



Universiteit  
Leiden

The Netherlands

## Inference and implicature

Terkourafi, M.; Haugh, M.; Kádár, D.Z.

### Citation

Terkourafi, M. (2021). Inference and implicature. In M. Haugh & D. Z. Kádár (Eds.), *Cambridge Handbooks in Language and Linguistics* (pp. 30-47). Cambridge: Cambridge University Press. doi:10.1017/9781108954105.004

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/3223054>

**Note:** To cite this publication please use the final published version (if applicable).

# 3

## Inference and Implicature

Marina Terkourafi

Whether or not it involves the use of a language or some other code, human communication is inferential communication

Sperber and Wilson (2008: 87)

### 3.1 Introduction

The notions of inference and implicature are central to pragmatics, so much so that they can be used to define the pragmatic enterprise itself. Meaning which does not come from the words themselves (is not encoded) but rather from how they are used (is inferred) is considered pragmatic meaning per excellence. Both inference and implicature can be used to describe the latter type of meaning. While similar, inference and implicature are not synonymous, yet they can be hard to distinguish and even harder to keep apart. This chapter presents an overview of the various understandings of these two terms, focusing especially on their intertwining as well as how and why we may (wish to) keep them apart. We begin by drawing three distinctions between the possible meanings of these terms at increasingly finer levels of granularity, which also allows us to begin to place them in various pragmatic frameworks. We then move on to discussing the main mechanisms involved in their generation according to these frameworks. The chapter ends with considering recent developments that challenge some of the earlier distinctions, pushing them in new directions.

### 3.2 Key Concepts and Theories: Definitional Issues

To help us begin to tease apart the various meanings of ‘inference’ and ‘implicature’, in this section I draw three distinctions, shifting from coarser to finer levels of granularity as I do this. The three distinctions are (a) between

inference in a broad sense and inference in a narrow sense, (b) between inference (in a narrow sense) and implicature and, finally, (c) between inference and implicature as both process and product. Of these, the first distinction concerns only inference, since this is the older and more general term, having been around since at least the late 1500s (OED, s.v. ‘inference’). Implicature, on the other hand, is a newer term, having been introduced only in the second half of the twentieth century as a term of art in the philosophy of language, a status which it retains to this day (OED, s.v. ‘implicature’).

### 3.2.1 Inference in a Broad and in a Narrow Sense

A first distinction can be drawn between inference in a broad sense, which refers to the process of linking premises with conclusions, and inference in a narrow sense, which refers to meaning that is not encoded in the speaker’s words but may be derived from them by following some reconstructible steps. Inference in a broad sense is part of the study of reasoning and is studied most prominently in psychology and in logic. Logic in particular aims to model particular types of inference: the classic modes of deduction (inference from what applies in general to specific instances), induction (inference from specific instances to what applies generally) and, more recently, abduction (inference to the best explanation) are all different types of inference.<sup>1</sup> It is also clear that inference in a broad sense can take as input any type of information (including visual, aural, and other stimuli), and this can further include stimuli that are non-agentively produced, in the sense that they are not intentional but can still be informative. The well-known joke of a mathematician, a physicist, and an engineer riding on a train through Scotland and seeing a black sheep can serve to illustrate this.

- (1) A mathematician, a physicist and an engineer are riding on a train through Scotland. The engineer looks out the window, sees a black sheep, and exclaims, “Hey! They’ve got black sheep in Scotland!” The physicist looks out the window and corrects the engineer, “Strictly speaking, all we know is that there is at least one black sheep in Scotland”. The mathematician looks out the window and corrects the physicist, “Strictly speaking, all we know is that one side of one sheep is black in Scotland”.

The engineer’s way of thinking is a case of inductive reasoning, the physicist is reasoning abductively, while the mathematician, at the other extreme, refrains from drawing any inferences from the perceptually

<sup>1</sup> Of these, deductive inference is truth preserving, while inductive and abductive inferences are both probabilistic (likelihood-enhancing in the former case and ignorance-preserving in the latter). On some accounts, entailment is also a type of (truth preserving) inference, although unlike these three modes, entailment describes things from the language’s (the system’s) perspective while deduction, induction and abduction do so from the user’s perspective. Implicature, on this account, is a type of defeasible (not necessarily truth preserving) inference (for more details, see Woods 2010: 218).

available information. The sheep's black fleece is of course not a case of the sheep trying to influence in any way what casual observers think and it is non-agentive in this sense; that is simply the way the sheep is. Information about its appearance can however be informative for observers, who can use it to draw conclusions about the world. This example also nicely illustrates the fact that observers may draw different inferences from the same perceptually available information.

