IMPRESSION MONITORING IN EVALUATION-ORIENTED DIALOG
The Role of the Listener's Assumed Expectations and Values in the Generation of Informative Statements

Anthony Jameson

Department of Social Psychology, University of Nijmegen
6500 HE Nijmegen, The Netherlands

ABSTRACT

A prototype dialog system is presented which specializes in responding to the questions of a user who is assumed to be attempting to form an evaluation of a given object. On the basis of explicit assumptions concerning the evaluator's standards and prior expectations, the system goes beyond the direct answering of the questions by selecting additional comments according to their anticipated impact on the evaluator's impressions of the object. The system may be positively *or* negatively biased in its selection of comments; taking into account the (possibly different) bias which it assumes the evaluator to ascribe to it, it anticipates how the fact that it has failed to make certain comments is likely to be interpreted. The system's central concepts are also used to quantify the notion of the relatedness of a given comment to a given topic and to guide the selection of connectives and sentential adverbs.

> A small boy in Texas had the habit of asking strangers where they were from. "You shouldn't ask them that," said his father. "If a man's from Texas, he'll tell you so himself. If he's from anywhere else, he'll be ashamed if you ask."

The research reported here was conducted in part at the Research Unit for Information Science and Artificial Intelligence at the University of Hamburg, which is supported by the German Ministry for Research and Technology. The author is indebted to the members of that group, as well as to Adam Mehrdad, for helpful discussions.

This anecdote illustrates three points which should be taken into account in the design of any dialog system which supplies the user with information which may be relevant to the evaluation of a particular person or object:

1. Human speakers frequently volunteer unsolicited comments which have consequences for the evaluation of an object under discussion.
2. The selection of such comments is influenced by the nature *of* the speaker's bias, e.g. a desire to present the object in a favorable light.
3. Listeners take these facts into account when interpreting such comments *(or* their absence) - so much so, in fact, that it is wise for the speaker to anticipate such interpretations.

The present paper describes IMP, a dialog system designed specifically as a restricted prototype which permits the exploration *of* these topics without the overhead associated with an NL system of broad capabilities. The system engages in what may be called evaluation-oriented dialog: It takes the role of an informant responding to the questions of an evaluator whose sole dialog goal is assumed to be the assessment of some object, e-g- with a view to making a decision concerning it (cf. the example dialog in Figure 1).

Dialogs of this general sort occur in many everyday situations - e.g. personnel selection interviews and discussions with salespeople. The evaluation-oriented dialogs which IMP can handle are quite restricted in terms of the number of speech act types and dialog goals involved, as can be seen from the description of its top-level control procedure in Figure 2.* This simplification makes it unnecessary for the system to reason explicitly about dialog goals and ways *of* realizing them (cf., e.g. Cohen, 1978), although the system does consider carefully what impact its utterances will have upon the listener.

1. What about facilities outside the room	Open-ended question; no direct answer possible.
THERE'S A KITCHEN	Availability of oven not important enough to warrant more complex comment.
AND THERE'S A BATHTUB	Prefaced with and because resulting impression shift is in same direction as previous one. Presence of washing machine not worth mentioning; it corresponds more or less to expectations.
AND BY THE WAY THE ROOM IS VERY LARGE	Though not directly related to question topic, worth mentioning because of large anticipated impression change. (If the system were not positively biased, it would have likewise spontaneously mentioned that the room is noisy.)
Is there a washing machine and a dryer	More specific question, direct answer possible.
NO. BUT THERE IS A WASHING MACHINE	Anticipated to produce slight downward impression shift.
	This comment now worth making because of narrower question scope, should shift impression back upward.
Is the room on a quiet street?	Direct answer does not betray how noisy room is.
NO. IN FACT UNFORTUNATELY THERE'S A LOT OF NOISE FROM THE STREET	System would prefer to make no additional comment but sees that silence would be inconsistent with its projected objectivity.
Is there a lot of noise from other rooms	Question concerning closely related topic.
WHY NO	Expression of surprise at question (cf. Lakoff, 1973): If this were the case, an objective informant would already have mentioned the fact when responding to the previous question. (To be sure, the system itself would not have done so, because of its positive bias.)

