

Part I The Domain of Pragmatics

1 Implicature

LAURENCE R. HORN

1 Implicature: Some Basic Oppositions

IMPLICATURE is a component of speaker meaning that constitutes an aspect of what is **meant** in a speaker's utterance without being part of what is **said**. What a speaker intends to communicate is characteristically far richer than what she directly expresses; linguistic meaning radically underdetermines the message conveyed and understood. Speaker S tacitly exploits pragmatic principles to bridge this gap and counts on hearer H to invoke the same principles for the purposes of utterance interpretation.

The contrast between the said and the meant, and derivatively between the said and the implicated (the meant-but-unsaid), dates back to the fourth-century rhetoricians Servius and Donatus, who characterized litotes – pragmatic understatement – as a figure in which we say less but mean more (“minus dicimus et plus significamus”; see Hoffmann 1987 and Horn 1991a). In the Gricean model, the bridge from what is said (the literal content of the uttered sentence, determined by its grammatical structure with the reference of indexicals resolved) to what is communicated is built through implicature. As an aspect of speaker meaning, implicatures are distinct from the non-logical inferences the hearer draws; it is a category mistake to attribute implicatures either to hearers or to sentences (e.g. *P and Q*) and subsentential expressions (e.g. *some*). But we can systematically (at least for generalized implicatures; see below) correlate the speaker's intention to implicate *q* (in uttering *p* in context *C*), the expression *p* that carries the implicature in *C*, and the inference of *q* induced by the speaker's utterance of *p* in *C*.

Subtypes of implicature are illustrated by (1a–c) (after Grice 1961: §3); the primed member of each pair is (in certain contexts) deducible from its unprimed counterpart:

- (1)a. Even KEN knows it's unethical.
a'. Ken is the least likely [of a contextually invoked set] to know it's unethical.

- b. [in a recommendation letter for a philosophy position]
Jones dresses well and writes grammatical English.
- b'. Jones is no good at philosophy.
- c. The cat is in the hamper or under the bed.
- c'. I don't know for a fact that the cat is under the bed.

Unlike an entailment or logical presupposition, the inference induced by *even* in (1a, a') is irrelevant to the truth conditions of the proposition: (1a) is true if and only if Ken knows it's unethical. The inference is not CANCELABLE without contradiction (#*Even Ken knows it's unethical, but that's not surprising*), but it is DETACHABLE, in the sense that the same truth-conditional content is expressible in a way that removes (detaches) the inference: *KEN knows it's unethical (too)*. Such detachable but non-cancelable aspects of meaning that are neither part of, nor calculable from, what is said are CONVENTIONAL implicatures, akin to pragmatic presuppositions (Stalnaker 1974). Indeed, along with connectives like *but*, the now classic instances of conventional implicature involve precisely those particles traditionally analyzed as instances of pragmatic presupposition: the additive component of adverbial particles like *even* and *too*, the "effortful" component of truth-conditionally transparent "implicatives" like *manage* and *bother*, and the existential component of focus constructions like clefts.

But in contrast with these non-truth-conditional components of an expression's conventional lexical meaning,¹ the inferences induced by (1b, c) are NON-conventional, i.e. calculable from the utterance of such sentences in a particular context, given the nature of conversation as a shared goal-oriented enterprise. In both cases, the speaker's implicature of the corresponding primed proposition is cancelable (either explicitly by appending material inconsistent with it – "*but I don't mean to suggest that . . .*" – or by altering the context of utterance) but non-detachable (given that any other way of expressing the literal content of (1b, c) in the same context would license the same inference).² What distinguishes (1b) from (1c) is the generality of the circumstances in which the inference is ordinarily licensed. Only when the speaker of (1b) is evaluating the competence of the referent for a philosophy position will the addressee normally be expected to infer that the speaker had intended to convey the content of (1b'); this is an instance of PARTICULARIZED conversational implicature.³ In (1c), on the other hand, the inference – that the speaker does not know in which of the two locations the cat can be found – is induced in the absence of a special or marked context. The default nature of the triggering in (1c) represents the linguistically significant concept of GENERALIZED conversational implicature. But in both cases, as with conventional implicature, it is crucially not the proposition or sentence, but the speaker or utterance, that induces the relevant implicature.

The significance of the generalized/particularized dichotomy has been much debated; cf. Hirschberg (1991) and Carston (1995) for skepticism and Levinson (2000a) for a spirited defense.⁴ Whatever the theoretical status of the distinction,

it is apparent that some implicatures are induced **only** in a special context (if Mr. Jones had been applying for a job as a personal secretary, Grice's remark in (1b) would have helped, rather than torpedoed, his candidacy), while others go through **unless** a special context is present (as in the utterance of (1c) as a clue in a treasure hunt). The contrast between particularized and generalized implicature emerges clearly in this scene from *When Harry Met Sally* (1989 screenplay by Nora Ephron). Harry (Billy Crystal) is setting up a blind date between his buddy Jess (Bruno Kirby) and his woman friend – but not (yet) girlfriend – Sally (Meg Ryan):

- (2) Jess: *If she's so great why aren't YOU taking her out?*
 Harry: *How many times do I have to tell you, we're just friends.*
 Jess: *So you're saying she's not that attractive.*
 Harry: *No, I told you she IS attractive.*
 Jess: *But you also said she has a good personality.*
 Harry: *She HAS a good personality.*
 Jess: *[Stops walking, turns around, throws up hands, as if to say "Aha!"]*
 Harry: *What?*
 → Jess: *When someone's not that attractive they're ALWAYS described as having a good personality.*
 Harry: *Look, if you were to ask me what does she look like and I said she has a good personality, that means she's not attractive. But just because I happen to mention that she has a good personality, she could be either. She could be attractive with a good personality or not attractive with a good personality.*
 Jess: *So which one is she?*
 Harry: *Attractive.*
 ⇒ Jess: *But not beautiful, right?*

Jess's first arrowed observation incorrectly reanalyzes a particularized implicature (S, in describing X to H as having a good personality implicates that X is not attractive) as generalized, to which Harry responds by patiently pointing out the strongly context-dependent nature of the inference in question. To see that this is no isolated example, consider a parallel dialogue from an earlier film, *The Shop Around the Corner* (1940 Ernst Lubitsch screenplay). Kralik (James Stewart) is describing his epistolary innamorata to his colleague Pirovitch (Felix Bressart):

- (3) Kralik: *She is the most wonderful girl in the world.*
 Pirovitch: *Is she pretty?*
 Kralik: *She has such ideals, and such a viewpoint of things that she's so far above all the other girls that you meet nowadays that there's no comparison.*
 → Pirovitch: *So she's not very pretty.*

Like Jess, Pirovitch (who, like Jess above, employs *so* to mark his pragmatic inference) misapplies the (here, tacit) inferential strategy to conclude from Kralik's impassioned (if unparsable) tribute to his love's virtues that she must be physically unprepossessing; in fact, Kralik believes (falsely) that he hasn't yet met her in the flesh, so no such implicature could have been made.

While the inferential step marked by the single arrows is indeed particularized and therefore context-dependent in the strong sense, the inference drawn by Jess at the double arrow is generalized, instantiating SCALAR IMPLICATURE, the upper-bounding of a weak predication ("X is attractive") to convey that the speaker was not in a position to assert any stronger counterpart ("X is beautiful"). The pattern exemplified by Jess's inference, and the reason why Jess is once again wrong to draw it, follow from our later discussion.