While logic is interested in relationships between propositions, cognitive psychology studies how people derive these relationships. Cognitive psychology is therefore also interested in a wide array of non-demonstrative inferences,<sup>2</sup> including those that result from categorization (*if one cat can see in the dark, then all cats can see in the dark*), encyclopaedic knowledge (*the janitor sweeps the floor [with a broom]*), and meaning generation guided by principles of rationality (Noveck 2010). This last sense is closer to how 'inference' is used within pragmatics, or what we will call 'inference in a narrow sense'. The term 'pragmatic inference' to characterize this last type of inference is therefore not redundant, since there are also other types of non-pragmatic (logical, encyclopaedic etc.) inference.<sup>3</sup>

### 3.2.2 Inference versus Implicature

Inference in a narrow sense is, as we just saw, the purview of linguistic pragmatics, which treats linguistic communication as a case of *ostensive-inferential* communication and treats utterances as ostensive stimuli. An ostensive stimulus is one that is produced intentionally by a communicator in order to bring about a change in the set of assumptions held by a recipient. Ostensive stimuli are couched within a communicative intention that serves to fix the meaning communicated by the stimulus itself. A handwave produced by a worker next to some roadworks can be a signal to passing drivers to keep moving forward and it is interpreted as such by them in virtue of attributing to the worker the intention to communicate precisely that meaning to them. The same gesture made by him to chase away a buzzing fly would not be communicative at all (although it may be mistaken as such). A crucial difference between the two is the communicator's producing this stimulus for the sake of some recipient: an ostensive stimulus is formulated with a particular audience in mind. The relevant distinction parallels one made by the philosopher H. P. Grice (1957) between natural and non-natural meaning ( $\text{meaning}_{\text{NN}}$ ). Grice's goal in proposing this was to differentiate cases where meaning attribution follows from natural necessity (*those spots mean measles, those dark clouds mean rain*) from cases where it is tied to a

<sup>2</sup> Non-demonstrative inferences are those in which the premises support the conclusion but they don't entail it. (The conclusion may still be false.)

<sup>3</sup> However, the modifier 'pragmatic' in front of 'implicature' is redundant, since 'implicature' as a term of art was introduced precisely to refer to a subcategory of defeasible inference.

communicator's intention to produce a certain effect in a hearer. Only the latter type of meaning is, according to Grice and later approaches to pragmatics, relevant to linguistic communication.

While the word 'inference' is in general use in the English language, 'implicature' is a technical term, introduced by Grice precisely to narrow down the object of investigation to a specific sub-case of meaning<sub>NN</sub>. In Grice's account of meaning, implicatures are tied to the speaker's intention. Grice ([1969] 1989: 92) defines the speaker's intention as follows:

'U[tterer] meant something by uttering [expression] x' is true if and only if, for some audience A, U uttered x intending:

- (1) A to produce a particular response r
- (2) A to think (recognize) that U intends (1)
- (3) A to fulfil (1) on the basis of his fulfilment of (2).

Note that what this definition requires is simply that the speaker intend all of (1)–(3); whether the audience in actual fact follows suit and produces the required responses has no impact on whether the speaker meant something or not. It is clear from this definition that the speaker's intention is itself a technical term referring to a special kind of intention that is fulfilled by its recognition, what has been called a reflexive intention (Bach and Harnish 1979: xiv–xv). Grice's delimitation of speaker's meaning with reference to the speaker's intention limits this notion only to what a speaker intended a hearer to recognize, or, in other words, what the speaker is willing to be held accountable for – even if there is further content that can be derived from the speaker's utterance and the hearer derives this. Speaker's meaning (or meaning<sub>NN</sub>) is further broken down, according to Grice, into what is said and what is implicated, which may in turn be implicated conventionally or conversationally (via the Cooperative Principle and the maxims of conversation; Grice [1975] 1989; Section 3.3), in normal circumstances (generalized conversational implicatures) or in specific contexts (particularized conversational implicatures). What all of these sub-categories of meaning<sub>NN</sub> have in common is that they are intended by the speaker to be so recognized by the hearer.<sup>4</sup> Grice was much less concerned with meaning that the hearer may derive from the speaker's utterance but was not intended (or else envisaged) by the speaker herself.<sup>5</sup> However, it is clear that such meaning exists and later accounts have been preoccupied with finding ways to talk about it.

Paradigmatic work in this vein has been carried out within Relevance Theory (Sperber and Wilson [1986] 1995). A first distinction drawn within RT that is relevant to our purposes is between explicatures and implicatures. An explicature is the hybrid product of decoding and inference, while an

<sup>4</sup> Or at least not consciously not intended by her. This qualification is important to accommodate the category of generalized conversational implicatures which are generated in normal circumstances without taking the speaker's intention into account although they may be canceled by it.

<sup>5</sup> For Grice this would simply not be meaning<sub>NN</sub>.

implicature is the product of inference alone. Thus, cases of disambiguation and reference assignment, for instance, which would have been part of ‘what is said’ for Grice, are explicatures in RT. Implicatures, on the other hand, are independent propositions derived by combining the explicature with elements of the context on the presumption that the speaker produced her utterance intending it to be optimally relevant (balancing cognitive effort against effects achieved) for the hearer. Thus, in the exchange

- (2) A: Have you been able to return to Mary the money you owe her?  
B: I haven’t been to the bank.

the proposition that B hasn’t used the services of a financial institution (including an ATM) is an explicature of B’s utterance (involving various processes of reference assignment to ‘I’, disambiguation of *bank* as ‘financial institution’ and subsequent adjustment of this concept to include facilities where one may obtain financial services such as an ATM, and narrowing of *be at* + *PLACE* to ‘being at a place for the purpose of undertaking the activities that are stereotypically undertaken there’), whereas the proposition that B hasn’t been able to return Mary the money that B owes her is an implicature of B’s utterance. While this way of distinguishing explicatures from implicatures amounts to redistributing meaning between ‘what is said’ and ‘what is implicated’ in the Gricean account, both explicatures and implicatures are intended by the speaker, and in this sense, they are both part of Gricean meaning<sub>NN</sub>.