figure 1. Unedited dialog in which IMP (output in capital letters) responds to questions concerning a particular room offered for rent. (The system's bias is positive here, but it attempts to maintain a neutral image.)

```
Engage in an evaluation-oriented dialog

Background:
• an object being evaluated ('the object[1]')
•a feiafi ('the actual bias') which characterizes
   the system
•a bias ('the projected bias') which the system
   assumes the evaluator to attribute to it and
   will attempt to maintain
Working Structures:
  Question, Answer: each a sequence of English
   words
Basic Method:
 •Keep repeating:
     •Assign the next input string as the
       question.
     •Try to determine a direct answer to it
     • If this fails, volunteer comments on the
       basis of the question.
     •If it succeeds then:
        •Update the record of the evaluator's
          specific expectations according to the
          answer's meaning [as described in
          Figure 6]
        •If this update results in no change, add
          'why' at the beginning of the answer
          [since it ought to have been known
          already].
        •Output the answer.
        •Volunteer additional comments on the
          basis of the question.
Classes:

   Describes: a direction in which an informant
    attempts to shift the Impressions of an
    evaluator
   All Instances: 'positive', 'negative',
    'objective'
```

Figure £. DL description of IMP's top-level control procedure.

THE REPRESENTATION OF VALUES AND EXPECTATIONS

IMP is supplied in advance of each dialog with an evaluation form (Figures 3 and 4), which can be viewed as a particularly explicit variant of a type of scale which one frequently sees used in practical contexts which require systematic evaluations (e.g. the refereeing of conference papers). The system behaves as if it believed that the evaluator possessed this same form and was asking questions in order to fill it in so as to be able to estimate a numerical rating for the object along the dimensions corresponding to the various scales. It is assumed that the evaluator will not in general ask questions about all of the items in the form, but will use the frequency information in it to make estimates concerning the scales not covered.

The basic idea underlying this metaphorical conception of the evaluator's judgmental processes is that the subjective evaluation of an object is an additive function of the perceived value and likelihood of its possession of various attributes.** This idea fits into a long tradition in normative and descriptive research on attitude and judgment (see, e.g., Fishbein & Aizen, 1975). It could be made more realistic if results concerning, e.g., the way people deal with small probabilities were taken into account but it is doubtful that such changes would noticeably improve the appropriateness of IMP's responses or increase its validity as a model of a naive informant.

* Winograd's (1983) language DL is used here (somewhat elliptically) to describe the essential workings of the system. Procedures whose internal structure is not relevant to the issues of interest here are not described; most of these are realized in IMP using ad hoc methods which make heavy use of specific hand-coded database entries.

Invocations of such procedures are underlined in the figures and marked with an & in the right margin; all other underlined concepts are introduced in the same figure or in the figure whose number appears to the right.

As actually implemented, the program comprises 50 LISP/FUZZY procedures and requires several seconds to respond to a question.

	Value	Frequency	Condition	Relative Importance
1. SIZE				.53
x A	+100	10%	More than 26 m2.	
B	+50	20%	More than 22 m2.	
C	0	40%	More than 17 m2.	
D	-50	20%	More than 13 m2.	
E	-100	10%	13 m2 or less.	
2. NOISE				37
2.1 NOISE FROM THE STREET				.32
A	+30	25%	Less than average.	
B	0	50%	Average.	
x C	-60	25%	More than average.	
2.2 NOISE FROM OTHER ROOMS				.19
A	+15	25%	Less than average.	
x B	0	50%	Average.	
C	-40	25%	More than average.	
3. FACILITIES OUTSIDE THE ROOM				.49
3.1 COOKING FACILITIES				.40
x A	+50	30%	Kitchen with oven.	
B	+30	10%	Kitchen without oven.	
C	0	20%	Cooking corner.	
D	-30	20%	Hotplate in the room.	
E	-60	20%	No cooking facilities.	
3.2 BATHING FACILITIES				.20
x A	+30	10%	Bathtub and shower.	
B	0	80%	No bathtub, but shower.	
C	-60	10%	Only a sink.	
3.3 LAUNDRY FACILITIES				.20
A	+30	5%	Washing machine and dryer.	
x B	+20	55%	Washing machine, no dryer.	
C	-10	20%	Only a laundromat nearby.	
D	-30	20%	None.	
4. BRIGHTNESS . . . [three further scales]				

