References

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topic, and point of view. In: Subject and topic. Edited by Charles Li. New York: Academic Press. 25-56.

Chomsky, Noam. 1977. Conditions on rules of grammar. Essays in form and interpretation. Amsterdam: North-Holland.

Gazdar, Gerald. 1979. Pragmatics. New York: Academic Press.

Higginbotham, James. 1980. Pronouns and bound variables. Linguistics Inquiry 11.679-708.

Jesperson, Otto. 1942. A modern English grammar. London: Allen and Unwin.

Katz, Jerrold. 1972. Semantic theory. New York: Harper.

Kempson, Ruth M. 1984. Anaphora, the compositionality requirement, and the semantics-pragmatics distinction. North-Eastern Linguistics Society Proceedings, vol. 14.

Sperber, Dan, and Deirdre Wilson. (forthcoming) Relevance: Foundations of pragmatic theory.

TOWARD A NEW TAXONOMY FOR PRAGMATIC INFERENCE: Q-BASED AND R-BASED IMPLICATURE

Laurence R. Horn
Yale University

1. The Principle of Least Effort (and the Principle of Sufficient Effort). Thirty-five years ago, George Kingsley Zipf set out to explain all of natural language (along with almost everything else in the human universe, from dreams, art, and ritual to war, schizophrenia, and the incest taboo) in terms of an overarching Principle of Least Effort. In the linguistic realm, however, Zipf (1949:20ff.) acknowledged two basic and competing forces. The Force of Unification, or Speaker's Economy, is a direct least effort correlate, a drive toward simplification which, operating unchecked, would result in the evolution of exactly one totally unmarked infinitely ambiguous vocable (presumably uhhhh). The antithetical Force of Diversification, or Auditor's Economy, is an anti-ambiguity principle leading toward the establishment of as many different expressions as there are messages to communicate. Given m meanings, the speaker's economy will tend toward 'a vocabulary of one word which will refer to all the m distinct meanings', while the hearer's economy will tend toward 'a vocabulary of m different words with one distinct meaning for each word'. As Zipf (1949:21) (under)states, 'The two opposing economies are in extreme conflict.'

It is in the crucible of this conflict, argue Zipf, Martinet, and allied functionalists, that language change is forged. As Martinet notes (1962:139):

> In order to understand how and why a language changes, the linguist must keep in mind two ever-present and antinomic factors: first, the requirements of communication, the need for the speaker to convey his message, and second, the principle of least effort, which makes him restrict his output of energy, both mental and physical, to the minimum compatible with achieving his ends.

In this paper, I seek to demonstrate that these same two antinomic forces--and the interaction between them--are largely responsible for

generating Grice's conversational maxims and the schema for pragmatic inference derived therefrom.

2. The maxims of conversation. In his ground-breaking work on language use and the logic of conversation, Grice (1975) suggests a procedure whereby participants in a conversational context may compute what was meant (by a given speaker's contributing a given utterance at a given point in the interaction) based on what was said (by that speaker, in that utterance, at that point). 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' (1975:45). Within this basic guideline, Grice establishes four specific subprinciples, the general, almost trivial-sounding, and presumably universal maxims of conversation which he takes to govern all rational interchange (Grice 1975:45-46; cf. Levinson 1983:100ff. for discussion).

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:
1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief. (Avoid unnecessary prolixity).
4. Be orderly.

The assumption that speaker and hearer are both observing the Cooperative Principle and its component maxims permits the exploitation of these maxims to generate conversational implicata, conveyed messages which are meant without being said.

The partial reductionist program I envision would retain the Maxim of Quality with its special character noted by Grice: unless Quality (or what Lewis 1969 has called a Convention of Truthfulness) obtains, the entire conversational and implicatural apparatus collapses. But the first Quantity maxim (henceforth Quantity_1, or simply Quantity) is essentially Zipf's Auditor's Economy, the need for the speaker to convey his message fully. Most if not all of the remaining Gricean rules respond to the Speaker's Economy, either directly (as consequences of the least effort principle) or indirectly (through the interaction of this principle with its antithesis). Notice in particular that the second quantity principle, as stated, is essentially akin to Relation (what would make a contribution more informative than required, except the inclusion of material not strictly relevant to and needed for the matter at hand?). Note also that Grice in effect builds in Relation when defining Quantity_1, and similarly

builds in Quantity_1 in defining Quantity_2. In section 3, I have tentatively boiled down the maxims (leaving aside Quality) to two fundamental principles, echoing the two functional economies of Zipf and Martinet.

3. Minding our Qs and Rs.

(1a) The Q Principle (Hearer-based):
MAKE YOUR CONTRIBUTION SUFFICIENT (cf. Quantity_1)
SAY AS MUCH AS YOU CAN (given R)
Lower-bounding principle, inducing upper-bounding implicata

(1b) The R Principle (Speaker-based):
MAKE YOUR CONTRIBUTION NECESSARY (cf. Relation, Quantity_2, Manner)
SAY NO MORE THAN YOU MUST (given Q)
Upper-bounding principle, inducing lower-bounding implicata

In the current coy style (cf. Chomsky 1982) which has given us D-structures and S-structures, LF and PF, R-expressions and A-binding (corresponding, more or less, to Deep and Surface structures, Logical and Phonetic form, Referring expressions, and Argument-binding, respectively), I am using Q and R to evoke Quantity (i.e. Quantity_1) and Relation while leaving open the extent to which my principles map onto these two maxims.

The hearer-based Q Principle is essentially a sufficiency condition. A lower-bounding law in terms of information structure, it may be (and systematically is) exploited to generate upper-bounding conversational implicata, as described by Grice (1975), Ducrot (1972), Horn (1972, 1973), and Gazdar (1979): a speaker, in saying '...p...', implicates that (for all she knows) '...at most p...'

The primary examples of generalized Q-based implicata arise from scalar predications. If I tell you that some of my friends are Buddhists, I license you to draw the inference that not all my friends are Buddhists. (If I knew they all were, and this knowledge was relevant to your interests, it would have been incumbent on me to obey the Q Principle and say so; the assumption that I am obeying Quantity allows you to infer that I did not know for a fact that the stronger predication --All of my friends are Buddhists--held.) Like all rules of pragmatic inference, Q-based implicature is context-dependent; as a generalized implicatum, the aforementioned inference goes through in unmarked contexts, but it may be cancelled--explicitly (cf. Some, {if not all/and perhaps all,} of my friends...) or implicitly (by establishing the appropriate context, in which all that is relevant, or can be known, is the lower bound).

Examples of Q-based scalar implicature are legion. The following sentences all assert (or entail) a lower bound ('at least ___', their 'one-sided' reading as Aristotle dubbed it), and characteristically implicate an upper bound ('at most ___'); the conjunction of the assertion and implicatum results in conveying the corresponding 'two-sided' understanding ('exactly __').

(2a) He ate 3 carrots < 1-sided: 'at least 3'; 2-sided: 'exactly 3'

(2b) You ate some of the cookies < 1-sided: 'some if not all'; 2-sided: 'some but not all'

(2c) It's possible she'll win < 1-sided: 'possible if not certain'; 2-sided: 'possible but not certain'

(2d) Maggie is patriotic or quixotic < 1-sided = inclusive or; 2-sided = exclusive or

(2e) I'm happy < 1-sided: 'happy if not ecstatic'; 2-sided: 'happy but not ecstatic'

(2f) It's warm < 1-sided: 'at least warm'; 2-sided: 'warm but not hot'

(cf. Horn 1972, 1973, to appear b for additional details)

This analysis, foreshadowed a century ago by DeMorgan (1847:100) and Mill (1867:501), allows us to preserve semantic parsimony by taking the infinitely many weak scalar predicates to be logically unambiguous while pragmatically ambivalent (cf. Hamilton 1860 and Kempson 1980 for alternative accounts in which the sentences of (2) are treated as logically ambiguous; these accounts are reviewed and rebutted in Horn 1973, to appear b).

If the Q-Principle corresponds to Quantity$_1$, the countervailing R-Principle collects not only Relation, but Quantity$_2$ and possibly all the manner maxims (although I shall not go through the arguments for that claim here). The R-Principle, mirroring the effect of the Q-Principle just discussed, is an upper-bounding principle which may be (and standardly is) exploited to generate lower-bounding implicata. A speaker who says '...p...' may license the Q-inference that he meant '...at most p...'; a speaker who says '..p...' may license the R-inference that he meant '...more than p...' The locus classicus here is the indirect speech act (Heringer 1972; Gordon and Lakoff 1975; Searle 1975): if I ask you whether you can pass me the salt, in a context where your abilities to do so are not in doubt, I license you to infer that I am doing something more that asking you whether you can pass the salt--I am in fact asking you to do it. (If I know for a fact that you can pass me the salt, the yes-no question is pointless; the assumption that I am obeying the Relation maxim allows you to infer that I mean something more than what I say.)

Grice notes that a speaker may 'quietly and unostentatiously violate a maxim' as well as exploit it to generate an implicature. Clark and Haviland (1977:2) have suggested that intentional covert maxim violations result in lies, while unintentional violations are simply misleading. In fact, what is crucial is just which sort of maxim or pragmatic principle is violated: intentional quality violations result in lies (another reason for the special status of quality; cf. Coleman and Kay 1981 for additional factors in defining lie), intentional violations of the Q-based sufficiency principle result in a speaker's misleading the addressee, and intentional violations of the R-based least effort principle are often simply unhelpful or perverse. A courtroom witness must swear to tell the whole truth and nothing but the truth, i.e. to obey quantity and quality, while violations of relevance lead only to a possible lawyer's objection or judge's scolding.

