

6.1. Context and Scalar Implicature

Recall that Gazdar's approach to the calculation of utterance meanings is to define functions which, for any utterance, will produce sets of entailments, potential presuppositions, and potential implicatures which that utterance might license. He then determines the actual meanings licensed by adding first entailments, then potential clausal quantity implicatures, potential value quantity implicatures, and potential presuppositions, to the discourse context. So long as each addition produces a consistent context, the context will be incremented and potential meanings will become actual meanings of the utterance. Meanings whose addition would produce an inconsistent (in the sense defined in Section 2.4.2.1) context Gazdar deems to be canceled by that context.

But this notion of context involves only prior assertions which are logically inconsistent with a potential implicature. In effect, if u_i can ever, under any circumstances, license p_j , Gazdar's system, as he has presented it, will predict that it will -- unless $\neg p_j$ or some sentences that logically imply it have previously been realized in the discourse.

Of course, it is not a simple matter to determine whether or not some expression will evoke an ordering or which ordering some expression will evoke. For example, in 218, B apparently perceives an ordering

- (218) A: Have you, me, and Ellen ever had dinner together?
B: We've had lunch.

defined over the set of meals as salient.¹³⁰ B implicates $\neg\text{BEL}(B, \text{we've had dinner})$ -- as well as $\neg\text{BEL}(B, \text{we've had breakfast})$ -- since *dinner* and *breakfast* are each alternate values with *lunch*. But B might perceive other orderings as salient from the same query, as in 219.

- (219) A: Have you, me, and Ellen ever had dinner together?
B: We've been to the movies.

Here the salient ordering appears to be defined on some set of social events. While one of the implicatures licensed by B's response will be the same as one licensed by 218 -- $\neg\text{BEL}(B, \text{we've had dinner})$ -- another might be $\neg\text{BEL}(B, \text{we've gone for drinks})$. But clearly, even if A and B both recognize some set of social events as salient, how will they know that it is the same

¹³⁰This ordering may be represented as follows:

| | | |
|-------------------|--------------------|----------------|
| {breakfast,lunch} | {breakfast,dinner} | {lunch,dinner} |
| {breakfast} | {lunch} | {dinner} |

such set? Can B really implicate $\neg\text{BEL}(B, \text{we've had drinks})$ unless s/he believes s/he and A both recognize not only the salient ordering metric but also the set of events over which it is defined? And, in 220, the salient ordering supporting B's implicature that $\neg\text{BEL}(B, \text{you, me, and Ellen have had dinner together})$ is not only different from the salient orderings evoked in 218 and 219, but is evoked by B's perception of a different expression -- (*you,me, Ellen*) -- as salient.

- (220) A: Have you, me, and Ellen ever had dinner together?
B: Ellen and I have.

Finally, 221 illustrates how S may evoke multiple salient orderings from multiple expressions in a single utterance:

- (221) A: Have you, me, and Ellen ever had dinner together?
B: Ellen and I have had lunch.

In this exchange, orderings defined over (*you,me, Ellen*) and (*breakfast,lunch,dinner*) are both evoked by B's response -- to implicate $\neg\text{BEL}(B, \text{you, me, and Ellen have had dinner together})$. However, how we might model other apparent implicatures, such as $\neg\text{BEL}(B, \text{you, me and Ellen have had lunch})$ and $\neg\text{BEL}(B, \text{Ellen and I have had dinner})$, is not so clear.

So, to calculate a scalar implicature from an utterance with semantic representation p_i , we need some way of determining which of all the expressions in p_i S is employing in licensing some scalar implicature (and, thus, must anticipate H will recognize) and which of potentially infinitely many possible ordering relations will be salient for those expressions. We must also determine how S and H can assume mutual recognition of the actual definition of such orderings.

6.2. Defining Posets in Discourse

Harnish, Horn, and Gazdar all implicitly assume that S and H share knowledge of a common set of ordering relations, each of which is automatically evoked when one of its values is referenced. But, of course, even when S and H perceive orderings that seem intuitively quite similar as being evoked by a given value, they may differ critically for the purpose of successful communication of scalar implicature. If S believes some ordering O_i is salient and relies upon H's perception of O_j as such for the interpretation of his/her implicature, but if H believes some distinct ordering O_k is salient instead, then S may not accurately anticipate the inference(s) that H will derive from S's utterance -- and thus H may infer meanings S has not intended. Or, H may fail to infer meanings that S has intended to convey. So, in anticipating H's general ability to 'work out' a scalar implicature, S must, in particular, anticipate whether H will be able to identify the ordering that relates S's utterance to other utterances s/he might have made and whether H will define this ordering as S does.

6.2.1. Sources of Divergent Definition

It seems reasonable to assume that *S* and *H* are more likely to disagree on the definition of domain-dependent and non-canonical orderings than on canonical orderings. Speakers' knowledge of canonical orderings such as the cardinals, quantifiers, and modals and many modifiers appears to stem from their linguistic competence or very general world knowledge and thus seems more likely to be shared among speakers. That is, knowledge of the relative positions of *one* and *four* or *some* and *all* or *pretty* and *beautiful* in their respective orderings should vary little from speaker to speaker. But even for such canonical orderings as the should vary little from speaker to speaker. But even for such canonical orderings as the quantifiers, the relative positions of *many*, *most*, and *quite a few* may not be agreed upon by *S* and *H*. And for orderings such as those which permit scalar implicature in 222, it would

- (222) A: Do you want a Rainbow?
 a. B: I want a 1200 band modem.
 b. B: I want an IBM PC.
 c. B: I want a piece of gum.

be most surprising if *S* and *H* could define the same set of salient entities evoked by each response.

While *S* and *H* may identify divergent orderings as salient under any circumstances, such divergence seems most likely when orderings are domain-dependent and when *S* and *H* possess unequal expertise in that domain. In 223, for example, the distinction between *pneumonia* and

- (223) A: You mean, I have pneumonia?
 B: A mild case.

mild pneumonia may not be apparent to a patient. However, such exchanges may convey to the less-expert partner the information that some term *H* had not viewed as a potential value might in fact be so viewed, i.e., that an ordering exists on which *pneumonia* and a *mild case of pneumonia* are values. Thus speakers need not always avoid the use of orderings their hearers may not recognize, but simply recognize their possible consequences — at best a request for clarification, at worst, miscommunication. And, by their use, *S* may actually convey to *H* that some previously unknown ordering in fact exists. So, in 223, while *A* may not actually understand the ordering *B* employs, *s/he* will be able to infer that there are 'degrees' of pneumonia.

Alternatively, if *H* has more expertise than *S*, *s/he* may derive implicatures *S* does not intend by interpreting some reference in a too technical light. In 224, if *A* interprets *B*'s response in a more technical sense than *B*, *s/he* may conclude — BEL(B), *the child is an idiot*), if *s/he* knows that idiots are technically classed as

- (224) A: Is that child handicapped?
 B: He's an imbecile.

having greater intelligence than imbeciles. In both cases of unequal expertise, some modeling of *S* or *H* is required to assess licensed implicatures properly. Such user modeling will be pursued here, but its utility should be noted for these cases.

6.2.2. Defining Poset Compatibility

Although above I have noted the potential for communication breakdown when *S* and *H* do not share a common definition for some ordering, and I have in previous chapters demanded that some particular ordering's salience be a mutual belief of *S* and *H* for scalar implicature, this constraint is in fact too strong. I will now propose a relaxation of this condition, expressed in terms of the metric σ and the set *E* that, together, define all orderings *O*.

From the examples given above it can be noted that two orderings O_1 and O_2 may differ from one another either in the set of expressions they are defined over or in the metrics which define those expressions — or, of course, in both. By our definition of orderings as posets, for any two posets O_1 and O_2 to be termed identical, they must be defined over identical sets of referents E_1 and E_2 by identical metrics σ_1 and σ_2 . So, for each ordered pair $e_1, e_2 \in O_1$, $e_1, e_2 \in O_2$, and for each $e_1, e_2 \in O_1$, $e_1, e_2 \in O_2$. That is, orderings are identical iff they are defined over the same set of referents and if the metrics ordering these sets order their members in exactly the same way. And orderings may differ if either E_1 and E_2 are different or if σ_1 and σ_2 are different. But note that the definition of a poset is independent of the label *S* and *H* place on either the set or the metric or the partial ordering itself. That is, σ_1 and σ_2 or E_1 and E_2 or O_1 and O_2 may be LABELED differently by *S* and *H* — yet still be identical according to this definition.

But even if *S* does not believe that the ordering O_1 , *s/he* believes to be salient will be recognized by *H*, *s/he* may still use O_1 to support a scalar implicature given that *s/he* believes certain less restrictive conditions hold. Suppose *S* believes O_1 to be defined by σ_1 over E_1 , but *s/he* also believes that *H* may only recognize a subset E_2 of E_1 , as ordered by σ_1 . For example, say in 225, *B* recognizes a set of desserts

- (225) A: Do you have jelly-o?
 B: We have pie.

ordering defined over the set {*pie*, *jelly-o*, *cake*, *ice-cream*, *fruit*}, but *A*'s set of desserts does not include *fruit*. Although the ordering by which *B* anticipates and *A* interprets scalar implicatures in this exchange will thus differ, the inferences *A* will draw given an ordering defined over {*pie*, *jelly-o*, *cake*, *ice-cream*} will be at least consistent with those *A* implicates via the larger set.

is, while B may intend also to convey \neg BEL(B, *we have fruit*) by 225, all those scalar implicatures that are understood by A will be scalar implicatures -- i.e., will be intended by B.

When O_1 and O_2 are identical except in that the set of referents over which O_2 is defined, is a proper subset of that over which O_1 is defined, E_1 , then O_2 will also be a proper subset h . When S believes that O_1 is salient but has reason to believe that H may recognize some h , then S will nonetheless believe that scalar implicatures inferred will be consistent with h implicatures intended, since the former will be a subset of the latter.

If instead S has reason to believe that H may recognize a larger set of salient referents in O_2 than S recognizes as salient in O_1 -- if E_1 is a proper subset of E_2 and, thus S's O_1 proper subset of O_2 -- then S may suppose that his/her knowledge will be accurately interpreted by H. Say, in 225, it is B, not A who fails to include fruit among his/her set of sets. Then, B may utter his/her response while believing that s/he does have fruit -- but A understands that \neg BEL(B, *we have fruit*). In such cases, when S has reason to believe that is a proper subset of O_2 , S may indeed anticipate that H may draw false conclusions from her implicature.

So, I will term S's salient ordering O_1 COMPATIBLE with any H's ordering O_2 if O_2 is a set of O_1 ; that is, O_2 is COMPATIBLE with O_1 if O_2 is either identical to O_1 or if O_2 is a proper set of O_1 . In other words, O_2 will only be incompatible with O_1 if O_1 is itself a proper subset of O_2 . S may then employ any O_1 to license scalar implicature if O_2 is compatible with the O_1 's) may recognize as salient.

Two orderings O_1 and O_2 may also differ in that, while they are defined over the same set E_2 of referents, two different metrics, e_1 and e_2 , order E_1 . That is, for some pair of expressions e_1 and e_2 , while $e_1 \notin E_1$, and $e_2 \notin E_2$, $e_1 \notin O_1$ but $e_2 \notin O_2$ or $e_1 \notin O_2$ but $e_2 \notin O_1$.¹³¹ So, for example, in 226, if B cannot anticipate that

(226) A: Is the Pacific Fleet on red alert?
B: The First Battle Division is.

will recognize that *Pacific Fleet* is a higher value than *First Battle Division*, she will not be able to implicate \neg BEL(B, *the Pacific Fleet is on red alert*) -- even if she does believe A will recognize that *Pacific Fleet* and *First Battle Division* appear in a whole/part ordering. For, if A instead believes that *Pacific Fleet* is a part of *First Battle Division*, then A will conclude that \neg BEL(B, *the Pacific Fleet is on red alert*) instead.

¹³¹ Again, remember that metrics e_1 and e_2 are identical just in case, for all $e \notin O_1$, $e \notin O_2$, no matter what the metrics are labeled.

However, like the sets of expressions over which orderings are defined, the metrics which define them need not be identical for S to believe that his/her implicature will be 'worked out' by H. If, for all $e \notin O_1$, $e \notin O_2$, then, again, we can say that O_2 is compatible with O_1 . For any relations O_1 and O_2 on a set E_1 , O_1 on E_1 COVERS O_2 on E_1 if $O_2 \subseteq O_1$. So, if O_1 covers O_2 , then, again, O_2 will be compatible with O_1 , and S may anticipate that any implicatures based upon H's recognition of O_2 will be inferred -- though not all those implicatures may be inferred. So, O_2 will be compatible with O_1 when $O_2 \subseteq O_1$. That is:¹³²

Ordering Compatibility:
 $O_2 \subseteq O_1 \Rightarrow \text{COMPATIBLE}(O_2, O_1)$

If S believes that H will recognize an O_2 that is compatible with the O_1 S recognizes as salient, S will at least believe that the inferences H draws will be consistent with those S has implicated. In such cases, we will say that S believes some ordering to be AVAILABLE for some H in some context C_h :

$(\text{BEL}(S, \text{SALIENT}(O_2, C_h)) \wedge \text{BEL}(S, \text{BEL}(H, \text{SALIENT}(O_2, C_h)))) \wedge \text{BEL}(S, \text{COMPATIBLE}(O_2, O_1)) \Rightarrow \text{BEL}(S, \text{AVAILABLE}(H, O_2, C_h))$

However, while we will want to demand that S believe O_2 to be compatible with all O_1 that H might infer as salient in its place, we will not want to demand that S believe that all orderings H believes salient in some context C_h are compatible with all orderings H believes salient -- e.g., that, in 221, the ordering of (*you, me, Ellen*) is compatible with the ordering of (*breakfast, lunch, dinner*). For this and other reasons which will become clearer in Section 6.3, I will redefine SALIENT(O_2, C_h) as SALIENT(e_2, O_2, C_h) -- that is, some ordering is salient in a context for some particular expression which appears in it. Then we can redefine the notion that S believes O to be available to H via mention of e_1 in C_h as follows:

Belief in Availability of Orderings:
 $(\text{BEL}(S, \text{SALIENT}(e_2, O_2, C_h)) \wedge \text{BEL}(S, \text{BEL}(H, \text{SALIENT}(e_2, O_2, C_h)))) \wedge \text{BEL}(S, \text{COMPATIBLE}(O_2, O_1)) \Rightarrow \text{BEL}(S, \text{AVAILABLE}(H, e_2, O_2, C_h))$

With ordering compatibility and availability to H defined, and with ordering salience redefined, we can now revise the salience condition presented in the definitions of the scalar implicature conventions -- BMB(S, H, SALIENT(O_2, C_h)) -- in terms of inference consistency as BEL(S, AVAIL(H, e_2, O, C_h)). So, one mutual belief requirement in our definition of conversational implicature can be relaxed.

¹³² As Bonnie Webber (Personal Communication) has suggested, in 226, if B believes that A considers *Pacific Fleet* and *First Battle Division* to be alternate values in some ordering, then s/he can also license the desired implicature. So, there seem to be other ways in which orderings are compatible in the sense described here of licensing only implicatures consistent with S's beliefs. However, in such a case, B should surely 'square away' the disparity between his/her beliefs and A's [Postal 82].

6.3. Saliency and Scalar Implicature

While studies of conversational implicature have readily acknowledged the overriding importance of 'context' in the generation and interpretation of implicature, little attempt has been made to define how the particulars of an exchange interact with the conventions that underly implicature. I noted in Section 3.2.2 that, of the authors who have studied quantity implicature, only Horn explicitly demands that the metric underlying a particular implicature must be 'relevant' in the discourse. Although Horn mentions that context, as well as explicit cancellation or suspension, can block implicatures, he goes no further than this. And, given his discussion of how implicatures involving positive poles must be inferred, it is difficult to see how the mention of some expression like *some* will not license *-all* in his theory -- unless it is explicitly canceled. Although Gazdar includes a more computational notion of cancellation in his theory, his too in effect precludes implicature from the semantic representation and explicit cancellation alone. His value quantity implicature is context-dependent only insofar as it can be canceled by prior utterances in the context. Otherwise, all possible implicatures which might be generated by an utterance are generated, regardless of context or speaker intention.

If S does not believe it relevant to affirm the highest value he truthfully can, surely s/he may -- cooperatively -- choose not to do so. For example, it seems equally likely that plausible that A not have implicated --BEL(A, you took a lot of computer courses) in 89, 133 even though in this exchange B has clearly inferred it. As O'Hair [O'Hair 69:41] notes (summarizing Grice) "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 could be conveyed by the latter." (My italics) Such factors justify the inclusion of a saliency condition as described above in the definition of AVALL. I will incorporate these notions of audience interest and relevance by claiming that the licensing of a scalar implicature is dependent upon whether the ordering that might support such an implicature is perceived as SALIENT in the discourse.

The notion of SALIENCE in AI and natural-language literature has so far been confined to discussion of the saliency of objects and their attributes (Carbounell 70, Conklin and McDonald 82, McCoy 85). The saliency of orderings or relations -- that is, of entities together with orderings on them -- examined in [Herskovits 82] and mentioned in recent work by Grosz and Sidner [Grosz and Sidner 85], is much less well understood. Psychological literature on human

¹³⁷That is, in

A: And you took some courses...

B: I've taken a lot of computer courses in high school.

propensities for viewing objects in terms of one relationship or another has been largely confined to children's ability to construct taxonomic hierarchies.

There are two possible approaches to the identification of the salient expressions and orderings necessary for the calculation of scalar implicatures in discourse: First, we might identify one or more expressions which denote items marked as salient in discourse and search for any posets they may appear in. Second, we might identify relationships marked as salient in the discourse, see if they represent partial orderings, and then identify any expression in the sentence under consideration which may appear in this poset.¹³⁴ I will examine each of these possible strategies below.

6.3.1. Identifying Salient Expressions

As Gazdar notes, identifying orderings evoked by an utterance is not a simple matter of identifying a lexical item in the surface representation of an utterance as a value in some ordering. However, neither is it a matter of identifying predefined values in some semantic representation -- as Gazdar and Horn in effect do. An expression may evoke an ordering in some contexts which it will not invoke in others -- even when it is part of an identical sentence. So, mention of *some* or *could* or *three* need not always convey *-all*, *-should*, and *-four*. However, it does seem plausible that domain independent orderings may more confidently be assumed recognizable by both S and H. So, all other things being equal, a domain independent ordering will be more likely to be assumed salient for the purpose of supporting scalar implicature than will a domain dependent ordering.

Identifying a broader range of relations that support scalar implicature, as I have done in Chapter 5, makes it even more difficult to determine whether or not some expression should be seen as evoking a salient relation -- i.e., whether the identification of some relation is necessary for the interpretation of the utterance -- by expanding the class of expressions which may evoke such orderings. In effect, any expression may license a scalar implicature, as suggested by 227.135

¹³⁴While it seems likely that, if relations are salient in the discourse, the items they order should also be salient. However, this is an open question.

¹³⁵Actual items corresponding to salient expressions are italicized.

- (227) A: Did the girl in the red dress spill a diet coke?
 a. B: She spilled a diet pepsi.
 b. B: She spilled a *regular* coke.
 c. B: She spilled a *glass of tomato* juice.
 d. B: *Jane* spilled a diet coke.
 e. B: The girl in the red *stacks* spilled a diet coke.
 f. B: The girl in the *green* dress spilled a diet coke.
 g. B: The girl in the *green stacks* spilled a diet coke.
 h. B: The *boy* in the red dress spilled a diet coke.
 i. B: The girl in the red dress *will spill* a diet coke.
 j. B: The girl in the red dress *drank* a diet coke.
 k. B: The girl in the red dress *spilled the* diet coke.

How salient expressions may be identified in some utterance is of course a hard problem, since salience in general is far from understood. Potential solutions to the problem at hand may lie in ongoing work in linguistic pragmatics and in natural-language processing on sentence TOPIC (or LOCAL FOCUS or CENTERING in the AI literature),¹³⁶ which may be signalled by syntactic, prosodic, or certain semantic-pragmatic cues. Among these, the best understood are syntactic and prosodic markings.