Figure 3. Part of a brief evaluation form for moderately priced rooms offered for rent in Hamburg. (The possibilities marked with x are those realized by the room of Figure 1's dialog.)

```
SCALE
  Describes: a scheme for rating objects within
   a given domain
  Kinds of Scale: Item, Composite scale
  Roles:
    Importance: a nonnegative integer
     reflecting the total initial uncertainty
     concerning the item(s) within the scale.
     If it is:
      •an item, then the uncertainty of the
        evaluator's initial impression
        concerning it
      •a composite scale, then the square root
        of the sum of the importances of its
        own scales
  Classes:

  ITEM
    Describes: a scheme for assigning a
     number of points to an object with
     respect to a particular aspect of it
    A Kind of: Scale
    Roles:
      Possibilities: a set of possibilities,
       exactly one of which must be realized
       by any object
      Remaining possibilities: the subset of
       the possibilities which have not yet
       been ruled out by the evaluator;
       initially all of them, updated during
       the dialog
      Importance: ... [see above]
    Classes:

    POSSIBILITY
      Describes: an evaluation-relevant
       condition
      Roles:
        Condition: a predicate
         applicable to objects in the
         relevant domain
        Value: a number of points to be
         added to the rating of the
         object if the condition is true
         of it
        Frequency: the relative
         frequency of objects in the
         relevant domain which satisfy
         the condition

  ... [A composite scale is basically just a
       set of scales.]
```

Figure 4. The structure of an evaluation form.

** The values specified are conceived as lying on a single interval scale. It is presupposed that the items are independent of each other in the sense that neither the values nor the frequencies associated with the possibilities for a given item depend on what possibilities are realized for other items.

The evaluation form of Figure 3 is an abbreviated version of one written largely intuitively by the author. The general values and expectations it expresses are ascribed by IMP to any evaluator who questions it on this topic and are not revised on the basis of the evaluator's behavior during the dialog. For use in future research, a wide variety of more empirical techniques are conceivable for obtaining scales for particular user groups or even for individual users, building, e.g. on the work of Rich (1979) or on techniques developed by decision analysts (surveyed by Slovic, Fischhoff, & Lichtenstein, 1977).

SELECTING COMMENTS - OVERALL STRATEGY

Figure 5 shows how in IMP the task of generating appropriate unsolicited comments is conceived as a search through the evaluation form for items which warrant some comment other than the pseudo-comment '•silence*'. This search could undoubtedly be made more efficient and plausible if additional heuristics were introduced to make it more selective, e.g. by making early reference to the items' overall importance, which is at present used only for the detailed rating of individual comments (cf. Conklin & McDonald, 1982).

```
Volunteer comments on the basis of a question

Background: a scale for evaluating objects in        4
 the domain in question ('the evaluation form')
Working Structures:
  Question topic: the scale within the
   evaluation form which the question concerns       X
Basic Method:
 • Order the items in the evaluation form
   according to their question-relatedness.
 • Step through the resulting sequence of items
   doing:
    • Select a comment on the item                   6
    • If the form of the result is not               8
      '*silence*' then add a connective or
      adverb to it and output the result.            X
Conditions: Except during the first invocation
 of this procedure, terminate the iteration
 through the items as soon as one is encountered
 whose question-relatedness is below a certain
 threshhold [i.e., no further comment other than
 '*silence*' is likely to be selected].
Procedures:

  Compute the question-relatedness of an item

  Basic Method: Return the ratio of
   • the importance of the question topic to         4
   • the importance of the lowest scale in the       4
     evaluation form which dominates both the
     question topic and this item.
```

Figure 5.

```
Select a comment on an item

Purpose: Produce a true comment on the item
 which affects the evaluator's impressions in a
 desirable way while remaining consistent with
 the system's projected bias
Basic Method:
 • Generate a set of possible comments on the         X
   item such that each one's literal                  6
   interpretation includes the possibility which      4
   is actually realized by the object.                X
 • Keep repeating:
   • Determine the comment with the highest           8
     rating for this possibility given the            10
     system's actual bias.                            2
   • Anticipate the pragmatic interpretation of
     this comment.
   • If this interpretation is empty [i.e. the
     comment is inconsistent with the projected       2
     bias], then:
     • If there are other possible comments
       left, continue with the comment with
       the next highest rating.
     • Otherwise admit that the actual bias is
       different from the projected bias and
       terminate the dialog.
   • If the interpretation is not empty, then:
     • Assign it as the set of remaining
       possibilities for the item [for future         4
       reference].
     • Return the comment as the result.
Procedures:

  Anticipate the evaluator's pragmatic
   interpretation of a comment (C)

  Purpose: Produce an interpretation which takes
   into consideration the alternative comments
   which might have been made.
  Background: a projected bias                        2
  Working Structures:
    Pragmatic interpretation: a set of
     possibilities, initially the literal             4
     interpretation of C                              6
  Basic Method:
   • For each possibility in the literal
     interpretation of C do:
     • Generate a set of possible comments which      X
       would be literally true for this
       possibility.
     • Determine the one with the highest rating      10
       for this possibility assuming the
       projected bias.
     • If this rating exceeds that of C for the
       same possibility and the same bias by a
       large interval [here: at least 30], then
       remove this possibility from the
       pragmatic interpretation.
   • Return the pragmatic interpretation.
```

Figure 6.