4. Q-based vs. R-based inference: Some early skirmishes. Like the antinomic economies from which they derive, the Q-based and R-based principles just outlined often directly collide. A speaker obeying only Q would tend to say everything she knows on the off-chance that it might prove informative, while a speaker obeying only R would probably, to be on the safe side, not open her mouth. In fact, many of the maxim clashes Grice and others have discussed do involve Quantity$_1$ vs. Relation. In delineating the operation of quantity to generate upper-bounding scalar implicatures in my thesis (Horn 1972: Chapters 1 and 2), I was bothered by the contrast between (3a), where Quantity is in force, and (3b), where the principle of relevance is apparently responsible for the implicatum (as Karttunen 1970 suggests).

(3a) It is possible that John solved the problem ⟶
(For all S knows) John didn't solve the problem (Q)

(3b) John was able to solve the problem ⟶
John solved the problem (R)

An application of quantity in the latter case would generate the opposite implicatum, viz. that John didn't solve the problem. Similarly, in (4a) quantity leads to an upper bound on the information communicated, while in (4b) an R-based inference renders the indefinite more informative than its logical form suggests (both examples from Grice 1975:56; cf. Harnish 1976).

(4a) X is meeting a woman this evening ⟶
The woman in question is not X's wife, sister, or close platonic friend (Q)

(4b) I broke a finger yesterday ⟶
The finger is mine (R)

5. Politeness rules, pragmatic competence, and maxim clashes. A parallel to these clashes emerges from work by Robin Lakoff on the maxims of politeness outlined in (5), and by Brown and Levinson in their reformulation of Lakoff's maxims in terms of face wants and needs:

(5) Rules of politeness (Lakoff 1973, Brown and Levinson 1979)

	Lakoff:	Brown and Levinson:
Rule 1.	Deference: 'Don't impose', 'Keep aloof'	Negative politeness, polite formality
Rule 2.	Give options (special case of Rule 1, or derived as theorem from Rules 1 and 3?)	Respect 'negative face', i.e. freedom from imposition and freedom of action.
Rule 3.	Be friendly (Camaraderie)	Positive politeness, polite friendliness Respect 'positive face', i.e. positive consistent self-image and approval/ appreciation by others.

Crucially for us, the Deference or Aloofness maxim, Rule 1, is an upper-bounding R-based constraint on one's actions, while Friendliness, Rule 3, is a lower-bounding Q-based imperative.

Lakoff has noted several instances in which Rule 1 Politeness seems to clash with Rule 3, e.g. belching after a meal, which may signal appreciation (via Rule 3) or may simply be taken as offensive (by Rule 1). More systematically, when we look at languages like those in the Romance, Germanic, Slavic, and Indic families of Indo-European, in which two different forms for the second person singular coexist but are associated with different presuppositions or conventional implicata (cf. Levinson 1983:128-29), we find that the use of a (familiar) T form rather than the more formal V form for the addressee may convey polite friendliness by Rule 3 or presumptuousness by Rule 1 (depending on the context and the interlocutors' assumptions and beliefs). Similarly, the use of a V form for the addressee may be polite by Rule 1, or aloof and unfriendly by Rule 3.

A particularly devastating kind of maxim clash which has been analyzed in similar terms is the communication mixup example from work by Tannen.

(6) Conversational breakdowns and marital breakups (Tannen 1975, 1979)
First exchange:
Wife: Bob's having a party. You wanna go?
Husband: OK.
Second exchange (later):
Wife: Are you sure you want to go?
Husband: OK. Let's not go. I'm tired anyway.
Post-mortem:
Wife: We didn't go to the party because you didn't want to.
Husband: I wanted to. You didn't want to.

On Tannen's gloss of this canonical interchange, one partner (the wife) is operating on a direct strategy utilizing Rule 3 politeness: if one had meant more, s/he would (and should) have said it. Her partner (or opponent), on the other hand, is employing an indirect, hint-seeking strategy which emanates from Rule 1 politeness: avoid saying too much when you can get it across by hints. As Tannen observes, each strategy may link up with a different pragmatic competence, the difference involving not the set of operative rules but the relative strength of opposing rules within that set.

Notice that Tannen's account of the clash instantiated in (6) superimposes directly onto an alternative gloss utilizing Grice's Quantity and Relation maxims (or Quantity_1 and Quantity_2), respectively. Crucially, in either version we have to deal with Q-based vs. R-based inference patterns.

A closely related example of maxim clash is cited by Keenan (1976). As she reports and interprets the facts, the Malagasy-speaking culture of Madagascar is a speech community in which the Cooperative Principle, and in particular the maxim of Quantity, is not observed. Informativeness is apparently absent as a working principle in conversation, and the participants in a talk-exchange tend not to draw what we have been calling Q-based inferences in situations where such inferences would be drawn in Western communities.

> [Interlocutors] regularly provide less information than is required by their conversational partner, even though they have access to the necessary information. If A asks B, 'Where is your mother?' and B responds, 'She is either in the house or at the market', B's utterance is not usually taken to imply that B is unable to provide more specific information needed by the hearer. The implicature is not made, because the expectation that speakers will satisfy informational needs is not a basic norm. (Keenan 1976:70)

In fact, however, as Keenan makes clear in her discussion, Quantity and Q-based inference do play a significant role in some conversational contexts. Where they do not, it is because Quantity is overridden by a countervailing R-based principle, in particular the imperative of avoiding tsiny (the responsibility, guilt, or other unpleasant consequences incurred by uttering claims which turn out to be false and/or offensive to other members of the society living or dead). Sex of speaker, and significance and accessibility of the information contributed, are other variables influencing the relative weights of Q-based and R-based principles and inference patterns in the Malagasy community. (Cf. Prince 1982 for a related critique of Keenan's conclusions.)

Both Tannen's and Keenan's cases involve situations in which different speakers practice different utilizations of essentially the same tools; pragmatic competence often differs across cultures--and across speakers within the same culture--in accordance with the assignment of relative weightings to different maxims or principles, and consequently with the inference patterns associated with the exploitation of those maxims.

6. Quantity vs. informativeness. The most detailed and careful discussion in the literature of Q vs. R clashes in English is due to Atlas and Levinson (1981) (cf. Levinson 1983: section 3.2 for related discussion). Atlas and Levinson begin by summarizing the evidence for the inference from Quantity for the scalar cases. Examples (after Horn 1972 and Gazdar 1979) include the implicata in (7a); the general principle for the Q-based cases is given in (7b).

(7) The inference from Quantity: What is communicated is more informative, more definite than what is said.

(7a) some $\longleftrightarrow$ not all
may $\longleftrightarrow$ may not
3 $\longrightarrow$ no more than 3
p or q $\longrightarrow$ not both p and q
a believes that p $\longrightarrow$ a does not know that p
(cf. discussion of example (3).)

(7b) 'Given that there is available an expression of roughly equal length that is logically stronger and/or more informative, the failure to employ the stronger expression conveys

that the speaker is not in a position to employ it.' (Atlas and Levinson 1981:38)

But, Atlas and Levinson point out, there is a substantial class of (somewhat disparate) cases for which Quantity gives exactly the wrong results. For these cases, including those in (8a), they invoke a Principle of Informativeness, (8b).

(8) The inference from Informativeness: What is communicated is more precise than (is a subcase of) what is said.

(8a) if p then q⟶if -p then -q (Geis and Zwicky's 1971 'invited inference' of conditional perfection)
p and q⟶p preceded q (cf. Grice 1975, Schmerling 1975 on 'asymmetric' conjunction)
(↓)
p caused q
a and b VP'd⟶ a and b VP'd together
a ate the cake⟶a ate the whole cake
a ate the apples⟶ a ate all the apples } (cf. Harnish 1976)
Do you know the time?⟶If you know the time, tell me what it is. (Indirect speech acts, cf. Cole and Morgan 1975)
I don't think that p⟶(I think that) not-p (Neg-raising, cf. Horn 1978b)
I have a new car and the windows don't close⟶...the windows of my new car.... (Bridging inferences, cf. Clark and Haviland 1976)

(8b) The Principle of Informativeness: 'Read as much into an utterance as is consistent with what you know about the world' (Levinson 1983:146-47)

Atlas and Levinson (1981:42) formulate their informativeness-based inference as an 'inference to the best interpretation'.

(8c) 'If a predicate Q is semantically nonspecific with respect to predicates P_i, $1 \leq i \leq n$, but for some j, $1 \leq j \leq n$, P_j is stereotypical of Qs, then in saying ⌜Q⌝ a speaker will convey ⌜P_j⌝.'

The key notion here is the restriction of a more general predicate to a stereotypical instance. Two examples of this inference to a salient subset or exemplar cited by Atlas and Levinson are those in (8d).

(8d) The secretary smiled.⟶The female secretary smiled.
John had a drink.⟶John had an alcoholic drink.

The class of indefinite descriptions provides a major source of clashes between Quantity-based inference (cf. (9a)) and Informativeness-based inference (cf. (9b)).

(9) Quantity vs. informativeness (Atlas and Levinson 1981:49ff.; cf. Harnish 1976)

(9a) Quantity in force:	(9b) Informativeness in force:
I slept on a boat yesterday⟶	I lost a book yesterday⟶
The boat was not mine	The book was mine.
I slept in a car yesterday⟶	I broke a finger yesterday⟶
The car was not mine	The finger was mine
Mort and David took a shower⟶	Mort and David bought a piano⟶
They took separate showers	They bought a piano together

A number of factors are involved in determining which principle takes precedence when the two are at odds; notice, for example, that the speaker in (9a) could have chosen the more precise genitive form (I slept on my boat, in my car) but did not do so, while the use of the genitive in (9b) (my book, my finger) might suggest wrongly that I have but one book or one finger. Atlas and Levinson cite two additional factors which are relevant to the weighting of Quantity vs. Informativeness: if an entailment-based 'Horn scale' (cf. Horn 1972) can be constructed on which the predicates in question can be readily ranked, Quantity is more likely to win out; if the application of Quantity tends to contradict our assumed 'Conventions of Noncontroversiality', Informativeness takes precedence.