To conclude our brief taxonomy of implicature, we should note that despite extensive investigation in work culminating with Karttunen and Peters (1979), conventional implicature remains a controversial domain. While it continues to be invoked to handle non-truth-conditional aspects of lexical meaning, this tends to constitute an admission of analytic failure, a label rather than true explanation of the phenomenon in question. It has on occasion been maintained that conventional implicature is a myth (Bach 1999b), and even for the true believers, the domain in which such implicatures have been posited continues to shrink, eaten away on one side by an increasingly fine-grained understanding of truth-conditional meaning and entailment⁵ (a trend begun in Wilson and Sperber 1979; see also Blakemore and Carston, this volume) and on the other by a more sophisticated employment of the tools of conversational implicature. While conventional implicature remains a plausible *faute de mieux* account of particles like *even* and *too*, whose contribution has not convincingly been shown to affect the truth conditions of a given utterance but is not derivable from general considerations of rationality or cooperation, the role played by conventional implicature within the general theory of meaning is increasing shaky.

2 Speaker Meaning, Inference, and the Role of the Maxims

Whether generalized or particularized, conversational implicature derives from the shared presumption that S and H are interacting rationally and cooperatively to reach a common goal. A speaker S saying *p* and implicating *q* can count on her interlocutor to figure out what S meant (in uttering *p* at a given point in the interaction) from what was said, based on the assumption that both S and H are rational agents. Speakers implicate, hearers infer. While work as distinct as that of Levinson (2000a) and Sperber & Wilson (1986a) often appears to assimilate implicature to non-logical inference, the two phenomena were quite distinct for Grice (1989) (see Bach 2001a and Saul 2002 for discussion). While successful communication commonly relies on implicature,

what a speaker implicates is often quite distinct from what her words imply or from what a hearer may be expected to take from them.

But it is *S*'s assumption that *H* will draw the appropriate inference from what is said that makes implicature a rational possibility. The governing dictum is the Cooperative Principle: "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" (Grice [1967]1989: 26).⁶ This general principle is instantiated by general maxims of conversation governing rational interchange (1989: 26–7):

- (4) **QUALITY:** 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 evidence.
- QUANTITY:**
 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.
- RELATION:** Be relevant.
- MANNER:** Be perspicuous.
 1. Avoid obscurity of expression.
 2. Avoid ambiguity.
 3. Be brief. (Avoid unnecessary prolixity.)
 4. Be orderly.

The fourfold set of macroprinciples has no privileged status, except as a nod to Kant's own categorical tetralogy. Note in particular that all maxims are not created equal. Following Grice himself –

The maxims do not seem to be coordinate. The maxim of Quality, enjoining the provision of contributions which are genuine rather than spurious (truthful rather than mendacious), does not seem to be just one among a number of recipes for producing contributions; it seems rather to spell out the difference between something's being, and (strictly speaking) failing to be, any kind of contribution at all. False information is not an inferior kind of information; it just is not information. (Grice 1989: 371)

– many (e.g. Levinson 1983, Horn 1984a) have accorded a privileged status to Quality, since without the observation of Quality, or what Lewis (1969) calls the convention of truthfulness, it is hard to see how any of the other maxims can be satisfied (though see Sperber and Wilson 1986a for a dissenting view).

But the role of the maxims is a more central problem. It is chastening to realize that for all the work inspired by the Gricean paradigm since the William James lectures first circulated in mimeo form among linguists and philosophers in the late 1960s, the nature of the enterprise stubbornly continues to be misunderstood. (See Green 1990 for an inventory of such misunderstandings.) Here is Exhibit A:

Communication is a cooperative effort, and as such should conform to certain definite rules, or maxims of conversation, which Grice enumerates. The maxims presuppose an almost Utopian level of gentlemanly conduct on the part of a speaker and an old-fashioned standard of truthfulness that George Washington might have found irksome.⁷ They remind one of the early Puritanism of the Royal Society. A speaker should give not too much but just enough information, hold his tongue about what he believes to be false, or for which he has insufficient evidence, be relevant, be brief and orderly, avoid obscurity of expressions and ambiguity. . . . Would we want to have dinner with such a person, such an impeccably polite maxim observer? (Campbell 2001: 256)

This passage is taken from Jeremy Campbell's natural history of falsehood, a treatise hailed by reviewers as "carefully researched," "enlightening," and "thought-provoking," an "almost breathless exercise in intellectual synthesis." But it is not just the laity who are at fault; professional linguists and ethnographers, following Keenan (1976), have at times concluded that Grice's maxims are trivial, naïve to the point of simple-mindedness, and/or culture-dependent (if not downright ethnocentric), and that they fail to apply to phatic and other non-information-based exchanges.

But neither the Cooperative Principle nor the attendant maxims are designed as prescriptions for ethical actions or as ethnographic observations.⁸ A more accurate approximation is to view them as default settings (or presumptions, à la Bach and Harnish 1979), the mutual awareness of which, shared by speech participants, generates the implicatures that lie at the heart of the pragmatic enterprise. Only if the speaker is operating, and presumes the hearer is operating, with such principles as defaults can she expect the hearer to recognize the apparent violation of the maxims as a source of contextual inference (see Grice 1989, Green 1996a, Levinson 2000a for elaboration). Further, as with presupposition (on the pragmatic account of Stalnaker 1974), conversational implicature operates through the mechanism of EXPLOITATION. Unlike syntactic and semantic rules, pragmatic principles and conventions do as much work when they are apparently violated – when speaker S counts on hearer H to recognize the apparent violation and to perform the appropriate contextual adjustment – as when they are observed or ostentatiously violated.

3 Scalar Implicature and Constraints on Lexicalization

For linguistic pragmatics, the core of the Gricean system is the first Quantity submaxim, which is systematically exploited to yield upper-bounding generalized conversational implicatures associated with scalar values (Horn 1972, 1989; Gazdar 1979; Hirschberg 1991). Under a variety of formulations, this principle and its explanatory potential have long been tacitly recognized,

especially for the interpretation of quantified sentences. Sir William Hamilton (1860: 254) distinguishes two senses of *some*, the INDEFINITE (*at least some*) and the SEMI-DEFINITE (*some but not all*), taking the latter as basic: “Some, if not otherwise qualified, means some only – this by presumption.” While acknowledging that such a presumption holds in “common language,” De Morgan (1847) offers a proto-Gricean argument for rejecting Hamilton’s thesis in favor of the standard practice of relegating the *some* → *not all* inference to an extralogical domain, as does Mill (1867: 501):

No shadow of justificaton is shown . . . for adopting into logic a mere sous-entendu of common conversation in its most unprecise form. If I say to any one, “I saw some of your children today”, he might be justified in inferring that I did not see them all, not because the words mean it, but because, if I had seen them all, it is most likely that I should have said so: even though this cannot be presumed unless it is presupposed that I must have known whether the children I saw were all or not.

Similarly, while disjunctions are naturally taken exclusively – “When we say A is either B or C we imply that it cannot be both” – this too cannot be a logical inference: “If we assert that a man who has acted in a particular way must be either a knave or a fool, we by no means assert, or intend to assert, that he cannot be both” (Mill 1867: 512).

Notice Mill’s epistemic rider in his *unless* clause: S’s use of the weaker *some* implicates that **for all S knows** the strongest operator on the same scale, *all*, could not have been substituted *salva veritate*. Mill’s tacit principle, with its epistemic condition, is independently invoked by later scholars:

What can be understood without being said is usually, in the interest of economy, not said . . . A person making a statement in the form, “Some S is P”, generally wishes to suggest that some S also is not P. For, in the majority of cases, if he knew that all S is P, he would say so . . . If a person says, “Some grocers are honest”, or “Some books are interesting”, meaning to suggest that some grocers are not honest or that some textbooks are not interesting, he is really giving voice to a conjunctive proposition in an elliptical way.