Figure 6 describes the processing performed for each item in the form that the system inspects. IMP has no interesting methods for constructing new comments on a given item; of interest is only the way in which it selects the most appropriate one from a set of candidates which is assumed to be representative of the infinite set of comments that

```
Choosing comment on Item 2.1, NOISE FROM THE STREET
Remaining        Value   Frequency  Condition              Uncertainty:     28   Relative Importance:    .32
Possi-      B      0       50%      Average.               Expected Value  -20   Question-Relatedness: 1.00
bilities: x C    -60       25%      More than average.
                                         Expected          1.80    x    + 1.80 x    - 1.00 x    =
       Possible                Literal   Value        Desirability   Uncertainty                  Rat-
       Comment              Interpretation After       of Shift       Reduction      Effort        ing
1. *SILENCE*                     (B C)    -20             0              0             0            0
2. (THERE'S A LOT OF NOISE FROM THE STREET)
                                   (C)    -60           -40             28            50          -72
   How would *SILENCE* be interpreted, assuming objectivity? . . .
      For possibility B they'd expect:
   (THERE'S NOT PARTICULARLY MUCH NOISE FROM THE STREET)
                                   (B)      0           +20             28            50           36
      For possibility C they'd expect:
   (THERE'S A LOT OF NOISE FROM THE STREET)
                                   (C)    -60           +40             28            50           72
      So it would be inconsistent with my projected bias.
   How would (THERE'S A LOT OF NOISE FROM THE STREET) be interpreted? . . .
      Literally.    OK.    Reducing the remaining possibilities from (B C) to (C) for Item 2.1.
>>> IN FACT UNFORTUNATELY THERE'S A LOT OF NOISE FROM THE STREET
```

Figure 7. Excerpt from IMP's trace for its third response in Figure 1, showing why it made an additional comment even though silence appeared at first to be the most attractive alternative. (Trace edited to increase legibility.)

could conceivably be made. Its overall strategy is first to select the comment which, if interpreted literally, would have the most desirable effect on the evaluator's impressions, and then to take into account the way this comment would in fact be interpreted. The system anticipates this pragmatic interpretation both in order to keep track of the evaluator's changing beliefs concerning the object and to help filter out comments which would betray its bias (as shown in Figure 7).

RATING POSSIBLE COMMENTS

Impression Changes

The kernel of IMP is the criteria which it uses to assign a numerical rating to a possible comment, assuming that the comment will be interpreted literally. The central factor which it takes into account is the comment's anticipated impact on the evaluator's corresponding impression of the object. (Figures 8 and 9). The motivation underlying the concepts of expected value and uncertainty introduced here can best be seen if one considers how changes in these quantities are related to the dialog goals of an informant.

On the one hand, any informant is likely to be interested in shifting the favorability of the evaluator's impression in some direction - either generally upward or generally downward (if the informant is biased) or in the direction of the truth (if the goal is to present an accurate profile of the object). In any case, the chance in expected value should enter into the formula for rating a comment (Figure 10).

On the other hand, even a comment which merely confirms that an expected value is in fact accurate may be considered worth making if considerable uncertainty previously existed with respect to the item. A reduction in uncertainty puts the evaluator in a better position to make decisions anticipates questions that might otherwise *have* to be asked and answered separately, and is generally felt to be a part of polite cooperative dialog behavior

COMMENT

Describes: a possible statement
Roles:
Form: a sequence of words or the symbol '*silence*'
Effort: a nonnegative integer, 0 for '*silence*'
Topic: the item it concerns — 4X
Literal interpretation: the subset of the item's remaining possibilities which are not ruled out when the comment is interpreted literally — 4 X
Impression before: an impression with item = the topic and possibilities considered = the topic's remaining possibilities
Impression after: an impression with item = the topic and possibilities considered = the literal interpretation
Classes:

IMPRESSION

Describes: an evaluator's expectations concerning a particular aspect of a specific object
Roles:
Item: the item corresponding to the relevant aspect of the object — 4
Possibilities considered: a subset of the item's possibilities — 4
Frequency distribution: a set of value-frequency pairs corresponding to those in the set of possibilities considered
Expected value: the mean of the frequency distribution
Uncertainty: the standard deviation of the frequency distribution

Figure 8.