Within the framework being explored here, Atlas and Levinson's inference from Quantity corresponds directly to our Q Principle. There is an equally strong parallel between Atlas and Levinson's inference from Informativeness and our R Principle; my only objection to their formulation concerns their terminology. First of all, Informativeness suggests the Q Principle to me more strongly than, as intended, the R Principle (note the language of Grice's Quantity_1 maxim, which Keenan 1976 indeed glosses 'Be informative'). Furthermore, Ducrot (1972), who independently proposes a 'Loi d'exhaustivité' to do the work of Grice's maxim of Quantity, also invokes a 'Loi d'informativité' of a rather different nature from Atlas and Levinson's Principle of Informativeness (cf. Ducrot 1972:132-34).[1] Thus, I feel justified in taking Atlas and Levinson's so-called Informativeness as an instance of the R Principle.

7. Negation as implicatum-canceller: Q-based vs. R-based inferences. An opposition as fundamental as that between the Q Principle and the R Principle would be expected to have major linguistic consequences, and indeed this expectation is not disappointed. We begin by considering the effect of negation on conversational implicata. What we find is that Q-based implicata can be readily cancelled by a negation which does not affect what is said (through what I have elsewhere termed 'metalinguistic' negation: cf. Horn to appear a), but R-based implicata cannot be cancelled by negation at all.

This is illustrated in what follows. First, some instances of negated scalar implicata are given in (10a-f). That negation is being used in a marked way to deny the Q-implicated upper bound in these examples can best be seen by considering the well-formed affirmative counterparts of

these sentences in (10a'-f') (cf. (2); also Horn to appear b, and references there).

(10a) He didn't eat 3 carrots--he ate 4 of them.
(10b) You didn't eat some of the cookies--you ate all of them.
(10c) It isn't possible she'll win--it's certain she will.
(10d) She isn't patriotic or quixotic--she's both.
(10e) I'm not happy--I'm ecstatic.
(10f) It isn't warm--it's downright hot.

(10a') He ate 3 carrots--in fact, he ate 4.
(10b') You ate some of the cookies--indeed, you ate all of them.
(10c') Not only is it possible she'll win--it's certain she will.
(10d') She's patriotic or quixotic--in fact, she's both.
(10e') I'm happy--indeed, I'm ecstatic.
(10f') It's warm--in fact, it's downright hot.

But now consider R-based implicata. To say that someone was able to solve the problem may R-implicate that she in fact solved it (Karttunen 1970). Similarly, to assert that someone was clever enough to do something generally implicates that he did it. As we have seen, my confiding to you that I broke a finger would normally be taken to refer to a finger of mine (given Atlas and Levinson's Conventions of Noncontroversiality) unless I knew you knew that I was an enforcer for the mob. And, as Wittgenstein and others have pointed out, an assertion of the form 'I believe that S' would normally be taken as an indirect assertion of S rather than merely a statement about my beliefs; this indirect speech act, too, works by R-inference. Yet none of these implicata can be cancelled by negation.

(11a) She wasn't able to solve the problem.
(≠ She was able to solve it, but didn't)
(11b) He wasn't clever enough to figure out the answer.
(≠ He was clever enough to do it, but he didn't)
(11c) I didn't break a finger yesterday.
(≠ I broke a finger, but it wasn't one of mine)
(11d) I don't believe the Yankees will win the pennant.
(≠ I believe they'll win the pennant, but I'm not [weakly] asserting that they will)

Why should this difference in cancellability exist? The answer lies in the logic of Q-based and R-based inference. Let S represent a given (stronger) proposition, and W the weaker proposition which it unilaterally entails and from which the relevant implicata are to be drawn. In the case of Q-based implicata, the assertion of 'W' Q-implicates -S. Where W is a scalar predicate truth-conditionally defined by its lower bound, the ordinary negation of W negates that lower bound, i.e. as 'less than W', and is hence incompatible with S; the assertion that he did not eat three carrots would be taken to amount to the assertion that he ate less than three (and hence not four, five, or more). But 'not W' uttered in a context where S is affirmed (as in (10a-f)) self-destructs on the

unmarked 'less than W' understanding and must therefore be sent back through, in effect--whence the marked, metalinguistic quality of this variety of negation.

In the case of R-based implicata, the assertion of 'W' R-implicates not -S but S: the proposition that she solved the problem unilaterally entails the proposition that she was able to solve it (S entails W), but the assertion that she was able to solve it may implicate that she in fact solved it ('W' R-implicates S). Once again, 'not W' signifies 'less than W' and hence licenses the inference of -S (via modus tollens from the original S ⊦W entailment). But crucially, there are no circumstances under which the implicatum S is cancelled and 'not W' cannot be interpreted consistently, as an ordinary descriptive negation. The negation in (11a-d) thus never gets sent back through to be interpreted metalinguistically, as an implicatum-canceller. Schematically, the situation we have is that in (12).

(12)

Q-based implicata:	R-based implicata:
S entails W	S entails W
'W' Q-implicates -S	'W' R-implicates S
normally, 'not W' = 'less than W', incompatible with S	normally, 'not W' = 'less than W'
'not W', asserted where S is given, reinterpreted as metalinguistic negation	'not W' ⊬ -S (modus tollens)
	'not W' never gets reinterpreted, since it's always compatible with -S (the denial of W's implicatum)

Now, R-based implicata can be cancelled without negation, simply by assigning the contradiction contour (cf. Liberman and Sag 1974) and stressing the implicatum-inducing element.

(13) She was able to solve the problem (but she didn't solve it).
I believe the Yankees will win the pennant (but I'm not saying they will).
I broke a finger today (but not one of mine).

Notice that we tend to get the opposite, Q-based implicatum in these contexts; the contour which cancels the R-based inference sets up a strong expectation for the kind of continuations exemplified above.

When we appear to get cancellation of an R-based implicatum by negation, the implicatum in question has, in fact, become conventionalized as part of literal meaning (cf. Grice 1975:58 and Morgan 1978 on the gradual conventionalization of conversational implicata). Thus, for example, predicate expressions which denote various personal relationships may take on a narrowed symmetric sense (cf. X and Y are {married/friends/lovers/in love}) but need not (cf. X and Y are spouses). When the symmetric sense of these predicates is intended, negation may leave the more general sense unaffected.

(14) They aren't {married/friends/in love}. (i.e. with or to each other)

To confirm the claim that only conventionalized R-based implicata can be cancelled by negation, we need only reconsider the pair of examples from Atlas and Levinson (1981) cited in (8d). Both speakers' intuition and lexicographers' practice suggest that the implicatum associated with drink ('alcoholic drink') has become fossilized into conventional meaning, while the implicatum associated with secretary ('female secretary') has not. Thus, in the terms of Horn (to appear b), drink represents an autohyponymous lexical item while secretary does not. In this light, negation behaves precisely as predicted.

(15a) My secretary didn't smile--I have a male secretary.
(15b) John didn't have a drink--that was a Shirley Temple.

A male secretary is still a secretary (although not one of the salient variety), but a nonalcoholic beverage may or may not count as a drink.

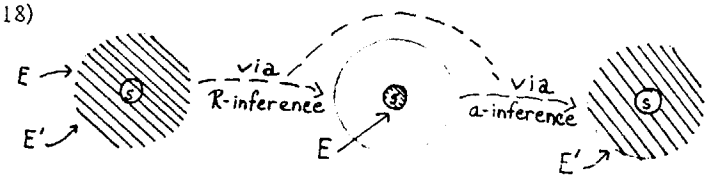
8. The division of pragmatic labor. While the Q Principle and the R Principle are diametrically opposed forces in inference strategies and language change, it is perhaps in the resolution of the conflict between them that they play their major role in both 'langue' and 'parole'. The most general pattern for this resolution, the synthesis of the two antitheses, is summarized in (16) and derived more explicitly in (17a-f).

(16) The use of a marked (relatively complex and/or prolix) expression when a corresponding unmarked (simpler, less 'effortful') alternate expression is available tends to be interpreted as conveying a marked message (one which the unmarked alternative would not or could not have conveyed).

(17a) The speaker used marked expression E' containing 'extra' material (or otherwise less basic in form or distribution) when a corresponding unmarked expression E, essentially coextensive with it, was available.
(17b) Either (i) the 'extra' material was irrelevant and unnecessary, or (ii) it was necessary (i.e. E could not have been appropriately used).
(17c) (17b(i)) is in conflict with the R Principle and is thus (ceteris paribus) to be rejected.
(17d) Therefore, (17b(ii)), from (17b), (17c) by modus tollendo ponens.
(17e) The unmarked alternative E tends to become associated (by use or--through conventionalization--by meaning) with unmarked situation s, representing stereotype or salient member of extension of E/E'. (R-based inference; cf. Atlas and Levinson, (8b), (8c).)
(17f) The marked alternative E' tends to become associated with the complement of s with respect to the original extension of E/E'. (Q-based inference; cf. (6.12))

The key steps in the argument are those sketched in (17a), (17e), and (17f); they represent a characteristic shift which can be schematized as in (18):

(18)



The result is an equilibrium which I shall call the division of pragmatic labor. (The equilibrium is, in fact, somewhat unstable; in particular, either the R-based inference represented in the second diagram or the Q-based inference in the third can become conventionalized, as we shall see later.) The remainder of this essay is devoted to rehearsing a number of instances of this pattern, varying in the details but essentially parallel in terms of the overall dynamics.

9. Avoid Pronoun (and unavoidable pronouns). In considering the near-complementary distribution between his abstract, phonetically unrealized PRO element (which usurps many of the functions of Equi-deletion sites in the earlier Standard Theory) and overt pronominals, Chomsky (1982:65) arrives at a general principle which he calls Avoid Pronoun, 'interpreted as imposing a choice of PRO over an overt pronoun where possible'. To illustrate, he cites the two cases of (19a-b).