6.3.1.1. Syntactic Cues

Syntactic marking is commonly seen as way of focussing an item. Topicalizations and other preposings ((228a)), *it*-clefts ((228b)), pseudo-clefts ((228c)), and there-insertion ((228d)) may serve to mark items syntactically.¹³⁷

- (228)
 a. A cat she's going to buy.
 b. It's a cat that she's going to buy.
 c. What she's going to buy is a cat.
 d. There's a cat she's going to buy.

If focus is an acceptable indicator of salience, then such clues may be included among the methods of identifying salient expressions for the purposes of interpreting scalar implicatures. However, syntactic marking too can accomplish only part of the task: In general, only one item can be marked in an utterance by syntactic means and not all salient expressions need be syntactically marked.

¹³⁶Below, I will use 'focus' to denote this general notion of S's centering attention on some particular aspect of the discourse. The item focused on serves as an index into the discourse structure, as well as an indication of which items inferences – such as scalar implicatures – are drawn about.

¹³⁷Among other things they do.

6.3.1.2. Intonational Cues

It has long been noted that intonation can serve to mark sentence FOCUS or TOPIC. How prosodic marking is actually interpreted has, however, been little studied. Recent work on so-called CONTRASTIVE STRESS [Culicover 83] and on FALL-RISE intonation [Ward 85a] suggests how these two prosodic features mark expressions as salient. For each, an expression including the semantic representation of the lexical item which contains the syllable receiving SENTENCE STRESS¹³⁸ would be a salient expression in our terms.

6.3.1.2.1. FALL-RISE Intonation

FALL-RISE intonation (See Note 61.) also helps to limit the range of salient expressions in a sentence uttered with this contour. Elsewhere [Ward 85a], Gregory Ward and I have proposed a procedure for identifying this range for a particular utterance made with FALL-RISE, which I will summarize below. Once this range has been identified, we can limit the set of possible posets from which S has chosen a salient ordering to those posets in which one of these potentially salient expressions appears.

In sentences uttered with the FALL-RISE contour, the syllable SYL receiving sentence prominence may mark a number of expressions as potentially salient. Consider the following example (recall that \SYL/ indicates SYL):

- (229)
 a. A: Did Liberals in southern Oregon vote for Hart?
 b. A: Did Liberals in the West vote for Hart?
 c. A: Did Liberals anywhere vote for Hart?
 d. A: Did anybody vote for Hart?
 B: Liberals in southern California voted for Hart.

In B's response in 229, SYL is contained in *California*, whose representation must therefore be included in any salient expression e_p . But only in B's response to (229a) does it seem that *California* constitutes e_p where an ordering defined over the set of U.S. states appears to be noted. If B's response is to (229b), however, then it appears that we should look at the expression denoting *southern California* to identify e_p and that *parts-of-the-western-U.S.* is an appropriate ordering here. In response to (229c), in *southern California* appears to represent e_p an ordering of locations of U.S. voters. Finally, in response to (229d), e_j represents a value in the ordering of U.S. voters, *Liberals in southern California*. So, the range of of potentially salient expressions includes at least these four. Note, however, that 230, where SYL is contained in *Liberals*,

- (230) B: *Liberals* in southern California voted for Hart

¹³⁸The most prominent syllable in a sentence.

cannot be a felicitous response to any of the queries in 229. Similarly, in 231, where SYL falls in *southern*, the only possible salient expression will in fact be *southern* itself, whatever

(231) B: Liberals in southern California voted for Hart

the context.

These observations can be accounted for by drawing upon Liberman and Prince's [Liberman 77] relational theory of stress, in which stress is analyzed as a binary feature with the values WEAK and STRONG. For example, Liberman and Prince's rules would assign the values STRONG (S) and WEAK (W) to the sentence *Liberals in southern California voted for Hart* (uttered with neutral intonation) as in Figure 6-1, where R labels the root of the tree.

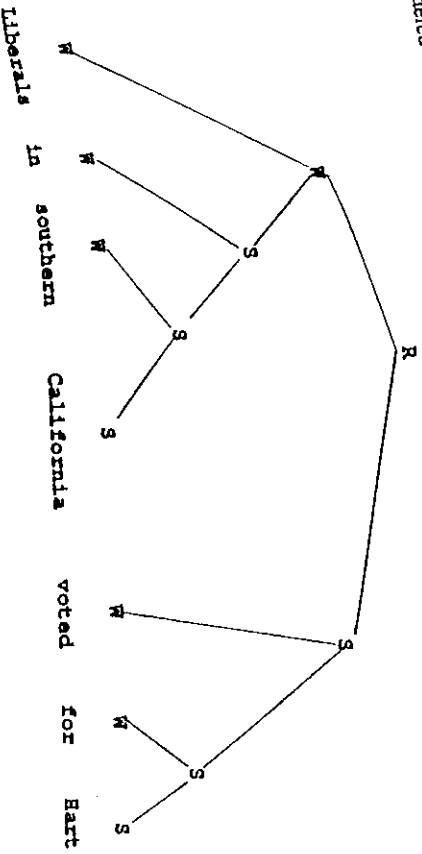


Figure 6-1: Metrical Stress Tree

Intuitively, when SYL occurs in a lexical item LEX which is dominated by a weak node in the prosodic tree for the corresponding utterance with unmarked intonation, only the expression including the representation of LEX can be marked as salient by FALL-RISE. Otherwise, larger expressions representing strings dominated by nodes in the prosodic tree dominating more than just LEX will also be candidates for expression salience.

Assuming a mapping between the semantic representation of an utterance u_i and its prosodic tree via the lexical items represented in each, this structure can be used to determine the set of potentially salient expressions E from such a prosodic tree as follows:

1. For any utterance u_i uttered with FALL-RISE, identify SYL and the lexical item it is included in, LEX.

2. Find the prosodic tree for u_i with unmarked/neutral intonation.
3. Traversing the shortest path from LEX to R in this tree, for each Node_i in this path (where i indexes the nodes from LEX to R):
 - A. If Node_i is weak (W), add the expression representing the leaves of the subtree rooted at Node_i to E and end the traversal.
 - B. If Node_i is strong (S), add the expression representing the leaves of the subtree rooted at Node_i to E and test Node_{i+1} .
4. Each of the expressions in E yields a possible salient expression for u_i .

Since *liberals* is immediately dominated by a W node in the tree above, in 230, where LEX is *liberals*, the traversal defined above will end with the testing of this first node on the path to R; thus only the subtree rooted at this node will be in E . In 231, only *southern* will be in E . However, in 229, no W node is encountered in the traversal until the trees represented by *California*, *southern California*, *in southern California*, and *liberals in southern California* have been added to E . So, four nodes are added to E for this utterance, as discussed above. By using an independently motivated relational theory of stress, we can provide a straightforward account of the range of potentially salient expressions for an utterance uttered with FALL-RISE — if prosodic information is available.

6.3.1.2.2. Contrastive Stress

In a study of the indirect speech acts performed by yes-no questions, Kieffer [Kieffer 80] proposes that certain yes-no questions which he terms FOCUSED actually function as *wh*-questions.¹³⁹ Focused yes-no questions for Kieffer are simply yes-no questions in which one constituent is marked by stress to indicate a BACKGROUND ASSUMPTION which S and H typically share.

(232a)-(232c) are focussed yes-no questions, with a different expression focussed in each, while (232d) contains no focussed element

- (232)
- a. Is John leaving for Stockholm TOMORROW?
 - b. Is John leaving for STOCKHOLM tomorrow?
 - c. Is JOHN leaving for Stockholm tomorrow?
 - d. Is John leaving for Stockholm tomorrow?

While any of the four may be answered with *yes* or *no*, it is also possible that if S asks (232a) she wants H to answer the question 'When is John leaving for Stockholm?'; if she asks (232b) she may want to know 'Where is John going tomorrow?'; and if she asks (232c) she may want to know 'Who is leaving for Stockholm tomorrow?'. To obtain these *wh*-questions, Kieffer

¹³⁹Similarly, [Hoopesman 83] claims that stress can indicate which element of a sentence is being queried.

constructs an open proposition from each focussed yes-no question by substituting a pro-word for the focussed element; this represents the query's background assumption. The questioner's request, then, is for an instantiation of the variable in this open proposition.

Although Kiefer does not recognize it, not all instantiations of his open proposition will be equally felicitous. (233a) is an odd

- (233) A: Is John leaving for Stockholm on Thursday?
 a. B: He's leaving for Stockholm on the train.
 b. B: He's leaving for good on Thursday.

response to (232a), and (233b), while a possible response to (232b), is a bit strange. However, since his variables are not typed, his theory will not preclude such infelicitous responses.

Kiefer also errs in claiming that only FOCUSED yes-no questions can elicit such indirect responses; nor does he recognize that more than one item in the query can be treated as queried. For example, all of the responses in 234 are possible responses to (232d).

- (234)
 a. B: Henry is.
 b. B: He's leaving for Stockholm on Wednesday.
 c. B: He's leaving for Paris tomorrow.
 d. B: He's leaving for Paris on Wednesday.
 e. B: Henry's leaving for Paris on Wednesday.

Finally, Kiefer's observations cannot account for the difference between responses that are appropriate to (235a) and those appropriate to (235b).

- (235)
 a. Is John leaving for the North Pole tomorrow?
 b. Is John leaving for the NORTH Pole tomorrow?
 c. He's leaving for Greenland.
 d. He's leaving for the South Pole.

While both (235c) and (235d) might be appropriate responses to (235a), only (235d) is an appropriate response to (235b).¹⁴⁰ As with FALL-RISE, contrastive stress on *pole* in (235a) may in fact focus several expressions, including *the North Pole* and *pole*; but stress on *north* can focus only *north* in (235b).

Culicover and Rochement's [Culicover 83] account of contrastive stress does accommodate the focusing of more than one expression via stress in one lexical item. They claim that the sentence FOCUS marked by contrastive stress can be determined directly from

¹⁴⁰That is, without some accompanying 'He's not going to either pole but...'

prosodic stress. However, their analysis would predict that any constituent that dominates the item containing the prominent syllable defines a possible focus constituent (1983:148). For contrastive stress, as for FALL-RISE, this account allows too broad a focal range. For example, Culicover and Rochement's analysis would predict that contrastive stress on *in* in ix could focus the entire expression *liberals in*

- (ix) A: Did Conservatives vote for Hart?
 B: #Liberals IN California voted for Hart.

California, and thus make it an appropriate response to A's query.

It seems likely that an algorithm similar to that presented in Section 6.3.1.2.1 can identify the range of possibly salient expressions marked by contrastive stress as well. Although such information would thus be quite useful for the calculation of scalar implicatures, it will only occasionally be available even in naturally occurring discourse. Also, intonational cues that can identify more than one salient expression in an utterance have not been studied. And, of course, computational recognition or generation of such cues is not yet possible.

6.3.1.3. Semantic and Pragmatic Cues

The determination of sentence focus from non-prosodic semantic and pragmatic cues has been proposed in studies of given/new information status, discourse coherence, thematic roles, definiteness and indefiniteness, reference and anaphora, and the goals of conversational participants.

6.3.1.3.1. Given and New Information

In distinguishing between GIVEN and NEW information,¹⁴¹ Chafe [Chafe 76] observes that mention of some class member can bring about the 'givenness' of the class. For example, in his 'bulldog' examples (reproduced in 236) the initial utterance might be followed by (236a), (236b), or (236c).

- (236) Yesterday I had my class disrupted by a BULLDOG.
 a. I'm beginning to DISLIKE bulldogs. (= [Chafe 76]'s 16b)
 b. I'm beginning to dislike dogs. (= [Chafe 76]'s 20b)
 c. I'm beginning to dislike animals. (= [Chafe 76]'s 22)

Chafe observes that, the higher the subsequent generic term in its classification hierarchy -- or, actually, the more distant from the initially mentioned item (e.g., *bulldog*, above), the more difficult it is for the generic to be treated as given (and deaccented, for example). That is, *animals* is less likely to be treated as 'in HF's consciousness' if S utters (236c) than would

¹⁴¹Classically, given information is information which can be assumed to be in HF's consciousness, through prior mention or some other means. It is often associated with linear position in an utterance; that is, given information is generally thought to come before new information.

bulldogs in (236a) or *dogs* in (236b); thus, moving up a hierarchy diminishes the implicit givenness that can be transferred. Likewise, moving down the taxonomic hierarchy does not produce a similar sense of givenness for *bulldog* in 237.

- (237) I'm beginning to dislike DOGS. (= [Chafe 76]s 23)
Yesterday I had my class disrupted by a BULLDOG.

Mention of an item higher in a classification hierarchy will not impart 'givenness' to lower items.

Chafe [Chafe 76:130] accounts for this phenomenon in terms of S's knowledge and awareness:

Whenever a speaker's knowledge is such that, for him, consciousness of X necessarily entails consciousness of Y, he will assume that the addressee's consciousness of X entails consciousness of Y also.

He notes that other orderings can impart a similar S assumption of mutual consciousness, as the whole/part ordering evoked in 238.

- (238) Yesterday my FANBELT broke. (= [Chafe 76]s 25)
I couldn't use the CAR all day.

Mention of *fanbelt* allows car to be treated as given.

If givenness can indeed be associated with salience, as has been claimed [Bock 77], then Chafe's observations might help identify salient expressions for the purpose of calculating scalar implicature. If mention of FANBELT can make *car* salient for a subsequent utterance, then presumably S and H can anticipate and identify this possibility. However, Chafe's notion fails to allow us to predict when items such as *bulldog* or *fanbelt* will not represent salient expressions -- i.e., when items which may be seen in some classification hierarchy or some part/whole relationship will not evoke these orderings. So, although these observations will prove useful below in identifying salient orderings for some salient expression, they do little to help us select among potentially salient expressions.

6.3.1.3.2. Centering and Discourse Coherence

Studies of discourse CENTERING seek to relate the coherence of a discourse to conceptual and positional relationships among items and references to them in a discourse and assumptions that can be made about H's awareness of those relationships. Work by Joshi, Weinstein, and Grosz [Joshi 81, Grosz 83] suggests a simple approach to the identification of salient expressions which does work for a restricted class of utterances.

Joshi et al. claim that a discourse will be more COHERENT when the BACKWARD-LOOKING CENTER of an utterance is either a member of the set of FORWARD-LOOKING CENTERS of the preceding utterance or is functionally dependent upon one such forward-looking center.

Forward-looking centers represent potential foci for subsequent utterances, and are defined tentatively in this work to include the set of noun-phrases in the current utterance.¹⁴² Since 'functional dependence' is both difficult to define and intuitively too narrow, Ward [Ward 85b] extends centering to encompass backward-looking centers which appear in partial orderings with some forward-looking center.

So, if coherent discourse is a goal of S and H, as is generally assumed, presumably we might restrict the range of possibly salient expressions for u_j to the forward-looking centers of its predecessor utterance u_i or to a set of preceding utterances in C_{u_i} . For utterances which license scalar implicature, salient expressions would have to represent a center that is functionally dependent upon a forward-looking center, or, following [Ward 85b], an expression in some partial ordering with an expression in the set of forward-looking centers. So, the location of sentence centers would seem to represent a method of locating our salient expressions.

[Joshi 81, Grosz 83]s sequential/conceptual relating of items in discourse does suggest a simple solution to the problem -- at least for certain types of exchanges. In any of the exchanges in 227 (repeated here for convenience),

- (1) A: Did the girl in the red dress spill a diet coke?
a. B: She spilled a diet *pepsi*.
b. B: She spilled a *regular* coke.
c. B: She spilled a *glass of tomato juice*.
d. B: *Jane* spilled a diet coke.
e. B: The girl in the red *slacks* spilled a diet coke.
f. B: The girl in the *green* dress spilled a diet coke.
g. B: The girl in the *green* slacks spilled a diet coke.
h. B: The *boy* in the red dress spilled a diet coke.
i. B: The girl in the red dress *will spill* a diet coke.
j. B: The girl in the red dress *drank* a diet coke.
k. B: The girl in the red dress spilled *the* diet coke.

for example, we might compare B's response with A's query to locate salient expressions as follows: Let P_j be the semantic representation of the DESIDERATUM¹⁴³ of A's query and P_j be the representation of the responder's answer. Then, if P_i and P_j represent simple expression alternants -- differing only in expressions e_i and e_j -- and if e_i and e_j appear together in some poset O , we can term $e_j P_j$'s backward-looking center, related to the forward-looking center, e_i . And, e_j will be a salient expression of P_j . For A's query in 227 and B's response of (227a), for

¹⁴²It is unclear whether they include noun-phrases from prior utterances as well.

¹⁴³A specification of the episodic state a questioner desires. Although centering is currently defined only for declaratives, this strategy will accomplish the desired transformation.

example, *pepsi* would represent a salient expression in this exchange and *set of sodas* might represent a potentially salient ordering, since *pepsi* and *coke* might be characterized as members of such a set. In this way it is possible to calculate expressions which B has taken to be salient -- and, thus, to calculate those scalar implicatures B may be licensing.

When more than one prior reference has been made to distinct values on some ordering, the selection of a salient ordering for a particular value reference should be even simpler. The more points specified on some ordering, the easier it is to identify it. For example, in 239, A's first query helps H to select among

(239) A: How many women in the company are managers?

B: Thirteen.

A: Do women employees take more days of sick-leave than male employees?

B: Non-management women do.

possible orderings on which *women employees* might appear to formulate a cooperative response to A's second query.

In fact, it does appear that a simple expression alternate condition alone is sufficient to identify salient expressions. Although the responses in 240 do appear

(240) A: Did the girl in the red dress spill a diet coke?

a. B: She spilled the bears.

b. B: The girl in the red car did.

intuitively less 'coherent' than the responses in 227, to make any sense of them A must postulate some set of *diet coke* and *bears* or *dress* and *car* -- and --BEL(B, *the girl in the red dress spilled a diet coke*) does seem to be a reasonable inference from both. However, when these values appear in more than one poset, other means must be used to identify which such poset is salient (See Section 6.3.2.3.).

Of course, the concept of centering which Joshi et al describe is so far defined only for entities, while salient expressions may refer to things other than entities, as noted in Section 5.1. Still, Joshi et al clearly intend some extensions to other items.

More serious for our purposes is that a simple expression alternate P_i immediately preceding P_j in the discourse will not always be easily discernible to permit the location of a salient expression in P_j , since scalar implicatures may be implicated when such a P_i is not the immediate predecessor of P_j or when no such P_i exists. In 241, for example, P_i is separated from

P_j by several intervening sentences. While the salience of *apples and pears* might be inferred by simple comparison of '*like(B, fruit)*' with '*like(B, {apples, pears})*', how one locates the former

(241) A: Do you like fruit?

B: Why, do you have some around?

A: I can get some.

B: Well, I like apples and pears a lot.

is not clear. So, the licensed implicature --BEL(B, *I like fruit other than apples and pears*) cannot be calculated simply by matching semantic representations of contiguous utterances. And for other utterances -- say 242 --

(242) Some people left early.

no P_j at all may be available for comparison. So, this procedure for locating salient expressions is of limited use.

Finally, although Joshi et al define discourse coherence in terms of forward- and backward-looking centers, they provide few insights on how to identify these centers (beyond the *post facto* 'what is pronominalizable') and no strategies for ordering forward-looking centers once they are found. So, although centering will allow A to identify salient expressions in B's response in 227, for example, it will not help B to choose the most salient expression from the forward-looking centers in A's query, '*the girl in the red dress, the girl, girl, the red dress, ... a diet coke ...*'. And, too, postulation of a single backward-looking center will not help us to identify scalar implicatures arising from the salience of more than one expression in an utterance.

6.3.1.3.3. Local Focus

In her computational account of LOCAL FOCUS, Sidner [Sidner 79, Sidner 83] presents a strategy for tracking the focussing process throughout a discourse and using information thus obtained to resolve anaphoric reference. Using syntactic features and thematic (case) roles of utterances, together with data provided by subsequent anaphoric reference, Sidner provides algorithms to hypothesize discourse and local foci, confirm or reject these hypotheses, and propose substitutions where necessary.