Though this is the usual manner of speech, there are circumstances, nevertheless, in which the particular proposition should be understood to mean just what it says and not something else over and above what it says. One such circumstance is that in which the speaker does not know whether the subcontrary proposition is also true; another is that in which the truth of the subcontrary is not of any moment. (Doyle 1951: 382)

The tacit principle to which Mill alludes, requiring S to use the stronger *all* in place of the weaker *some* when possible and licensing H to draw the corresponding inference when the stronger term is not used, later resurfaces within Grice’s program as the first Quantity maxim, which is systematically

exploitable to yield upper-bounding generalized conversational implicatures associated with scalar operators. Quantity-based scalar implicature – e.g. my inviting you to infer from my use of *some . . .* that for all I know *not all . . .* – is driven by our presumed mutual knowledge that I expressed a weaker proposition in lieu of an equally unmarked utterance that would have expressed a stronger proposition. Thus, what is said in the use of a weaker scalar value like those in boldface in the sentences of (5) is the lower bound (. . . *at least n . . .*), with the upper bound (. . . *at most n . . .*) implicated as a cancelable inference generated by (some version of) the first maxim of quantity. What is communicated in the default case is the TWO-SIDED UNDERSTANDING that combines what is said with what is implicated.

(5)	ONE-SIDED →	TWO-SIDED
a. Pat has 3 children.	"...at least 3..."	"...exactly 3..."
b. You ate some of the cake.	"...some if not all..."	"...some but not all..."
c. It's possible she'll win.	"...at least \diamond ..."	"... \diamond but not certain..."
d. He's a knave or a fool.	"...and perhaps both"	"...but not both"
e. It's warm .	"...at least warm..."	"...but not hot"

The alternative view, on which each scalar predication in (5) is lexically ambiguous between one-sided and two-sided readings, is ruled out by the general metatheoretical consideration that Grice dubs the Modified Occam's Razor principle: "Senses are not to be multiplied beyond necessity" (1989: 47).

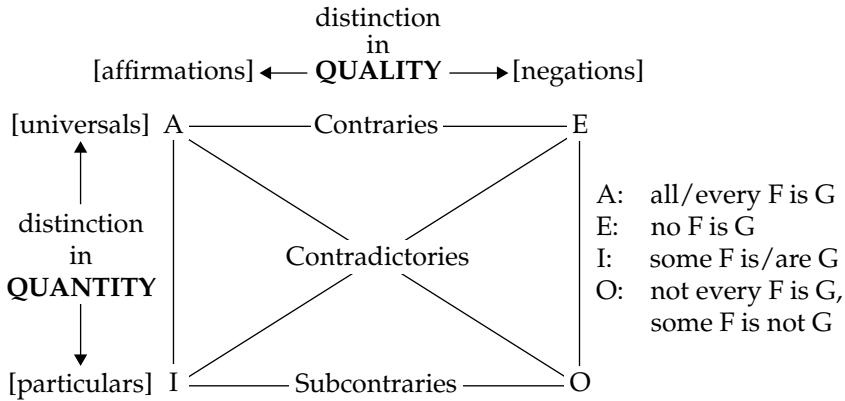
Negating such predications denies the lower bound: to say that something is not possible is to say that it's impossible, i.e. less than possible. When it is the upper bound that appears to be negated (*It's not possible, it's NECESSARY*), a range of syntactic, semantic, and prosodic evidence indicates the presence of the METALINGUISTIC or echoic use of negation, in which the negative particle is used to object to any aspect of an alternate (actual or envisaged) utterance, including its conventional and conversational implicata, register, morphosyntactic form or pronunciation (Horn 1989: Chap. 6; Carston 1996). If it's hot, it's (a fortiori) warm, but if I know it's hot, the assertion that it's warm can be echoed and rejected as (not false but) insufficiently informative:

- (6)a. It's not WARM, it's HOT!
 b. You're right, it's not warm. It's HOT!

As seen in (6b), the metalinguistic understanding typically requires a second pass and the effect is typically that of an ironic "unsaying" or retroactive accommodation (Horn 1992).

The central role played by scalar implicature in natural language is illustrated by a systematic pattern of lexical gaps and asymmetries. Consider the post-Aristotelian square of opposition, defined by the logical relations definable between pairs of quantified expressions (ranging over non-empty sets):

(7) Square of Opposition



- (7') • Corresponding A and E statements are CONTRARIES; they cannot be simultaneously true (though they may be simultaneously false).
 • Corresponding A and O (and I and E) statements are CONTRADICTIONARIES; members of each pair cannot be true OR false simultaneously.
 • An I statement is the SUBALTERN of its corresponding A statement (and O of E); a subaltern is unilaterally entailed by its corresponding superaltern.
 • Corresponding I and O statements are SUBCONTRARIES and cannot be simultaneously false (though they may be simultaneously true).

Note in particular that the assertion of either of the two subcontraries Quantity-implicates the negation of the other. While what is said in *Some men are bald* and *Some men are not bald* is distinct, what is communicated is typically identical: *Some men are bald and some aren't*. Given that languages tend not to lexicalize complex values that need not be lexicalized, particularly within closed categories like quantifiers, we predict that *some . . . not* will not be lexicalized, and this is precisely what we find.

In a wide variety of languages, values mapping onto the southeast, O corner of the square are systematically restricted in their potential for lexicalization (Horn 1972, 1989: 4.5). Thus alongside the quantificational determiners *all, some, no*, we never find an O determiner **nall*; corresponding to the quantificational adverbs *always, sometimes, never*, we have no **nalways* (= "not always," "sometimes not"). We may have equivalents for *both, one (of them), neither*, but never for **noth* (= "not both," "at least one . . . not"); we find connectives corresponding to *and, or*, and sometimes *nor* (= "and not"), but never to **nand* (= "or not," "not . . . and"). The story of **O* extends to the modals and deontics, as illustrated by the fact that the inflected negative in *He can't go* and the orthographic lexicalization in *He cannot go* only allows wide-scope (E vertex) negation, while the unlexicalized counterpart *He can not go* is ambiguous. The

relation of mutual quantity implicature holding between positive and negative subcontraries results in the superfluity of one of the two subcontraries for lexical realization, while the functional markedness of negation assures that the unlexicalized subcontrary will always be **O**.

4 Q-based and R-based Implicature: Clash and Resolution

The earliest discussions of scalar quantity implicature were based on the informative content associated with values whose lexical semantics defined the relevant scale: *necessarily p* entails *possibly p* and not vice versa, whence the implicature from the utterance of the latter to the negation of the former. But as Fauconnier (1975b) and especially Hirschberg (1991) have eloquently shown, scales must be essentially pragmatic in nature. Indeed, Hirschberg has demonstrated that not just scales as such but any **POSET** (partially ordered set) can in principle define a quantity implicature in the right context. Thus if Robin is traveling westward from New York to California, my utterance *Robin has made it to Chicago* will implicate that Robin hasn't made it to Denver, but will not implicate that she hasn't yet reached Cleveland. As usual, such implicatures can be cancelled (*Not only has Robin made it to Chicago, but to Denver*). If Robin were traveling eastward, the facts would be reversed. (See Hirschberg 1991 for extensive elaboration.)

M. Walker (1994) extends Hirschberg's results to show how quantity implicature functions to implicitly reject a proposition consistent with the context (cf. also Horn 1989: 410). Thus, in response to your question, "Is Smith honest and ambitious?" or to your assertion, "Smith is honest and ambitious," my assertion, "He's ambitious" will convey my belief that he's not honest; this proposition is, in Walker's terms, rejected by implicature. (See Ward and Hirschberg 1985, Horn 1989, and M. Walker 1994 on the role of intonation in such examples.) An attested example of the same phenomenon was provided in the exchange in (8) from the Senate investigation of President Clinton. Senator Ed Bryant is interrogating Monica Lewinsky on her affidavit in the Paula Jones case:

- (8) Mr. Bryant: "Were portions of it false?"
Ms. Lewinsky: "Incomplete and misleading."