(19a) John would much prefer [his going to the movie]
(19b) John would much prefer [his (own) book]

PRO may appear in the frame of (19a) in place of the pronoun (cf. John would much prefer going to the movie), and so his is taken here as non-coreferential with John; in (19b), PRO cannot appear and the overt pronominal his may be (and with own present must be) taken as coreferential with John. As Chomsky notes, the Avoid Pronoun principle

> might be regarded as a subcase of a conversational principle of not saying more than is required, or might be related to a principle of deletion-up-to-recoverability, but there is some reason to believe that it functions as a principle of grammar. (Chomsky 1982:65)

Three important points emerge from this brief passage: the link between Avoid Pronoun and analogous principles applying to deletion, the grounding of the principle(s) in least effort (i.e. our R Principle), and the fossilization or grammaticization of the functional principle into conventional rules and constraints within particular grammars.

The transderivational nature of Avoid Pronoun is clear from Chomsky's discussion of (19a-b) and an analogous contrast in French (Chomsky 1982:146).

(20a) Je veux qu'il vienne. 'I want him to come'
(20b) Il_i veut qu'il_j vienne. 'He_i wants him_j to come' (where $i \neq j$)
(20c) *Il_i veut qu'Il_i vienne. 'He_i wants $him(self)_i$ to come'

Avoid Pronoun, in ruling out (20c), must refer to the availability of a corresponding derivation, that of (21).

(21) Il veut venir. 'He wants to come'

In (21) the subject of 'come' is realized as an abstract PRO argument of the embedded infinitive rather than an overt pronominal subject of an embedded subjunctive, as in (20a-c).

The applicability of Avoid Pronoun is also partially dependent on ambiguity avoidance, a fact not discussed by Chomsky. Thus, in the English versions of (20a-c) and similar sentences, coreference is ruled out totally only in third person cases (cf. Morgan 1970:387).

(22a) *He_i wants him_i to win. (no coreference possible)
(22b) (?)I_i want me to win. (coreference possible in contrastive contexts)
(22c) (?)You_i want you_i to win.

(Technically, Avoid Pronoun should be activated in (22c), since one can be speaking to different addressees within a single sentence, distinguishing them by pointing or eye gaze. This situation, however, is too marginal to trigger the ambiguity avoidance condition on Avoid Pronoun, and the second person case thus works like the first person case in (22b).)

The literature on reflexives offers further grist for a suitably upgraded Avoid Pronoun mill. As is well known, it is usually (although not always) the case that a nonreflexive pronoun can be interpreted as coreferential with a given antecedent just in case a reflexive (bound by that antecedent) could not have appeared in that position. Many instances of this pattern respond to the Disjoint Reference principle suggested in Postal (1974) and discussed further within the EST (cf. Chomsky 1982). Thus, compare (23a) and (23b).

(23a) He_i likes $\{himself_i/*him_i\}$.
(23b) He_i said that she likes $\{*himself_i/him_i\}$.

Here again, as with the 'Equi' cases, ambiguity avoidance is relevant in the determination of when coreference is possible.

(24a) He's voting for him. (no coreference possible)
(24b) I'm voting for me. (OK contrastively)

In 'pro-drop' languages, PRO may appear in place of overt NPs freely in a wide range of positions, including subject and object; generally, verb agreement and/or pragmatic context permits recoverability of the missing referent. As expected, the scope of the Avoid Pronoun principle is commensurately widened in these languages. In her analysis of one pro-drop language, Turkish, Enç (1982) shows that semantically

redundant pronouns (i.e. those whose referent would be recoverable) tend to be interpreted as expressing contrast--including topic-change, which she argues is a subtype of contrast. (Notice that in English, too, the occurrence of an overt nonreflexive pronoun in a frame which permits a referentially equivalent gap or reflexive, e.g. (22b) or (24b), is likewise interpreted contrastively.) Enç argues that the interpretation of such marked pronouns, whose appearance seems to violate parsimony, is based on a conventionalization of a quantity-based inference--but she is referring here to Grice's $Quantity_2$ which we have previously subsumed under our R Principle (i.e. least effort).

In fact, while Avoid Pronoun, or whatever more general principle ultimately includes it, is basically a least effort correlate (as Chomsky observes), the division of labor we ultimately arrive at (in which abstract pronouns, i.e. PROs, are interpreted one way and real pronouns another) requires reference to both R and Q Principles, in the manner outlined in section 8.

A related instantiation of this division of pragmatic labor makes this clear. Chomsky proposes a 'general discourse principle' for R(eferring)-expressions, i.e. names and descriptions (1982:227).

(25a) 'Avoid repetition of R(eferring)-expressions, except when conditions warrant.'
(25b) 'When conditions warrant, repeat.'

Note that this discourse principle comes in two parts, one (25a) R-based and one (25b) Q-based; furthermore, as with Grice's two submaxims of quantity, each of Chomsky's subprinciples is explicitly bounded in its applicability by the other (as the subordinate clause material indicates).

The choice among referentially equivalent referring expressions for picking out a given individual is also subject to R- and Q-based considerations, as Prince shows in working out her Scale of Assumed Familiarity (1981:245). The matter is rather complex, however, and I shall not pursue it here.

One last example of the division of pragmatic labor emanating from the domain of reference is presented by Levinson (1983:75):

> The deictic words yesterday, today, and tomorrow pre-empt the calendric or absolute ways of referring to the relevant days. Thus the following, said on Thursday, can only be referring to next Thursday (or perhaps some more remote Thursday), otherwise the speaker should have said today:
>
> I'll see you on Thursday.
>
> The same holds if it is said on Wednesday, due to the pre-emptive tomorrow.

10. The division of pragmatic labor and the lexicon. Aronoff has shown (1976:43ff.) that the existence of a simple lexical item can block the formation of an otherwise expected affixally derived form synonymous with it. In particular, the existence of a simple abstract nominal underlying a given -ous adjective blocks or prevents the formation of an -ity nominalization based on that adjective.

(26a) fury furious *furiosity
*cury curious curiosity
(26b) fallacy fallacious *fallacity
*tenacy tenacious tenacity

Aronoff's blocking phenomenon is the limiting case of a more general pattern independently observed and discussed by McCawley (1978), Kiparsky (1982), and indeed Bréal (1900); a pattern which directly reflects the division of pragmatic labor sketched in section 8 and exemplified in section 9: unmarked forms tend to be used for unmarked situations (via R) and marked forms for marked situations (via Q).

Kiparsky (1982) begins by noting that Aronoff's blocking paradigm is both too strong and too weak. Contrary to Aronoff's predictions, productive derivational processes are sometimes but not always blocked by the existence of a less productive corresponding form: decency and aberrancy block *decentness and *aberrantness, but gloriousness and furiousness survive alongside glory and fury. Blocking may also extend to inflectional processes as well, although again inconsistently: *mans/men, *goed/went, but kneeled/knelt, dreamed/dreamt. Aronoff's formulation of blocking, limited as it is to less than fully productive derivational processes, has nothing to say about these obviously related cases.

Kiparsky suggests a reformulation of Aronoff's blocking as a subcase of the (Pāṇini-Anderson-Kiparsky) Elsewhere Condition ('Special rules block general rules in their shared domain'--specifically, in morphology, irregular forms preclude regular forms). Kiparsky notes that 'blocking can be partial in that the special [less productive] affix occurs in some restricted meaning and the general [more productive] affix picks up the remaining meaning'. (These two components of the blocking process correspond to the second and third diagrams in our representation of the division of pragmatic labor, (18).) He cites as examples of partial blocking refrigerant/refrigerator, informant/informer, contestant/contester; full blocking results when there is no meaning 'left over' for the more productive form to pick up (*borer/bore$_N$, *inhabiter/inhabitant).

To handle these and other cases, Kiparsky formulates a general condition which he calls Avoid Synonymy.

(27) Avoid Synonymy: 'The output of a lexical rule may not be synonymous with an existing lexical item'.

This principle applies to both derivation and inflection, if we assume a level-ordered morphology. Its transderivational nature allows blocking between morphologically unrelated stems: thief blocks *stealer (but cf. base-stealer, with a noncompositional meaning), cutter ≠ knives, scissors, etc.

While there is something right about this principle, it is still too strong, as Kiparsky concedes. For one thing, we need to define a notion of 'corresponding item' relativized to a given speech level or register; words like fridge (underived), icebox (derived by compounding), and refrigerator (derived by affixation) can coexist within a single idiolect despite their referential equivalence. Indeed, the principle is almost

self-falsifying: the doublets synonymy and synonymity strike me as perfect synonyms!

Working independently of the Aronoff-Kiparsky line, McCawley (1978) collects a number of examples where the appropriate use of a given expression formed by a relatively productive process (including syntactic formations) is restricted by the existence of a more 'lexicalized' alternative to this expression (i.e. one formed by a relatively nonproductive process). One case in point is originally due to Householder: the collocation pale red is (in the language of Aronoff and Kiparsky) fully or partially blocked by the lexical alternative pink. For some speakers pale red is simply anomalous (or at least nonoccurring); for others, it picks up whatever part of the 'pale' domain of 'red' pink has not preempted. In either case, pale red is limited in a way that pale blue and pale green are not.

In the same way, McCawley observes, the distribution of productive causatives (in English, Japanese, and other languages) is restricted by the existence of a corresponding lexical causative. Lexical causatives (e.g. (28a)) tend to be restricted in their distribution to the stereotypic causative situation: direct, unmediated causation through physical action.

(28a) Black Bart killed the sheriff.
(28b) Black Bart caused the sheriff to die.