The relative weight assigned to these two factors by a given informant can in principle vary greatly as a function of various features of the dialog situation. At present, they are given equal weight, so that they both have noticeable effects on the system's behavior. Their sum is regarded as the 'benefit' which would result from making the comment in question.

Effort

If the system is not to break into tedious monolog after each question, it must have some notion of the effort involved in making and understanding a comment. The simplest solution would be for it to subtract from the rating of every comment except '"silence*" some number such as 50 (the exact one chosen depending *on* the degree of volubility desired). But this would not do justice to the fact that some utterances are more detailed, complex, roundabout, or unnatural than others (cf. Grice's [19751 Maxims of Quantity and Manner) and must thus be considered to be associated with more effort. IMP makes no contribution to the difficult problem of defining and quantifying such a notion (see, e.g., McCawley, 1978); it simply uses stored ratings which associate different degrees *of* effort with the various possible comments.

Question-Relatedness

Even when volunteering comments which have not been specifically requested, the system should give preference to comments which are directly or indirectly related to the topic of the evaluator's latest question (cf. Grice's Maxim of Relation). There are a variety of reasons why speakers tend to

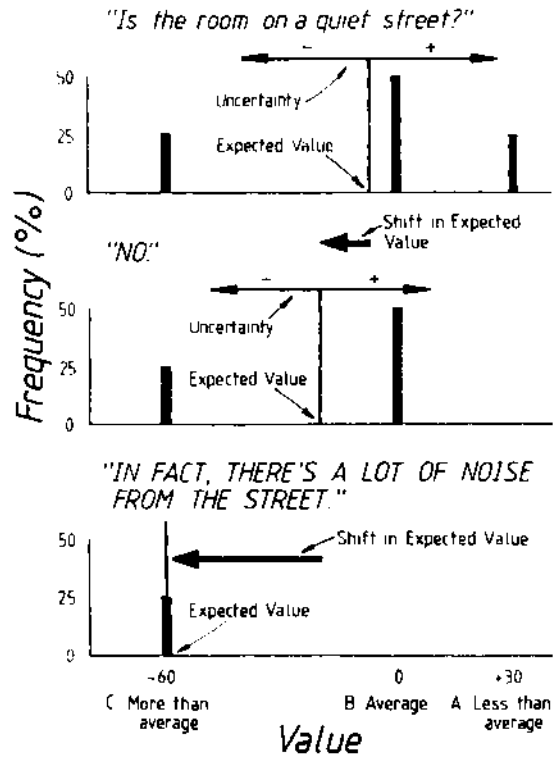

Figure 9. Illustration of the impression changes concerning the 'noise from the street' item which IMP expects to result from its third response in Figure 1.

Rate a comment (C) assuming a particular realized possibility and a bias

Purpose: Produce an integer reflecting the net benefit for the informant of making C assuming a literal interpretation

Working Structures:

AV: the actual value, i.e. the y_alii£ of the possibility assumed realized

UR: theuncertainty reduction, i.e. the arithmetic difference between the respective uncertainties of C's impression before and impression after

DIS: the desirability of the impression shift, an integer describing the relationship between the expected yaliffi of C's impression before (EV1) and impression after (EV2): If the bias is;

- ••Positive', then EV2 - EV1 [upward shifts are desirable]
- •'Negative', then EV1 - EV2 [downward shifts are desirable]
- • 'Objective', then lEV1 -AVI - EV2 - AVI [reductions in inaccuracy are desirable]

Effort: the effort associated with C.

Rel-Imp: the relative importance of C's topic, i.e. the ratio of its importance to that of the entire evaluation form.

QR: the question-relatedness of C's topic

Basic Method:

Return the following value:

$$\sqrt{\frac{QR}{\text{Rel-Imp}}} \times (UR + DIS) - \frac{\text{Effort}}{\sqrt{QR}}$$

Figure 10.

do this. For example, when a question has been asked about a given topic which, according to the evaluation form, is relatively unimportant, this fact suggests that the present evaluator may attach more importance to this topic than was originally expected. Although this need not be the case, it may be worthwhile to devote increased attention to such topics in order to take the possibility into account.

IMP's criteria for choosing comments are designed in such a way that the system's ratings will satisfy several constraints which appear generally reasonable in view of considerations such as these.