Noun-phrases in an initial utterance are preferred for the DEFAULT EXPECTED FOCUS LIST in the following order:¹⁴⁴

- the subject of a there-insertion clause or an 'isa' sentence;
- the THEME of the sentence, which Sidner defines as the verb relation that indicates the property of being affected by the action of the verb;¹⁴⁵
- the other case roles present in the sentence, with a weak preference for GOAL or for any role filled by an indefinite, with AGENT ordered last;

¹⁴⁴Sidner notes that prosody is not well enough understood to figure in these calculations.

¹⁴⁵Presumably, the patient in other thematic typologies.

- the sentence's VP, which may be specified by a nominalization of the verb.
- an item selected by this means may be rejected when unacceptable coreference is predicted after in the discourse and another member of the default focus list substituted as discourse focus.

Sidner's theory accommodates FOCUS SHIFT during the discourse to any term in a subsequent utterance. Potential foci for such shifts may be anticipated on the same basis as the choice of the default focus list: actual shift is confirmed or rejected by subsequent anaphoric reference. Provision is also made for a return to previous foci -- FOCUS POPPING -- in the discourse, through the maintenance of a FOCUS STACK. Potential referents are collected in a focus stack, created by ranking NP's from most to least likely to represent the 'focus of attention' in the discourse; This stack will be searched after the discourse focus and potential focus list.

Sidner's restriction of possible foci to noun- and verb-phrases, like [Josh 81, Gross 83]'s restriction of centering to noun-phrases, clearly represents a serious problem for our purposes. Since any expression in a sentence can represent a salient expression for the purpose of supporting scalar implicature, any constituent's semantic representation may serve as a salient expression. The commonly accepted notion that contrastive stress may mark focus provides independent support that any item can be focussed, as in 243, where UNDER seems obviously to be the item focussed.

(243) I left the book UNDER the table.

Similarly, in 229, identifying the sentence's focus with the NP

(244) I left the RED book under the table.

the *red* book does not capture the intuitive notion that it is not the entity but one of its attributes that is being attended to. Exclusion of all but NP's and VP's from consideration as foci thus limits the use of existing algorithms for the definition of salient expressions.

6.3.1.4. Communicative Dynamism

Since several expressions may be salient in a given sentence, an ideal focus algorithm for our purposes would also provide us with information about the degree to which items may be 'in focus' in a discourse. Work in COMMUNICATIVE DYNAMISM (CD) by the Prague School has long studied the 'distribution of various degrees of communicative dynamism (CD) over the elements within the sentence' [Firbas 65:170]. The degree of CD for any element is determined by the extent to that element contributes to the development of the communication and is related to the amount of 'givenness' of the element. For example, the lowest degree of CD of a sentence is associated with its 'theme' and the highest with its 'rheme'. Other elements are

associated with communicative 'transition' and carry an intermediate degree of CD. Again, this research has far to go before it can predict reliable orderings of focussed elements.

6.3.1.5. Defining Salient Expressions Using Focus

Limited though our current understanding of focus may be, research on this topic holds out hope of an (eventual) independent source of the salience information required to calculate scalar implicatures. Although I will neither propose a new theory of focus assignment nor commit myself to an existing focus algorithm, I will assume in the remainder of this work that some such algorithm is available to rank expected foci. In this section I suggest how focus information might be used to predict salient expressions for the generation and interpretation of scalar implicature.

Assuming the availability of information about can we make of knowledge about the degree of speaker/hearer attention each item is receiving in the discourse for the purpose of anticipating or interpreting scalar implicatures? An obvious approach is to say that the more in focus some item is the more likely it is figure in a scalar implicature. Say *S* wishes to convey a scalar implicature via evocation of some ordering *O*. Then, if *S* believes that *s/he* and *H* mutually believe that some e_i in *O* denotes an item that is already salient in the discourse, *S* will anticipate that *his/her* use of a sentence containing e_i to license a scalar implicature will be likely to be correctly interpreted by *H*. The justification for this claim is as follows: If some item v_i represented by e_i is salient in the discourse, it seems plausible to assume that items it is related to will also be salient, along with the metric by which these items are related to v_i . Kiefer's account of focussed yes-no questions supports this strategy. In effect, the focussed item is the item most likely to be mentioned in a subsequent utterance or 'replaced' with another value in a salient ordering. Of course, when an item is linked to a focussed item by functional dependence, a poset relation [Ward 85b], or some other relationship [Gross 77], most studies count this a form of focus maintenance.

An alternate strategy might also seem plausible: that *S* seeks to maintain discourse coherence by continuing to mention focussed items and by 'replacing' items not in focus by values in some ordering. However, this hypothesis seems intuitively less satisfying than the approach I have taken. Consider the oddness of B's response in (245a), when A has marked *George* as

- (245) A: Did GEORGE burn the cake?
 a. B: He burned the cookies.
 b. B: Harry did.

focus by stress. 146 (245b), on the other hand, seems much more appropriate. So, focussed items do seem to be salient items for the purposes of licensing or interpreting scalar implicature: that is, when S licenses some scalar implicature, that scalar implicature will either draw upon the salient of some item v_i already focussed in the discourse by mentioning some value in an ordering in which v_i appears, or S will focus attention upon some v_j in his/her utterance which she intends to license a scalar implicature.

A look at some naturally occurring discourse illustrates this strategy: In 140 (repeated here for convenience, with stressed items are in upper case), A first queries a lower value in an ordering

- (1) B: I as a result of my extensive investigations and especially on an international basis -- I was involved in industrial and governmental espionage so I had to be very careful about --
 A: CIA stuff?
 B: Pardon me?
 A: CIA stuff?
 B: We (PAST) HAVE worked WITH them.
 A: I see.
 B: I (PRES) don't work FOR them.
 A: I see.
 B: Uh I (PRES) work FOR my corporation.
 A: Uh huh.
 B: I (PAST) have also worked with security organizations, the FBI --
 A: Uh huh.
 B: And the intelligence groups. But what happened was I had to remain silent about my inheritance because uh we were threatened.

of industrial and governmental espionage employers. B first evokes tense and VP orderings, past present and work with work for by stressing them, to clarify her relationship with the focussed item CIA, while maintaining CIA as the topic of the discourse. She implicates -BEL(B, We now work for the CIA), as she states in the next utterance. B then refers to A's ordering of industrial and governmental espionage employers by mentioning my corporation. In the same utterance, however, she also refers to the previously evoked tense and prepositional orderings. In her next utterance, B refers to other values in the same three orderings, mentioning tense, work with, and security organizations in general and the FBI in particular. In this utterance she implicates -BEL(B, I now work for security organizations) and -BEL(B, I now work for the FBI).

Without a more reliable independent test of focus, it is impossible to test this association

149] use stress here to mark focus since it seems the least ambiguous method.

between focus and salient expressions more objectively. Tentatively, however, I will propose that, given any focus algorithm, the ordering of salient expressions for the purpose of inferring or anticipating the inference of scalar implicatures will be the order of focussed items -- from most to least in focus. However, given an ordering of possible foci, we must then determine the point at which items become 'too little in focus' to be counted as salient. This remains an open question.

Clearly, however, if some expression e_i is marked as salient by some focus algorithm but if S cannot discern or does not believe that H can discern any ordering in which that expression occurs, S will not make reference to e_i to license a scalar implicature. Alternatively, if more than one such ordering in which e_i appears exists for S -- and may be assumed discernible by H -- S will need to choose that ordering which H is most likely to discern to ensure that his/her implicature is properly understood -- i.e., S must select the most salient ordering in which e_i appears.

6.3.2. Identifying Salient Orderings

Although the salience of relations in discourse has been less studied than the salience of entities and attributes, [Grosz and Sidner 85, Conklin and McDonald 82] there has been considerable discussion of those relations which may hold among propositions in the discourse to account for discourse structure. [Hobbs 77, McKeown 82, Mann and Thompson 83, Mann 84] However, these authors have been more concerned with identifying the range of possible relations between discourse units than with how speakers select from this range. Even if salient expressions can be identified in the discourse, knowledge of which relations are also salient will be needed to select among possible orderings for salient expressions. In addition, where independent information about salient expressions is ambiguous or unavailable, salient orderings may themselves point to salient expressions: That is, if an ordering O_i is salient in the discourse and if some expression e_j appears both in the current sentence and in O_i , the salience of O_i might indicate the salience of e_j . Thus, the salience information necessary for the generation and interpretation of scalar implicature might be acquired by looking first for a salient expression or by first locating a salient poser.

It should be evident from examples such as 218 and 219, 222, 225, and other exchanges presented above that a given expression may evoke a number of different orderings. A cooperative S should seek not only an ordering that can be identified by H -- and defined as S defines it -- but the most salient such ordering for the particular exchange. Of course, S may instead attempt to make an ordering salient for H which s/he does not believe is already salient for H -- by evoking it via mention of some value on it -- but only if s/he thinks H will recognize both the ordering and the process by which s/he has evoked it.

While the need for S to reason about which ordering *H* may perceive as salient for a given expression is easiest to see where domain-dependent orderings are involved, even expressions used to evoke the canonical, domain-independent orderings like the modals may instead evoke some other ordering, as in the following hypothetical exchanges between a hospital visitor and patient in 246:

- (246) A: Can you get out of bed?
 a. B: I'm supposed to at least once a day.
 b. B: I can sit up.

Since *can* may have either a modal or an 'ability' reading, it may evoke a deontic ordering ((246a)) or an ability ordering ((246b)).

Temative and partial answers to the question of how one locates the most salient relation for a particular expression in a particular sentence may nonetheless be sought in this domain independent/ dependent distinction as well in information about which orderings have been salient previously in the discourse, in cognitive science studies of human perception of relations, and in clues that particular metrics provide as to the salience of relations they order.

6.3.2.1. Selecting Salient Classification Hierarchies

To date, the major studies of human choice among potentially salient ordering relations have focussed upon children's ability to classify. [Inhelder and Piaget 64, Rosch 76, Markman 83, Gelman and Bailargeon 83, Sugarman 83, Waxman 85] That is, they have been mainly concerned with identifying the development of children's ability to classify taxonomically as opposed to imposing thematic or idiosyncratic organizations upon groups of items. However, one segment of literature on adult classification does suggest how to identify salient classification hierarchies when some such hierarchy is found to be salient.

Scalar implicatures licensed via a classification ordering present peculiar problems for the selection of a salient ordering, since classification hierarchies are potentially infinite. That is, we might represent the following classification of pets finitely, as below:

| | | | |
|-------------|------------|-----|-----|
| lilac point | blue point | | |
| Siamese | Burmese | | |
| | | cat | |
| | | | dog |
| | | | |
| | | | pet |

But, for any classification hierarchy, the most specific levels in that hierarchy may always be

further specified, even if that specification is somewhat idiosyncratic, as say, *round-faced lilac points*, *round-faced lilac points with double paws*, and so on.

While it is clear that B may license the implicature \neg BEL(B, *I have a Siamese*) by his/her response in 247, it is not clear that we want to say all the (potentially infinite) higher values in this and similar hierarchies will also be the subject of implicatures via this utterance -- e.g., \neg BEL(B, *I have a lilac-point Siamese*) and \neg BEL(B, *I have a round-faced lilac-point Siamese*).

- (247) A: Do you have a Siamese?
 B: I have a cat.
 And, in 248, we would not want to say that B's response licenses

- (248) A: Do you have a pet?
 B: I have a cat.

the implicature \neg BEL(B, *I have a Siamese*), although *Siamese* is a higher value in the hierarchy than the mentioned *cat* and the ordering is clearly salient in this exchange. Also, in 249, B's response may indeed license \neg BEL(B, *I have a Siamese*), just as,

- (249) A: Do you have a Siamese?
 B: I have a pet.

in 250, B's affirmation of *pet* may implicate \neg BEL(B, *I have a cat*).

- (250) A: Do you have a cat?
 B: I have a pet.

But it seems odd to say that B may implicate \neg BEL(B, *I have a lilac-point Siamese*) in either 249 or 250. How then do we explain the limits on scalar implicature licensed in 247-250? A classification hierarchy is clearly salient in these, but it does not appear to be the same in each. And for none of these responses would we want to say that the salient ordering is an infinite one -- but how can we bound such orderings in a principled way?

6.3.2.1.1. Basic and Entry Levels Defined

Studies of human classification have long debated whether or not certain levels of abstraction in classification hierarchies have particular psychological significance. BASIC categories have been defined by Rosch [Rosch 73, Rosch 76] as those categories in classification hierarchies which carry the most information; are most differentiated from other members of the hierarchy at their level; are the most inclusive categories whose members have a significant number of attributes in common; have similar motor programs and similar shapes; and can be most easily identified from the averaged shapes of members of the class. While Rosch's first examples were of color and form, her findings were soon generalized to biological classifications and to hierarchies such as *food* and *furniture*.

Rosch was particularly interested in identifying prototypical members of these classes.

Empirical studies lead her in [Rosch 75] to adopt Wittgenstein's 'family resemblance' notion, that prototypical members of categories are those which possess the most attributes in common with other members of their category and the fewest in common with members of other categories. So, robins are prototypical birds because they exhibit a high proportion of those attributes common to all birds and fail to exhibit attributes which might link them with competing categories, such as *fish* or *reptile*. Bats are atypical exemplars of the category *bird*, since, while they have wings, they have fur instead of feathers. Thus they share important attributes of the class *mammal*.

In a standard classification hierarchy entities are assigned to one of four levels, ranked on the basis of decreasing generality, from SUPERORDINATE to INTERMEDIATE to BASIC to SUBORDINATE level. An example of such a hierarchy is:



Intermediate levels are often difficult to determine. The basic category *dog* in this hierarchy, like other basic categories, can be identified as 'basic' by a number of generally convergent tests.¹⁴⁷ For example, a basic category often corresponds to the highest-frequency item of the hierarchy in vocabulary counts -- as does *dog*. Also, the word for the basic category tends to be earned earlier than words expressing other levels. [Anglin 79] And this category is often the most abstract member of its hierarchy that is monomorphemically lexicalized, as is *dog* or *chair* in the hierarchy *furniture/ things to sit on/ chair/ armchair*. For some domains, such as color and shape, it has even been argued that this latter feature is universal: for example, few or no languages have a single word for 'a washed-out pinkish-red' or have polymorphemic 'red'. [Berlin 69]

Basic categories are also natural candidates for ostensive introduction: [Fodor 81] that is, one cannot teach *poodle* ostensively to child who has not already been taught *dog*, and similarly with *pale red* and *red*. Basic categories are typically the most abstract members of their hierarchies which differentiate individuals of approximately similar appearance [Rosch 76] While subjects can draw a dog, for example, they cannot draw an animal; similarly for *chair*

¹⁴⁷Where particular basic categories are identified in examples in this section, they have been verified by (others') empirical research along the lines presented below.

and *piece of furniture*. Objects belonging to same basic category have a greater degree of shape overlap (when considering canonical depictions) than do objects belonging to the same superordinate category, but not to the same basic category; and objects belonging to the same basic category do not have significantly less overlap than do objects belonging to the same subordinate-level category.¹⁴⁸ There is also considerable evidence that perceptual identifications which involve access of basic categories are faster than those involving 'non-basic categories'. [Trentham 81] For example, when subjects were asked to find a picture of a woman walking a dog they accomplish this task more rapidly than when they are asked to find a picture of a woman walking a poodle. Subjects can also name objects or match names with pictures faster at basic level [Brownell 78, Hutcheon 70, Rosch 76, Segui 68, Smith 78] These results suggest that basic categorizations may be made on the basis of the visual properties of objects.

Basic categories also yield INFORMATION PEAKS when subjects are asked to list properties of categories: They can list more properties of basic categories than of more abstract categories and very few more properties for more specific categories than for the basic level. Also, more subjects tend to list the same attributes of basic categories than they do of other levels. Subjects were also asked to describe motor behaviors that they would emit in the presence of specific objects; more common motor behaviors were listed for basic categories than for other level categories. It has been claimed that basic categories encode the most information per unit of judgment [Fodor 83]; thus, by Grice's maxims of Quantity and Manner, this observation predicts that, *ceteris paribus*, basic categories are the natural ones to use for describing things. [Fodor 83] claims that the basic level thus represents the preferred or default level of description even where more information is known -- unless special circumstances indicate otherwise. That is, basic categories are phenomenologically given, providing the natural level for describing things to oneself as well as to others.¹⁴⁹

While accepting the basic notion of a psychologically privileged level, [Jolicoeur 84] contends that a more sophisticated notion of ENTRY LEVEL is required to accommodate subjects' responses to some of the tests of this level when atypical members of basic categories

¹⁴⁸When subjects were presented with drawings created by averaging the outlines of two shapes and asked to name the category of the object, they could perform the task when shapes were drawn from two members of a basic category or from two members of a subordinate category, but had great difficulty when shapes of different basic categories belonging to a given superordinate were averaged; e.g., it was hard to identify the average of apple and banana as fruit.

¹⁴⁹Of course, the basic level for particular hierarchies may be influenced by culture and individual experience. So, if I have grown up in a fishing village, fish -- generally considered an intermediate category with *mammal* and *reptile* -- may be a basic category for me.

are involved. Their empirical studies reveal that, for atypical members of a basic category (as for *chicken* an atypical member of the basic category *bird*) the atypical subordinate (e.g., *chicken*) will pass the tests for 'basic level' discussed above. That is, the 'entry' level to a classification hierarchy will be a function both of an item's position in the classification hierarchy and of its typicality at that level. Atypical members of a basic category tend to have their privileged or entry level at the subordinate level, as in (251a), while typical members have

(251) A: Did you see any birds?

- a. I saw a chicken.
b. I saw a robin.

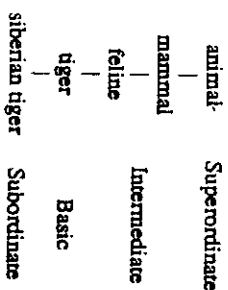
their entry level at the basic level, as in (251b). That is, while accepting the notion of a privileged level in classification hierarchies, (Jolicoeur 84) find that, for atypical members of a basic category, the diagnostics for basic level such as those described above are actually satisfied more often at the subordinate level.

6.3.2.1.2. Bounding Hierarchies by Entry Level

Now, returning to the original problems identified in 247-250, I propose to account for the apparent bounding of the classification hierarchies salient in these exchanges in terms of a hierarchy's entry level. In the pet hierarchy illustrated above, *pet* will represent the superordinate level, *cat* the basic level, and *Siamese* the subordinate level, with *Ilac* point a member of some sub-subordinate level. *Siamese* is not an atypical member of the type *cat*, so the entry level for this hierarchy should be its basic level, *cat*.

Note that, while items more specific than the entry level for the pet hierarchy may indeed be mentioned to license scalar implicature, where all other things are equal the entry level will pose a default bound upon licensed implicatures. So, when a classification metric is salient, we can normally take the resultant salient ordering to be bounded from (above in my representation but below in standard classification hierarchies) by the hierarchy's entry level. So, in 248, for example, B's response will not license —BEL(B, *I have a Siamese*) because this level will not be a part of the exchange's salient classification hierarchy; that is, while B's affirmation of *cat* will license his/her lack of commitment of higher values, *Siamese* will not be a higher value in a 'pet' classification hierarchy bounded (for *Siamese*) by the basic level *cat*. Of course, if some more specific level of the hierarchy is mentioned, a larger ordering must of course be taken as salient. In 247 and 249, thus, the salient ordering will include the subordinate level. However, more specific levels will be added to the salient hierarchy only if they, or still more specific levels, have been explicitly evoked by mention of a member — as, *Siamese*. So, in such cases, scalar implicatures will be licensed only up to and including the most specific evoked level of a classification hierarchy. B's response in 247, for example, will license only the implicature —BEL(B, *I have a Siamese*). In effect, then, scalar implicatures will not be licensed 'above' (at a more specific level than) the basic level or above a more specific evoked level. Some further examples should illustrate these findings.

Say we propose a path in a classification hierarchy as follows:



where the levels are as indicated, with an intermediate level at either *feline* or *mammal*. Then, if B responds to A's query in 252 with (252a), *s/he* can implicate —BEL(B, (252b)), since *tiger* represents this hierarchy's basic level and is a higher value in a type/ subtype ordering than *feline*. Since *tiger* represents a

(252) A: Was that an animal over there?