In implicating (but not saying) that no portions of her affidavit were technically false, Lewinsky, in the words of *New York Times* reporter Francis X. Clines (February 6, 1999), "exhibited a Clintonian way with the meaning of words."

Other questions arising in early work on implicature concern the nature and scope of implicature. While the utterance of a weaker scalar value . . . $p(i)$. . . tends to implicate that the speaker was not in a position to assert the correspondingly stronger value . . . $p(j)$. . . (thereby implicating against the stronger

value), this tendency is subject to a variety of constraints. The inheritance or projection properties of conversational implicature is a complex matter; it appears (Horn 1989: 234) that scalar implicature is blocked in precisely those environments where “scale reversal” applies, in the context of downward-entailing operators like negation and other negative polarity item (NPI) triggers, whence the disappearance of the upper-bounding implicature (*possible* +> *not certain*) in *If it’s possible that it will rain I’ll bring an umbrella*. (Chierchia 2001 argues from this correlation for the semantic status of scalar implicature; cf. Sauerland 2001.) Levinson (2000a: 80), on the other hand, has noted that if scale reversal is taken seriously, implicature need not be extrinsically blocked in such environments; rather, it will predictably be associated with the opposite scale, given the generalization that the negation of a weak positive value will constitute a strong value on the corresponding negative scale, and vice versa for the corresponding strong positive, for example (using the standard ⟨STRONGER-VALUE, WEAKER-VALUE⟩ notation), ⟨*certain, possible*⟩ vs. ⟨*not possible, not certain*⟩.

The significance of the first Quantity maxim for the form and function of natural language reflects its status as one of two cardinal principles regulating the economy of linguistic information. Setting Quality aside as unreducible, we can collapse the remaining maxims and submaxims into two fundamental principles corresponding to Zipf’s “speaker’s and auditor’s economies” (1949: 20ff.; cf. Horn 1984a). The Q principle is a lower-bounding hearer-based guarantee of the sufficiency of informative content (“Say as much as you can, modulo Quality and R”); it collects the first Quantity maxim along with the first two “clarity” submaxims of manner and is systematically exploited (as in the scalar cases discussed above) to generate upper-bounding implicature. The R principle, by contrast, is an upper-bounding correlate of the Law of Least Effort dictating minimization of form (“Say no more than you must, modulo Q”); it collects the Relation maxim, the second Quantity maxim, and the last two submaxims of Manner, and is exploited to induce strengthening implicature.⁹ Q-based implicature is typically negative in that its calculation refers crucially to what could have been said but wasn’t: H infers from S’s failure to use a more informative and/or briefer form that S was not in a position to do so. R-based implicature involves social rather than purely linguistic motivation and is exemplified by indirect speech acts and negative strengthening (including so-called neg-raising, i.e. the tendency for *I don’t think that ϕ* to implicate *I think that not- ϕ*).

R-based implicature, while calculable, are often not calculated on line; a specific form of expression may be associated with a given pragmatic effect while an apparently synonymous form is not. Thus the question *Can you close the window?* is standardly used to convey an indirect request while *Are you able to close the window?* is not; *I don’t guess that ϕ* allows a strengthened “neg-raised” understanding in a proper subset of the dialects for which *I don’t think that ϕ* does. These are instances of SHORT-CIRCUITED CONVERSATIONAL IMPLICATURE OR STANDARDIZED NON-LITERALITY (cf. Morgan 1978, Bach and Harnish 1979, Bach 1987b, Horn 1989).

The Zipfian character of the implicata generating indirect speech acts was recognized by Searle in his proposal for a condition on directives that “It is not obvious to both S and H that S will do A in the normal course of events”:

I think this condition is an instance of the sort of phenomenon stated in Zipf’s law. I think there is operating in our language, as in most forms of human behaviour, a principle of least effort, in this case a **principle of maximum illocutionary ends with minimum phonetic effort**, and I think [this] condition is an instance of it. (Searle 1965: 234–5, my emphasis)

Similar cost/benefit or minimax principles have been proposed by Paul, Zipf, and Martinet (see Horn 1993 for references and discussion) and by Carroll and Tanenhaus (1975: 51): “The speaker always tries to optimally minimize the surface complexity of his utterances while maximizing the amount of information he effectively communicates to the listener.” Indeed, the interplay of perspicuity (or clarity) and brevity was a key issue for classical rhetoricians, who advanced their own minimax guidelines:

If it is prolix, it will not be clear, nor if it is too brief. It is plain that the middle way is appropriate . . . , saying just enough to make the facts plain. (Aristotle, *Rhetoric*, 3.12–3.16)

Brevis esse laboro; obscurus fio. ‘I strive to be brief; I become obscure’. (Horace, *Ars Poetica*, l. 25)

Personally, when I use the term brevity [*brevitas*], I mean not saying less, but not saying more than the occasion demands. (Quintilian, *Institutio Oratio*, iv.ii.41–43)

While the bilateral brevity of Quintilian may seem quirky, it is no more so than current redefinitions of relevance as a minimax equilibrium of effort and effect:

Human cognitive activity is driven by the goal of maximizing relevance: that is . . . to derive as great a range of contextual effects as possible for the least expenditure of effort. (Carston 1995: 231)

The two antinomic Q and R forces interact definitionally and dynamically, each referencing and constraining the other.¹⁰ Grice himself incorporates R in defining the primary Q maxim (“Make your contribution as informative as is **required**” [here and below, emphasis added]), while Quantity₂ is constrained by Quantity₁¹¹ and essentially incorporates Relation: what could make a contribution more informative than required, except the inclusion of contextually irrelevant material? This interdependence was noted by Martinich (1980: 218), who urged collapsing Q₁ and Q₂ into a joint maxim dictating that the speaker “contribute as much as, but not more than, is required (for the current

purposes of the exchange),” while rejecting the broader Relation as a “marauding maxim.”

The role of relevance and clarity in constraining the informative strength of the Q principle emerges in its various incarnations, beginning with Strawson (1952: 178–9), who credits Grice for his “general rule of linguistic conduct”: “One should not make the (logically) lesser, when one could truthfully (**and with greater or equal clarity**) make the greater claim.” Grice’s (1961: 132) own “first shot” at the relevant rule is bound by a similar rider – “One should not make a weaker statement rather than a stronger one **unless there is a good reason for so doing**” – as are later versions of the principle constructed in the wake of the maxim of quantity:

Make the strongest possible claim **that you can legitimately defend!**

RULE OF STRENGTH (Fogelin 1967: 20–2)

Unless there are outweighing good reasons to the contrary, one should not make a weaker statement rather than a stronger one **if the audience is interested in the extra information that would be conveyed by the latter.**

(O’Hair 1969: 45)

Make the strongest relevant claim **justifiable by your evidence.**

MAXIM OF QUANTITY-QUALITY (Harnish 1976: 362)

The “good reason” for avoiding the stronger scalar value thus may be either qualitative, constrained by truth (S doesn’t know that the stronger value is applicable), or quantitative, where both relevance and brevity enter the picture (S doesn’t believe the extra information is justified in terms of H’s interests or S’s own efforts in uttering it). Telling you that my wife is either in the kitchen or the bedroom, I will (*ceteris paribus*) Q-implicate that I don’t know that she’s in the kitchen – but I can tell you “The kitchen is a mess” without implicating that the bedroom isn’t. If you tell me X is possibly true, I will infer you don’t know it’s true, but if you tell me X is true (e.g. that all bachelors are unmarried), I will not infer you don’t know it’s necessarily true. The use of a weak I or O proposition licenses the inference that the speaker was not in a position to use the basic unquantified, unmodalized proposition that unilaterally entails it, as the Q principle predicts, but the use of the basic propositional form does not Q-implicate the negation of its strong counterpart, A or E respectively. Since there is no quantity- or information-based distinction between these (sub)subalternations, we must seek the source of the asymmetry elsewhere.

