than Grice thought, and is unlikely to be bridged simply by assigning values to referential expressions. Thus, consider (4a)–(4b):

(4)   a. The sea is *too cold*.
         b. That book is *difficult*.

Even after disambiguation and reference assignment, sentences (4a) and (4b) are semantically incomplete: in order to derive a complete, truth-evaluable proposition, the hearer of (4a) must decide what the speaker is claiming the sea is too cold for, and the hearer of (4b) must decide whether the speaker is describing the book as difficult to read, understand, write, review, sell, find, etc., and by comparison with what. It is quite implausible that these aspects of truth-conditional content are determined by purely linguistic rules or conventions, and fairly implausible that they are determined merely by assigning values to linguistically specified variables. Given an inferential system rich enough to disambiguate, assign reference and derive implicatures, it is more natural (and parsimonious) to treat the output of semantics as a highly schematic logical form, which is fleshed out into fully propositional form by pragmatic inferences that go well beyond what is envisaged on a literalist approach. The result is a division of labour in which semanticists deal with decoded meaning, pragma-ticists deal with inferred meaning, and pragmatic inference makes a substantial contribution to truth-conditional content.

In fact, the contribution of pragmatic inference to the truth-conditional content of utterances goes much further than examples (3)–(4) would suggest. Consider (5a)–(5c):

(5)   a. I'll bring a *bottle* to the party.
         b. I'm *going to* sneeze.
         c. If you leave your window open and a burglar *gets in*, you have no right to *compensation*.

Whereas in (4a)–(4b) inferential enrichment is needed to complete a fragmen-tary sentence meaning into a fully propositional form, in (5a)–(5c), inferential

enrichment of a fully propositional form is needed to yield a truth-conditional content that satisfies pragmatic expectations (e.g. the presumption of optimal relevance from section 1.1). Thus, the speaker of (5a) would normally be understood as asserting not merely that she will bring some bottle or other, but that she will bring a *full* bottle of *alcohol*; the speaker of (5b) would normally be understood as asserting not merely that she is going to sneeze at some time in the future, but that she is going to sneeze *very soon*; and the speaker of (5c) would normally be understood as asserting that if a burglar gets in *through the window as a result of its being left open by the hearer*, the hearer has no right to compensation *for any consequent loss*. Enrichments of this type are surely driven by pragmatic rather than semantic considerations. They argue for a contextualist approach to semantics, combined with an inferential pragmatics which makes a substantial contribution to the proposition expressed.

From a radical literalist perspective, on which the semantics–pragmatics borderline should coincide with the borderline between saying and implicating, examples such as (4)–(5) show unexpected 'intrusions' of pragmatic inference into the domain of semantics. As Levinson (2000: 195) puts it, 'there is no consistent way of cutting up the semiotic pie such that "what is said" excludes "what is implicated"'. Literalists see this as a problem. Levinson's solution is to abandon Grice's view that saying and implicating are mutually exclusive. From a contextualist perspective, on which the semantics–pragmatics distinction coincides with the borderline between decoding and inference, examples such as (4)–(5) come as no surprise. An obvious way of handling these cases is to abandon the assumption that sentences are the primary bearers of truth conditions, and to break down the assignment of truth conditions to utterances into two theoretically distinct phases. In one phase of analysis, utterances of natural-language sentences would be seen as decoded into schematic logical forms, which are inferentially elaborated into fully propositional forms by pragmatic processes geared to the identification of speakers' meanings.[10] These propositional forms would be the primary bearers of truth conditions, and might themselves provide input, in another phase of analysis, to a semantics of conceptual representations (what Fodor calls 'real semantics') which maps them onto the states of affairs they represent. On this approach, there is no pragmatic 'intrusion' into a homogeneous truth-conditional semantics. Rather, there are two distinct varieties of semantics – linguistic semantics and the semantics of conceptual representations – of which the first, at least, is contextualist rather than literalist.[11]

1.3   Explicit and implicit communication

In much of contemporary philosophy of language and linguistics, the notions of saying and literal meaning are seen as doing double duty, characterising, on the