- a. B: It was something feline.
b. B: It was a tiger.
c. B: It was a Siberian tiger.

typical animal, then it seems reasonable that B has considered this and other typical animals but is not able to affirm some particular one of them. But, if B responds with (252b), it seems unlikely that *s/he* intends to implicate that the animal is not a *Siberian tiger* or some other more specific variety. That is, it seems that the entry level of a classification hierarchy establishes its 'appropriate level of detail' — not only for simple description but for implicit meanings such as scalar implicature. Of course, B may indeed choose to respond at a more specific level than the entry level, as in (252c). Now, however, while such a response may alter the 'default' appropriate level of detail — in this case to the subordinate level — no scalar implicatures will be drawn beyond this mentioned level. That is, A will not infer that B cannot commit him/herself to some more specific description of the Siberian tiger, such as *blue-eyed Siberian tiger*.

I propose that a classification hierarchy's entry level will establish a default boundary beyond which scalar implicatures will not be drawn — in effect, creating finite type/ subtype orderings from potentially infinite ones. While S may evoke a larger ordering by explicit mention of a more specific level, this mention will only serve to raise the default boundary to the mentioned level.

S may alter this default appropriate level of detail in (at least) one other way: Barcelona [Barcelona] notes that, when S says 253,

(253) I'm thinking of something and it's not a dog.

H will infer that *S* is thinking of 'car' say, over choices like 'parkette' or 'hamburger'. We might account for this preference by noting that *car* is intuitively 'at the same level' as *dog* in a classification hierarchy. Furthermore, we could interpret this licensed inference and those licensed by similar utterances as scalar implicatures: In denying a value in a classification hierarchy (or, type/ subtype ordering), *S* may license -BEL(*S*, -*alternate values*), by Imp₂, where *dog* and *car* are alternate values. This observation suggests another way by which *S* may modify the default salient level of a classification hierarchy. *Car* and *dog* are both basic categories, and, thus, an exchange like 254 will license no implicatures about values more specific than *dog* in the pet

(254) A: Do you own a dog?

B: I have a car.

hierarchy. However, when *S* mentions a value in a classification hierarchy which is alternate to a value already salient in the discourse - i.e., a value at the same level in that hierarchy, even when that level is not the entry level for that path in the hierarchy, it appears that *S* may set the default level of the hierarchy at that level.

Consider, for example, the following hierarchy of household goods, from the superordinates *furniture* and *cookware* down to the subordinate *armchair*.

| | | | |
|----------|------------------|-----------------|------------------|
| | furniture | household goods | |
| | things to sit on | | cookware |
| | | | baking equipment |
| chair | | bench | |
| armchair | | | |

exchange in 255. While *cookware* is not at the basic level for

(255) A: Does Jane need furniture?

B: Cookware.

this hierarchy, its mention does not seem to license scalar implicatures about more specific (higher) members of the hierarchy, such as *baking equipment*. Yet, clearly, part of this hierarchy is salient, since B can implicate -BEL(B, *Jane needs furniture*) by the response. So, it seems that *S* may alter the default appropriate level of detail by reinforcing a more general level previously salient in the discourse. In 255, A has made the superordinate level salient by his/her mention of *furniture*, and B responds at the superordinate level with *cookware*. In such a case, the reinforced level will represent the salient level for the discourse, and the hierarchy will be bounded at that level.

In sum, there are at least two ways *S* may alter the default level of a classification hierarchy and thus bound these potentially infinite orderings for the interpretation of scalar implicatures. The default boundary for any classification hierarchy is the entry level for any path in the hierarchy. *S* may reset this default either by mentioning a more specific item than the entry level or by mentioning a less specific item at the same level (alternate to) as an item previously salient in the discourse. Whatever action *S* chooses to take, scalar implicatures will not be licensed beyond (higher or more specific than) this revised level.

These observations help determine how to define the portion of a classification hierarchy that will be salient in an exchange - given that some classification hierarchies may generally be examples (See Section 6.3.1.3.1) suggest that classification hierarchies may generally be assumed salient, since mention of some value in them imparts 'givenness' to other values. That is, *S* may anticipate that, if *H* does not already view such orderings as salient, *s/he* will be able to recognize their evocation. So, *S* may anticipate that *H* will be able to work out implicatures which rely for their interpretation upon perception of a classification hierarchy (type/ subtype relationship).¹⁵⁰

6.3.2.1.3. Extending 'Entry Level' Beyond Classification

While no similar investigation of human perception of other ordering relations has been made, it does seem reasonable that other orderings too might have their 'appropriate level of detail'. The general problem has been recognized in the literature. In discussing his example of clash between the maxims of Quantity and Quality (illustrated in 256, Grice observes that B's response is clearly not appropriately detailed, although it is as detailed as B can truthfully make it: Recall that A is planning his French holiday. B knows A wants to visit C.

(256) A: Where does C live?

a. B: Somewhere in the south of France.

b. B: In Marseille.

c. B: On Rue de Janvier in Marseille.

d. B: In a tiny one-bedroom apartment in a three-story brick building on Rue de Janvier in Marseille.

By saying (256a), Grice claims that B implicates *s/he* does not know which town C lives in. For, if *s/he* did, *s/he* should have mentioned it, since A will need that information to locate C. The question is, if B had responded (256b), could she have implicated -BEL(B, (256c))? Or, if she had responded with (256c), could she have implicated -BEL(B, (256d))? So, again, we have potentially infinite hierarchies - in this case defined in terms of specificity of information. However, it is less clear how to assign *a priori* to any such hierarchy a privileged

¹⁵⁰It is also interesting to note that, if licensable scalar implicatures can be associated with entry levels in the way described above, scalar implicature may provide yet another test for entry level in classification hierarchies.

level of informativeness. Consider how the 'appropriate level of detail' in 256 will change if, instead of planning to visit C, A is merely inquiring as to what has become of his/her old friend. So, when generalization/ specialization relationships are salient, while it does seem both appropriate and even essential to assign some 'entry level' into the hierarchy, it seems less easy to assume that this level will be independent of context.

3.2.2. Salient Orderings from Speakers' Goals

Linguistic and AI research on speech acts, planning, and goal inference has convincingly argued that speakers' goals can determine what is relevant or salient in a discourse. [Morgan 8, Hobbs 79b, Allen 80, Pollack 86] For example, depending upon the goals of the exchanges in 257, B might choose to evoke one of several possible orderings.

(257) A: Are the Yankees in town this week?

- a. B: Giddy is.
- b. B: The Mets are.
- c. B: They're in town next week.

Each of the responses in 257 will be appropriate in the corresponding contexts in 258.

(258)

- a. A: I need to interview a pitcher about this new baseball rule. Are the Yankees in town this week?
- b. A: I want to take some clients to a game this afternoon. Are the Yankees in town this week?
- c. A: Maybe we can see Henderson break the record. Are the Yankees in town this week?

If A's goal is *interviewing a pitcher*, then the subset of Yankee pitchers will be salient after his/her query. If A's goal is finding a ball game to take clients to, then the set of baseball clubs in New York will be salient. And, if A's goal is to see X break a particular record, then a temporal scale will be salient. Although it is easy to analyze these exchanges 'after the fact', however, it is not clear how S goals can be translated into salience information in general -- and no information about salient items and relationships in particular.

[Gross 77, Hobbs 79a] propose that, in task-oriented domains, the stages of the process by which that task can be accomplished will be relevant, since S's goal is to accomplish a process in order to accomplish a task. If this assumption is correct, then we might propose that process orderings may be supposed to represent salient relations in such domains. So, in Hobbs 79a]'s example (reproduced in 259), where the

(259) A: Have you disconnected the air line?

B: I loosened it.

ask is to assemble an air compressor, *disconnecting the air line* is a relevant process since B

must accomplish it in order to accomplish the task as a whole. So, we can say that stages in this process are salient here and, thus, B can anticipate this recognition on A's part when A interprets the implicature B licenses that *-BEL(B, I disconnected it)*. Of course, even in non-task-oriented domains, such implicatures may also be generated and understood, so this explanation is partial at best.

Pollack [Pollack 86] proposes a more general model of plan inference which also permits simple location of some types of salient relations for our purposes. Distinguishing GENERATION relations between actions from ENABLEMENT relations, following studies in the philosophy of action, Pollack defines speaker goals in terms of plans which are built up from such relations. In simple terms, when one action α generates another action β , then an agent need only do α and β will automatically be done also. So 'turning on the light' by 'flipping the switch' represents a generation relation between the mentioned actions; 'flipping the switch' generates 'turning on the light'. When α enables β , then an agent needs to do something more than α to guarantee that β will be done. 'Knowing where the switch is' enables 'turning on the light by flipping the switch'. The plans Pollack infers are in effect partial orderings of actions related by generation or enablement. Thus they support scalar implicature, as in 260. Implicatures based on similar orderings

(260) A: Did you turn on the light?

B: I flipped the switch.

have been discussed in Section 5.1.9 as implicatures deriving from the perception of process orderings or prerequisite orderings as salient. Since plans inferred from a discourse can plausibly be assumed salient for S and H in that discourse, these plans will thus represent salient relations for our purposes.

6.3.2.3. Selecting Between Duals

A problem in discerning salient relations by identifying salient expressions which refer to them involves differentiating between duals, or, posers Q_1 and Q_2 which differ from one another only in that, where the metric Q_1 orders any pair of expressions e_1 and e_2 as $e_1 < e_2$, the metric Q_2 orders e_1 and e_2 as $e_2 < e_1$. This is the notion discussed in Section 5.1.4.1.3 that some orderings appear to 'reverse' other orderings defined over the same set of expressions.

Consider orderings of temperature degrees, ordered by either *is-colder-than* or its dual, *is-warmer-than*. Each mention of a temperature may thus evoke at least two posers. Suppose that, after a cold winter the weatherman has predicted relief from the bitter cold, B returns from getting the paper to be greeted by A, who is just getting up.

(261) A: Is it warm outside?

B: It's no 60 degrees out there.

Here B appears to license the implicature $\neg\text{BEL}(B, \text{it's warm})$; so, the salient temperature ordering for this exchange is apparently .../ 50/ 60/ 70/ ..., ordered by *is-colder-than*. By denying 60 degrees, B appears to confirm some lower value on the scale. Alternatively, after a blazing summer the weatherman has promised relief from the heat, B's response in 262 appears to license the implicature $\neg\text{BEL}(B, \text{it's cool})$ by implicitly conveying commitment to some lower value

(262) A: Is it cool outside?

B: It's no 60 degrees out there.

on a 'reversed' temperature defined by *is-warmer-than*, .../ 70/ 60/ 50/ ...

In cases such as these we might propose some appropriate 'script' to explain how A and B understand why one scale is appropriate in one context and another in another. Shared knowledge of a prior weather forecast or of the reliability of weather forecasters in general might prove useful in this strategy. However, a strategy less dependent upon such broad contextual cues is also possible: In these and similar discourses, it appears that prior salience (explicit mention in 261 and 262) of a value in the set of referents under consideration may indicate the orientation of the salient ordering. When some such value e_i is salient in prior discourse, e_i may establish the orientation of an ordering in which it appears (or which includes values e_j can characterize) in that the pole closest to e_i (or values associated with it) will be the positive pole of the salient ordering. So, in 261, mention of *warm* favors an ordering which places 'warmer' values at the positive pole, while, in 262, mention of *cool* favors the dual to this ordering.

In some of the naturally occurring data examined for this work an interesting phenomenon emerged: Speakers referred alternately to one poset and then to its dual, apparently to negotiate a range of values which both could agree to. For example, in 263, A first denies a value in

(263) A: It's not cool outside.

B: Well, it's not warm.

a temperature ordering *hot/warm/tepid/cool/cold* to license $\neg\text{BEL}(A, \neg\text{it's } X \text{ outside})$ where $X < \text{cool}$. Then, B denies a value in *cold/cool/tepid/warm/hot* to license $\neg\text{BEL}(B, \neg\text{it's } Y \text{ outside})$ for all $Y < \text{warm}$. In effect, A's and B's implicatures agree only at the value *tepid* – the intersection of their implicatures. Similar exchanges occur for modifiers like *young/old, bad/good, and happy/sad*.

The problems of selecting salient orderings should by now appear considerable – and the solutions proposed tentative and partial. While much remains to be explored in the

incorporation of 'salience' into the calculation of scalar implicature, some progress has been made. First, as noted in Section 6.2, it is possible to relax the condition that S and H identify the same ordering as salient and demand only that orderings be compatible. So, where several orderings are potentially salient for some expression, it may not always be necessary to select the 'most salient', so long as S believes all orderings H might see as salient to be compatible with S 's ordering. Second, it should be possible to use research conclusions from studies of focus to identify salient expressions in an utterance, even if, to date, no satisfactory focus algorithm is available. As we learn more about 'what is attended to', we will be better able to test the claim made here that focussed items identify salient expressions for the purpose of generating and interpreting scalar implicature. Finally, while even less is known about the focussing of relationships in discourse than about focussed entities, salient orderings may be identifiable in other ways: For some orderings metrics, it is possible to choose from among several candidate orderings on the basis of properties particular to the defining metric: For classification hierarchies, we can propose a principled mechanism for bounding hierarchies. Where the choice is between dual orderings, we can propose a principled way to select one over another. Knowledge of the domain or of the intentions of conversational participants also may provide clues to the salience of orderings.

While these observations are very far from a comprehensive account of how salient relations may be identified by S and H , they do suggest fruitful avenues for future research. For the implementation described in Chapter 7, salient expressions and orderings are identified interactively.

6.4. Calculating Scalar Implicatures

In contrast to Gazdar's two-stage calculation of implicature – first potential, then actual – I have proposed that actual scalar implicatures be calculated directly – but that this calculation be based upon additional information about speaker and hearer beliefs and about the discourse context. Given an utterance, information about speaker intention and mutual speaker-hearer belief about speaker cooperativity, together with contextual information about salient entities and relationships in the discourse, we can use the definition of S 's conversational implicature and of H 's inference of conversational implicature presented in Chapter 2, the scalar implicature conventions discussed in Chapter 3.3 and refined in Chapters 4, the definitions of utterance ranking presented in Chapter 5, and the conditions on ordering compatibility and ordering salience presented above to compute those scalar implicatures a speaker licenses as well as to calculate inferences which a hearer is entitled to draw.

Recall from Chapter 2 that p_j can be said to represent a conversational implicature of S in a context C_a iff:

- S intends to cause H to believe p_j by an utterance u_i (Condition 1);
- S believes that his/her own cooperativity in C_H is a mutual belief of S and H (Condition 2);
- S belief that s/he and H mutually believe that the saying of u_i in C_H must license H's belief in p_j given belief in S cooperativity (Condition 3);
- p_j is cancelable (Condition 4), non-detachable (Condition 5), and reinforceable (Condition 6).

Recall also that for scalar implicature, Condition 3 can be satisfied by the satisfaction of the LICENSE predicate, which can be accomplished by S cooperativity plus the truth of any of the scalar implicature conventions introduced in Chapter 3.3 (and modified in Chapters 4 and above). For convenience, I will repeat these here:

- Imp₁*: $3O \text{ BEL}(S, \text{AVAIL}(H, e_i, O, C_H)) \wedge \text{REALIZE}(u_i, \text{AFFIRM}(S, e_i, \text{BEL}(S, p_j))) \wedge (\text{HIGHER_SENT}(p_i, p_i, O) \vee \text{ALT_SENT}(p_i, p_i, O))$
 $\Rightarrow \text{SCALAR_IMP}(S, H, u_i, \neg \text{BEL}(S, p_j), C_H)$
- Imp₂*: $3O \text{ BEL}(S, \text{AVAIL}(H, e_i, O, C_H)) \wedge \text{REALIZE}(u_i, \text{DENIAL}(S, e_i, \text{BEL}(S, \neg p_j))) \wedge (\text{LOWER_SENT}(p_i, p_i, O) \vee \text{ALT_SENT}(p_i, p_i, O))$
 $\Rightarrow \text{SCALAR_IMP}(S, H, u_i, \neg \text{BEL}(S, \neg p_j), C_H)$
- Imp₃*: $3O \text{ BEL}(S, \text{AVAIL}(H, e_i, O, C_H)) \wedge \text{REALIZE}(u_i, \text{IGN}(S, e_i, \neg \text{BEL}(S, p_j))) \wedge (\text{LOWER_SENT}(p_i, p_i, O) \Rightarrow \text{SCALAR_IMP}(S, H, u_i, \neg \text{BEL}(S, \neg p_j), C_H))$

\vee
 $(\text{HIGHER_SENT}(p_i, p_i, O) \Rightarrow \text{SCALAR_IMP}(S, H, u_i, \neg \text{BEL}(S, p_j), C_H)) \vee$
 $(\text{ALT_SENT}(p_i, p_i, O) \Rightarrow \text{SCALAR_IMP}(S, H, u_i, \text{BEL}(S, p_j), C_H))$

Then, a scalar implicature p_j can be said to be licensed by S's utterance of u_i iff S is known or assumed to intend to convey p_j , S and H mutually believe S to be cooperative in uttering u_i , if S is obeying the maxims of QUANTITY and QUALITY, if one of the scalar implicature conventions holds, and if the meaning licensed is cancelable, nondetachable, and reinforceable. One of the scalar implicature conventions will hold when, roughly speaking, p_j represents a higher, lower or alternate sentence with respect to the p_i realized by u_i , via the presence in p_i and p_j of expressions e_i and e_j which appear in some ordering O believed by S to be salient in the discourse and compatible with all the possible orderings S believes H might find salient. The following algorithm encodes the process by which the scalar implicature conventions can be tested for a given utterance:

- Given S's u_i with the semantic representation P_i :
1. decide whether any e_i in P_i evokes a salient O_i ;
 2. determine whether P_i represents the affirmation, denial, or declaration of S's ignorance about some proposition with semantic representation P_k which contains

- e_i if P_k is simple with respect to e_i ;¹⁵¹
3. create an open sentence P_x by replacing e_i in P_k with x ;
 4. locate other values e_j appearing with e_i in O_i ;
 5. for each e_j create a sentence P_j by replacing x with e_j in P_x -- so that P_j and P_k are simple expression alternatives (See Section 4.1.3.);
 6. determine whether P_j satisfies any of the scalar implicature conventions.

Returning to examples 218-221 (repeated here for convenience), we can now account for licensed implicatures in terms of this algorithm. Recall that, in 218,

- (1) A: Have you, me, and Ellen ever had dinner together?
 B: We've had lunch.

assuming B's intentions and cooperativity, we can say that $\neg \text{BEL}(B, \text{we have had dinner together})$ is a scalar implicature of B's response, so long as B believes an ordering defined by set inclusion on the set of meals (= {breakfast, lunch, dinner}) is salient and that A will recognize either this ordering or one compatible with it -- say, perhaps just the explicitly mentioned items {lunch, dinner}. Since *lunch* and *dinner* represent proper (singleton) subsets of this set (See Note 130.), they represent alternate values in a salient ordering. By *Imp₁*, affirmation of some value v_i (denoted by an expression e_i) may license the belief that alternate values v_j are false or unknown. So, the queried *dinner* -- and other alternates such as *breakfast* -- can be implicitly marked as false or unknown in this way.

In terms of the algorithm presented above, this implicature might be analyzed as follows: Let P_k the representation of B's response, be 'BEL(PastCard(B, A, Ellen), {lunch}), $\{v_0\}$ '.¹⁵² In this sentence, the expression *lunch* evokes a salient ordering of meals, so $e_i = \text{lunch}$ and $O = \text{the set of meals}$. P_i represents the affirmation by B of P_k . 'PastCard(B, A, Ellen), {lunch}), $\{v_0\}$ ' contains e_i . We can create the open sentence 'PastCard(B, A, Ellen), X), $\{v_0\}$ ' from P_k by replacing e_i with a variable X. Then for all alternate values in the above representation -- {breakfast} and {dinner} -- we can create simple sentence alternatives 'PastCard(B, A, Ellen), {breakfast}), $\{v_0\}$ ' and 'PastCard(B, A, Ellen), {dinner}), $\{v_0\}$ '. Since these instantiations of X do represent alternate values in the set of meals, by the scalar implicature conventions, we can

¹⁵¹ L_{e_i} contains no negation with scope over e_i . See Section 4.1.3.