The crucial distinction here relates not to the content (what is said) but to the form (how what is said is said). Because the basic forms are not only more informative but briefer than their I/O counterparts, the use of the latter will strongly implicate against the former. But the strong values, while more informative than their unmodified counterparts, are also more prolix, so Quantity is offset by Manner and potentially by Relation: the Q principle of

informative sufficiency yields to the R principle of least effort. The richness of the pragmatic framework allows us to predict not just what can be implicated but what will be implicated in a given context.

When degree of lexicalization is not a factor, scalar implicature is normally generated. Thus, each of the ordered n-tuples of items in (9)

- (9) *<always, usually, often, sometimes>*, *<and, or>*, *<certain, likely, possible>*, *<cold, cool, lukewarm>*, *<excellent, good, OK>*, *<the, a>*

constitutes a Q-relevant scale in that the affirmation of any weak or intermediate value will implicate (*ceteris paribus*) that – for all the speaker knows – the value(s) on its left could not be substituted *salva veritate*.

But when the stronger value is less economical than the weaker one, no Q-implicature is triggered. Thus the apparent symmetry of the relevant scales – *<x and y won, x won, x or y won>*, *<a must be F, a is F, a may be F>* – is misleading. This extends to non-quantitative “scales” of items differing in informative strength. Thus, while the use of *finger* typically conveys “non-thumb,” it does not convey “non-pinky (finger),” nor does the use of *toe* convey “non-big toe,” although the big toe is an analogue of the thumb. Crucial here is the status of *thumb* (as opposed to *pinky*) as a lexicalized alternative to *finger*. In the same way, *rectangle* conveys “non-square” (i.e. “non-equilateral rectangle”) given the availability of the lexicalized alternative *square*, while *triangle* does not convey “non-equilateral triangle” – indeed, the prototype triangle IS equilateral – because of the non-existence of a lexicalized counterpart.

One robust linguistic phenomenon involving the interaction of Q and R principles is the DIVISION OF PRAGMATIC LABOR. Given two expressions covering the same semantic ground, a relatively unmarked form – briefer and/or more lexicalized – tends to be R-associated with a particular unmarked, stereotypical meaning, use, or situation, while the use of the periphrastic or less lexicalized expression, typically more complex or prolix, tends to be Q-restricted to those situations outside the stereotype, for which the unmarked expression could not have been used appropriately.¹² Thus consider the following pairs:

- (10)a. He got the machine to stop.
He stopped the machine.
- b. Her blouse was pale red.
Her blouse was pink.
- c. She wants her to win.
She wants PRO to win.
- d. I am going to marry you.
I will marry you.
- e. My brother went to the church (the jail, the school).
My brother went to church (jail, school).
- f. It’s not impossible that you will solve the problem.
It’s possible that you will solve the problem.

- g. That's my father's wife.
That's my mother.

The use of the periphrastic causative in (10a) implicates that the agent achieved the effect in a marked way (pulling the plug, throwing a shoe into the machine), *pale red* in (10b) suggests a tint not pre-empted by *pink*, the choice of a full pronoun over null PRO in (10c) signals the absence of the coreferential reading associated with the reduced syntax, the periphrastic form blocks the indirect speech act function of promising conveyed by the modal in (10d), the full Det-N versions of (10e) imply literal motion toward the specified location without the socially stereotypic connection that is R-associated with the corresponding institution on the anarthrous version, the double contradictory negation in (10f) signals a rhetorical effect absent from the direct positive, and the more complex description in (10g) suggests that the more basic and lexicalized alternative could not have been used appropriately (the referent is probably the speaker's stepmother). When a speaker opts for a more complex or less fully lexicalized expression over a simpler alternative, there is a pragmatically sufficient reason, but which reason depends on the particular context. (See Horn 1991a, 1993, Levinson 2000a, and Blutner and Traugott, this volume, for references and related discussion.)

A particularly rich explanatory vein lies in the realm of anaphora, in which the choice of an overt pronoun over controlled PRO in infinitivals in both English object raising (ECM) and Romance subjunctive constructions can be attributed to the division of pragmatic labor, as can switch-reference phenomena and the use of an overt subject in a pro-drop (null-subject) language like Turkish or Catalan, in which the overriding of "Avoid Pronoun" will often implicate change of topic. Valuable cross-linguistic studies of the neo-Gricean pragmatics of anaphora, with copious references, are provided in Levinson (2000a: Chapter 4) and Huang (2000a, this volume).

5 Implicature, Explicature, and Pragmatic Intrusion

Where the model we have been exploring retains two antinomic principles Q and R along with an unreduced maxim of Quality, and where the related model of neo-Gricean pragmatics urged by Levinson (2000a) contains the three Q, I, and M heuristics, a more radical simplification has been urged in the framework of relevance theory, in which a reconceptualized Principle of Relevance is taken to be the sole source of pragmatic inference.¹³ At the heart of this program is a reworking of the architecture of the theory of logical form and utterance interpretation (Sperber and Wilson 1986a; cf. also Carston 1998b, this volume; Wilson and Sperber, this volume).

Even for Grice, propositional content is not fully fleshed out until reference, tense, and other indexical elements are fixed. But, taking their lead from earlier

work by Atlas (1979), relevance theorists have argued that the pragmatic reasoning used to compute implicated meaning must also be invoked to flesh out underspecified propositions in which the semantic meaning contributed by the linguistic expression itself is insufficient to yield a proper accounting of truth-conditional content. Thus, to take one example, when a pundit observed, as the jury retired to consider their verdict in the O. J. Simpson murder trial, that “It will take them some time to reach a verdict,” the proposition he communicated (that it will take a long time) seems intuitively false, a fact hard to reconcile with a strict Gricean analysis on which the time communicated by S is merely an implicatum read off the underspecified content contributed by linguistic meaning alone, i.e. a trivially true existential proposition. Apparently the pragmatically strengthened communication contributes to, or intrudes upon, the propositional content.

A classic example of such apparent intrusion is illustrated by the temporal and causal asymmetry of conjoined event-denoting VPs and sentences. The logical “&” is a symmetric truth function; “p & q” is true if p and q are both true and false otherwise (as, of course, is “q & p”). Strawson (1952: 80) pointed to the apparent contrast in meaning exhibited by pairs like (11a, b)

- (11)a. They got married and (they) had a child.
- b. They had a child and (they) got married.
- c. They got married and then (they) had a child.

as *prima facie* counterexamples to this thesis, since the former appears to amount to the statement in (11c). (I add the parenthetical pronoun to render these sentences closer to the corresponding logical conjunctions, although that renders the asymmetric understanding less inevitable.) Similarly, Ryle (1954) famously observed that to get on one’s horse and ride away is not the same as to ride away and get on one’s horse. For Urmson (1956: 9–10), however, the truth-functional picture, while incomplete, is not *ipso facto* incorrect:

In formal logic, the connectives “and” and “or” are always given a minimum meaning . . . such that any complex formed by the use of them alone is a truth-function of its constituents. In ordinary discourse the connectives often have a richer meaning; thus “he took off his clothes and went to bed” implies temporal succession and has a different meaning from “he went to bed and took off his clothes”. Logicians would justify their use of the minimum meaning by pointing out that it is the common element in all our uses of “and.”

For the classical Gricean approach, an assertion of the conjunction in (11a) will implicate (11c) by virtue of the Manner submaxim “Be orderly” (Grice 1981: 186). Indeed, Grice’s approach was prefigured in the observation that “Events earlier in time are mentioned earlier in the order of words than those which occurred later,” one of the eight “natural principles” that influence word order in the inventory of Dionysius of Halicarnassus, *Peri syntheseos*

onomaton (*On the Juxtaposition of Words*), first century BC, cited in de Jonge (2001).