However, RT also draws a further distinction between those implicatures that are strongly communicated and those that are weakly communicated. An implicature is strongly communicated if drawing it is necessary to make the speaker’s utterance optimally relevant. So, for instance, in example (2) above, the implicature that B hasn’t paid Mary back is strongly communicated by B’s reply or else B’s utterance would fail to provide an answer to A’s question. An implicature is weakly communicated in case it is one among many that would make the speaker’s utterance optimally relevant. So, for instance, when a wife tells her husband over breakfast:

- (3) “I have to work late tonight”.

she could be implicating any and all of the following:

- (4) a. ‘Don’t wait up for me’.  
b. ‘Don’t make dinner for me’.  
c. ‘Go ahead and have dinner without me’.  
d. ‘Please feed the kids’.  
e. ‘Please put the kids to bed’.  
f. ‘Don’t go worrying that something bad happened to me’.  
g. ‘Don’t call looking for me’.  
h. ‘I will miss our favorite show on TV’.  
i. ‘Please record it for me’.  
j. ‘...’

Items (4.a–j) are all weak implicatures of the speaker's utterance, which it would be uneconomical (not to mention tedious) for her to list one by one. As this list may be open-ended (this is what (4.j) stands for), an exhaustive listing may not even be possible; in this case, her utterance is more like an invitation for him to draw further inferences that she may be currently forgetting but is happy to subscribe to. To the extent that her husband derives at least some of these inferences, her utterance will have achieved its goal, in other words, it will have been optimally relevant. That implicatures can have this property of indeterminacy was acknowledged also by Grice ([1975] 1989: 40). However, it is only within RT that the full implications of this were drawn.

Correlating implicature strength with the extent to which an implicature is determinate, Wilson and Sperber (1986: 253) note: "The weaker the implicature, . . . the weaker the speaker's responsibility for its truth, up to the point where the implicature disappears altogether and the responsibility for the assumptions used and the conclusions drawn lies solely on the side of the hearer". What this statement makes clear is that all assumptions derived by combining a speaker's utterance with a context are inferences from the speaker's utterance but only those that are intended by the speaker herself to be recognized by the listener as so intended are implicatures of her utterance. This might be schematically represented as in (5):

(5) implicature = inference + speaker's (communicative) intention

Or, as we may say, paraphrasing an older adage:<sup>6</sup> Speakers implicate, hearers infer.

In the same essay, Wilson and Sperber (1986: 252) write: "sometimes, a speaker can observe the principle of relevance without having any idea of the sort of context in which the utterance will be processed, or the sort of conclusions that will be derived. In these cases, the utterance will have no implicatures at all". This, however, does not mean that it will not lead to any inferences either. Listeners can (and, according to RT, will, since the principle of relevance is more of a basic communicative instinct than a social convention or something that is learnt; cf. Carston 2002) always draw inferences from what a speaker said (sometimes also if they are not among the speaker's intended audience or there is no such audience).<sup>7</sup>

There are a few reasons why it can be useful to maintain a strict inference versus implicature distinction along these lines, with only the latter being connected to the speaker's intention. First and foremost, as we just noted, inferences not intended by the speaker are not part of Gricean meaning<sub>NN</sub>. As such, Grice does not offer an account of how they are derived. However, as a framework approaching communication from the hearer's perspective, RT

<sup>6</sup> "Propositions imply, people infer", cited in Woods (2010: 218).

<sup>7</sup> Incidentally, that is why the worker's chasing away a buzzing fly may be mistaken as a communicative handwave in the example at the start of this section.

can account for all the inferences a hearer draws from a speaker's utterance so long as the balance between cognitive effects and effort is maintained.

Another reason why it can be useful to maintain a strict inference versus implicature distinction is that implicature (as speaker intended) and inference (as hearer derived) can overlap but they are not always co-extensive. A speaker may implicate a proposition (intend the hearer to think something) that the hearer doesn't end up thinking: this does not mean that the speaker did not implicate it – although it does mean that this aborted meaning will not become part of the conversational record or 'common ground' (Clark 1996) between interlocutors. If the speaker becomes aware of this and really wants the hearer to get it, she will have to try harder – possibly by repeating her utterance or making this content more explicit.

Conversely, a hearer may infer a proposition that the speaker did not intend: that does not make the speaker responsible for this meaning that she had not foreseen – although if it does become apparent to her that the hearer has derived this and it is a meaning she would like to distance herself from, she should take steps to do so, lest the inference becomes part of the conversational record that is taken for granted in later discourse. This captures the truism that speakers and listeners may not always end up entertaining the same meaning out of an utterance, which is a major advantage of an inferential model of communication over a code one. In a code model of communication (such as Shannon and Weaver 1963), any discrepancies between the meaning intended by the speaker and the one derived by the listener are considered exceptional and attributed to noise in the channel. However, in an inferential model (such as Sperber and Wilson [1986] 1995), such discrepancies are expected and attributed to the fact that communicating partners can only make hypotheses about each other's cognitive environment and take this into account within the limits of their own abilities and preferences.