This restriction can be viewed as a straightforward R-based conversational implicatum--an inference 'to the best interpretation', in the language of Atlas and Levinson (1981). The use of the relatively marked, morphologically more complex periphrastic causative (e.g. (28b)) will then Q-implicate that the unmarked situation does not obtain. Thus, (28b) suggests that (28a) could not have been used appropriately, possibly because Bart caused the sheriff's gun to backfire by stuffing it with cotton, or arranged for scorpions to be placed in the room of the sheriff (who is known to have a weak heart), etc. Similarly, the use of the unmarked lexical causative in (29a) R-implicates that the action was brought about in an unmarked way (presumably by stepping on the brake pedal), while the choice of the periphrastic (29b) correspondingly Q-implicates that some unusual method was employed (pulling the emergency brake, telekinesis, etc.).

(29a) Lee stopped the car.
(29b) Lee got the car to stop. (Lee made the car stop.)

McCawley's account of the division of labor between lexical and periphrastic causatives, like Chomsky's Avoid Pronoun and Kiparsky's Avoid Synonymy (or indeed, Grice's Avoid Prolixity maxim, which they all seem to reflect), is explicitly transderivational, and thus runs into some of the same problems of overgeneralization. Consider the data in (30a-d), which McCawley borrows from Heringer (1976:207).

(30a) John made the plate move. (indirect)
(30b) John moved the plate. (direct)
(30c) John made the plate fall. (indirect or direct)
(30d) *John felled the plate.

Here, the periphrastic make versions, 'normally used only in a situation involving indirect causation', as in (30a) whose use is restricted (by the existence of (30b)) to situations involving some sort of supernatural power on John's part, may generalize (as in (30c)) to a wider domain of causative situations 'just in those cases where no lexical causative exists to express direct causation' (Heringer 1976:207) (cf. (30d)). But, as Heringer and McCawley fail to notice, there is a simple lexical form corresponding to (30c), (30d), viz.

(30e) John dropped the plate.

If drop does not count as a corresponding form for make...fall because it involves an unrelated stem, then we cannot legitimately invoke (28a) to predict the limited distribution of (28b).

Other problems for the analysis of McCawley (1978) are discussed in Horn (1978a), where it is argued that the general 'least effort' principle employed by McCawley is simply too powerful and, more specifically, that we need to develop a notion of 'corresponding item' which will take into account such variables as morphological relatedness, markedness, speech register, and inherent complexity. For example, lexical causative verbs may indeed be unmarked in English with respect to their periphrastic make or cause counterparts, but this is not the case for lexical causative adjectives. Thus, the distribution of the periphrastic forms in (31a-c) is not constrained by the existence of the corresponding 'simple' forms in (32a-c).

(31a) That sort of behavior really makes me angry.
(31b) I didn't know that teasing your dog would get you so upset.
(31c) Wild horses couldn't make me stay away.

(32a) That sort of behavior really angers me.
(32b) I didn't know that teasing your dog would upset you so (much).
(32c) Wild horses couldn't keep me away.

Similarly, as a variety of linguistic and psycholinguistic evidence demonstrates, incorporated negation remains relatively complex, and so the distribution of (33a-b) is not constrained by the existence of the superficially simpler examples of (34a-b).

(33a) I persuaded Bill not to date many girls.
(33b) It's not likely that your coin will land heads.

(34a) I dissuaded Bill from dating many girls.
(34b) It's unlikely that your coin wil land heads.

In particular, (33a) is neutral with respect to the (conventionalized) R-implicatum associated with the use of (34a), viz. that Bill had previously intended to date many girls, and (33b) is unspecified for the implicatum associated with (34b), i.e. that it is likely your coin will not land heads (the so-called neg-raised interpretation). No Q-implicata are triggered in (31a-c) or (33a-b), and hence the working out of the division of pragmatic labor is not complete. (Cf. Horn 1978a for further examples and related discussion.)

While we must therefore refine and/or reevaluate the transderivational mechanisms invoked by Aronoff, McCawley, and Kiparsky for describing the division of labor between the use (or existence) of a simple lexical form and that of its more complex (lexical or phrasal) counterpart--in particular, by developing the tools to predict just when a given form counts as a counterpart to some other given form--the insight behind their various accounts is real, and it is essentially a single insight: the unmarked form is used for a stereotypical, unmarked situation (via R-implicature) and the marked counterpart for the situations 'left over' (via Q-implicature). This is the division of pragmatic labor outlined in section 8.

11. Division of pragmatic labor: Additional cases. Two more instances of the division of labor are worth mentioning here. First, consider the realm of indirect speech acts (cf. Searle 1975, Gordon and Lakoff 1975, and other articles in Cole and Morgan 1975). We find that modal auxiliaries which can be associated with indirect speech acts (ISAs) tend to become conventionally associated with the ISAs they convey. If (following Searle) we derive the ISA by an exploitation of the maxim of Relation (cf. section 3), we must nevertheless account for the fact that (35a-b) are conventionally used to convey the request in (36), while (35c-e)--which may (very indirectly) convey that request--are not conventionally used to do so.

(35a) Can you (please) close the window?
(35b) Could you (please) close the window?
(35c) Are you able to (?please) close the window?
(35d) Do you have the ability to (*please) close the window?
(35e) It's (**please) cold in here.
(36) (Please) close the window.

Searle observes (1975:76) that 'there can be conventions of usage that are not meaning conventions' by which 'certain forms will tend to become conventionally established as the standard idiomatic forms for indirect speech acts', including the can you/could you forms of (35a-b) for indirect requests. Morgan (1978) takes these usage conventions to represent 'short-circuited conversational implicatures' which, though calculable, are no longer (after short-circuiting) actually calculated in normal conversation. It is this short-circuiting of the R-based implicatum which licenses the preverbal please in (35a-b); cf. (35c-e) where no short-circuiting has taken place. (A parallel account of the neg-raising phenomenon and the triggering of negative polarity items is offered in Horn and Bayer to appear.)

But if the modal auxiliaries, idiomatic and semantically versatile as they are, become associated through a convention of usage or short-circuited implicature with the ISAs they may be used to convey, their periphrastic counterparts tend (predictably) to be interpreted literally. Thus, compare the behavior of the modals in examples (37a), (38a), (39a), and (40a) (where the implicatum to the ISA is short-circuited) with that of their periphrastic equivalents in (37b), (38b), (39b), and (40b) (where the indirect reading is not a conventional use of the expression and may be totally unavailable).

(37a) Can you pass the salt? (request)
(37b) Are you able to pass the salt? (literal question)

(38a) Here, I can help you with that. (offer)
(38b) (?Here,) I am {able/allowed} to help you with that. (not an offer)

(39a) Will you join us? (invitation)
(39b) Are you going to join us? (literal question)

(40a) I will marry you. (promise)
(40b) I am {going to/willing} to marry you. (only very indirect promise)

Noting the tendency for the most colloquial expressions (typically, as we have seen, the modal auxiliaries) to serve as the conventional forms for conveying indirect speech acts, Searle invokes a new neo-Gricean maxim, 'Speak idiomatically unless there is some reason not to'. He observes (in language mutually intelligible with that spoken by Kiparsky 1982 and McCawley 1978),

> In general, if one speaks unidiomatically, hearers assume that there must be a special reason for it, and in consequence,...the normal conversational assumptions on which the possibility of indirect speech acts rests are in large part suspended. (Searle 1975:76-77)

The pattern discerned here by Searle responds not only to our division of pragmatic labor, but also to the set of statistical correlations at the core of Zipf's analysis of the Principle of Least Effort and its linguistic reflexes (Zipf 1949). Zipf's Law of Abbreviation posits an inverse relation between the length of a word and the frequency of its tokens in an arbitrary text. His Principle of Economic Versatility stipulates a direct correlation between a word's frequency and its semantic versatility (i.e. the number of discrete senses or meanings it allows). The Principle of Economic Specialization states that the age of lexical item in the language correlates inversely with its size and directly with its frequency. By these measures, the relative simplicity of (35a), (35b) as against (35c) is directly confirmed: can and could are historically older than their periphrastic counterparts, phonologically simpler and shorter, more frequent in text tokens, and certainly more versatile semantically, as the contrasts in (41a-c) make clear.

(41a) I {can/am able to} stand on my nose.
This knife {can/?is able to} cut the salami. (cf. ...is capable of cutting it)
The salami {can be cut/*is able to be cut/is capable of being cut} by this knife.
(41b) Can it really be raining out?/Is it possible that it's raining out?
(41c) Can Billy come out and play?/Is Billy permitted to come out and play?

It is significant that modals, the constituents par excellence of least effort, figure so prominently among those forms conventionalized in English and other languages as conveyors of indirect speech acts, while their more expensive and complex counterparts do not.

Finally, we come to the case of logical double negation. What we find here is that the two negatives of the form not-(not-p) do not cancel out functionally even when they do semantically: they convey a positive which is characteristically weaker than the corresponding simple affirmative. In Jespersen's words,

> The two negatives [in not common, not infrequent] do not exactly cancel one another so that the result is identical with the simple common, frequent; the longer expression is always weaker; 'this is not unknown to me' or 'I am not ignorant of this' means 'I am to some extent aware of it', etc. The psychological reason for this is that the detour through the two mutually destructive negatives weakens the mental energy of the listener and implies...a hesitation which is absent from the blunt, outspoken common or known. (Jespersen 1924:332; cf. Horn 1978c: section 3.1 for discussion)

Rather than appealing, with Jespersen, to the metaphysical (and somewhat Victorian) notion of double negation sapping the listener's mental energy, we can more plausibly ascribe the weakening effect to the same general tendency we have already observed: the use of a marked expression when there is a shorter and less 'effortful' alternative available signals that the speaker felt s/he was not in a position to employ the simpler version felicitously.

With double negation, as with indirect speech acts, we see an especially clear correlation between the stylistic naturalness of a given form, its relative brevity and simplicity, and its use in stereotypic situations (via R-implicature). The corresponding periphrastic forms, stylistically less natural, longer and more complex, are restricted (via Q-implicature) to those situations outside the stereotype, for which the unmarked expression would have been inappropriate.