On this Dionysian/Gricean line, the distinction in meaning between (11a, b) need not be laid at the doorstep of an ambiguous *and* operator. For those who would semanticize temporal asymmetry, such a lexical ambiguity must be invoked for the fact that a non-sequential interpretation is available not only for non-eventive sentences (*They are tall and they are rich*) but even for (11a) in the appropriate context (“What stressed them out last year?”). Arguments against a lexical ambiguity for *and* (“and also” vs. “and then”) include the following:

- 1 On the two-*and* theory, conjunction in (almost?) every language would just happen to be similarly ambiguous.
- 2 No natural language contains a conjunction *shmand* that would be ambiguous between “and also” and “and earlier” readings so that *They had a baby shmand they got married* would be interpreted either atemporally or as “They had a baby and, before that, they got married.”
- 3 Not only temporal but causal asymmetry would need to be built in, as a variety of apparent strengthenings of the conjunction arise in different contexts of utterance.
- 4 The same “ambiguity” exhibited by *and* arises when two clauses describing related events are juxtaposed *asyndetically* (*They had a baby. They got married.*)

However, if conjunctions are semantically univocal while Manner- (or R-) implicating that the events occurred in the order in which they were described, the impossibility of the conjunction *shmand* can be attributed to the absence of any maxim enjoining the speaker to “Be disorderly.” As with scalar implicature, the asymmetric implicatum may be canceled or suspended: *They had a baby and got married, but not necessarily in that order.*

But if the “and then” reading comes in only as an implicature, it is hard to explain its apparent contribution to truth-conditional meaning in embedded contexts, and in particular the non-contradictory nature of (12a–c) as pointed out by Cohen (1971) and Wilson (1975):

- (12)a. If they got married and had a child, their parents will be pleased, but if they had a child and got married their parents will not be pleased.
 - b. They didn’t get married and have a child; they had a child and got married.
 - c. It’s more acceptable to get married and have a child than to have a child and get married.

One possible conclusion is that while pragmatically derived, the enriched meaning is an EXPLICATURE, corresponding to what is said rather than to what is (merely) implicated¹⁴ (see Carston, this volume); another is that we must

revisit the architecture of Gricean theory to allow implicature to help determine propositional content (Levinson 2000a: chapter 3).

The explicature view also yields a re-evaluation of the traditional view of scalar predications, so that both one-sided and two-sided understandings of sentences in (5) will now be directly represented at the level of logical content. While such scalar predications are now all taken to be ambiguous, the ambiguity is situated not at the lexical but at the propositional level: what is **said** in an utterance is systematically underdetermined by the linguistic content of what is **uttered**. In particular, it does not seem possible to maintain the original Gricean line on the meaning of cardinal operators (lower-bounded by meaning, upper-bounded by implicature).

However, while a strong case can be made for an enrichment analysis of the meaning contribution of the cardinals, it does not generalize straightforwardly to the inexact scalar values. Evidence for this asymmetry (summarized in Horn 1992) comes from the contextual reversibility of cardinal scales and the non-implicating (“exactly *n*”) reading of cardinals in mathematical, collective, and elliptical contexts, none of which applies to the scalar operators in, for example, (5b–e). Note also the contrast in the exchanges below:

- | | |
|---|--|
| (13) A: Do you have two children? | (14) A: Did many of the guests leave? |
| B ₁ : No, three. | B ₁ : ?No, all of them. |
| B ₂ : ?Yes, (in fact) three. | B ₂ : Yes, (in fact) all of them. |

Further, a bare negative response to (13A) is compatible with an “exactly *n*” reading in an appropriate context (if B believes A is interested in precisely how many children B has, rather than in B’s candidacy for a subsidy), while an unadorned negative response to (14A) can only be understood as conveying “fewer than many.” In the same way, there is a sharp contrast between the “game-playing” nature of (15a), with ordinary scalar *like*, and the straightforward (15b), with cardinal values:

- (15)a. #Neither of us liked the movie – she adored it and I hated it.
 b. Neither of us has three kids – she has two and I have four.

Similarly, if (5e) were truly propositionally ambiguous, there is no obvious reason why a “No” response to the question “*Is it warm?*” should not be interpretable as a denial of the enriched, two-sided content and thus as asserting that it’s either chilly or hot, or why the comparative in “*It’s getting warmer*” cannot denote “less hot” instead of “less cold.” This suggests the need for a mixed theory in which cardinal values may well demand an enriched-content analysis, while other scalar predications continue to warrant a standard neo-Gricean treatment on which they are lower-bounded by their literal content and upper-bounded, in default contexts, by Q-implicature.

Standard critiques (e.g. Carston 1988, Recanati 1989) of traditional Gricean accounts of scalar implicature can be countered if this distinction between

cardinals and other scalar values is maintained. Nor is it surprising to see the same distinction surfacing as significant in early childhood, as has been supported by recent work in developmental psycholinguistics (Papafragou and Musolino 2003).

6 Implicature vs. Implicature: “What is said” Revisited

But are we really dealing with post-semantic implicature here in the original Gricean sense, or with a different aspect of what isn’t said? The arguments we have been reviewing rest on the tacit assumption that whatever is communicated but not said must be implicated. Some (e.g. Levinson 2000a) have argued from this assumption that implicatures can affect (“intrude on”) truth-conditional meaning after all, given cases like the asymmetric conjunction in (11); others have argued instead for the notion of explicature, i.e. pragmatically determined content. But what if not all implicit components of communicated meaning are implicatures? As stressed by Bach (1994a, 2001a), some aspects of speaker meaning need not be considered either part of what is implicated or of what is said. Thus consider the following utterances with the typically conveyed material indicated in curly brackets:

- (16)a. I haven’t had breakfast {today}.
 b. John and Mary are married {to each other}.
 c. They had a baby and they got married {in that order}.
 d. Robin ate the shrimp and {as a result} got food poisoning.
 e. Everybody {in our pragmatics class} solved the riddle.

In each case, the bracketed material contributing to what is communicated cannot be derived as a Gricean implicature (*pace* Levinson 2000a: chapter 3), given that it is truth conditionally relevant, but neither can it be part of what is said, since it is felicitously cancelable:

- (17)a. John and Mary are married, but not to each other.
 b. They had a child and got married, but not necessarily in that order.

Bach has proposed that in such cases the enriched material may be regarded instead as an *IMPLICITURE*, an implicit weakening, strengthening, or specification of what is said. This permits an intuitive characterization of propositional content, a conservative mapping from syntactic structure to what is said, and an orthodox Gricean conception of implicature, albeit as a more limited construct than in much neo-Gricean work. While Levinson (2000a) bites the bullet and, accepting the relevance theorists’ arguments for pragmatic intrusion into propositional content, concludes that implicatures must feed truth-conditional

interpretation, Bach retains a neo-classically Gricean semantic characterization of what is said,¹⁵ along with a post-semantic understanding of conversational implicature: it is implicatures, not implicatures, that can determine the relevant truth conditions in such cases. Furthermore, it is misleading to take the expansions in (16) to be explicatures, since there is nothing explicit about them, and indeed the cancelability of such expanded understandings supports their status as implicit. At the same time, the standard view that every sentence expresses exactly one proposition must be abandoned, as it is typically and in some cases ONLY the implicature – the expanded proposition that the speaker communicates but does not directly express – that is plausibly assessed for truth or falsity.¹⁶

Others have reached similar conclusions by different routes. Taylor (2001) stresses the role of beliefs about the world to explain why enrichment proceeds differently in contexts like *I haven't had breakfast* vs. *I haven't had sex*, although this too could (predictably) change in a culture in which it is expected that one has sex (but not necessarily breakfast) each morning. Saul (2002) has argued persuasively that the (neo-)Gricean and relevance theoretic conceptions of meaning are not as incompatible as it may appear, given that Grice's concerns lay in an account of speaker meaning (of which implicature constitutes a proper subpart), while relevance theorists have been primarily concerned with developing a cognitive psychological model of utterance interpretation, which does not address the question of how and why the speaker, given what she wants to convey, utters what she utters. Inevitably, the two goals must part company, as Saul demonstrates in some detail. While there is a natural tendency to characterize Grice's project in terms of the plausible interpretation of utterances (whence Levinson's 2000a depiction of generalized conversational implicatures as default inferences), it must be resisted, as Bach and Saul have argued.