¹⁵²The algorithm is independent of any particular representation. For this example I will employ a higher order modal logic which permits representation of tense (here, 'Past(x, y)' indicates that x held at some time before time y and quantification over sets. The latter is necessary here to provide a dummy solution to the well-known problem of representing *together*. That is, utterances such as *Jane and Bill are eating lunch together* are not properly represented by 'eat(lunch, jane) & eat(lunch, bill)'.

culate that $-BEL(B, \text{Past}(\text{eat}(B, A, \text{Ellen})), \{\text{breakfast}\}, \{q_0\})$ $-BEL(B, \text{Past}(\text{eat}(B, A, \text{Ellen})), \{\text{dinner}\}, \{q_0\})$ may be licensed in 218.

In 219, however, B perceives an ordering of *social events* to be salient. So, we might calculate this response by representing P_2 as

- (1) A: Have you, me, and Ellen ever had dinner together?
B: We've been to the movies.

Let P_2 be $\text{Past}(X(B, A, \text{Ellen}), Y, \{q_0\})$ and implicatures e_i of the set $\{X(B, A, \text{Ellen}), Y\} \in \text{social events}$ that are alternate values to $\text{got}(B, A, \text{Ellen}, \{\text{movies}\})$.

Finally, in 220, where the ordering supporting B's implicature is defined over the set of *you, me, Ellen*, the implicature $-BEL(B, \text{you, me, and Ellen have had dinner together})$ may be calculated as follows:

- (1) A: Have you, me, and Ellen ever had dinner together?
B: Ellen and I have.

Let P_2 be $\text{Past}(\text{eat}(B, \text{Ellen}), \{\text{dinner}\}, \{q_0\})$ and P_2 be $\text{Past}(\text{eat}(X, \{\text{dinner}\}), \{q_0\})$. Then, for e_i that are alternate expressions to the set $\{\text{Ellen}, B\}$ or that are higher expressions in the licit ordering, B may license the implicature $-BEL(B, \text{Past}(\text{eat}(e_i, \{\text{dinner}\}), \{q_0\}))$. Hence, for a higher expression $\{B, A, \text{Ellen}\}$, B licenses $-BEL(B, \text{Past}(\text{eat}(B, A, \text{Ellen}), \{\text{dinner}\}), \{q_0\})$.

So, given an utterance, an expression from its semantic representation, and a salient ordering, it is a fairly straightforward task to calculate the scalar implicatures which may be licensed via this utterance. However, the algorithm presented above will not accommodate the calculation of scalar implicatures when more than one expression is salient in this utterance, as in 21 (repeated below).

- (1) A: Have you, me, and Ellen ever had dinner together?
B: Ellen and I have had lunch.

In this exchange, orderings defined over the subsets of $\{\text{you, me, Ellen}\}$ and $\{\text{breakfast, lunch, dinner}\}$ are both evoked by B's response to implicate $-BEL(B, \text{you, me, and Ellen have had dinner together})$. While P_2 is $\text{Past}(\text{eat}(B, \text{Ellen}), \{\text{lunch}\}, \{q_0\})$ and P_2 is $\text{Past}(\text{eat}(X, Y), \{q_0\})$, it is not clear how licensed implicatures should be represented -- or even just what they are. Intuitively, B seems to be licensing $-BEL(B, \text{Past}(\text{eat}(B, A, \text{Ellen}), \{\text{dinner}\}), \{q_0\})$, but does he also license $-BEL(B, \text{Past}(\text{eat}(B, A, \text{Ellen}), \{\text{breakfast}\}), \{q_0\})$? $-BEL(B, \text{Past}(\text{eat}(B, A, \text{Ellen}), \{\text{dinner}\}), \{q_0\})$? These more complex implicatures will require some revision of the algorithm presented above.

6.4.1 Multiple Salient Expressions

When several expressions in a sentence are salient, the calculation of potential and actual implicatures is much more complex than has previously been recognized. A major weakness of Gazdar's method for calculating potential implicature is that implicatures are calculated separately for each expression recognized in a sentence. That is, for salient expressions e_i, e_j contained in a sentence P_2 with associated orderings O_i and O_j , implicatures licensed via e_i will be calculated by replacing e_i in P_2 by higher¹⁵ values on O_i . Then, implicatures licensed via e_j will be calculated -- also by 'replacing' e_j with higher values in O_j in the original P_2 . Implicatures that may involve, say, higher values on O_i in addition to higher values on O_j will not be predicted by Gazdar's method -- or by the theory I have presented so far.

So, for example, applying these methods to B's response 221 would predict that B may license -- for any P_2 from (264a)-(264d) -- but neither (264e), nor (264f), nor (264g). If indeed the orderings stipulated are salient, then

(264)

- Ellen, you, and me have had lunch.
- Ellen and you have had dinner.
- Ellen and I have had dinner.
- Ellen and I have had breakfast.
- Ellen, you, and me have had dinner.
- Ellen, you, and me have had breakfast.
- Ellen, you, and me have had breakfast and dinner.

It seems clear that (264e) and (264f) should be considered scalar implicatures as well.

A simple addition to the definitions of higher, lower, and alternate sentences permits calculation of implicatures licensed when one or more expressions are salient in a sentence: Recall that sentences are defined as higher, lower, or alternate to other sentences when they differ in only a single expression as follows:

Higher Sentences:

$$\text{HIGHER_SENT}(P_i, P_j, O) \leftrightarrow \exists O_i \exists e_j (\text{HIGHER}(e_i, e_j, O) \wedge \text{SIMPLE_EXP_ALT}(P_i, P_j, e_i, e_j))$$

Lower Sentences:

$$\text{LOWER_SENT}(P_i, P_j, O) \leftrightarrow \exists O_i \exists e_j (\text{LOWER}(e_i, e_j, O) \wedge \text{SIMPLE_EXP_ALT}(P_i, P_j, e_i, e_j))$$

Alternate Sentences:

$$\text{ALT_SENT}(P_i, P_j, O) \leftrightarrow \exists O_i \exists e_j (\text{ALTERNATE}(e_i, e_j, O) \wedge \text{SIMPLE_EXP_ALT}(P_i, P_j, e_i, e_j))$$

¹⁵Gazdar does not deal with lower or alternate values.

individually, we want to extend the notions of utterance ranking so that one sentence P_i is higher than another P_j , for at least one pair of differing expressions in which P_i and P_j differ. P_i ranks higher than P_j and for no other pair of differing expressions does P_j outrank P_i . So, (264g) will represent a higher-ranked sentence than B's response in 221, as well as (264a)-(264f). Similarly, (264e) includes and thus is higher than the expression {Ellen, B} in (264c) in an ordering defined on the non-null subsets of the original exchange 221 -- 'Ellen and I have had lunch.' (264e) and {dinner} represent alternates in the ser-of-meats ordering, even though (264e) and the response in 221 differ in these expressions (264e) will still be ranked higher via the {Ellen, A, B} ordering. So, B may implicate (264e) in 221.

Then P_i should rank lower than P_j whenever the dual holds: for at least one ordering P_i ranks lower than P_j while nowhere else ranking higher than P_j . B's responses in 221-(264f) are, ranked *lower* than P_j whenever else ranking higher than P_j . B's responses in 221-(264f) are, ranked *lower* than P_j while nowhere else ranking higher than P_j . Sentences should be alternates whenever every expression in the {Ellen, A, B} ordering. So, B may implicate (264e) in 221.

Then P_i should rank lower than P_j whenever the dual holds: for at least one ordering P_i ranks lower than P_j while nowhere else ranking higher than P_j . B's responses in 221-(264f) are, ranked *lower* than P_j while nowhere else ranking higher than P_j . Sentences should be alternates whenever every expression in the {Ellen, A, B} ordering. So, B may implicate (264e) in 221.

Then P_i should rank lower than P_j whenever the dual holds: for at least one ordering P_i ranks lower than P_j while nowhere else ranking higher than P_j . B's responses in 221-(264f) are, ranked *lower* than P_j while nowhere else ranking higher than P_j . Sentences should be alternates whenever every expression in the {Ellen, A, B} ordering. So, B may implicate (264e) in 221.

Then P_i should rank lower than P_j whenever the dual holds: for at least one ordering P_i ranks lower than P_j while nowhere else ranking higher than P_j . B's responses in 221-(264f) are, ranked *lower* than P_j while nowhere else ranking higher than P_j . Sentences should be alternates whenever every expression in the {Ellen, A, B} ordering. So, B may implicate (264e) in 221.

Then P_i should rank lower than P_j whenever the dual holds: for at least one ordering P_i ranks lower than P_j while nowhere else ranking higher than P_j . B's responses in 221-(264f) are, ranked *lower* than P_j while nowhere else ranking higher than P_j . Sentences should be alternates whenever every expression in the {Ellen, A, B} ordering. So, B may implicate (264e) in 221.

Then P_i should rank lower than P_j whenever the dual holds: for at least one ordering P_i ranks lower than P_j while nowhere else ranking higher than P_j . B's responses in 221-(264f) are, ranked *lower* than P_j while nowhere else ranking higher than P_j . Sentences should be alternates whenever every expression in the {Ellen, A, B} ordering. So, B may implicate (264e) in 221.

$$\begin{aligned} \text{EXT_LOWER_SENT}(P_i, P_j) &\leftrightarrow (\text{SO_LOWER_SENT}(P_i, P_j, O_i)) \vee \\ & (\exists P_k \text{SO_LOWER_SENT}(P_i, P_k, O_j) \vee \text{ALT_SENT}(P_i, P_k, O_j)) \wedge \\ & \text{EXT_LOWER_SENT}(P_i, P_k) \\ \text{EXT_ALT_SENT}(P_i, P_j) &\leftrightarrow (\text{SO_ALT_SENT}(P_i, P_j, O_i)) \vee (\exists P_k \text{SO_ALT_SENT}(\\ & O_j) \wedge \text{EXT_ALT_SENT}(P_i, P_k)) \end{aligned}$$

Then the resulting version of the scalar implicature conventions, incorporating these extended notions of sentence ranking, will be:

$$\begin{aligned} \text{Imp}_1: & \text{SO}(\text{BEL}(S, \text{AVAIL}(H, e_i, O, C_n)) \wedge \text{REALIZE}(u_i, \text{AFFIRM}(S, e_i, \text{BEL}(S, \\ & P_i))) \wedge (\text{EXT_HIGHER_SENT}(P_i, P_k, O) \vee \text{EXT_ALT_SENT}(P_i, P_k, O))) \\ & \Rightarrow \text{SCALAR_IMP}(S, H, u_i, -\text{BEL}(S, P_i), C_n) \\ \text{Imp}_2: & \text{SO}(\text{BEL}(S, \text{AVAIL}(H, e_i, O, C_n)) \wedge \text{REALIZE}(u_i, \text{DENIAL}(S, e_i, \text{BEL}(S, \\ & -P_i))) \wedge (\text{EXT_LOWER_SENT}(P_i, P_k, O) \vee \text{EXT_ALT_SENT}(P_i, P_k, O))) \\ & \Rightarrow \text{SCALAR_IMP}(S, H, u_i, -\text{BEL}(S, -P_i), C_n) \\ \text{Imp}_3: & \text{SO}(\text{BEL}(S, \text{AVAIL}(H, e_i, O, C_n)) \wedge \text{REALIZE}(u_i, \text{IGN}(S, e_i, -\text{BEL}(S, \\ & P_i)))) \Rightarrow \\ & ((\text{EXT_LOWER_SENT}(P_i, P_k, O) \Rightarrow \text{SCALAR_IMP}(S, H, u_i, -\text{BEL}(S, \\ & -P_i), C_i)) \vee \\ & (\text{EXT_HIGHER_SENT}(P_i, P_k, O) \Rightarrow \text{SCALAR_IMP}(S, H, u_i, -\text{BEL}(S, P_i), \\ & C_i)) \vee \\ & (\text{EXT_ALT_SENT}(P_i, P_k, O) \Rightarrow \text{SCALAR_IMP}(S, H, u_i, \text{BEL}(S, P_i), C_i))) \end{aligned}$$

Then the algorithm for calculating scalar implicatures must be altered to accommodate multiple salient expressions as follows: Given S 's u_i with the semantic representation P_i :

1. determine whether any expressions e_i in P_i evoke salient orderings O_i and, if so, associate each such e_i with the appropriate O_i ;
 2. determine for each e_i whether P_i represents the affirmation, denial, or declaration of S 's ignorance about some proposition with semantic representation P_k which contains e_i (i.e., if P_k is simple with respect to e_i ;¹⁵⁴
 3. create an open sentence P_k by replacing each e_i in P_k with a (distinct) variable x_i ;
 4. locate other values e_j appearing with each e_i in the appropriate O_i ;
 5. for every P_j resulting from the replacement of one or more x_i in P_k by some e_j such that P_j and P_k are simple expression alternatives,¹⁵⁵ determine whether P_j represents a scalar implicature by the scalar implicature conventions.
- These revisions complete the theory of scalar implicature presented in this thesis.

¹⁵⁴See Section 4.1.3.

¹⁵⁵See Section 4.1.3.

5. Summary

In this chapter I have presented some limited solutions to the problem of how context might be incorporated into the calculation of scalar implicature. I have proposed a number of ways in which salience information might be derived from focus information and from knowledge of relations that may be salient in the discourse. Finally, I have identified certain revisions which permit the calculation of calculate scalar implicatures licensed when utterances include one or more salient expressions and have presented an algorithm which can be used to calculate these implicatures.

While the calculation of licensed implicatures should be important to any attempt at natural-language generation or interpretation, permitting the anticipating or understanding of pragmatic inferences derivable from any natural-language output or input, it is helpful to show how the computation of scalar implicatures can be turned to particular computational use. In Chapter 7, I will describe how knowledge of scalar implicature permits the generation of one form of cooperative response in question/ answer systems.

CHAPTER VII Scalar Implicature and Question-Answering

I don't want you, but I need you.
I don't like you, but I love you.

Eddy Rabbir

Studies of question-answering by philosophers, linguists, and computer scientists have proposed numerous accounts of speakers' propensity for providing more or less information than has been requested by a questioner. In this chapter I add to this literature. A large class of indirect and modified or qualified direct responses to yes-no questions can be explained as attempts by speakers to block scalar implicatures which hearers might otherwise be expected to infer. First, I situate this proposal in previous studies of cooperative question-answering and theoretical studies of yes-no questions. Next, I describe how scalar responses can be explained as attempts to block potential inferences of scalar implicatures. Finally, I describe QUASI, a question-answering system that provides cooperative responses to yes-no questions by calculating the scalar implicatures users might be entitled to infer from direct responses to these questions. While it would be foolish to interpret or to generate responses to any class of questions solely on the basis of a single phenomenon such as scalar implicature, QUASI does demonstrate the feasibility of calculating scalar implicatures and one area in which their calculation is useful. In Chapter 8, I will speculate about some further ways in which their calculation of scalar implicatures should be important to more general studies of discourse.

7.1. Appropriate Responses to Yes-No Questions

Cooperativity in question-answering has been defined by a variety of disciplines in a variety of ways: Philosophical studies of question-answering have defined the notion of ANSWERHOOD: the conditions under which a response counts as an answer to a natural-language query. Students of natural-language processing have identified types of cooperative speaker behavior from studies of naturally occurring dialogues and from introspection about the sort of

behavior computer-human interaction should support. In particular, AI's approach to question-answering has focussed on identifying when systems should provide more information than their users have requested and, of course, what that information should be. Anticipation of follow-up questions, explanation of questioners' violated expectations, provision of information relevant to questioners' inferred goals, and correction of misconceptions perceived in a query are only a few types of behavior which AI systems have been or are being designed to support.

7.1.1. Theoretical Accounts of Yes-No Questions

Theorists of question-answering commonly define questions in terms of the set of all their possible (true) answers [Kiefer 80, Hamblin 71]. Yes-no questions are generally seen as a disjunction of all the possible response to them. Traditional accounts portray them as propositional questions (7P) or as a special type of alternative question (7P \vee ?Q) in which the second alternative is simply the negation of the first (7P \vee ?-P). So, 'I ask you whether x' or 'Tell me which of the following is true, x or -x' [Katz 64] or 'Bring it about that I know that x or -x' [Hintikka 78], where x is the DESIDERATUM¹⁵⁶ of the query, have all been employed as representations. The 'meaning' of these questions then is the set of all their possible answers, i.e., the set of propositions which might truly be asserted in response.

However, both theoretical work and empirical studies of naturally occurring question-answer exchanges have shown this approach to be inadequate: *Yes* and *no* or even *yes*, *no*, and *unknown* form only a small portion of the set of all appropriate responses to a yes-no question. Furthermore, for some yes-no questions, none of these simple direct responses alone is appropriate.

I have previously noted (See Section 6.3.1.2.2.) Kiefer [Kiefer 80]'s claim that certain yes-no questions actually function as *wh*-questions and indicated its limitations. Aqvist [Aqvist 77] also notes that representing yes-no questions as alternative questions does allow proper treatment of negative queries. In Bolinger's [Bolinger 78] view the term YES-NO QUERY has hypnotized scholars into assuming that, simply because a class of question can be answered by a *yes* or *no*, these alternatives are critical, and every yes-no question is intended to elicit one or the other. He proposes instead that yes-no questions be viewed as hypotheses put forward for confirmation, amendment, or disconfirmation — in any degree. Thus, in his example 265, the

(265) Q: Do you like Honolulu?

R: Just a little.

¹⁵⁶A specification of the epistemic state desired by the questioner.

Q's hypothesis 'you like Honolulu' is amended by the R in a response which is neither *yes* nor *no* but somewhere in between. In his example 266, Q's hypothesis 'it is

(266) Q: Is it difficult?

R: It's impossible.

difficult' is confirmed by R's assertion of a more positive response than a simple *yes*.

While Bolinger makes a good case for the inadequacy of standard views of yes-no questions, the revision he proposes is itself too limited. 'It's impossible', in 266, does more than simply present a strong affirmation of the hypothesis 'it is difficult' — it provides new and unrequested though pertinent information. In fact, 'strong affirmation' might better be provided by a response such as 'I am absolutely sure it's difficult' than by the response he suggests. And there are equally appropriate responses to the queries in 265 and 266 that are not easily explained in terms of degree of hypothesis confirmation, as shown in 267 and 268.

(267) Q: Do you like Honolulu?

a. R: I don't like Hawaii.

b. R: I like Hilo.

(268) Q: Is it difficult?

a. R: It could be.

b. R: Mike says so.

An alternative account of certain yes-no questions from a COMMON-SENSE REASONING point of view is presented by Sadock [Sadock 77] in his work on MODUS BREVIS. He suggests that exchanges like 269

(269) A: Is a bat a bird?

B: Well, it's got wings.

in fact represent forms of fallacious reasoning — in 269, the fallacy of *asserting the consequent*. So, in this exchange, A argues for inclusion of *bats* among *birds* by the implicit 'Well, if something is a bird, it's got wings.' Similarly, according to Sadock, in 270, B invokes a *modus tollens* argument by his/her response, which also is fallacious:

(270) A: Do you have Pepsi?

B: We have Coke.

That is, B assumes the premise 'We have Pepsi or we have Coke' and incorrectly assumes the exclusivity of *or*.

Numerous additional examples are presented in further evidence of such 'informal' reasoning; Sadock concludes that speakers' failure to make their premises and conclusions explicit makes it possible 'to cover up logical abuses that might be more flagrant if the argument were stretched out' [Sadock 77:547]. However, it is not clear that B is necessarily

conveying an affirmative response in 269. While Sadock's explanation is ingenious, the theory of scalar implicature presented in this thesis seems to present a simpler explanation of much of his data without imputing wide-spread fallacious reasoning to speakers.

7.1.2. Computational Approaches to Yes-No Questions

A number of natural-language question-answering systems have recognized the desirability of supporting the sort of indirect responses to yes-no questions examined here. However, to date, few have proposed independent principled bases for the behavior they model. The 'cooperativity' of the additional information provided must largely be assumed and often is driven solely by the fact that the information has already been calculated; I will classify such systems as 'availability-driven' with regard to their provision of additional information to queries. Also, of the systems discussed below, only the HAM-ANS system is particularly concerned with yes-no questions.