### 3.2.3 Inference and Implicature as Process and Product

A third distinction can be drawn between inference and implicature as both process and product. As Lyons (1995: 35) makes clear, many terms in linguistics are used in this dual sense, without clear signposting of which of the two senses is meant each time. For instance, in the discussion of RT above, inference is used to refer to both a process of deriving new propositions by combining a speaker's utterance with a context as well as the products of that process, the actual propositions inferred. Implicature, on the other hand, is reserved only for a subset of those products, namely those propositions that the hearer infers and which are actually intended by the speaker. That is also why inference (as a process) can also contribute to explicatures. More generally, in RT, inference as a process can lead to implicatures as products or not. The term 'inferential process' is sometimes used to distinguish inference as process from inference as product.



Mindful of this distinction, Grice ([1975]/1989: 24) introduced alongside ‘implicature’, that is, the act of implying, the noun ‘implicatum’, that is, what is implied and was careful to maintain the distinction between the two throughout his writing. For instance, when talking about the indeterminacy of conversational implicatures, he makes clear that this is a property “that many actual implicata do in fact seem to possess” (40). Others have not always emulated this careful usage and over time the distinction has faded. The distinction between implicature and implicatum can nevertheless be useful to keep in mind, as it is possible that the same process can lead to different products or, conversely the same product can be the result of different processes. So much has become apparent in recent experimental pragmatics research, which has granted researchers the opportunity to observe the interpretation process unfold in real time. What such research, drawing among others on recordings of event-related brain potentials, has revealed is that there can be qualitative differences in the interpretation patterns of different individuals, even if the overall quantitative trends remain the same (Franke and Degen 2016; Tanner et al. 2018).

### 3.3 Matters of Process

That a proposition must be capable of being worked out from the information available in the speaker’s utterance and the context following some reconstructible steps<sup>8</sup> is a requirement for pragmatic inference and in this regard, Grice once again set the scene, with his Cooperative Principle (henceforth CP) and the maxims of conversation.<sup>9</sup> While the maxims

<sup>8</sup> This is of course the property of calculability of conversational implicatures, which, alongside cancellability, and non-detachability, set conversational implicatures apart from conventional ones (Grice [1975] 1989: 30). Later scholars have added reinforceability, universality and non-conventionality to this list (Huang 2007). Because conventional implicatures are not calculable but have to be learnt they are not a type of pragmatic inference as this is understood in this chapter (although they can be considered a type of encyclopedic inference having to do with the lexical meaning of words).

<sup>9</sup> The Cooperative Principle reads: “Make your conversational 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”. This subsumes four categories of maxims which spell out what it means to be conversationally cooperative. These are:

The maxims of Quantity

- (i) Make your contribution as informative as is required (for the current purposes of the exchange).
- (ii) Do not make your contribution more informative than is required.

The maxims of Quality

Supermaxim: Try to make your contribution one that is true.

- (i) Do not say what you believe to be false.
- (ii) Do not say that for which you lack adequate evidence.

The maxim of Relation

Be relevant.

The maxims of Manner

(falling under the four categories of Quantity, Quality, Relation and Manner) are typically honoured in the breach, meaning it is their (real or apparent) non-observance by the speaker that licenses the derivation of an implicature by the listener, the CP is always tacitly assumed to be in operation (unless explicitly opted out of; Grice [1975] 1989: 30) and provides the general background that justifies the listener thinking that the speaker meant more than what they said.

In describing how a listener goes about figuring out this additional meaning, Grice outlined five sources of information on which the listener relies (Grice [1975] 1989: 31). These are (1) the conventional meaning of the words used, together with the identity of any references that might be involved (essentially, his notion of ‘what is said’); (2) the Cooperative Principle and its maxims; (3) the context, linguistic or otherwise, of the utterance; (4) other items of background knowledge and (5) the fact (or supposed fact) that all relevant items falling under the previous headings are available to both participants and both participants know or assume this to be the case. Of these, items (3) (=context, including the discourse and physical context) and (4) (=background knowledge) are especially interesting from a sociopragmatic point of view. Their inclusion among the sources of information that determine the outcome of the inferential process essentially opens up the possibility that different listeners may draw different implicatures from the same utterance – something speakers themselves may purposefully exploit, as in the case of coded communication, which is frequent in politicians’ and social media discourse (Marwick and boyd 2011: 123).<sup>10</sup>

More often than not, however, discrepancies in world knowledge or, more generally, in what interlocutors take for granted, result in miscommunication. Some often cited examples of this come from exchanges between women and men, such as the example in (6) (from Tannen 1993: 168):

- (6) (Context: A male and female co-worker are walking between buildings on a cold day)
- |                  |                      |
|------------------|----------------------|
| Woman co-worker: | “Where’s your coat?” |
| Male co-worker:  | “Thanks mom”.        |

Supermaxim: Be perspicuous.

- (i) Avoid obscurity of expression.
- (ii) Avoid ambiguity.
- (iii) Be brief (avoid unnecessary prolixity).
- (iv) Be orderly. (Grice [1975] 1989: 26–7).

<sup>10</sup> Coded communication refers to using a stimulus which speakers know different audiences will interpret differently to activate these different interpretations in the respective audiences’ minds. Examples include pop singer Madonna’s early image (which signaled empowerment to young women and sexual appeal to young men) and references to Christian texts in former US President George W. Bush’s speeches (which appealed to his base without alienating others). In the case of coded communication, all of these inferences are implicatures of the speaker’s utterance, since they are all equally intended by the speaker (formulated with different audiences in mind).