As for pragmatic intrusion into propositional content and the determination of truth conditions, it should be noted that the Cohen-type argument for the intrusion of temporal asymmetry into the compositional meaning of conditionals (as in (18a) vs. (18b)) can be paralleled by other cases suggesting that all natural language epistemic conditionals are *ceteris paribus* claims; the statements in (19b–d) are no better candidates for valid inferences from (18a) than is (19a).

- (18)a. If Annie got married and had a baby, her grandfather will be happy.
 b. If Annie had a baby and got married, her grandfather will not be happy.
- (19) If Annie got married and had a baby
 a. but in the opposite temporal order
 b. but her baby was born a week after the wedding
 c. but her husband was not the father of the baby
 d. but she married Sue and had the baby by artificial insemination her grandfather will be happy.

Similarly, consider the conditionals in (20), in which an explicature theorist would build the stronger (bilateral) meaning (e.g. *some but not all*, *warm but not very warm*) into what is said:

- (20)a. If some of my friends come to the party, I'll be happy – but if all of them do, I'll be in trouble.
 b. If it's warm, we'll lie out in the sun. But if it's {very warm/hot}, we'll go inside and sit in front of the air-conditioner.
 c. If you're convicted of a felony, you'll spend at least a year in jail. And if you're convicted of murder, you'll be executed.
 d. If you're injured, the paramedics will take you to the nearest trauma center. But if you're fatally injured, you'll be taken to the morgue.

In each of these contexts, it's only when the stronger scalar is reached that the earlier, weaker one is retroactively accommodated, as it were, to incorporate an upper bound into its semantics, with, for example, "some" being REinterpreted as expressing (rather than merely communicating) "some but not all." This reinterpretation is facilitated by the obligatory focus on the relevant scalar operators (*some*, *warm*, etc.).

The same issues arise for other applications of the pragmatic intrusion argument. Thus, Levinson (2000a: 210) extends the classic Cohen–Wilson argument from conditionals like (18) to the *because* clauses of (21):

- (21)a. Because he drank three beers and drove home, he went to jail.
 b. Because he earns \$40,000, he can't afford a house in Palo Alto.
 c. Because he's such a fine friend, I've struck him off my list.
 d. Because the police recovered some of the missing gold, they will later recover it all.

But these examples are heterogeneous. (21a) sports the familiar temporal strengthening, while (21b) involves a cardinal, which as we have seen is plausibly reanalyzed as involving an adjustment of what is said. The example of "such a fine friend" in (21c), on the other hand, involves conventionalization of the sarcastic meaning; cf. *?Because he's so considerate, I fired him*. The *all* in the second clause of (21d) forces the reprocessing of the *some* in the first clause as "some but not all" (a reprocessing again triggered by the focal stress on *some*); in the other examples, the general context alone is sufficient to force the narrowed interpretation. Without the *all* or a similar context-forcing continuation, this narrowing appears to be impossible:

- (22) Because the police recovered some of the gold, the thieves are expected to return later #(for the rest).

In general, such *because* cases are quite constrained, in particular for the non-cardinal scalar cases in which the implicated upper bound is taken to be the

reason for the truth of the second clause (as in (22)) and in which no reprocessing is forced by the affirmation of a stronger value. Thus consider:

- (23)a. #Because it's warm out [i.e. because it's warm but not hot], you should still wear a long-sleeved shirt.
b. #Because you ate some of your spinach [i.e. and not all], you don't get dessert.

Of course, a move from *warm* or *some* to *only warm* or *just some* render these causals impeccable, but then the scales have been reversed.

7 Implicature, Cooperation, and Rationality

As we have seen, Paul Grice's pragmatic framework in general, and the elaboration of conversational implicature in particular, are founded on the Cooperative Principle. But while cooperation is a key notion, the role of an even more general principle has not always been fully appreciated. Describing the maxims of conversation, Grice cites the basis of rationality as the reason his program extends beyond communication to non-linguistic interchanges:

As one of my avowed aims is to see talking as a special case or variety of **purposive, indeed rational behavior**, it may be worth noting that the specific expectations or presumptions connected with at least some of the foregoing maxims have their analogues in the sphere of transactions that are not talk exchanges. (Grice 1989: 28; emphasis added)

As Smith (1999: 15) has noted, the Cooperative Principle need not be stipulated as an arbitrary convention (cf. Lewis 1969), but rather constitutes "a deduction from the general principle that we expect others to behave as best suits their goals."¹⁷ The role of rationality in pragmatics has been stressed by Kasher (1982: 32), whose PRINCIPLE OF EFFECTIVE MEANS stipulates "Given a desired end, one is to choose that action which most effectively, and at least cost, attains that end, *ceteris paribus*." It will be noted that Kasher's principle incorporates the minimax of effort and cost that also underlies models as diverse as the apparently monoprincipled relevance theory (Sperber and Wilson 1986a), the dual Q- and R-based approach of Horn (1984a, 1993), and the tri-heuristic Q/I/M theory of Levinson (2000a).

In particular, the speaker's and hearer's joint (though tacit) recognition of the natural tendency to avoid unnecessary effort, and the inferences *S* expects *H* to draw from the former's efficient observance of this tendency, are more explicable directly from rationality than from cooperation as such. While Grice (1989: 28) describes how the maxims apply to cooperative ventures outside of language (baking a cake, fixing a car), collaboration need not be present, much

less communication, at least for the quantity maxims. It seems plausible to assume that the generalized forms of both Q and R principles – “Do enough; Don’t do too much” – govern ANY goal-oriented activity: a person brushing her teeth or working out a problem in the philosophy of language, a dog digging a hole to bury a bone. In this way, the maxim of quantity, in both its opposed (Q and R) subforms, is a linguistic instantiation of these rationality-based constraints on the expenditure of effort. Of course, as Grice recognized, the shared tacit awareness of such principles to generate conversational implicatures is a central property of speaker meaning within the communicative enterprise.

With a fuller understanding of the interaction of pragmatics and propositional content, we see that while the explanatory scope of conversational implicature may have been reduced from the heyday of the classical Gricean program, his framework and the pragmatic principles motivating it – rationality, common ground, and the distinction of implicit vs. explicit components of utterance meaning – continue to play a key role in the elaboration of dynamic models of context. As recent work on language acquisition (Noveck 2001, Chierchia et al. 2001, Papafragou and Musolino 2003)¹⁸ and on lexical change (Traugott and Dasher 2002; Traugott, this volume) has further demonstrated, a suitably refined and constrained notion of conversational implicature remains at the heart of linguistic pragmatics.¹⁹

ACKNOWLEDGMENTS

Thanks to Barbara Abbott, Kent Bach, Betty Birner, Yasuhiko Kato, Benjamin Smith, J. L. Speranza, and Gregory Ward for helpful comments on some of this material.