7.1.2.1. Availability-Driven Responses

An early example of systems which provide additional information essentially as a side-effect of other system processes is Lehner [Lehner 77]'s QUALM, a question-answering system which answered questions about stories conceptually represented in the system KB. QUALM included a verification option which could, under certain conditions, give the effect of providing modified direct response to yes-no questions by verifying affirmative responses. When QUALM was in *talkative* mode and the question category was *verification*, and the direct response to a question was affirmative, dialogues like 271 could result:

- (271) Did John go to New York?
 Yes, John went to New York by bus.
 Did John eat?
 Yes, John ate lasagna.
 Did someone pick John's pocket?
 Yes, a thief picked John's pocket.
 Did John pay the check?
 Yes, John paid the bill.

QUALM's elaboration mechanism, in generating verifications from conceptual representations in its KB, output 'Yes, *X*' where *X* was the conceptualization found in the story representation that matches the question concept. Where this conceptualization did not exactly duplicate the question concept matched to it, additional information (in effect, QUALM's own representation of the concept) was presented.

[Kaplan 79]'s COOP system provided early evidence that pragmatic phenomena could

help in the selection of system responses: COOP defined cooperative responses to wh-questions in terms of their PRESUMPTIONS, which Kaplan derived from the pragmatic notion of presupposition.¹⁵⁷ While the correction of false presuppositions was the heart of the COOP system, Kaplan also suggested a parallel account of indirect responses to yes-no questions in his FOCUS responses, in which respondents identify information likely to be requested in a subsequent query with the focus of the current query, as in 272, and his SUGGESTIVE indirect responses

- (272) Q: Is there a mailbox on this block?=[Kaplan 79]'s 1)
 R: No, but there's one down the street.
 as in 273. In the latter, which are clearly included in

- (273) Q: Is John a senior?=[Kaplan 79]'s 8)
 R: He's a junior.

my class of scalar responses, Kaplan claims that speakers phrase questions to avoid a negative or trivial response like 'none' or 'zero'. If respondents are asked questions eliciting such responses, they are justified in assuming that some (perhaps rather weak) expectation the speaker had has been violated. So, a suggestive indirect response like that in 273 is appropriate. However, this claim is not fully developed or independently motivated.

The most significant existing system to address the problem of providing modified direct responses to yes-no questions directly is the HAM-ANS system [Hoepfner 84b, Hoepfner 84c, Hoepfner 84d]. Disclaiming any attempt at linguistic or cognitive justification [Hoepfner 83], the designers of this system provide EXTENDED RESPONSES to yes-no questions based upon a variety of heuristics, such as finding explicit partial answers to each conjunct of conjoined NPs to avoid an uncooperative *no* when at least one conjunct is successful, as in 274; 'modifying' universal

- (274) Is there an easy-chair or a large chair in the room?
 A large chair is not available, but an easy-chair.

quantifiers or cardinals when a queried quantifiers cannot be satisfied to a one which can be, as in 275;¹⁵⁸ and filling optional deep case slots in the case frame of a verb used in the query, as in 276.

¹⁵⁷Kaplan defined *LOADEN* questions as those which indicate a questioner presumes something to be true about the domain which is in fact false. So, in x, if Q knows that CIS 110 has not been offered this year, s/he could be assumed

(x) Q: How many students failed CIS 110 this year?
 R: CIS 110 wasn't offered this year.

to know the direct response to his/her query, *none*. By asking the query, Q reveals a false presumption, which R cooperatively corrects.

¹⁵⁸Note that the direct response *yes* is not itself appropriate here.

(275) User: Is there a lamp next to every bed?

HAM-ANS: Yes, next to almost every one.

(276) Did a yellow car and a green stationwagon turn off earlier?

No. A yellow car turned off Harungstrasse onto Schluenerstrasse, but the green stationwagon did not turn off first.

The quantity of additional information provided is limited by other heuristics and by the amount of information needed to compute the response.¹⁵⁹ While HAM-ANS results often appear quite 'natural', the lack of independent linguistic or cognitive motivation behind them can produce some inconsistency; the amount and nature of information provided is ultimately dependent upon the case frame of the verb or the presence of a few particular lexical items. So, questioner's use of one verb may result in the inclusion of an additional modifier in the response, where use of another will not. The critical link between the behavior produced and its appropriateness is not established.

1.2.2. Goal Inference and Indirect Responses

A more principled approach to indirect responses to yes-no questions was signalled by early work on indirect responses in task-oriented domains by Hobbs and Robinson (Hobbs 79a), in a classic essay, Hobbs and Robinson distinguish three classes of appropriate responses which, though indirect, nonetheless addressed the goals of the question asked: those which indeed answer the question asked indirectly, those which deny a presupposition of the question, and those which address a higher goal of the question. Problems with identifying responses in and those which address a higher goal of the question. Problems with identifying responses in the first and third categories purely from linguistic evidence lead Hobbs and Robinson to restrict the utterances under consideration to those contained in task-oriented dialogues where speaker and hearer goals can be predetermined; their tokens are drawn from dialogues between an expert and an apprentice trying to repair an air compressor. Example 259 (repeated here) is thus explained by (Hobbs 79a) in terms of particular

E: Have you disconnected the air line?

A: I loosened it.

domain goals: Since Expert and Apprentice both believe that the goal of their discourse is the completion of the air compressor repair, then if A is obeying the Maxim of Quantity, he will give as much information about his progress toward fulfilling this goal as he can. If to discoverer means to cause not to be attached and to loosen means to cause not to be tightly attached, then, since E can assume A has done no more toward achieving the goal of

¹⁵⁹SCALIBRE uses this means more simplistically to guide the provision of additional information in responses to wh-questions, outputting all the information accessed to compute the direct response.

disconnection than he has stated, he can conclude that, although the air line is not attached tightly it is still attached and the literal answer is no. Such answers are thus appropriate because, even if they do not provide the information requested directly, they provide some information that allows the questioner to achieve the goal of his/her query.

While (Hobbs 79a)'s account of 259 is intuitively satisfying, it lacks the generality: Domain goals may indeed provide an important clue to the appropriateness of indirect responses, but it is not clear how one might abstract beyond the particular examples provided to derive useful generalizations. However, research more directly concerned with goal inference does show how inferred goals can guide the provision of additional cooperative information. Allen and Perrault (Allen 80) examine the provision of additional information to direct and indirect responses based upon questioners' inferred goals in their. Since they limit possible goals to either taking or meeting a train, however, abstraction to broader classes of indirect responses is again problematic. More recent work by Carberry (Carberry 83) on goal tracking and by Pollack (Pollack 86) on inferring queries which questioners may themselves be unaware of promises a broader range of goal inference. In Section 6.3.2.2 I have proposed how such goal inference might facilitate the calculation of scalar implicatures by providing information on salient expressions and orderings.

The detection and correction, as well as the prevention of misconceptions forms a broad category of domain-independent goals for indirect and modified direct responses. Joshi (Joshi 82) has noted the importance of *squaring away* mutual beliefs in question-answering when respondents perceive misconceptions in an exchange — lest the questioner's misconceptions be implicitly confirmed. In human-machine interaction even more than in natural discourse, Joshi suggests that people may expect such cooperative behavior, since the possible processing restraints which make human cooperativity difficult may not be expected to apply to more resourceful machines.¹⁶⁰ This assumption underlies Kaplan's work, as well as related work on presupposition failure by Mays (Mays 80a, Mays 80b), more recent work on utterance presupposition by Mercer and Rosenberg (Mercer 84), and work on the correction of object-related misconceptions by McCoy (McCoy 85).

Joshi's redefinition of the Maxim of Quality (See Section 2.3.) suggests that, even beyond correcting perceived misconceptions, cooperative speakers have an obligation to anticipate and prevent misconceptions which their otherwise truthful utterances might engender. Preliminary

¹⁶⁰For example, *H* may not expect a human *S* to consider all the possible consequences of his/her utterances. Such consideration would clearly require super-human computational power. But in interaction with a computer, human may expect more, since they know/believe that machine processing is easier. So, the user of a question-answering system may expect that system to be more cooperative than another human (Joshi 82:191).

studies by Joshi, Webber, and Weischedel [Joshi 84a, Joshi 84b] have examined exchanges in which cooperative respondents are obligated to block misleading inferences. If a questioner Q expects that an expert respondent R would inform him of some proposition p_i related to his/her query if in fact that R knew p_i to be true, then Q may interpret the R's silence regarding p_i as implying $\neg p_i$. So, if R knows p_i to be true, his/her silence may lead to Q's being misled. So, R should block such inferences, as in:

Q: Is Sam an associate professor?
R: Yes, but he doesn't have tenure.

The authors suggest that their phenomenon might be represented in Reiter's [Reiter 80] default logic. The actual characterization of those p_i to be represented, however, awaits further study.

7.1.3. Blocking Potential False Scalar Implicatures

Many of the approaches described above provide useful information for the planning of text to be generated in response to yes-no questions. I will now discuss yet another approach which subsumes some of those described above and complements others.

A large class of indirect and modified direct responses to yes-no questions, exemplified in many of the examples presented throughout this thesis, may be analyzed as attempts by cooperative speakers to block potential false inferences which hearers might otherwise (wrongly) infer to be implicatures arising from a direct response -- while also providing information from which the direct response can be derived. While no claims to cognitive reality are made for this explanation, it does appear to account for the data. It is plausible (particularly if Joshi's claims are correct), and it does provide a principled and linguistically motivated basis on which to support much of the behavior supported by the systems described in Section 7.1.2.1.

Recall that a central claim of work on quantity implicature in general has been that speakers derive considerable information from what is not said in comparison to what is said. Similarly, the appropriateness of indirect responses such as that provided in 277 depends in large part upon the inappropriateness of alternative responses available to B.

(277) A: Has Marvin had his medication?
B: He's taken the exadin.

Suppose that B (a simple question-answering system) can only generate simple direct responses to yes-no questions even when more information is available, as in 278:

Such a respondent forces A to play Twenty Questions to elicit the information provided in a single indirect response in 277. While such limited response capability is obviously tedious, it may have more serious consequences. For, if A does not recognize that B can provide only yes

(278) A: Has Marvin had his medication?

B: No.

A: Do you mean he hasn't had any of it?

B: No.

A: Well, has he had his cortisone?

B: No.

A: Exadin?

B: Yes.

A: How much of that has he had?

B: 5 cc.

A: Has he taken the naprin?

or no responses, she may not ask for clarification after B's initial *no*. So, she will not learn that in fact Marvin has taken part of his medication. Thus, simple direct responses may be uncooperative, if technically truthful. And note that, in 277, A may infer the direct response (*No, Marvin has not taken all of this medication*) via the scalar implicature conventions, if she understands that B has said as much as s/he truthfully can.

In addition to facilitating the brief conveyance of relevant information, knowledge of scalar implicature can be used more directly to avoid block certain false inferences I will term POTENTIAL FALSE SCALAR INFERENCES. These inferences may be identified computationally by calculating the scalar implicatures a speaker (system) might license given that certain expressions and posets are deemed salient in the discourse and by comparing these licensable inferences to the speaker (system's) knowledge base.

Assume that B believes that Madge has five children. Let X = the set of propositions { \neg BEL(B, Madge has 4 children), \neg BEL(B, Madge has 5 children)}. Then, given the scalar implicature conventions, we know that a simple direct response (279a) to A's query in 279 would

(279) A: Does Madge have three children?

a. B: Yes (Madge has three children).

b. B: She has five.

allow B to implicate X -- if B believes s/he and A mutually believe the cardinals (or some compatible ordering) salient in the discourse. While, by the definition of conversational implicature presented in Chapter 2, B will only implicate X if s/he intends to implicate X, another necessary condition on conversational implicature is B's anticipation of A's ability to 'work out' X. This process is itself defined in Section 2.4.4. Even if B does not intend to convey X, then, s/he may still anticipate that A may infer X, if the conditions under which B might implicate X hold. If B does anticipate that A may infer X -- but if X is inconsistent with B's beliefs, then we may explain B's alternative response of (279b) as the blocking of a potential inference which is inconsistent with B's beliefs.

So, where a simple *yes* or *no* might be interpreted as licensing implicatures inconsistent with *S*'s beliefs, *S* may provide a scalar response to block those inferences. More generally, for any speaker *S*, a sentence P_j represents a potential false implicature of an utterance u_j iff, in the relevant context C_H , *H* will be entitled to infer that *S*'s u_j licenses P_j as a scalar implicature -- but P_j is inconsistent with *S*'s beliefs.

Of course, even with this admittedly over-simplified and stylized analysis, there may be more than one scalar response which will not license potential false inferences of the sort described above. For example, if Marvin's

(280) A: Has Marvin taken his medication?

a. B: He's taken the exadin.

b. B: He hasn't taken the naprin.

medication consists of *exadin* and *naprin*, and if the appropriate mutual beliefs hold between A and B, then (280a) and (280b) will both represent 'cooperative' responses in the sense that they will not license potential false inferences of the sort described above (if an ordering of *medication* with its subparts *exadin* and *naprin* is salient). B may deny the lowest value *she* can truthfully deny or affirm the highest -- in this example lowest and higher are in fact alternate values in the ordering. Such examples illustrate the dangers of a 'single theme' approach to question-answering: While the notion of blocking false potential implicatures can indicate that either (280a) or (280b) is to be preferred over a simple *no* when *parts-of-medication* is salient, it cannot help to select between the two responses.

So, this approach to providing cooperative responses to yes-no questions should be seen only as providing one test of the appropriateness or cooperativity of any response. Utterances that license no potential false scalar inferences may be inappropriate on other grounds -- perhaps they fail to address the user's goals in some way or provide more information than the user desires. Other aspects of the discourse -- prior 'cancellation' for one -- might override the inappropriateness of utterances that do license such false inferences. So, in an ideal world, an understanding of scalar implicature should be used for cooperative question-answering only in conjunction with other criteria. With this caveat, I will proceed to describe the QUASI system implemented to demonstrate the calculation of scalar implicature.

7.2. The QUASI System

QUASI is a question-answering system which proposes cooperative responses to yes-no questions about a mail system from its users. Using the representation and algorithms described in previous chapters, QUASI calculates the potential scalar implicatures which simple direct or alternative responses might license when certain expressions and posets are salient. It tests these possible inferences for consistency with its own knowledge base, and, when inconsistency is detected, proposes alternate responses that do not license inconsistent scalar implicatures. The system has been implemented in C-Protolog on a VAX 11-785. In this section, I will provide an overview of the system, examine its major constituents, and illustrate particular system features.

The goal of this implementation is to demonstrate how the formalization of scalar implicature can aid in the calculation of licensable implicatures and to suggest a simple use for such information. Not all aspects of the theory presented in the thesis have been implemented. First, in the absence of a user model, QUASI makes the simplifying assumption that the user believes the system cooperative and that the user can 'work out' any inferences the system can calculate. Since QUASI's goal is not the licensing of particular inferences -- but rather the generation of responses that will not license false inferences -- these assumptions are actually conservative: the system will err on the side of caution if it anticipates inferences the user would not be in a position to draw. Second, QUASI does not adopt an independent strategy for determining when and which expressions and posets are salient, but must rely upon interaction with the user to select from expressions appearing on posets known in the system. Third, QUASI does not maintain or assume information from a discourse history, which might modify its decisions on when a false inference is likely -- i.e., when the user has already demonstrated knowledge of some licensable inference. However, the core of the thesis, the scalar implicature conventions and algorithms for the calculation of scalar implicatures when one or more expressions are salient, has been incorporated into QUASI.

7.2.1 Overview of the System

The implemented system is depicted schematically in Figure 7-1: Major system components are identified by capital letters, information available in the system by small capitals, user inputs by italics, outputs by bold italics, and information passed through the system by normal font.

QUASI's semantic translation module creates a semantic representation of a user query which its direct response retrieval module uses both to retrieve and to represent the direct response in terms of system beliefs. This module also determines whether or not the query is a yes-no

entered to implicate $\neg\text{BEL}(S, (p_i \text{ and } p_j))$.⁹² Since, in standard first-order logic, ' $p_i \wedge p_j$ ' entails $p_i \vee p_j$ Horn claims that *and* semantically entails *or*. The lower value in his *or* and scale, ' p_i or p_j ', may be uttered to implicate that the higher value, ' p_i and p_j ', is false or unknown, as in 99.

- (99) A: Are you serving tea and coffee?
B: Tea or coffee.

And, in 100, A rejects the implicature she infers from B's '*hell or high water*'.

- (100) A: Have a happy Memorial Day Weekend.
B: Thanks. Same to you. I just wish I could enjoy it. It'll be just like the last weekend, and the last, and the last...
A: But I hope not the next, and the next, and the next...
B: No way. Come hell or high water, it will be done.
A: Come hell and/or high water...

Prince [Prince 82a] suggests that this \vee/\wedge ordering may also account for scalar implicatures arising from the utterance of conditional sentences. She notes that, when a conditional sentence, ' $p_i \Rightarrow p_j$ ', is transformed into its disjunctive equivalent ' $\neg p_i \vee p_j$ ',⁹³ this \vee/\wedge ordering can account for the implicature $\text{KNOW}(S, p_j \Rightarrow p_i)$ ⁹⁴ as follows: By asserting the equivalent of ' $\neg p_i \vee p_j$ ' S may implicate $\text{KNOW}(S, \neg(\neg p_i \wedge p_j))$ – the falsity of conjunction. This implicature is itself equivalent to $\text{KNOW}(S, p_j \Rightarrow p_i)$: $\neg(\neg p_i \wedge p_j) \Leftrightarrow p_i \vee \neg p_j \Leftrightarrow p_j \Rightarrow p_i$. Thus the fallacy of affirming the consequent can be explained in terms of quantity implicature via the *or* and scale. Note also that, since ' $p_j \Rightarrow p_i$ ' is equivalent to ' $\neg p_i \Rightarrow \neg p_j$ ', this manipulation also accommodates the fallacy of 'denying the antecedent'. (And see Section 3.1.1.2). So, Prince's analysis would predict that the utterance of (101a) may convey (101b).

- (101)
a. If Bill goes then Mary will go.
b. If Mary goes then Bill will go.

However, this suggestion raises an interesting problem. With equal plausibility, one might define an ordering from the logical connectives, $\Rightarrow/\leftrightarrow$, since $p_i \leftrightarrow p_j$ entails $p_i \Rightarrow p_j$. Then the utterance of a conditional $p_i \Rightarrow p_j$ may also convey $\text{KNOW}(S, \neg(p_i \leftrightarrow p_j))$ via the $\Rightarrow/\leftrightarrow$ ordering. However, by this analysis, the utterance of (101a) could convey $\text{KNOW}(S, \neg(102))$ – and also the meaning

⁹²Note that, while the use of disjunction may be taken as exclusive it need not be, so ' p_i or p_j ' cannot always be translated $(p_i \vee p_j) \wedge \neg(p_i \wedge p_j)$.

⁹³This translation itself is controversial, of course [McCawley 81:49ff].

⁹⁴Prince follows Gazdar's [Gazdar 79b] view of the epistemic force of quantity implicature and represents these implicatures as $K_S \neg(p_i)$.

(102) If Bill goes then Mary will go and if Mary goes then Bill will go.
 $\text{KNOW}(S, \neg(101b))$. -- where Prince would predict it could convey $\text{KNOW}(S, (101b))$. (That is, if S asserts ' $p_i \Rightarrow p_j$ ' and implicates ' $\neg(p_i \leftrightarrow p_j)$ ', ' $\neg(p_j \Rightarrow p_i)$ ' follows.) Thus, implicature licensed via one ordering contradicts implicature predicted by an alternate analysis.

Cases such as this recall Gazdar's injunction that the semantic representation from which conversational implicatures are derived must be rather more 'surfacey' than a simple first order representation, the substitution of equivalent schemata during the computation of implicatures is risky business indeed and will not be permitted for the calculation of scalar implicatures.