12. Q-based and R-based processes in language change. Perhaps the clearest lexical correlates of the clash and resolution of the R vs. Q conflict are in the area of diachrony. The most obvious R-based effects in language change are the reflexes in 'langue' of the well-known 'fast speech phenomena' in 'parole': contraction, clipping or truncation, assimilative shifts, and so on. Clipping (truncation) is a direct corollary

of the aforementioned Law of Abbreviation: the more frequently a word or expression is used (within a given speech community), the shorter it will tend to become.

Among the contemporary examples of clipping cited by Stern (1931:258) in his standard work on lexical change are pram (from perambulator), specs (< spectacles), rep (<repetition or reprobate); it is striking that the same clipped forms persist even when their conventional value has shifted--specs is now likely to abbreviate specifications, and rep would probably be taken to denote representative (or perhaps reputation, as in the somewhat dated bad rep). More clearly entered in our current lexicon are such conventional truncations as TV (or Brit. telly), phone, bus (<omnibus), and the conventionalized output of an extremely productive process of acronym formation (e.g. US(A), EST (with at least two senses for east-coast syntacticians), UCLA, OSU--'Oregon State University', 'Oklahoma State University', or 'Ohio State University', depending on the shared assumptions of the members of the speech community).

In discussing the development of the truncated forms, Stern (1931:257-58) warns that 'the demands of the speech functions must set a limit to the economic tendency'--i.e. the Q Principle constrains the power of the R Principle. (He also disparages 'the use of pronouns, or of generic words, to save mental effort' as 'especially characteristic of undeveloped minds, unintelligent or immature'. Avoid Pronoun or else!)

A more complex area of language change, involving the interaction of R-based and Q-based processes, is that of lexical shifts. Two traditional categories of lexical change (discussed by Paul 1909 and Bréal 1900, inter alia) worth examining in this light are narrowing (or reduction) of meaning and broadening (or expansion) of meaning.

Narrowing generally involves an R-based shift from a set denotation to a subset (or member) of that set, representing the salient or stereotypical exemplar of the general category. Examples cited in the standard works (cf. Breal 1900, Stern 1931) which fit this definition include Greek alogon (lit. 'speechless one') for 'horse', Latin fēnum (orig. 'produce') for 'hay', and English poison (cognate with potion), liquor (cf. liquid), undertaker (from 'one who undertakes' to 'mortician'), and corn (used for whatever grain is the most important cereal crop of a particular region, e.g. wheat in England, oats in Scotland, maize in Australia or the New World).

In these instances, the shift has become virtually complete (although the original, broader extension may persist in marginal uses). Other cases manifest the intermediate stage of 'autohyponymy'[2] (Horn to appear b), in which the basic, general sense survives in privative opposition with a specific sense derived from it. Autohyponyms which have developed their specific meaning (indicated in parentheses) through R-based narrowing include the following:

(42) color (for 'hue', i.e. the range of colors excluding blacks, whites, and grays): in color, color TV
temperature (for 'fever'): The baby has a temperature
number (for 'integer'): Pick a number from 1 to 10
drink (for 'drink alcohol'): I don't drink. Cf. the second example of (8d)
smell (for 'stink'): Something smells here
Frau (Ger.), femme (Fr.), mujer (Sp.) (for 'wife' as well as 'woman')

Dismissing a contemporary account of this variety of lexical shift in the evolution of homo 'man' from a source akin to humus 'earth', the first century grammarian Quintilian asks rhetorically, 'Are we to believe that homo comes from humus, because man is born of the earth, as if all animals had not the same origin?' (cited by Bréal 1900:114). Yet the derivation is correct, as noted by Bréal, who points out that alogon can likewise designate 'horse' in modern Greek without implying that no other animal lacks the faculty of speech.[3] Like those logical fallacies based on real pragmatic inference patterns (e.g. 'denial of the antecedent', 'post hoc ergo propter hoc', and 'secundum quid' (Horn 1973:212-13), all responding to the workings of the R Principle as exemplified in (8a)), the R-based inference from a set to a salient or stereotypical member is as linguistically plausible as it is logically invalid.

In other cases, lexical narrowing is Q-based, typically resulting is autohyponymy. Kempson (1980:15-16) offers a characterization of the general process:

> If for some general term, representing a lexical field, there is a gap in the sub-parts of that field, with only one more narrowly specified lexical item, then the gap may be filled by a more specific use of the general term.

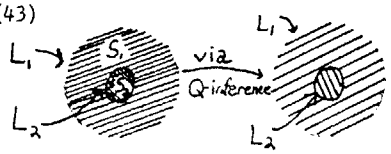
Thus, for example, the existence of bitch 'female dog', in the absence of a sex-specific mate, leads the general term dog to develop a hyponymic sense designating the male of the species.

The governing principle for such cases is given as follows (Kempson 1980:15):

> If a lexical item L_1 has as its extension a set S_1 which includes the set S_2 which a second lexical item [L_2] has as its extension..., then the lexical item L_1 may be used to denote that subset of S_1 which excludes S_2.

We thus have a partial reconstruction of the division of pragmatic labor, in that the existence of a more informative, marked term (L_2), together with the choice by a (fully informed) speaker to employ a less informative, unmarked term (L_1) (in a context where the additional information would have been relevant), licenses the Q-inference that the speaker was not in a position to employ the more informative term. We can represent this state of affairs schematically as in (43).

(43)



(L_1 may be used to denote S_1-S_2, i.e. the complement of S_1 with respect to S_2)

Thus, we obtain a pragmatic resolution of the semantic asymmetry inherent in the original state diagram. At the same time, L_1 may (and usually does) retain its general application (so as to include S_2) in contexts where there is no contrast at issue with L_2: it is only when the sex of the beast is relevant that bitches do not count as dogs.

Noncanine examples of autohyponymy (i.e. private polysemy) deriving from Q-based narrowing include the following (cf. Kempson 1980, Horn to appear b):

(44) cow (excluding bulls)
rectangle (excluding squares)
finger (excluding thumbs)
gay (excluding lesbians)
player (excluding pitchers, in baseball)

I have argued elsewhere (Horn to appear b) that these cases are not homogeneous in nature, differing in degree of conventionalization of the relevant Q-based inference, and that (contra Kempson) they do not in any case represent the sole source of autohyponymy, given the examples of R-based narrowing cited earlier and those illustrating R-based broadening to be discussed further on.

The results of R-based and Q-based narrowing may be synchronically indistinguishable. As against the dog/bitch (and lion/lioness) variety, the narrowing of man seems to have preceded the development of its counterpart woman, and thus to represent an R-based rather than Q-based shift, males presumably being reckoned as the salient members of the species. More recent instances of what feminists have appropriately dubbed the masculine usurpation of the generic include mankind, chairman, and poet. Here, it is clearly the prior (R-based) specialization of the general term which created the perceived need for--and conscious innovation of--the corresponding feminine form; it is not the existence of sex-specific womankind, chairwoman, or poetess which led to a (Q-based) restriction on the extension of the general terms.

The specific L_2 term which triggers a Q-based restriction on the meaning or use of the general L_1 term must be sufficiently natural and stylistically unmarked, or (as we have observed in our earlier survey of Q-based restrictions) it will not count as a 'corresponding' item. Thus, Blackburn (1983:495) observes that animal may (or may not) be used so as to exclude humans, and it may likewise be taken in the appropriate context to exclude birds and/or fish, but there is no use of animal which excludes mammals--despite the obvious fact that mammalia constitute just as valid a subgrouping of animals as do birds and fish. Crucially, however, mammal (unlike man, bird, fish) does not correspond to a basic

level category in the sense of Rosch (1977) and so cannot trigger the division of pragmatic labor illustrated in (43).

We have seen that narrowing of a lexical item may be either R-based (the spontaneous delimitation of a general term to a sense representing a salient exemplar of the category denoted by that term) or Q-based (the motivated specialization of a general term triggered by the prior existence of a hyponym of that term). The converse process--lexical broadening or expansion--is always R-based: the generalization of a term for a species to cover the encompassing genus, from genus to phylum, from subset to superset. Thus, Latin pecūnia, originally denoting 'property or wealth in cattle' (cf. pecū 'livestock, cattle'), generalizes to signify 'wealth' and eventually 'money', a shift paralleled in the English cognate fee (<OE feoh 'cattle'→'property'). As noted by Bréal (1900) and Ullman (1959), broadening is often accompanied by 'semantic impoverishment' resulting from the attrition of a qualifying context, as in the expansion of (assumed) Late Latin adripare, arripare, 'come to shore' into French arriver 'arrive (tout court)', or the generalization of panarium 'bread basket' into panier 'basket' (Ullman 1959:209).

Broadening tends to apply regularly with place and origin names, as political entities grow and mutate; examples include the expansion of Lat. romanus (or Eng. New Yorker) to designate someone or something from the empire (state) at large, rather than specifically from its major city. An even more productive source of lexical broadening involves trade names which have lost their capital letters and become generics (cf. Mason and Pimm 1982, Horn to appear b), including those in (45).

(45)

xerox	jello	good humor
kleenex	thermos	toll-house cookies
scotch tape	vaseline	hoover (Br., 'vacuum cleaner')

As in many of the instances of narrowing discussed earlier, the net result (or at least temporary equilibrium state) here is autohyponymy: the broadened term retains its original specific meaning in at least some contexts. We may even end up with multiple hyponymy, as in the case of the enthogeographic label Yankee, with its three semantically nested extensions cited by McCawley (1981:9-10) and standard lexicographers,[4] 'native or inhabitant of New England; or, more generally, of northern U.S.; or, more generally, of U.S.'