Constraint 1. The less important the topic of a question is, the more strongly the benefits associated with a directly related comment should be 'magnified' before being weighed against the effort involved in making the comment. The relevant notion of the general importance of a topic can be captured quite naturally as in Figure 4 above in terms of the uncertainty of the evaluator's initial impression concerning it, since this reflects the likely impact that knowledge concerning the topic will have on the evaluation of an arbitrary object.

Constraint 2. This question-dependent magnification should not be so strong as to cancel out all differences in the intrinsic importance of topics. If the magnification factor used were simply the reciprocal of the topic's relative importance, the result would be be that about as many comments would be volunteered in response to unimportant questions as in response to important ones. So that the system will instead take both factors into account, the square root of this ratio is used instead (more generally, the kth power, where k lies between 0 and 1 and reflects the system's responsiveness to the question topic.

Constraint 3. The less closely a comment is related to the question topic, the less strongly its benefits should be magnified. The degree of question-relatedness of the topic of a comment can be captured quantitatively with reference to the most specific topic that the two both have some bearing on, as in Figure 5. It can be seen as reflecting the extent to which the question asked would have to be broadened before it would include the comment within its scope. Multiplying the magnification factor for the comment by its question-relatedness has the consequence that there will be no magnification at all for comments which have no specific relationship to the question.

Constraint 4. The effort associated with a comment should be considered greater the less closely it is related to the question. Although Constraint 3 takes into account the fact that unrelated topics have not been singled out for attention by the questioner, it does not do justice to the fact that such a comment shifts the evaluator's attention away from the latest topic introduced. This shift is conceived in IMP as affecting the associated effort multiplicatively, with the reasonable consequence that the system is ever less likely to choose a comment requiring a large amount of effort for a topic which is not directly related to the question.

CONNECTIVES AND SENTENTIAL ADVERBS

Whenever a sequence of evaluation-relevant statements is made, it is usual to preface each one (except perhaps the first), with words which announce to the listener in advance what sort of impression change is about to occur. The rules that IMP uses for the selection of expressions like but, by the way, and unfortunately make reference to a comparison of the shift in expected value produced by the present comment with that produced by the preceding statement, taking into account the question-relatedness of their respective topics. Some conceptual apparatus such as the one introduced above for independent reasons is required if the use of such expressions in evaluation-oriented statements is to be explicated with a satisfactory degree of precision (cf. Weydt,

CONCLUDING REMARK

When IMP's actual and projected biases diverge, as in the examples given here, its behavior smacks of deception, and it may be asked what practical consequences the deve lopment of systems of this sort might have. But as dialog systems continue to be accorded more initiative in their interactions with users, questions concerning their dialog motivation will have be investigated explicitly and publicly, whether the aim is to formulate realistic norms or simply to prevent discrepancies from arising between the systerns' actual motivation and their users' image of it.

REFERENCES

Cohen, P.R. On knowing what to say: Planning speech acts (Technical Report No. 118). Unpublished doctoral dissertation, Univ. of Toronto, Dept. of Computer Science, January 1978.

Conklin, E.J., & McDonald, D.D. Salience: The key to the selection problem in natural language generation. Proceedings of the 20th Annual Meeting of the Association for Computational Linguistics, Toronto, June 1982, pp. 129-135.

Fishbein, M., & Ajzen, I. Belief, attitude, intention and behavior. Reading, Mass.: Addison-Wesley, 1975.

Grice, H.P. Logic and conversation. In P. Cole & J.L. Morgan (Eds.), Syntax and semantics, Vol. 3: Speech acts. New York, Academic, 1975, pp. 41-58.

Lakoff, R. Questionable answers and answerable questions. In B.B. Kachru, R.B. Lees, Y. Malkiel, A. Pietrangeli, & S. Saporta (Eds.), Issues in linguistics: Papers in honor of Henry and Renée Kahane. Urbana: Univ. of Illinois, 1973, pp. 453-467.

McCawley, J.D. Conversational implicature and the lexicon. In P. Cole (Ed.), Syntax and semantics, Vol. 9: Pragmatics. New York: Academic, 1978, pp. 245-259.

Rich, E. User modeling via stereotypes. Cognitive Science, 1979, 3, 329-354.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. Annual Review of Psychology, 1977, 28, 1-39.

Weydt, H. "Immerhin". In H. Weydt (Ed.), Die Partikeln der deutschen Sprache. Berlin: de Gruyter, 1979, pp. 335-348.

Winograd, T. Language as a cognitive process. Volume 1: Syntax. Reading, Mass.: Addison-Wesley, 1983.