According to Tannen's analysis of this example, the man's response frames the woman's utterance as parental advice, reflecting a (male) worldview in which conversational exchanges aim at domination. The woman's utterance, however, she argues, may have been simply motivated by friendly concern as is common among peers, reflecting a (female) worldview where actions are driven by solidarity. The differences in worldviews between speaker and hearer result in different inferences being drawn by them from the same utterance. Which of these was actually intended by the speaker (and therefore an implicature of her utterance) is a question best left to psychologists or sociologists to answer. For so long as our account of meaning can reconstruct the steps by which different interpretations can be arrived at, our work will have been done.

Or will it? Grice ([1975] 1989: 31) summarizes the process of implicature derivation as follows:

[The speaker] has said that *p*; there is no reason to suppose that he is not observing the maxims, or at least the CP; 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*.

One important question this account leaves unanswered is how we get from *p* or 'what is said' by the speaker's utterance to the specific *q* that the speaker thereby implicates. As Wilson and Sperber (1986: 244) note, out of any proposition, myriad others may be justifiably drawn. To give but one quick example, in (6) above, some of the (more readily accessible) inferences from the female co-worker's utterance include that she assumes that her addressee owns a coat and that he has brought it to work that day. How does the listener know which of these inferences are actually intended by her, making them implicatures of her utterance?

What Wilson and Sperber are highlighting here is the well-known frame-problem in artificial intelligence research (McCarthy and Hayes 1969).<sup>11</sup> This is more generally the problem of delimiting the relevant domain or universe of discourse in which to interpret the speaker's utterance and has led to various taxonomical classifications of world knowledge from general (knowledge we have because we are human and live on earth) to specific (including cultural and other routinized knowledge, and even knowledge pertaining to specific individuals), such as Schank and Abelson's (1977) four levels of themes-goals-plans-scripts. Episodic memory plays an important role in these classifications and the organization of knowledge is based on experience, making it possible for interlocutors to be operating with

<sup>11</sup> Rather than as the frame problem, Wilson and Sperber (1986) discuss this as the problem of hypothesis formation and relate it back to Fodor's (1983) discussion of global and local processes.

different knowledge structures (or frames) in mind (at least at the lower, culture-specific levels). What specific information is stored in a frame or other stereotypical data structure,<sup>12</sup> how this is acquired and how it is accessed during utterance interpretation remain major problems of artificial intelligence research.

Wilson and Sperber's (1986: 249) solution to this problem is to argue that the concepts encoded in the speaker's utterance provide entry points into the relevant encyclopaedic entries, thus helping the listener to delimit the appropriate search space for hypothesis confirmation and disconfirmation.<sup>13</sup> In this process of delimitation, the listener is guided by the presumption of optimal relevance, which states that "(a) the ostensive stimulus is relevant enough for it to be worth the addressee's effort to process it [and] (b) the ostensive stimulus is the most relevant one compatible with the communicator's abilities and preferences" (Sperber and Wilson [1986] 1995: 270). The listener then engages in a comprehension strategy that enjoins them to "(a) consider interpretations (disambiguations, reference assignments, enrichments, contextual assumptions etc.) in order of accessibility (i.e. follow a path of least effort in computing cognitive effects) and (b) stop when the expected level of relevance in reached" (Carston 2002: 143).

A similar intuition about the importance of world knowledge is captured in Neo-Gricean frameworks (Horn 1984; Levinson 2000). These frameworks recast Grice's maxims in a smaller number of heuristics or principles (three for Levinson, two for Horn),<sup>14</sup> which are activated automatically by specific lexical items or constructions contained in incoming utterances, amplifying their content in predictable directions. Levinson's I-heuristic, in particular, is activated by describing things in a normal, unmarked way and makes specific reference to knowledge of stereotypical relations (*They unpacked the picnic. The beer was warm*, which I-implicates 'The beer was part of the picnic'), indicating how world knowledge can be brought into the utterance interpretation process. This is less so for his other two heuristics, Q and M, which are said to be metalinguistic in nature (relying on knowledge of language rather than knowledge about the world; for instance, knowledge about scalar

<sup>12</sup> Levinson's (1992) notion of 'activity types' is another.

<sup>13</sup> Note, however, that their later claim that concepts can be adjusted during the utterance interpretation process (Sperber and Wilson 2008) complicates this picture considerably, possibly to the point of making it circular.

<sup>14</sup> The three Levinsonian heuristics are:

Q[quantity]: 'What is not said is not the case'

I[Informativeness]: 'What is simply described is stereotypically and specifically exemplified'

M[manner]: 'Marked descriptions warn "marked situation".'

Horn's two principles are:

The Q[quantity]-principle (hearer-based): Say as much as you can (given R).