5.1.3.1. Disjunctive Assertions

According to [Kempson 75, McCawley 78], assertion of a disjunction may also implicate that S is not able to assert either disjunct alone, as in 103.

- (103) A: Do you think she's ataxic or she's weak?
B: Pause...sigh...I think she's ataxic [...].

That is, if S says ' p_i or p_j ', she may implicate $\neg\text{BEL}(S, p_i)$ as well as $\neg\text{BEL}(S, p_j)$. If S can affirm p_i but instead asserts ' p_i or p_j ', she will violate the Maxim of Quantity. Note the oddness of 104, in which the humor of B's

- (104) A: What do you think, Grandpa. Will it be a boy or a girl?
B: Probably.

response rests upon just such an implicature -- $\neg\text{BEL}(B, \text{it will be a boy}) \wedge \neg\text{BEL}(B, \text{it will be a girl})$. For such scalar implicatures, the ordering $p_i \vee p_j/p_i$ must be salient. However, a different ordering must be seen as salient to explain the scalar implicature licensed in 105 -- $\neg\text{BEL}(B, \text{I want to go out to dinner})$.

- (105) A: Do you want to go out to dinner or find a movie to see?
B: A movie sounds good.

Since $\neg p_j$ is cancelable, it does not follow from logical implication, i.e., ' $p_i \vee \neg p_j$ ' cannot be seen as an exclusive disjunction here. Such implicatures are better explained by postulating the evocation of a set member relationship among various alternatives (See Section 5.1.1.10). By affirming a member of that set of disjuncts, as in 105, S may implicate that other members (alternate values) are false or unknown.

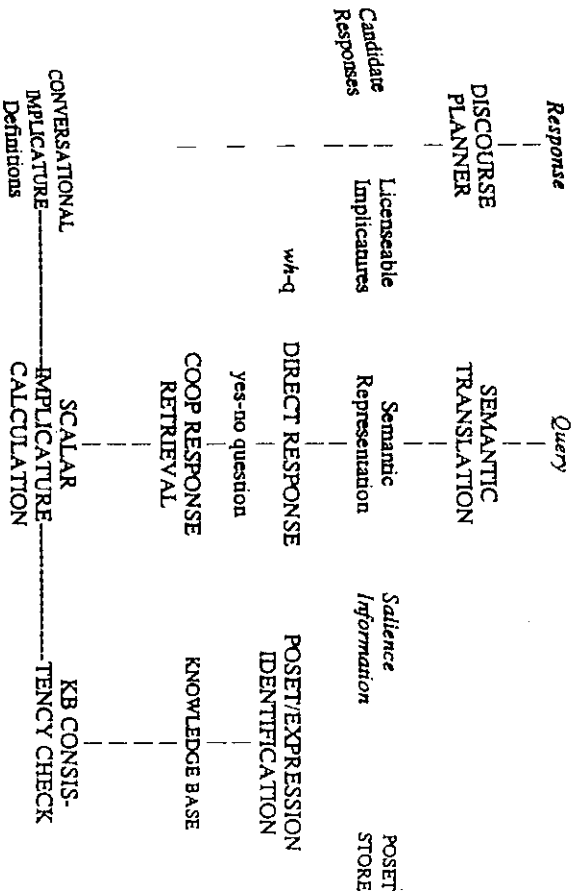


Figure 7-1: The QUASI System

question. If so, it is passed to another module which guides the generation of one or more candidate cooperative responses. Representations of responses to *wh*-questions are passed directly to the discourse planner.

The cooperative response retrieval module (CRR) first acquires information about salient expressions and posets from the poset/ expression identification module, which provides the user with a list of possible expression/ poset pairs from which to choose and returns those choices to CRR. CRR passes this information and the representation of the direct response to a scalar implicature calculator, which determines licenseable scalar implicatures from it. If implicatures are licenseable, CRR sends them to the KB consistency checker. If this check discovers licenseable implicatures which are inconsistent with the system KB, CRR will propose alternate responses which will themselves be checked for consistent implicatures. When all true responses which license only scalar implicatures which are consistent with the KB have been found, these are output to the discourse planner as candidate responses.

QUASI also provides some help facilities as well as tracing, which will be demonstrated in examples below.

7.2.2. Domain

The domain chosen to demonstrate the QUASI system was a mail system currently in use at the University of Pennsylvania.¹⁶¹ Currently, the system KB only includes information about boolean switches in this mail system and how they are used, including the mail commands which can set them, the arguments these commands may take, default switch values, and values that are associated with each switch. For example, the switch 'auto-archive' has the default setting 'off' and takes an associated value which is the name of an archive file; the default setting for this value is the filename 'archive.msg'. Both the switch value and that of its associated values can be set with 'setswitch' or 'unsetswitch'. The latter command cannot be used in maildir files (i.e., noninteractively) and is employed for temporary switch-setting.

This domain was chosen because, while it is small and well-defined, its conceptual model naturally reflects a wide variety of partial orderings discussed in Chapter 5, including sets, whole/ part relations, entity/ attribute orderings, and simple taxonomic hierarchies, as well as the canonical orderings, as well as process or prerequisite relationships. Although not all these

¹⁶¹This system was written by Sharon E. Peil.

orderings have in fact been represented in the current implementation,¹⁶² the potential for extensibility is clear. Also, while QUASI currently operates only on information about switches in the mail domain, the KB could readily be extended to other aspects of the mail system. Additionally, while switches make up an important component of this mail system, they are little understood by the mail user community. So, a question-answering system in this domain has potential for practical use.

7.2.3. Knowledge Representation

7.2.3.1. Epistemic Representation

Because of Prolog's closed-world assumption, it was necessary to make a major simplification of the epistemic force of licensed implicatures. Since Prolog, like most systems, assumes 'negation as failure', ignorance cannot be represented in it; it supports only a standard two-valued logic. Everything provable in the KB is true and all else is assumed false. This assumption is easily translated in terms of system beliefs as: the system believes everything provable in the KB and it believes that it believes -- otherwise it believes things false.

```
bel(system, Sent1) :- Sent1.
bel(system, ~Sent1) :- \+Sent1.163
```

Since ignorance is not representable in the system KB, the representation of scalar implicatures as epistemic disjunctions of belief or ignorance must be simplified to simple belief. Where scalar implicatures are represented as disjunctions in a three-valued logic -- S believes $p_i \vee S$ does not know whether p_i or S believes $\neg p_i \vee S$ does not know whether $\neg p_i$ -- they are in effect reduced to their first conjuncts in a two-valued logic. So $\neg\text{bel}(\text{Speaker}, \text{Sent1})$ is true iff Sent1 is false -- just as $\text{bel}(\text{Speaker}, \sim\text{Sent1})$, and $\neg\text{bel}(\text{Speaker}, \sim\text{Sent1})$ is true iff Sent1 is true -- as $\text{bel}(\text{Speaker}, \text{Sent1})$. While a more expressive system would permit a more accurate reflection of the theory, in fact, little is lost in this translation: QUASI returns representations of system beliefs which then may be simplified in their interpretation for a given system -- or not. Too, since few systems support a three-valued logic,¹⁶⁴ it seems reasonable to test the theory in a more restricted environment.

¹⁶²Of the canonical orderings, modals and ordinals are not represented. Also, there is no way to represent temporal orderings, and no special support for the representation of 'entity' attribute or 'whole' part orderings.

¹⁶³ Sent1 is not provable.

¹⁶⁴See [Lavesque 84] for a recent proposal.

7.2.4. Representing Posets

Although a new domain was modeled for QUASI, the system as a whole has been designed to operate on an existing KB. The chief expense involved in transferring QUASI to other domains is the need to define particular ways in which posets may be derived from a particular knowledge representation. For standard Prolog databases, however, such redefinition will be minimal, since QUASI includes its own representations of the canonical orderings and derives posets from relations explicit or implicit in the domain by its own poset derivation rules. Although some meta-level translation must take place when, for example, existing KB's include more sophisticated knowledge representations which will facilitate poset inference. Also, if any of the canonical orderings are not supported by a new KB, it will be misleading to include these among user options when salient expressions and posets are selected.

7.2.4.1. Canonical Orderings

Posets are represented explicitly in a domain-independent poset store or are derived from the domain KB by rules represented in that store. Explicitly represented posets include the canonical orderings such as the quantifiers, cardinals,¹⁶⁵ and logical connectives, as described in Chapter 5. These orderings are represented as 4-tuples: a poset label, which is used to identify the ordering uniquely for system purposes and for interaction with the user when appropriate posets are chosen during the query analysis; an arbitrarily complex specification of the set of referents in the poset; the metric partially ordering this set of referents, which must itself be defined in the system; and a keyword specifying the method of derivation for the poset. For predefined canonical orderings this is simply 'canonical'. Posets are accessed and manipulated with predicates that 'specify' (or test) the membership of expressions in their set of referents, their ordered pairs, and the incomparability of items in the poset. These definitions are used to define higher, lower, and alternate sentences as described in Chapter 4.

7.2.4.2. Derived Posets

However, the approach to scale outlined in Section 3.3 should support domain independence; in particular the scale derivation rules proposed there should allow interface to existing knowledge bases without extensive recoding. That is, no specific knowledge representation should be required for this extension. Relationships and values present in the knowledge base should be interpretable as scales as required by a given query. While initial translation from such relationships to scales might cut response time for given queries, it seems

¹⁶⁵Unlimited here arbitrarily to 1-5.

likely that considerations of space and initial cost will indicate that a lazy evaluation of scales will be preferable.

Posets are derived from the KB by means of a set of poset derivation rules stored in the poset store. The essence of poset derivation is simply the identification of partial orderings from relations represented explicitly or implicitly in the KB. While only some poset derivation rules have been implemented in QUASI, I will discuss others which might easily be added.

In any representation, certain distinguished predicates must always define relations which, as I have noted in Chapter 5, represent posets. The predefined predicate *setof* in most Prolog implementations will always define posets ordered by set inclusion. In the main switches KB used in this application, an *isa* predicate identifies classification hierarchies, which also represent partial orderings of their members. Similar meta-level predicates might be defined for other knowledge representations and added to the poset derivation rules in the poset store.

Posets can be derived from any KB predicate of two or more arguments as follows: If a predicate *P* defines a relation on any two of its arguments that is *irreflexive*, *asymmetric*, and *transitive* or that is *reflexive*, *antisymmetric*, and *transitive*, then *P* defines a poset on these arguments. Posets defined by set inclusion can also be derived from the range of values a single argument of a predicate which does not satisfy these conditions may take on. Boolean combinations of predicates for which the ranges of some argument intersect may also define similar orderings.

7.2.4.3. Storing Identified Posets

Once a poset implicit in the KB has been identified, its representation tuple is stored in the poset store for later access -- to minimize redundant poset inference. Modules which access the poset store look first at posets which have already been derived. It seems likely too that orderings deemed salient should be so marked by the discourse planner in a more sophisticated application, for future identification of salient posets. This is not done in QUASI, since salience information is provided by the user. The availability and compatibility of posets is testable in QUASI but results are not used in the absence of a user model.

7.2.5. Input and Semantic Processing

QUASI accepts either *wh* or *yes-no* questions in English as input, although only the latter can generate 'cooperative' responses. The ability to handle *wh*-questions was provided to support more natural behavior in the system.

Legal queries are given a semantic representation in first-order predicate calculus using a definite-clause grammar extended from that presented in [Pereira 80].¹⁶⁶ At present, the grammar considerably limits the range of queries that can be handled by QUASI: for example, conjunction is only available at the sentence level.

The parse produces a simple first-order semantic representation of the query which corresponds to the notion of SENTENCE discussed in Chapter 4. So, a query like (281a) will be represented as (281b).

- (281)
 a. Are any switches boolean?
 b. *exists(X, isa(switch, X) & boolean(X))*

In trace mode, QUASI produces the following:

| : | : Are any switches boolean?

The semantic representation of your query is:

exists(_247, isa(switch, _247) & boolean(_247))

Singular definites and indefinites are represented as existentials and plural definites are universally quantified, following [Pereira 80, Warren 81] and cardinals are also treated as determiners, following [Jackendoff 68]. Also, the first successful representation for each query is accepted: So, for example, no scope ambiguities are recognized; in effect, the first quantifier in the surface order of the input query will have scope over subsequent quantifiers.

7.2.6. Direct Response Retrieval

Once the semantic representation of a query has been determined, QUASI proceeds to determine a direct response to the query and then to describe that response in terms of the system's beliefs, e.g., '*BEL(system, P_i)*', where *P_i* is the direct response to the user's query. For *wh*-questions, QUASI simply returns this information to the discourse planner. For *yes-no* questions,¹⁶⁷ QUASI proceeds to identify a cooperative response, passing the direct response to a cooperative-response retrieval module. With the direct response plus information about salient expressions and posets, this module controls the generation and testing of possible alternate responses. Salience information is acquired interactively.

¹⁶⁶This type of grammar permits the generation of a semantic representation directly, without constructing an intermediate parse tree.

¹⁶⁷Since *how* and *why* queries are not currently supported by the grammar, QUASI treats non-*wh*-questions as *yes-no* questions.

2.7. Identifying Salient Expressions and Posets

From the semantic representation of the query, QUASI identifies all the legal bexpressions of the query's semantic representation. As I noted in Section 4.1.2.2, a bexpression of a sentence P_i can be any constant, predicate, logical operator (including the systemic operators), connective, or quantifier symbol of P_i or any wff contained in P_i that is not identical to P_i . So, the following subexpressions are contained in the sentence 'exists(Y, (a(switch, Y) & boolean(Y))), and will be so identified

```
exists
  &
  isa(switch, Y) & boolean(Y)
  &
  isa(switch, Y)
  isa
  switch
  boolean(Y)
  boolean
```

Y QUASI's posexpression identification component. In effect, the module excludes the sentence itself, variables and non-wffs other than predicates, constants, and logical symbols from the set of possible subexpressions of a sentence.

Once subexpressions are identified, QUASI checks each for membership in the set on which some partial ordering explicit or implicit (See Section 7.2.4.) in the poset store or KB is defined. For the above query, QUASI will currently identify the following expression/ poset pairs:

```
Please choose salient expressions and a salient ordering for
each from the following list, forming a list of lists (e.g.,
[exists, quantifiers], [3, ucardinals]]):
```

```
Expression      Poset
&               [andor]
exists          [quantifiers]
| : | : [exists, quantifiers]]
```

QUASI currently asks the user to select one or more expressions from those identified as associated with posets known to the system. For each selected expression, the user must choose a single salient poset, as above, forming a list of lists. This list of lists will be passed on to the scalar implicature calculation module as the list of salient expression/ poset pairs. Salient

¹⁶⁸U'cardinals' designate the cardinals ordered by '<'.

expression/ poset pair lists might also be input from the discourse planner if QUASI were incorporated into a sophisticated natural-language processing system.

7.2.8. Calculating Cooperative Responses

The generation of cooperative responses is driven by the calculation of potential scalar implicatures and the testing of these implicatures against the system KB. Where a candidate response might license inferences about system beliefs which are inconsistent with the system's KB, alternate responses are identified and themselves tested for the scalar inferences they might license. A simple example of this process of testing and retesting with accompanying trace information from QUASI follows:¹⁶⁹

```
| : TRACE.
| : | : ARE ANY SWITCHES BOOLEAN?
```

The semantic representation of your query is:

```
exists(_247, isa(switch, _247) & boolean(_247))
```

The direct response is: true

Please choose salient expressions and a salient ordering for each from the following list, forming a list of lists (e.g., [[exists, quantifiers], [3, ucardinals]]):

```
Expression      Poset
&               [andor]
exists          [quantifiers]
| : | : [exists, quantifiers]]
```

If the following expressions and orderings are salient,

```
Expression      Poset
exists          [quantifiers]
```

The scalar implicatures that could be licensed by the assertion of:

```
bel(system, exists(_247, isa(switch, _247) & boolean(_247)))
```

¹⁶⁹User input has been altered to upper case for clarity.

```

are:
[-bel(system, all(_247, isa(switch, _247) => boolean(_247)))]
The following licensable implicature is inconsistent with the
system KB:
-bel(system, all(_247, isa(switch, _247) => boolean(_247)))
The scalar implicatures that could be licensed by the assertion
of:
bel(system, all(_247, isa(switch, _247) => boolean(_247)))
are:
[]
This assertion is consistent with the system KB.
Output this cooperative response to the query:
bel(system, all(_247, isa(switch, _247) => boolean(_247)))

```

In this sample query analysis, the user first puts QUASI into trace mode and then enters a query. QUASI first returns the semantic representation of the input query and then identifies the direct response. It then asks the user to choose a list of salient expressions and posets from those expressions identified in the query for which QUASI can locate posets from the poset store or poset derivation rules. Here the user chooses the quantifiers as salient. By accessing the scalar implicature conventions in its definition of conversational implicature, QUASI determines that, if the quantifiers are salient, the scalar implicatures that could be licensed by the assertion of the direct response to the user's query (that the system does indeed believe that some switches are boolean), would be the implicature that the system does not believe that all switches are boolean. QUASI then determines that this belief is inconsistent with its KB -- for, in fact, it does believe that all switches are boolean. So, it tests an alternate response, derived by replacing the salient expression in the response it is currently testing -- *exists* -- with another values aim the salient ordering -- *all*. It finds that no scalar implicatures whatsoever will be licensed by the assertion that the system believes all switches to be boolean, given that only the quantifiers are salient; *all* is in fact the highest value in this ordering, and affirmation of a value can license scalar implicatures only about still higher values. QUASI finds that asserting *bel(system, all(_247, isa(switch, _247) => boolean(_247)))* is consistent with the KB; it is true itself and it licenses no (and thus, no false) scalar implicatures. So, it proposes this cooperative response to the discourse planner:

Note that, if no expressions are chosen as salient, the cooperative response will be identical with the direct response, as in the following interaction where the same query elicits a different 'cooperative' response:

```

| : | ARE ANY SWITCHES BOOLEAN?
The semantic representation of your query is:
exists(_4211, isa(switch, _4211) & boolean(_4211))
The direct response is: true
Please choose salient expressions and a salient ordering for
each from the following list, forming a list of lists (e.g.,
[exists, quantifiers], [3, cardinals])):
Expression      Poset
&               [andor]
exists          [quantifiers]
| : | : []

```

If the following expressions and orderings are salient,
 Expression Poset

```

The scalar implicatures that could be licensed by the assertion
of:
bel(system, exists(_4211, isa(switch, _4211) & boolean(_4211)))
are:
[]

```

This assertion is consistent with the system KB.

Output this cooperative response to the query:

```

bel(system, exists(_4211, isa(switch, _4211) & boolean(_4211)))

```

Since the user chooses no expression/poset pairs from the proffered list, no scalar implicatures can be calculated from the assertion of the direct (affirmative) response. Licensable implicatures are, then, trivially, consistent with the KB, so the direct response is deemed cooperative by QUASI.

2.8.1. Responses When Multiple Expressions are Salient

When multiple expression/poset pairs are selected as salient, QUAST identifies scalar implicatures as described in Section 6.4.1, as illustrated in the following exchange:

: DO ANY COMMANDS TAKE 2 ARGUMENTS?

re semantic representation of your query is:

```
lists(_11127, isa(command, _11127) & numberof(_11141,
sa(arg_type, _11141) & take(_11127, _11141), 2))
```

re direct response is: true

lease choose salient expressions and a salient ordering for each from the following list, forming a list of lists (e.g., [exists, quantifiers], [3, ucardinals]):

```
expression      Poset
                [cardinals, ucardinals]
                [and/or]
lists           [quantifiers]
                [cardinals]
```

: | : [EXISTS, QUANTIFIERS], [2, DCARDINALS]].