In these examples, broadening results when a specific term representing a salient exemplar (often the salient exemplar) of a wider class generalizes to denote that wider class; lexical expansion thus constitutes a perfect mirror image of R-based narrowing from a set to a salient exemplar of it. Once the new value for the term is entered in the lexicon--alongside or in place of the original specific value--semantic shifts (often culturally triggered) may ensue, obscuring the original set/salient member relation. It was the Romans' use of livestock as the medium of exchange which led to the broadening of pecūnia, a derivative of the term for 'livestock', into a term denoting 'property'--and it was the subsequent abandonment of barter in favor of a monetary system that led to the later loss of the etymological component altogether, to the point where pecūnia simply stood for 'wealth'.[5] So too, in Mayan

languages studied by Lounsbury, a word referring to 'serpent' or 'jaguar'--animals with a particular ritual significance--became generalized as the standard term for 'animal', retaining its original specific meaning in compounds. The manufacturers of the items in (45) should therefore find a silver lining in the legal cloud of their copyright loss: their products obviously represent the prototypes or epitomes of their respective kind, even if this results in their brand name coming to stand for that kind.

Broadening and narrowing often operate in tandem within a given language, or in complementary fashion across related languages. Thus, German Tier has broadened from 'wild animal' to include domestic livestock and pets as well as man (especially in compounds like Tierwelt 'animal kingdom'). At the same time, its English cognate has turned autohyponymous (and eventually unambiguous once more) through narrowing. OE dēor, ME deer originally designated beasts in general, especially 'objects of chase' (Stern 1931:416), then became restricted (initially in the lexicon of hunters) to single out the object of chase par excellence, fam. Cervidae. By the early modern period, the general use of deer had been largely supplanted by the Romance loans beast, brute, and animal, although it continued to retain a marginal application to the class of quadrupedal mammals, undifferentiated for species (as in Shakespeare's reference to rats and mice and such small deer). Eventually, only the specialized hyponymic sense survived, spreading from hunters' use into the general speech community.

The adjustment in the extension of deer vis-à-vis animal reflects a general tendency in language insightfully described by Bréal (1900: 27ff.). Bréal's principle, in effect a diachronic precursor of Kiparsky's Avoid Synonymy, is the Law of Differentiation, governing

> the intentional [!], ordered process by which words, apparently synonymous, and once synonyms, have nevertheless taken different meanings, and can no longer be used indiscriminately...either they are differentiated, or else one of the two terms ceases to exist. (Bréal 1900:27-28)

Typically, an older word for a given referent is retained, but limited to a specialized (often humble, 'degraded', or 'trivial') domain. Once again, the marked form is limited to a marked use.

Thus, Bréal informs us, the Swiss patois word for 'room', païlé, is restricted to the meaning 'garret' after standard French chambre is adopted as the unmarked term. The general use of Oscan popina 'kitchen', displaced by its Latin cognate coquina, eventually comes to denote 'tavern'. And now 'the Savoyard uses the names of père and mère for his parents, while he keeps for his cattle the old words pâré and mâré' (Bréal 1900:29).

Among English examples of Bréal's Law of Differentiation we might reckon brethren (whose restricted use is motivated by the adoption of the standard unmarked plural brothers), the deer/animal case already discussed, and the related and notorious hound/dog affair.

Once again, we find broadening and narrowing operating hand-in-hand (or paw-in-paw) until the eventual division of labor is arrived at. OE dogca, referring to a particular breed of dog (it is not entirely clear

which one), represented a hyponym of the general term hound, then denoting the entire kind 'dog' (as its German littermate Hund continues to do). Sometime around the fourteenth century, when Chaucer's warning 'It is nought good a slepyng hound to wake' was turning into Heywood's 'It is evyll wakyng of a sleepyng dog', dog and hound were presumably both autohyponyms, with different specific understandings. Eventually, hound was totally displaces by dog in its general application but, in accordance with the Law of Differentiation, continued to retain its specialized use (originally developed via R-based narrowing in the vocabulary of hunters, for whom hounds were the salient representatives of the species, dogs par excellence).

Thus, narrowing and broadening, separately and in conjunction, reflect the centrality of R-based and Q-based shifts in the development of the lexicon. In R-based and Q-based narrowing, in R-based broadening, and in instantiations of Bréal's Law of Differentiation, as in the dynamics of use and meaning described by Chomsky (1982), Aronoff (1976), Kiparsky (1982), McCawley (1978), and Searle (1975), we descry the recurring patterns of our two general pragmatic principles and of the division of labor resulting from their interaction.

13. Varia and concluding remarks. Other apparent reflexes of the R vs. Q dynamic cannot be discussed here in detail, but are worth mentioning as possible topics for further investigation. First, there is the class of rhetorical figures of speech, including synecdoche, metonymy, and litotes, representing the 'parole'-based equivalent of conventionalized narrowing (part for whole) and broadening (whole for part).

Second, there is the privative relation between the meanings (and uses) associated with nominative and ergative case marking for subjects of intransitive verbs in 'fluid' or 'active' systems (cf. Dixon 1979 and especially Holisky 1983). In Bats (a.k.a. Batsbiy and Tsova-Tush), a Caucasian language investigated by Holisky, the facts appear to be as follows:

(46) With intransitive verbs which are typically agentive and intentional (e.g. 'get dressed', 'wash', 'hide', 'run', 'bump into'), NOM(inative) case is semantically marked for nonagentivity and ERG(ative) is unmarked.

(47) With intransitives which are typically nonagentive (e.g. 'die', 'burn', 'become poor', 'forget', 'drown', 'go crazy'), ERG is marked as agentive and NOM is neutral.

(48) With intransitives which allow both agentive and nonagentive interpretations equally freely (e.g. 'fall asleep', 'fall down', 'lose weight', 'be late'), ERG and NOM are equally natural and no markedness relation obtains.

In the first two sets of examples, we are dealing with a privative opposition (Zwicky and Sadock 1975, Horn to appear b), in which 'the marked member conveys its meaning truth-functionally, while the neutral member does so by implicature' (Holisky 1983:5). The existence of the more informative, marked form, together with the speaker's choice of the unmarked, semantically neutral form, allows the addressee to construct

a Q-based implicature: the inference from the use of ERG with the verbs of (46) that an agent was involved, and the corresponding inference from the use of NOM with the verbs in (47) that the nonagentive interpretation was intended. Once again, we arrive at a division of pragmatic labor, in which the marked form is used for the marked situation (relativized to the semantics of the verb in question), and the unmarked form for the unmarked situation.

Finally, one more possible locus of the dynamics and resolution of the R/Q conflict is the range of 'switch-reference' constructions (cf. e.g. Finer 1984). As I read the data, there seem to be some languages (including Seri and Washo) in which the presence of a DS (different-subject) marker indicates that an embedded clause has a different subject from the main clause, while the lack of a marker is semantically unspecified for same vs. different referent, but tends to be interpreted as indicating that the subject is the same in contexts where the distinction is relevant and no further disambiguating factors are available. The asymmetry involved here is apparently analogous to that just touched on in Bats case-marking (as well as other examples discussed earlier), and thus similarly reflective of the use of Q-based implicature to complete the division of labor, but further investigation is required to sharpen the account of switch-reference and situate it more clearly within the proposed framework.

We have surveyed (all too cursorily) a wide range of linguistic phenomena, both synchronic and diachronic, both lexical and syntactic, both 'parole'-based and 'langue'-based, from conversation implicature and politeness strategies to the interpretation of pronouns and gaps, from blocking and distributional constraints on lexical items to indirect speech acts, from lexical change to case marking. If I am right, these apparently diverse and unrelated domains are all motivated and governed by the same functional dynamic, the ongoing Zipfo-Gricean dialectic between the Q-based Sufficiency Principle and the R-based Principle of Least Effort.

Notes

1. Ducrot's model of pragmatic inference shares with Grice's the crucial feature of indeterminacy. Given an utterance like (i),

(i) La situation n'est pas excellente. 'The situation isn't excellent'

Ducrot notes (1972:132), an addressee may infer that the speaker intended to convey (ii),

(ii) Elle est franchement mauvaise. 'It's pretty bad'

through the exploitation of the R-based 'Loi d'informativité' (the assumption that the hearer does not already know the information the speaker is conveying); since he has not said (ii), however, the speaker can always retreat to the literal meaning of what he has said, i.e. the weaker (i). On the other hand, given the Q-based 'Loi d'exhaustivité' (the principle which demands that the speaker provide the strongest possible information which he possesses and which he believes may interest the hearer; cf. Ducrot 1972:134), someone who utters (i) may, in the appropriate context, implicate that the situation is pretty good. This indeterminacy is perhaps more apparent when the negation applies to a semantically negative predication, as pointed out by Stern (1931:312):

> Not bad, taken literally, leaves a large latitude, from indifferent to excellent, and may mean [sic] either, depending on the intonation used and the circumstances.

2. Following Lyons (1977:9.4), A is called a 'hyponym' of B iff the extension of A is properly included in that of B: Labrador retriever is a hyponym of retriever, retriever of dog, dog of mammal, and so on. But some words are hyponyms of themselves: dog and bitch are (sex-differentiated) co-hyponyms of dog, lion and lioness of lion, etc. In these cases, we can call the unmarked term (dog or lion) an 'autohyponym'. Autohyponymy thus represents privative polysemy or ambiguity within a single lexical item (cf. Zwicky and Sadock 1975 on privative opposition).

3. Bréal (1900:108) suggests that in fact alogon [aloγo] came to stand for the horse simply because 'the rider, speaking of his mount, was accustomed to say "the animal"'; similarly, homines were so-called not merely because of man's preeminent position among the creatures of earth, but because of the intended opposition between the earthbound human race as against 'the inhabitants of the sky Dii or Superii' (Bréal 1900:114).