The R[relation]-principle (speaker-based): Say no more than you must (given Q).

alternates relies on a relationship of semantic entailment which is independent of context).<sup>15</sup>

While both Relevance theorists and Neo-Griceans acknowledge the importance of world knowledge to the outcome of the utterance interpretation process (the actual inferences drawn), there is considerable disagreement as to the temporal parameters of this process – namely at what point during the utterance interpretation process is world knowledge, including information about the context of utterance, taken into account? For Relevance Theory, all inference is local, meaning world knowledge is present at all times and comprehension draws on a set of assumptions activated by previous discourse and by the information conveyed by the incoming utterance. On this account, there are no default interpretations attached to specific expressions out of context. For Neo-Griceans, on the other hand, pre-contextual defaults (such as the interpretation ‘not all’ attached to utterances containing *some*) can be attached to expressions in the lexicon and filtered out in particular contexts. A middle-of-the-road solution here is that some parameters of context (such as the language spoken by the speaker, their age, gender, ethnicity and so on) may be presumptively fixed prior to interaction (while remaining open to revision later on), with other elements (intonation, gestures etc.) affecting interpretation as they become available parallel with the speech signal itself (Terkourafi 2005). These questions are currently being addressed in experimental pragmatics, from where new insights have also originated (e.g. constraint-based pragmatics; Degen and Tanenhaus 2019).

### 3.4 Critical Overview and Future Directions

We said earlier that for Grice, speaker’s meaning is a case of meaning<sub>NN</sub>, which crucially relies on the speaker having a certain type of communicative (reflexive) intention when producing an utterance. Whether an inference is an implicature of the speaker’s utterance or not, then, depends on whether the speaker is happy to be credited as the originator of this meaning. Speakers can only take responsibility for those inferences they intended the hearer to recognize as intended by them and only those are implicatures of their utterances. While RT acknowledges the possibility of inferences that were not intended by the speaker and offers an account of how they are derived (see Section 3.2.2), it retains the notion of communicative intention as central to communication: communication requires a change in the communicators’ mutual cognitive environment (i.e. in the set

<sup>15</sup> Nevertheless, recent work on scalar diversity (Van Tiel et al. 2016; McNally 2017) and work on non-entailment scales more generally (Hirschberg 1991) has shown that sentence as well as situational context can affect which expressions are considered alternates, blurring the distinction between knowledge of language and knowledge about the world in this respect.

of assumptions that they not only share but are mutually aware that they share) and inferences that are not implicatures of the speaker's utterance fall outside its scope.

This take on implicature can be problematic for two reasons. The first is that intentions can be hard to verify: as mental states, intentions are in speakers' heads so they cannot be observed directly, let alone subsequently confirmed. Indeed, deniability is one of the main reasons (though not the only one) why a speaker may choose to lead a hearer to understand something without telling them as much (Pinker et al. 2008). Also, it is not always the case that a speaker has a specific intention in mind that they have fully worked out in advance and want the hearer to recognize; speakers can also have proto-intentions and their utterances can be invitations to hearers to help clarify these (Terkourafi 2014). For this reason, the Gricean and post-Gricean accounts' reliance on intention makes the notion of implicature hard to implement empirically. This has prompted, in conversation analytic accounts, a move away from speakers' intentions towards evidence of such intentions in the conversational transcript. That evidence is not necessarily verbal: other types of cues such as gaze or even absence of a behavioural cue (silence) can be evidence of a speaker's internal mental state and stance on the situation. The notion of speaker accountability (Haugh and Jaszczolt 2012: 110, and references therein; see also Chapter 4) has been productively explored in this respect.

The second reason why tying the notion of implicature to the speaker's intention is problematic is because intention-based accounts have been criticized as Western-centric (Ameka and Terkourafi 2019). As anthropologists have been quick to point out, "even if it is true that the capacity for inferring the mental states of others is a generically human one, and plays a part in communication everywhere, it does not follow that all language ideologies will give it equal prominence, or even allow it to be openly recognized or actualized in speech" (Robins and Rumsey 2008: 414). What the experience of other cultures suggests on this point is that cultural groups vary in the extent to which they consider the speaker's intention to be the primary determinant of the meaning that the speaker must take responsibility for, another important determinant of this being the meaning encoded in the words themselves. In an older article, Matsumoto (1989) makes the point that Gricean principles apply better to English than to Japanese, where social context is encoded in obligatory grammatical choices, with the result that all the available forms carry additional meanings – what she calls "interactional implicatures" (1989: 210) – that the speaker cannot help but communicate. Similarly, based on her work with the Mopan Maya, Danziger (2011: 123) has argued for a

symptomatic set of understandings about language . . . in which sign form is taken to be necessarily related to sign content through indexical relations of cause and effect. . . . Under such a philosophy, the hearer

need seek no assistance for the task of interpretation in the context of any utterance's production – certainly not in “what is in the mind of the speaker”. Utterance effect is instead believed to be achieved regardless of such circumstances... Adherence to a Symptomatic philosophy corresponds to a belief that signifiers are related ‘naturally’ and of necessity to their signifieds, in a way that can be ideologically contrasted with the “non-natural” [Grice 1957] relations of the arbitrary Symbol.

Despite coming from a non-Western perspective, the idea that words can produce meanings directly regardless of how the speaker may have intended them should not be so exotic to Western ears either, if we consider cases where speakers have had to apologize for what their words meant, even if they were adamant that they did not mean it themselves (the 2007 case of radio show host Don Imus is a rather (in)famous example of this; see Allan 2016: 219).