: the following expressions and orderings are salient.

```
expression      Poset
                [quantifiers]
lists           [ucardinals]
```

re scalar implicatures that could be licensed by the assertion

```
;;
n1(system, exists(_11127, isa(command, _11127) &
numberof(_11141, isa(arg_type, _11141) & take(_11127,
_1141), 2)))
```

re:

```
-bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
)) , -bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
)) , -bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
```

```
4))) , -bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
5))) , -bel(system, exists(_11127, isa(command, _11127) &
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
3))) , -bel(system, exists(_11127, isa(command, _11127) &
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
4))) , -bel(system, exists(_11127, isa(command, _11127) &
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
5))) ]
```

The following licensable implicature is inconsistent with the system KB:

```
-bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
2)))
```

The scalar implicatures that could be licensed by the assertion of:

```
bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
2)))
```

are:

```
[ -bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
3))) , -bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
4))) , -bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
5))) ]
```

The following licensable implicature is inconsistent with the system KB:

```
-bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
3)))
```

The scalar implicatures that could be licensed by the assertion of:

```
bel(system, all(_11127, isa(command, _11127) =>
numberof(_11141, isa(arg_type, _11141) & take(_11127, _11141),
3)))
```

```

are:
[-bel(system, all(11127, isa(command, 11127) =>
  numberof(11141, isa(arg_type, 11141) & take(11127, 11141),
  4)), ~bel(system, all(11127, isa(command, 11127) =>
  numberof(11141, isa(arg_type, 11141) & take(11127, 11141),
  5)))]

```

This response is consistent with the system KB.

Output this cooperative response to the query:

```

bel(system, all(11127, isa(command, 11127) =>
  numberof(11141, isa(arg_type, 11141) & take(11127, 11141),
  3)))

```

The questioner selects two expression/poset pairs as salient for the analysis of this query, the quantifiers and the cardinals ordered by '<', termed here the 'cardinals'. QUVASI first identifies all the implicatures licensable when both exists and 2 are treated as salient expressions. Checking these potential implicatures against its KB, QUVASI finds the first inconsistent implicature '-bel(system, all(6278, isa(command, 6278) => numberof(6292, isa(arg_type, 6292) & take(6278, 6292), 2))' and attempts to find a more cooperative response. This process of checking potential responses and finding an inconsistent scalar implicature is carried out once more for the candidate response 'bel(system, all(6278, isa(command, 6278) => numberof(6292, isa(arg_type, 6292) & take(6278, 6292), 2))'. Finally, QUVASI identifies a candidate response which licenses no false scalar inferences. 'bel(system, all(6278, isa(command, 6278) => numberof(6292, isa(arg_type, 6292) & take(6278, 6292), 3))'. So, QUVASI outputs this candidate cooperative response to the discourse planner.

Note that changing the order in which expression/poset pairs are considered produces no change in QUVASI's output:

1: DO ANY COMMANDS TAKE 2 ARGUMENTS?

The semantic representation of your query is:

```

exists(18636, isa(command, 18636) & numberof(18650,
  isa(arg_type, 18650) & take(18636, 18650), 2))

```

The direct response is: true

Please choose salient expressions and a salient ordering for

each from the following list, forming a list of lists (e.g.,

```

[exists, quantifiers], [3, ucardinals]]):
Expression      Poset
2               [cardinals, ucardinals]
&              [andor]
exists          [quantifiers]

```

1: 1: [[2, UCARDINALS], [EXISTS, QUANTIFIERS]].

If the following expressions and orderings are salient,

```

Expression      Poset
2               [ucardinals]
exists          [quantifiers]

```

The scalar implicatures that could be licensed by the assertion of:

```

bel(system, exists(18636, isa(command, 18636) &
  numberof(18650, isa(arg_type, 18650) & take(18636, 18650),
  2)))

```

are:

```

[-bel(system, all(18636, isa(command, 18636) =>
  numberof(18650, isa(arg_type, 18650) & take(18636, 18650),
  2))), ~bel(system, all(18636, isa(command, 18636) =>
  numberof(18650, isa(arg_type, 18650) & take(18636, 18650),
  3))), ~bel(system, all(18636, isa(command, 18636) =>
  numberof(18650, isa(arg_type, 18650) & take(18636, 18650),
  4))), ~bel(system, all(18636, isa(command, 18636) =>
  numberof(18650, isa(arg_type, 18650) & take(18636, 18650),
  5))), ~bel(system, exists(18636, isa(command, 18636) &
  numberof(18650, isa(arg_type, 18650) & take(18636, 18650),
  3))), ~bel(system, exists(18636, isa(command, 18636) &
  numberof(18650, isa(arg_type, 18650) & take(18636, 18650),
  4))), ~bel(system, exists(18636, isa(command, 18636) &
  numberof(18650, isa(arg_type, 18650) & take(18636, 18650),
  5)))]

```

The following licensable implicature is inconsistent with the system KB:

```

~bel(system, all(18636, isa(command, 18636) =>
  numberof(18650, isa(arg_type, 18650) & take(18636, 18650),

```

2)))
The scalar implicatures that could be licensed by the assertion

```
of:
del(system, all(_18636, isa(command, _18636) =>
numberof(_18650, isa(arg_type, _18650) & take(_18636, _18650),
2)))
```

```
are:
[-bel(system, all(_18636, isa(command, _18636) =>
numberof(_18650, isa(arg_type, _18650) & take(_18636, _18650),
3)))
-bel(system, all(_18636, isa(command, _18636) =>
numberof(_18650, isa(arg_type, _18650) & take(_18636, _18650),
4)))
-bel(system, isa(arg_type, _18650) & take(_18636, _18650),
5)))]
```

The following licensable implicature is inconsistent with the

```
system KB:
-bel(system, all(_18636, isa(command, _18636) =>
numberof(_18650, isa(arg_type, _18650) & take(_18636, _18650),
3)))
```

The scalar implicatures that could be licensed by the assertion

```
of:
del(system, all(_18636, isa(command, _18636) =>
numberof(_18650, isa(arg_type, _18650) & take(_18636, _18650),
3)))
```

```
are:
[-bel(system, all(_18636, isa(command, _18636) =>
numberof(_18650, isa(arg_type, _18650) & take(_18636, _18650),
4)))
-bel(system, all(_18636, isa(command, _18636) =>
numberof(_18650, isa(arg_type, _18650) & take(_18636, _18650),
5)))]
```

This response is consistent with the system KB.

Output this cooperative response to the query:

```
del(system, all(_18636, isa(command, _18636) =>
numberof(_18650, isa(arg_type, _18650) & take(_18636, _18650),
3)))
```

The scalar implicatures identified in the above exchange and the candidate responses proposed are identical to those of the previous exchange. This demonstrates that the calculation of scalar implicatures is not sensitive to the ordering of salient expressions.

7.2.8.2 Changing Salient Expression Changes Output

Although the ordering of salient expressions does not change the calculation of scalar implicatures, the selection of one expression as salient over another certainly does alter the implicatures that will be identified and the cooperative responses suggested by QUASI. Consider:

! : DO ANY COMMANDS TAKE 2 ARGUMENTS?

The semantic representation of your query is:

```
exists(_21330, isa(command, _21330) & numberof(_21344,
isa(arg_type, _21344) & take(_21330, _21344), 2))
```

The direct response is: true

Please choose salient expressions and a salient ordering for each from the following list, forming a list of lists (e.g., [[exists, quantifiers], [3, ncardinals]]):

```
Expression      Poset
2                [cardinals, ncardinals]
&                [andor]
exists           [quantifiers]
```

! : [[EXISTS, QUANTIFIERS]].

If the following expressions and orderings are salient,

```
Expression      Poset
exists           [quantifiers]
```

The scalar implicatures that could be licensed by the assertion

```
of:
del(system, exists(_21330, isa(command, _21330) &
numberof(_21344, isa(arg_type, _21344) & take(_21330,
_21344), 2)))
```

are:

```
[-bel(system, all(_21330, isa(command, _21330) =>
  numberof(_21344, isa(arg_type, _21344) & take(_21330,
    _21344), 2)))]
```

The following licenseable implicature is inconsistent with the system KB:

```
-bel(system, all(_21330, isa(command, _21330) =>
  numberof(_21344, isa(arg_type, _21344) & take(_21330,
    _21344), 2)))]
```

The scalar implicatures that could be licensed by the assertion of:

```
bel(system, all(_21330, isa(command, _21330) =>
  numberof(_21344, isa(arg_type, _21344) & take(_21330,
    _21344), 2)))]
```

are:

```
[1]
```

This assertion is consistent with the system KB.

Output this cooperative response to the query:

```
bel(system, all(_21330, isa(command, _21330) =>
  numberof(_21344, isa(arg_type, _21344) & take(_21330,
    _21344), 2)))]
```

When a quantifier ordering is salient, QUASI will suggest 'bel(system, all(_21330, isa(command, _21330) => xnumberof(_21344, isa(arg_type, _21344) & take(_21330, _21344), 2))' as a cooperative response. However, when the cardinals are salient, QUASI will calculate a different cooperative response:

```
! : DO ANY COMMANDS TAKE 2 ARGUMENTS?
```

The semantic representation of your query is:

```
exists(_26492, isa(command, _26492) & xnumberof(_26506,
  isa(arg_type, _26506) & take(_26492, _26506), 2))
```

The direct response is: true

Please choose salient expressions and a salient ordering for each from the following list, forming a list of lists (e.g., [[exists, quantifiers], [3, ucardinals]]):

```
Expression      Poset
2               [dcardinals, ucardinals]
&              [and/or]
exists          [quantifiers]
```

```
! : [[2, UCARDINALS]].
```

If the following expressions and orderings are salient,

```
Expression      Poset
2               [ucardinals]
```

The scalar implicatures that could be licensed by the assertion of:

```
bel(system, exists(_26492, isa(command, _26492) &
  xnumberof(_26506, isa(arg_type, _26506) & take(_26492,
    _26506), 2)))]
```

are:

```
[-bel(system, exists(_26492, isa(command, _26492) &
  xnumberof(_26506, isa(arg_type, _26506) & take(_26492,
    _26506), 3)))]
& xnumberof(_26506, isa(arg_type, _26506) & take(_26492,
  _26506), 4))]
-bel(system, exists(_26492, isa(command, _26492)
  & xnumberof(_26506, isa(arg_type, _26506) & take(_26492,
    _26506), 5)))]
```

The following licenseable implicature is inconsistent with the system KB:

```
-bel(system, exists(_26492, isa(command, _26492) &
  xnumberof(_26506, isa(arg_type, _26506) & take(_26492,
    _26506), 3)))]
```

The scalar implicatures that could be licensed by the assertion of:

```
bel(system, exists(_26492, isa(command, _26492) &
  xnumberof(_26506, isa(arg_type, _26506) & take(_26492,
    _26506), 3)))]
```

are:

```
[-bel(system, exists(_26492, isa(command, _26492) &
```

```

xnumberof(_26506, isa(arg_type, _26506) & take(_26492,
_26506), 4))) , ~bel(system, exists(_26492, isa(command, _26492,
& xnumberof(_26506, isa(arg_type, _26506) & take(_26492,
_26506), 5))))]

```

The scalar implicatures that could be licensed by the assertion
of:

```

bel(system, exists(_26492, isa(command, _26492) &
xnumberof(_26506, isa(arg_type, _26506) & take(_26492,
_26506), 3)))

```

are:

```

[~bel(system, exists(_5451, isa(command, _5451) &
xnumberof(_5465, isa(arg_type, _5465) & take(_5451, _5465),
4))) , ~bel(system, exists(_5451, isa(command, _5451) & xnum
berof(_5465, isa(arg_type, _5465) & take(_5451, _5465), 5)))]

```

This assertion is consistent with the system KB.

Output this cooperative response to the query:

```

bel(system, exists(_26492, isa(command, _26492) &
xnumberof(_26506, isa(arg_type, _26506) & take(_26492, _26506),
3)))

```

7.2.8.3. When Several Responses are Cooperative

In some cases, more than one response to a query may be cooperative, in the sense that QUASI predicts it will license no scalar implicatures inconsistent with the KB. One such case is illustrated below:¹⁷⁰

```

! : IS MESSAGES NONINTERACTIVE AND IS SWITCH NONINTERACTIVE?
The semantic representation of your query is:
noninteractive(messages) & noninteractive(setswitch)

```

The direct response is: false

Please choose salient expressions and a salient ordering for each from the following list, forming a list of lists (e.g., [[exists, quantifiers], [3, cardinals]]):

¹⁷⁰Where '#' designates logical disjunction and '&', conjunction.

| | |
|-------------|-------------------|
| Expression | Poset |
| & | [andor] |
| [messages] | [switch_commands] |
| [setswitch] | [switch_commands] |

! : ! : [[&, andor]].

If the following expressions and orderings are salient,

| | |
|------------|---------|
| Expression | Poset |
| & | [andor] |

The scalar implicatures that could be licensed by the assertion
of:

```

bel(system, ~(noninteractive(messages) &
noninteractive(setswitch)))

```

are:

```

[~bel(system, ~(noninteractive(messages) &
noninteractive(setswitch)))]

```

This response is consistent with the system KB.

The scalar implicatures that could be licensed by the assertion
of:

```

bel(system, noninteractive(messages) &
noninteractive(setswitch))

```

are:

```

[~bel(system, noninteractive(messages) &
noninteractive(setswitch))]

```

This response is consistent with the system KB.

Choose one of these candidate cooperative responses:

```

bel(system, noninteractive(messages) &
noninteractive(setswitch))
bel(system, ~(noninteractive(messages) &
noninteractive(setswitch)))

```

this exchange, where the user has chosen the logical connectives as salient. QUASI finds that no candidate responses license no misleading scalar implicatures and outputs both. In each case, QUASI presents those implicatures that are licensable via the utterance of each.

2.9. Outputs

Examples of system output when QUASI is in trace mode have been presented above. Then not in trace mode, QUASI returns only a set of candidate cooperative responses to the minimal.¹⁷¹ The first example, when QUASI is not in trace mode, thus runs as follows:

```

: ARE ANY SWITCHES BOOLEAN?
Please choose salient expressions and a salient ordering for
each from the following list, forming a list of lists (e.g.
[exists, quantifiers], [3, uecardinals])):
expression      Poset
lists            [andor]
                 [quantifiers]
: | : [EXISTS, QUANTIFIERS]].
al(system, all(_218, isa(switch, _218)) => boolean(_218)))

```

Given the expressions and posets selected as salient, each member of this set satisfies the criterion of not licensing misleading implicatures. From the returned set of candidate appropriate responses plus information external to QUASI, it is intended that the discourse manager select an actual response. The system also outputs the scalar implicatures which each response (if output) might license. If QUASI were embedded in a larger natural-language question-answering system, the choice of one of these outputs should trigger the marking of the corresponding set of licensable implicatures as communicated to the user implicitly.

2.10. Future Research

As noted above, some aspects of the theory presented in previous chapters are not implemented in QUASI. I have mentioned some of these in the discussion above. Too, although classification hierarchies are represented in the switches domain, no attempt has been

¹⁷¹Other information is of course available to the discourse planner.

made to limit salient hierarchies using the 'entry level' notion discussed in Section 6.3.2.1, in large part because the hierarchies modeled here are not infinite and because there is no empirical evidence allowing us to identify entry levels in an electronic mail domain. At present, QUASI generates only semantic representations of candidate responses, since this module is not actually suggested as a stand-alone natural-language interface. However, for demonstrational purposes, it might be desirable to return actual English output.

7.2.11. Summary

The system described above demonstrates that the theory of scalar implicature presented in this thesis is both computational and of some potential use for natural-language generation. Whether or not the approach to response generation suggested above has any cognitive reality, it provides a convenient fiction which accounts for naturally occurring data in a reasonable way. However, the real computational benefit which results from the calculability of scalar implicature is a more general one: the calculation of what may be conveyed and what may be inferred — beyond the literal propositional content of generated or interpreted text.

In this application, I have chosen to focus on the use of a theory of scalar implicature in natural-language generation and have limited this application to the generation of cooperative responses to yes-no questions. However, the formalism and, indeed, the core of the implementation may easily be adapted to other uses in generation — and to the interpretation of user input. The anticipation of potential scalar implicatures licensable via system output exactly parallels the analysis of implicatures possibly licensed by user input. Even in systems that do not support natural-language interaction, the notion of scalar implicature proposed here should prove useful, since these inferences are analyzed from formal language representations of natural language input.

CHAPTER VIII

Conclusions

In this thesis I have outlined a computational approach to the study of conversational implicature. I have identified one type of generalized conversational implicature, scalar implicature, and demonstrated how these implicatures can be calculated. I have further shown how the calculation of conversational implicatures can be applied to a particular problem in natural-language question-answering, the specification of cooperative responses to yes-no questions.

In Chapter 2, I described obstacles to a definition of conversational implicature, but argued for a set of necessary and sufficient conditions on this phenomenon based upon speaker intention, mutual belief in speaker cooperativity and in hearer ability to 'work out' an inference, as well as the cancellability, nondetachability, and reinforcingability of that inference. In Chapter 3, I proposed a reinterpretation of the class of GENERALIZED CONVERSATIONAL IMPLICATURES, which include what have previously been identified as generalized quantity implicatures. I then examined attempts to define such quantity implicatures by Horn, Hamish, and Gazdar and the obstacles which have stymied them. I showed that p_j represents a quantity implicature when the 'working out' of p_j is dependent upon speaker's and hearer's perception of some salient metric ordering speaker's utterance with other potential utterances. Finally, I introduced scalar implicature and distinguished it from these earlier definitions of quantity implicature. In particular, I showed how scalar implicatures are supported by perception of a wider variety of orderings, including set inclusion, whole/part, entity/attribute, and type/subtype, as well as orderings, including set inclusion, whole/part, entity/attribute, and type/subtype. Scalar implicature is defined not only for the affirmation of values in such an ordering, but for the denial or assertion of ignorance regarding such values. Scalar implicatures may represent inferences about higher, lower, or alternate values to those mentioned. From an intuitive description of this phenomenon, I proposed a set of scalar implicature conventions which, for a given utterance and given salient orderings, permit the anticipation or interpretation of scalar implicatures.

In Chapters 4-6, I showed how these conventions may be used to calculate licensable scalar implicatures. Chapter 4 dealt largely with representational issues: I adapted Gazdar's syntactic approach to the ranking of utterances via the presence in their semantic representations

of certain rankable expressions and discussed the epistemic force appropriately assigned to licensed implicatures. In Chapter 5, I first described the varied types of ordering relations that can support the ranking of utterances for the licensing of scalar implicature. I next argued that all and only those relations that can be modeled as partial ordering relations or POSSETS support scalar implicature. Partial orderings of higher, lower, and alternate expressions can inurectly order semantic representations of utterances similarly, so that the appropriate scalar implicature convention can be accessed.

In Chapter 6 I related the conventional aspects of scalar implicature to the interpretation of particular inferences in context. I proposed a definition of ordering COMPARABILITY and suggest how this concept may be used to relax the requirement that speaker believe some partial ordering is mutually believed salient for speaker and hearer. To identify salient expressions and salient partial orderings in a discourse, I proposed use of FOCUS/TOPIC/CENTERING information and suggested how current work on these phenomena might be adapted/extended to this end. I examined certain strategies for marking and identifying focus, including syntactic, prosodic, and semantic-pragmatic cues. I also proposed ways to limit potentially infinite orderings to locally salient orderings using concepts of BASIC and ENTRY LEVEL for classification hierarchies and ways to distinguish between duals when one is believed salient. I also suggested possibilities for future research in the identification of salient orderings. I extended the scalar implicature formalism to accommodate scalar implicatures arising from utterances in which more than one expression is salient.

In Chapter 7, I described an application of the theory defined in previous chapters to the calculation of indirect responses to yes/no questions. I described QVASI, a natural-language interface to a database which generates indirect and qualified direct responses to yes/no questions about switches in a computer mail system. Given a semantic representation of the desideratum of a yes/no question, QVASI retrieves direct responses from the knowledge base and calculates the scalar implicatures that could be licensed by their generation. Where any such inferences are inconsistent with the system's knowledge base, QVASI calculates alternative responses that will not license such false inferences.

While I have demonstrated one practical use for the calculation of one type of conversational implicature, the undemonstrated uses are clearly even more significant. If we can calculate conversational implicature, then we can access a heretofore inaccessible level of utterance meaning computationally. For natural-language understanding systems, this means that systems can understand far more from user input than they have previously understood. For natural-language generation systems, it means they can anticipate far more accurately the effect that system output will have on its audience. And, in less practical terms, it means that we are one step closer to understanding utterance meaning.