NOTES

- | | |
|--|---|
| <p>1 To say that an implicature (conventional or conversational) makes a non-truth-conditional contribution to an expression’s meaning is not to say that the implicatum itself (= what is implicated) lacks truth conditions, but rather that the truth conditions of the original expression are not affected by the truth or falsity of the implicatum.</p> | <p>2 Beyond cancelability and non-detachability, another proposed criterion for conversational implicature is non-redundant reinforceability. Sadock (1978) argues that an inference can be non-redundantly reinforced just in case it can be canceled without contradiction, viz. when it is a conversational implicature (see also Morgan 1969, Horn 1972). Thus we</p> |
|--|---|

have the contrast between

(i) and (ii):

(i)a. Some but not all men are chauvinists.

[non-contradictory]

b. Some men are chauvinists; indeed all are.

[non-redundant]

(ii)a. #It's odd that dogs eat cheese, and they don't.

[contradictory]

b. #It's odd that dogs eat cheese, and they do. [redundant]

But concession/affirmation structures *can* be felicitous even when informationally redundant provided the two clauses involved are rhetorically opposed – whence the adversative *but*:

(iii)a. It's (#not) odd that dogs eat cheese, but they do.

b. I #(don't) know why I love you, but I do.

Thus, contra Sadock, and Hirschberg (1991), semantically inferrable (entailed or presupposed) material may be felicitously reinforced. (See Horn 1991b for details.)

- 3 Although the “Gricean letter of recommendation” in (1b) has become legendary, it appears not to be legal in the very state in which Grice taught:

If an employer chooses to provide a reference or recommendation, the reference giver must include factual negative information that may be material to the applicant's fitness for employment in addition to any positive information. Campus managers and supervisors who provide employment references on current or former employees must be aware that untrue, incomplete or misleading information may cause a

different liability – negligent referral.

The court in *Randi M. v. Livingston Union School District*, 1995 . . . found that: “**A statement that contains only favorable matters and omits all reference to unfavorable matters is as much a false representation as if all the facts stated were untrue.**”

[Emphasis added; *gratia* Bill Ladusaw]

- 4 For Davis (1998: 21), a particularized implicature reflects speaker implicature, while a generalized implicature is sentence implicature: “what speakers using the sentence with its regular meaning would commonly use it to implicate” (Davis 1998: 6). See Saul (to appear) for commentary.
- 5 Horn (to appear) argues for a distinction between what is entailed and what is asserted; entailed material that is not asserted (like the positive component of *Bush barely carried any northern states* or *Only Chris has ever been to Bhutan*) is ASSERTORICALLY INERT and plays no role in NPI licensing. On this account, scopal patterns taken to be diagnostic for conventional implicature or pragmatic presupposition are reanalyzed as diagnostics for non-assertion. See also Abbott (2000).
- 6 Grice (1989: 30–1) characterizes conversational implicature as follows: “A man who by saying that *p* has implicated that *q*, provided that (1) he is to be presumed to be observing the conversational maxims, or at least the Cooperative Principle; (2) the supposition that he is aware that, or thinks that, *q* is required in order to make his saying . . . *p* consistent with this presumption; and (3) the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp

- intuitively, that the supposition mentioned in (2) is required." Many such implicatures will constitute non-literal or indirect speech acts overlaid on what is said; see Bach and Harnish (1979), Bach (this volume), and Sadock (this volume) for discussion, and Davis (1998) for vigorous critique.
- 7 Washington in fact promulgated his own set of maxims with close parallels to Grice's (see Horn 1990), but the father of his country did not account for his countrymen's ability to exploit these maxims to generate implicatures, while the father of pragmatics did.
 - 8 As Smith (1999) points out, Keenan's central critique (1976: 79) that for Grice "the conversational maxims are not presented as working hypotheses but as social facts" should be reversed, with a twist: the maxims are indeed working hypotheses, but for the speaker (and indirectly the hearer), rather than for the philosopher, linguist, or anthropologist. Keenan's depiction of cases where the maxim of quantity is overridden by cultural taboos in fact supports rather than refutes the Gricean narrative, since her evidence shows that it is just when the maxims are predicted to be in operation that they can be exploited to generate implicata; cf. Prince (1983), Brown and Levinson (1987: 288–9) for further discussion.
 - 9 In Levinson's work (Atlas and Levinson 1981; Levinson 1983, 2000a), the counterpart of the R principle is the I (for "Informativeness") heuristic; see Huang (this volume) for a definition and application to the characterization of anaphoric relations.
 - 10 Recent work has incorporated the dialectic of Q- and R-based implicature and the division of pragmatic labor into models of bidirectional Optimality Theory and game theory; cf. Blutner (1998; this volume) and van Rooy (to appear a).
 - 11 Consider the boldened portion of the two submaxims of quantity –
 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.

– in light of the fact that (as noted in Horn 1972) an equative of the form *X is as A as Y* (e.g. *Robin is as tall as Sandy*) will Q_1 -implicate that (for all I know) *X is not A-er than Y* (e.g. *Robin is not taller than Sandy*), given the ⟨more A than, as A as⟩ quantity scale. Thus, the utterance of Q_1 as stated will (auto-)implicate Q_2 . As Gregory Ward points out, a similar auto-implicature can be detected in Martinich's duplex quantity maxim.
 - 12 Levinson's (2000a) version of the Division of Pragmatic Labor involves not Q-narrowing but what he calls the M(anner) heuristic. He argues that the notion of minimalism involved in the inference from *some* to *not all* is defined in terms of an informational measure rather than complexity of production or processing; because of the apparent role of Manner in the latter case, Levinson refers to the Division of Labor as M-based (Q/M in Levinson 1987a), with Q reserved for pure scalar cases. As he acknowledges, however, the two patterns are related, since each is negatively defined and linguistically motivated: H infers from S's failure to use a more informative and/or briefer form that S was not in a

- position to do so. R/I-based implicature is not negative in character and tends to be socially rather than linguistically motivated.
- 13 As noted above, relevance theory is predicated on a minimax or cost/benefit relation which takes the goal of communication as maximizing contextual effects while minimizing processing effort, and the Principle of Relevance is itself couched in terms of this trade-off of effort and effect. In this sense, relevance theory is a dialectic model as much as that of Horn (1984a, 1993), although the former model associates effort with the hearer rather than the speaker.
 - 14 While (12b) may be attributed to metalinguistic negation (Horn 1989: 373), this analysis is unavailable for (12c).
 - 15 Bach (2001a) adopts the SYNTACTIC CORRELATION CONSTRAINT, based on the position of Grice (1989: 87) that what is said must correspond to “the elements of [the sentence], their order, and their syntactic character”; typical aspects of enriched content that are not directly linked to the utterance cannot be part of what is said.
 - 16 Those enrichments constituting necessary conditions for the expression of truth-evaluable propositions involve what Recanati (1989, 2002a) has called saturation. In such cases, there is a “bottom-up” process triggered by such linguistic elements as genitives (*John’s car* – the one he owns? is driving? following? painting? repairing?), unspecified comparison sets (*Chris is tall* – for an adult (fe)male? adult American? human?) or other expressions with free variable slots: *Kim is ready* (for what?), *Robin is too short* (for what?). See Bach (1994a, 1994b) and Carston and Recanati (this volume) for related discussion.
 - 17 Kent Bach points out the plausible invocation here of the reformulation of the Cooperative Principle (CP) as a communicative presumption: when people converse, they do so with an identifiable communicative intention (Bach and Harnish 1979: 7). The role of rationality and cooperation is also addressed in McCafferty (1987).
 - 18 One interesting result from this work is that children may be more adept than adults at distinguishing the contributions to overall speaker meaning contributed by what is said vs. what is implicated.
 - 19 Davis (1998) offers a wide-ranging attack on the theory of implicature; cf. Saul (to appear) for an assessment. See also R. Walker (1975), Wilson and Sperber (1981), Neale (1992), and Matsumoto (1995) for further critical commentary, and Levinson (2000a) for a conspectus and comprehensive bibliography.