4. As noted in Horn (to appear b), there is an even more narrowly defined sense of Yankee, the sense in which the Kennedys are disqualified from true Yankee status by their Irish Catholic heritage. On this ultrarestrictive interpretation, a Yankee is someone from New England who approximates to a sufficient degree the prototype WASP of the Pepperidge Farm commercials. (We may need to invoke a Rosch (1977)-style prototype theory in any case to explain why a Vermont farmer or a Maine lobsterman is more of a Yankee than is a Greenwich stockbroker.)

5. The mirror image relation between R-based broadening (from salient subset to superset) and R-based narrowing (from superset to salient subset) is highlighted when we juxtapose the development of pecūnia with the opposite shift exemplified by ktimata in Greek, from the general 'possessions' to the specific 'cattle' (Bréal 1900:109).

References

Aronoff, M. 1976. Word formation in generative grammar. Cambridge, Mass.: MIT Press.

Atlas, J., and S. Levinson. 1981. It-clefts, informativeness, and logical form. In: Cole, ed. (1981:1-61).

Blackburn, W. 1973. Ambiguity and non-specificity: A reply to Jay David Atlas. Linguistics and Philosophy 6.479-98.

Bréal, M. 1900. Semantics. Trans. Mrs. H. Cust. New York: Henry Holt.
Brown, P., and S. Levinson. 1979. Universals in language use: Politeness phenomena. In: Questions and politeness: Strategies in social interaction. Edited by E. Goody. Cambridge: Cambridge University Press. 56-311.
Chomsky, N. 1982. Lectures on government and binding, 2nd ed. Dordrecht: Foris.
Clark, H., and S. Haviland. 1977. Comprehension and the given-new contract. In: Discourse production and comprehension. Edited by R. Freedle. Hillside, N.J.: Erlbaum. 1-40.
Cole, P., ed. 1978. Syntax and semantics 9: Pragmatics. New York: Academic Press.
Cole, P., ed. 1981. Radical pragmatics. New York: Academic Press.
Cole, P., and J. Morgan, eds. 1975. Syntax and semantics 3: Speech acts. New York: Academic Press.
Coleman, L., and P. Kay. 1981. Prototype semantics: The English word lie. Lg. 57.26-44.
DeMorgan, A. 1847. Formal logic. [Reprinted London: Open Court Co., 1926.]
Dixon, R. M. W. 1979. Ergativity. Lg. 55.59-138.
Ducrot, O. 1972. Dire et ne pas dire. Paris: Hermann.
Enç, M. (to appear) Topic switching and pronominal subjects in Turkish. In: Studies in Turkish linguistics. Edited by K. Zimmer.
Finer, D. 1984. The formal grammar of switch reference. University of Massachusetts dissertation.
Gazdar, G. 1979. Pragmatics. New York: Academic Press.
Geis, M., and A. Zwicky. 1971. On invited inferences. Linguistic Inquiry 2.561-65.
Gordon, D., and G. Lakoff. 1975. Conversational postulates. In: Cole and Morgan, eds. (1975:41-58).
Hamilton, Sir W. 1860. Lectures on logic. Edinburgh: Blackwood and Sons.
Harnish, R. M. 1976. Logical form and implicature. In: An integrated theory of linguistic ability. Edited by T. Bever, J. Katz, and D. T. Langendoen. New York: Crowell. 464-79.
Heringer, J. 1972. Some grammatical correlates of felicity conditions. Ohio State University dissertation. Distributed by Indiana University Linguistics Club.
Heringer, J. 1976. Idioms and lexicalization in English. In: Syntax and semantics 6: The grammar of causative constructions. Edited by M. Shibatani. New York: Academic Press. 205-16.
Holisky, D. 1983. The privative nature of agent-marking in Bats (and English). Paper presented at annual meeting of the Linguistic Society of America, Minneapolis.
Horn, L. 1972. On the semantic properties of logical operators in English. University of California at Los Angeles dissertation. Distributed by Indiana University Linguistics Club, 1976.
Horn, L. 1973. Greek Grice: A brief survey of proto-conversational rules in the history of logic. In: Chicago Linguistic Society 9.205-14.

Horn, L. 1978a. Lexical incorporation, implicature, and the least effort hypothesis. In: Papers from the Parasession on the Lexicon. Chicago: Chicago Linguistic Society. 196-209.
Horn, L. 1978b. Remarks on neg-raising. In: Cole, ed. (1981:129-220).
Horn, L. 1978c. Some aspects of negation. In: Universals of human language, vol. 4: Syntax. Stanford, Calif.: Stanford University Press. 127-210.
Horn, L. (to appear a) Metalinguistic negation and pragmatic ambiguity. Lg.
Horn, L. (to appear b) Ambiguity, negation, and the London School of Parsimony. In: Proceedings from New England Linguistic Society 14. Edited by C. Jones and P. Sells.
Horn, L., and S. Bayer. (to appear) Short-circuited implicature: A negative contribution. Linguistics and Philosophy.
Jespersen, O. 1924. The philosophy of grammar. [Reprinted New York: Norton, 1976.]
Karttunen, L. 1970. The logic of English predicate complement constructions. Unpublished MS, University of Texas.
Keenan, E. O. 1976. The universality of conversational postulates. Language in Society 5.67-80.
Kempson, R. 1980. Ambiguity and word meaning. In: Studies in English linguistics. Edited by S. Greenbaum, G. Leech, and J. Svartvik. London: Longmans. 7-16.
Kiparsky, P. 1982. Word-formation and the lexicon. In: Proceedings of the 1982 Mid-America Linguistic Conference, University of Kansas. Edited by F. Ingemann.
Lakoff, R. 1973. The logic of politeness; or, minding your P's and Q's. In: Chicago Linguistic Society 9.292-305.
Levinson, S. 1973. Pragmatics. Cambridge: Cambridge University Press.
Lewis, D. 1969. Convention. Cambridge, Mass.: Harvard University Press.
Liberman, M., and I. Sag. 1974. Prosodic form and discourse function. In: Chicago Linguistic Society 10.416-27.
Lyons, J. 1977. Semantics 1. Cambridge: Cambridge University Press.
McCawley, J. 1978. Conversational implicature and the lexicon. In: Cole, ed. (1978:245-59).
McCawley, J. 1981. Everything that linguists have always wanted to know about logic. Chicago: University of Chicago Press.
Martinet, A. 1962. A functional view of language. Oxford: Clarendon Press.
Mason, J., and D. Pimm. 1982. Generic examples: Seeing the general in the particular. Unpublished MS, The Open University, Milton Keynes, England.
Mills, J. S. 1867. An examination of Sir William Hamilton's philosophy, 3rd ed. London: Longmans.
Morgan, J. 1970. On the criterion of identity for noun phrase deletion. In: Chicago Linguistic Society 6.380-89.
Morgan, J. 1978. Two types of convention in indirect speech acts. In: Cole, ed. (1978:261-80).

Paul, H. 1909. Prinzipien der Sprachgeschichte, 4th ed. Halle: M. Niemeyer.
Postal, P. 1974. On raising. Cambridge, Mass.: MIT Press.
Prince, E. 1981. Toward a taxonomy of given-new information. In: Cole, ed. (1981:223-55).
Prince, E. 1982. Grice and universality: A reappraisal. Unpublished MS, University of Pennsylvania.
Rosch, E. 1977. Human categorization. In: Advances in cross-cultural psychology. Edited by N. Warren. New York: Academic Press. 1-72.
Schmerling, S. 1975. Asymmetric conjuction and rules of conversation. In: Cole and Morgan, eds. (1975:211-31).
Searle, J. 1975. Indirect speech acts. In: Cole and Morgan, eds. (1975:59-82).
Stern, G. 1931. Meaning and change of meaning. Bloomington: Indiana University Press.
Tannen, D. 1975. Communication mix and mixup, or how linguistics can ruin a marriage. San Jose State Occasional Papers in Linguistics.
Tannen, D. 1979. Ethnicity as conversational style. Sociolinguistics Working Paper No. 55, Southwest Educational Development Laboratory, Austin, Texas.
Ullman, S. 1959. The principles of semantics, 2nd ed. Glasgow: Jackson, Son and Co.
Zipf, G. K. 1949. Human behavior and the principle of least effort. Cambridge: Addison-Wesley.
Zwicky, A., and J. Sadock. 1975. Ambiguity tests and how to fail them. In: Syntax and semantics 4. Edited by J. Kimball. New York: Academic Press. 1-36.

INTENSITY

William Labov
University of Pennsylvania

At the heart of social and emotional expression is the linguistic feature of intensity.[1] It is a difficult feature to describe precisely. Intensity by its very nature is not precise: first, because it is a gradient feature, and second, because it is most often dependent on other linguistic structures. Most discussions of intensity involve specialized prosodic contours and a set of adverbs that code intensity directly: really, so, and very in English. But intensity is most often expressed through linguistic forms that are normatively devoted to logical relations and conceptual categories. The use of such forms to signal intensity can lead to changes in the subsystems involved. If grammatical descriptions don't take social and emotional expression into account, and their effect on the underlying system, they will be incomplete and even misleading for language learners.

1. Intensity in adverbs: Cognitive zeroes. Emotion is often expressed through peripheral, gradient systems: by prosody, vocal qualifier, and gesture. Information on emotional states can be conveyed by the central grammatical apparatus, completely verbalized in propositional form, but with lower chances of success. Any imaginable emotional state can be stated as a proposition: 'I am moderately angry with you', or 'I'm entirely committed to this line of action'. But we are all familiar with situations where listeners refuse to accept these words at their face value. On the other hand, De Groot's principle (1949) tells us that whenever the message conveyed by the intonational system is at variance with the information contained by the words, the intonational message will be understood as the one intended. It would follow that the peripheral systems are the primary means of conveying social and emotional information, and the grammatical mechanism is the primary means for conveying referential and cognitive information.[2]

Yet, some elements of the grammatical system are specifically devoted to emotional expression, and the most common of these are adverbs that signal intensity. 'Intensity' is defined here as the emotional