A different yet related critique, again from a non-Western angle, is discussed by Haugh (2002), this time relating to the distinction between explicatures and implicatures within RT. Using examples from Japanese, Haugh argues that the line between explicit and implicit meaning, and therefore between explicatures and implicatures, is blurred in this language by the existence of utterances that ‘trail off’. What speakers do with such utterances is to leave something unsaid while simultaneously indicating that they are leaving it unsaid. Haugh discusses two possibilities in this regard: one in which the speaker does not trail off their utterance (thereby not indicating that they are leaving something unsaid) and one in which they do, using a discourse marker (*kara*, lit. ‘so, therefore’) that indicates that the hearer is to draw a conclusion from what the speaker said that they are hesitant to articulate themselves. Haugh furthermore argues that the second possibility is perceived as more polite than the first. Although, as Haugh convincingly argues, it is hard to tell whether the meaning that is left unsaid is communicated implicitly in the first case and explicitly in the second (where a discourse marker encoding procedural meaning guides its derivation), distinguishing explicatures from implicatures from a process angle, by considering the former to be the product of decoding + inference and the latter to result from inference alone (see Section 3.2.2), may solve this problem. Moreover, considering that at least part of the meaning comes from decoding in the second example could explain why it is considered more polite: the explicit guidance offered by the discourse marker in this case lessens the imposition on the listener's cognitive resources (Blum-Kulka 1987) reducing costs, or, conversely, shows that the speaker is at least trying to be relevant (provide an answer to a previous question and not just avoiding the topic) while avoiding any disaffiliating effects their utterance might have. While the challenge to RT may thus not be so great as envisaged, one may still find it worrisome that the distinction between explicatures and implicatures cannot be defended on intuitive



grounds but only intra-theoretically. In other words, what is the status of this distinction outside the theory itself?

The final critique I would like to raise relates to the closing lines of the quote from Danziger (2011) above. There Danziger claims, based on her observations of Mopan Maya discourse, that at least on some occasions, signifiers can be linked to their signifieds in a more direct and ‘natural’ way than is envisaged within Grice’s notion of meaning<sub>NN</sub>. This point is also acknowledged by pragmaticians and philosophers of language, who are increasingly paying attention to the fact that all kinds of behaviour and not just that produced intentionally can give rise to inferences and that theories of meaning should also account for these types of inferences, since they constitute a type of meaning occasioned by the speaker’s behaviour. A recent attempt at describing this is as “organic meaning”, a type of meaning proposed by Green (2019) as meeting some, though not all, of Grice’s conditions on meaning<sub>NN</sub> and thus lying halfway between natural and non-natural meaning in this respect.<sup>16</sup> As an example of organic meaning in human behaviour (most of his examples come from the animal kingdom whereby theory of mind cannot be taken for granted), Green cites uptalk – ending one’s statements with rising intonation as in “My name is Jenny Smith?” – which in contemporary American English can convey an attitude of being accommodating and non-aggressive (and, I would add, potentially of a certain generation or age, gender, social class and from a certain part of the country). These meanings, which are all part of uptalk’s “indexical field” (Eckert 2008) and are conveyed in specific situations, can depend on the speaker as much as on the addressee: as we have already mentioned several times throughout this chapter, addressees may well read meanings into a speaker’s utterance that a speaker never intended to convey. What is relevant to our purposes is that the respective meanings can arise without the speaker wanting their signal to be taken in any particular way, or in fact thinking of their behaviour as a signal at all. I am reminded here of an incident from my student days at the centre of the Greek capital, Athens, where, after I placed my order at a fast-food restaurant, one worker turned to the other repeating my order and adding: ‘make it a good one for the lass from Crete’. The island of Crete is indeed where I was brought up but I had no clue how the worker could have known that – until I realized I had used the dialectal pronunciation [oi] for ‘no’ (Standard Greek /oxi/). It is these unintended yet consequential aspects of our behaviour that Green’s term “organic meaning” aims to capture. Increasing attention to these aspects means that not only the line between inference and implicature, but also that between natural meaning and non-natural meaning, and correspondingly between inference in the broad sense and in the narrow sense, is being scrutinized and potentially redrawn.

<sup>16</sup> For a related attempt from an RT angle, see Wharton (2003, 2009).



As this brief overview has hopefully shown, despite being foundational pillars of all pragmatics research, the notions of inference and implicature continue to be hotly debated and their boundaries are active loci of research in pragmatics and the philosophy of language. Considering that much of the research cited in this section comes from a sociopragmatics perspective, an important contribution of sociopragmatics to the study of these notions has been to question their generality: are these really culturally neutral and/or asocial, or are they as culturally imbued as Foucault (1978) has claimed all theorizing to be? With much of this research currently conducted within sociopragmatics, this is also an area whence new developments and insights can be expected to emerge.

## References

- Allan, K. (2016). The reporting of slurs. In A. Capone, F. Kiefer and F. Lo Piparo, eds., *Indirect Reports and Pragmatics: Interdisciplinary Studies*. Cham, Switzerland: Springer, pp. 211–32.
- Ameka, F. K. and Terkourafi M. (2019). What if . . . ? Imagining non-Western perspectives on pragmatic theory and practice. *Journal of Pragmatics*, 145, 72–82.
- Bach, K. and Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.
- Blum-Kulka, S. (1987). Indirectness and politeness in requests: Same or different? *Journal of Pragmatics*, 11, 145–60.
- Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Danziger, E. (2011). Once more with feeling: A forbidden performance of the Great Speech of the Mopan Maya. *Anthropological Quarterly*, 84(1), 121–40.
- Degen, J. and Tanenhaus, M. K. (2019). Constraint-based pragmatic processing. In C. Cummins and N. Katsos, eds., *Handbook of Experimental Semantics and Pragmatics*. Oxford: Oxford University Press.
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), 453–76.
- Fodor, J. A. (1983). *Modularity of Mind*. Cambridge MA: MIT Press.
- Foucault, M. (1978). *The History of Sexuality*. New York: Pantheon Books.
- Franke M. and Degen, J. (2016). Reasoning in reference games: Individual- vs. population-level probabilistic modeling. *PLoS ONE*, 11(5), e0154854. doi:10.1371/journal.pone.0154854
- Green, M. (2019). Organic meaning: An approach to communication with minimal appeal to minds. In A. Capone, M. Carapezza and F. Lo Piparo, eds., *Further Advances in Pragmatics and Philosophy. Part II Theories and Applications*. Cham, Switzerland: Springer, pp. 211–28.

- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66, 377–88. Reprinted in: Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, pp. 213–23.
- Grice, H. P. (1969). Utterer's meaning and intentions. *The Philosophical Review*, 78, 147–77. Reprinted in: Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, pp. 86–116.
- Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan, eds., *Syntax and Semantics. Vol. III: Speech Acts*. New York: Academic Press, pp. 41–58. Reprinted in: Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, pp. 22–40.
- Haugh, M. (2002). The intuitive basis of implicature: Relevance Theoretic implicitness versus Gricean implying. *Pragmatics*, 12(2), 117–34.
- Haugh, M. and Jaszczolt, K. (2012). Speaker intentions and intentionality. In K. Allan and K. M. Jaszczolt, eds., *Cambridge Handbook of Pragmatics*. Cambridge: Cambridge University Press, pp. 87–112.
- Huang, Y. (2007). *Pragmatics*. Oxford: Oxford University Press.
- Hirschberg, J. (1991). *A Theory of Scalar Implicature*. New York: Garland.
- Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin, ed., *Meaning, Form, and Use in Context*. Washington, DC: Georgetown University Press, pp. 11–42.
- Levinson, S. C. (1992). Activity types and language. In P. Drew and J. Heritage, eds., *Talk at Work: Interaction in Institutional Settings*. Cambridge: Cambridge University Press, pp. 66–100.
- Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.
- Lyons, J. (1995). *Linguistic Semantics*. Cambridge: Cambridge University Press.
- Marwick, A. and boyd, d. m. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media and Society*, 13(1), 114–33.
- Matsumoto, Y. (1989). Politeness and conversational universals: Observations from Japanese. *Multilingua*, 8, 207–21.
- McCarthy, J. and Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4, 463–502.
- McNally, L. 2017. Scalar alternatives and scalar inference involving adjectives: A comment on van Tiel, et al. 2016. In J. Ostrove, R. Kramer and J. Sabbagh, eds., *Asking the Right Questions: Essays in Honor of Sandra Chung*. Santa Cruz: University of California Santa Cruz, Linguistics Research Center, pp. 17–28.
- Noveck, I. (2010). Inferential comprehension. In L. Cummings, ed., *The Pragmatics Encyclopedia*. London: Routledge, pp. 220–22.
- Pinker, S., Nowak, M. A. and Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, 105(3), 833–8.
- Robbins, J. and Rumsey, A. (2008). Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropological Quarterly*, 81(2), 407–20.

- Schank, R. C. and Abelson, R. P. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Erlbaum.
- Shannon, C. E. and Weaver, W. (1963). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Sperber, D. and Wilson, D. ([1986] 1995). *Relevance: Communication and Cognition*. 2nd ed. Oxford: Blackwell.
- Sperber, D. and Wilson, D. (2008). A deflationary account of metaphors. In R. Gibbs, ed., *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press, pp. 84–105.
- Tannen, D. (1993). The relativity of linguistic strategies: Rethinking power and solidarity in gender and dominance. In D. Tannen, ed., *Gender and Conversational Interaction*. Oxford: Oxford University Press, pp. 165–88.
- Tanner, D., Goldshtein, M. and Weissman, B. (2018). Individual differences in the real-time neural dynamics of language comprehension. *Psychology of Learning and Motivation*, 68, 299–335.
- Terkourafi, M. (2005). Pragmatic correlates of frequency of use: The case for a notion of ‘minimal context’. In K. Nikiforidou, S., Marmaridou and E. Antonopoulou, eds., *Reviewing Linguistic Thought: Converging Trends for the 21st Century*. Berlin: de Gruyter, pp. 209–33.
- Terkourafi, M. (2014). The importance of being indirect: A new nomenclature for indirect speech. *Belgian Journal of Linguistics*, 28(1), 45–70.
- van Tiel, B., van Miltenburg, E., Zevakhina, N. and Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, 33, 137–75.
- Wharton, T. (2003). Natural pragmatics and natural codes. *Mind and Language*, 18, 447–77.
- Wharton, T. (2009). *The Pragmatics of Non-Verbal Communication*. Cambridge: Cambridge University Press.
- Wilson, D. and Sperber, D. (1986). Inference and implicature in utterance interpretation. In T. Myers, K. Brown and B. McGonigle, eds., *Reasoning and Discourse Processes*. London: Academic Press, pp. 241–63.
- Woods, J. (2010). Inference. In L. Cummings, ed., *The Pragmatics Encyclopedia*. London: Routledge, pp. 218–20.