Natural Language Access to Data Bases: Interpreting Update Requests[1]

James Davidson and S. Jerrold Kaplan[2]

Computer Science Department
Stanford University
Stanford, CA 94305

For natural language data base systems to operate effectively in practical domains, they must have the capabilities required by real applications. One such capability is understanding and performing update requests. The processing of natural language updates raises problems not encountered in the processing of queries. These difficulties stem from the fact that the user will naturally phrase requests with respect to his conception of the domain, which may be a considerable simplification of the actual underlying data base structure. Updates that are meaningful and unambiguous from the user's standpoint may not translate into reasonable changes to the underlying data base. Update requests may be *impossible* (cannot be performed in any way), *ambiguous* (can be performed in several ways), or *pathological* (can be performed only in ways that cause undesirable side effects).

Drawing on work in linguistics and philosophy of language, we have developed a domain-transparent approach to identifying and performing "reasonable" changes in response to a user's update request, using only knowledge sources typically present in existing data base systems. A simple notion of "user model" and explanation with respect to the user's state of knowledge are central to the design. This paper describes a prototype system PIQUE *(Program for Interpretation of Query/Update in English)*, which implements this approach.

1. Introduction

Natural language is a desirable access mechanism for data base systems because it frees the user from the task of understanding the details of the data base structure. A number of systems have provided natural language query capabilities (for example, Hendrix et al. 1978); however, few of these allow the user to perform updates (changes) to the data base using natural language. (For an example of one that does allow simple updates, see Henisz-Dostert and Thompson 1974.)

The provision of update capabilities introduces problems not seen in handling queries. These problems arise because the user is phrasing his requests with respect to his view of the data base, which may be a simplification or transformation of the actual data base structure. While a well-formed query expressed in terms of the user's view of the data base will always result in the same answer, regardless of how the query may be mapped into the actual data base structure for execution, this is not the case for an update expressed on a view.

Since updates request modification of the content of the data base, different mappings of the update request into the actual data base structure may result in different effects. Some of these effects may be undesirable or unanticipated. Specifically, the user may make requests that are *impossible* (cannot be performed in any way, due to hidden constraints on the data base), *ambiguous* (can be performed in several ways), or *pathological* (can be performed only in ways that cause unanticipated side effects). While human

[1] This work is part of the Knowledge Base Management project (DARPA contract #N00039-82-C-0250), which is investigating the application of artificial intelligence techniques to issues in the field of data base management. The views and conclusions contained in this document are those of the authors and should not be interpreted as representative of the official policies, either expressed or implied, of DARPA or the U.S. Government.

[2] Authors' current address: Teknowledge Inc., 525 University Avenue, Palo Alto, CA 94301

0362-613X/83/020057-12$03.00

speakers would intuitively reject these unusual readings, a computer program may be unable to distinguish them from more appropriate ones.

For example, a simple request to "Change the teacher of CS345 from Smith to Jones" might be carried out by altering the number of a course that Jones already teaches to be CS345, by changing Smith's name to be Jones, or by modifying a "teaches" link in the data base. While all of these may literally carry out the update, they may implicitly cause unanticipated changes such as altering Jones's salary to be Smith's.

Our approach to this problem is to treat updates as requesting that the data base be put into a self-consistent state in which the request is satisfied; the problem is then to select the most desirable of (potentially) several such states. The most desirable such state is considered to be the "nearest" one to the current state (in the sense that it involves the least disruption). A set of domain-independent heuristics is used to rank the potential changes along these dimensions.

This process may be guided by various linguistic considerations, such as the difference between transparent and opaque readings of the user's request, the distinction between the sense and reference of referring expressions, and the interpretation of counterfactual conditionals.

This paper describes a system, *PIQUE*, which implements this approach by retaining a model of the user's view and considering possible methods of performing the update in light of the model. Given an update request, the system generates the set of possible changes to the underlying data base that will literally fulfill the request. These candidate changes are then evaluated as to their effects on the user's view, the underlying data base, and the data base constraints. If possible, an appropriate one is selected; otherwise an informative message is presented to the user.

2. The Problem

As a hypothetical example of the problems that can arise during updates, consider a relational data base of employees, salaries, departments, and managers, consisting of two relations:

```
Employees:
  EMP     SAL  DEPT
  Adams   30   Invntry
  White   35   Mkting
  Brown   25   Sales
  Smith   25   Sales
  Pullum  25   Sales
```

```
Departments:
  DEPT     MGR
  Sales    Jones
  Mkting   Baker
  Invntry  Fisher
```

Consider the following dialogue:

```
Q1: List the employees and their
    managers.
R1: EMP     MGR
    Adams   Fisher
    White   Baker
    Brown   Jones
    Smith   Jones
    Pullum  Jones
Q2: Change Brown's manager from Jones
    to Baker.
R2: Done.
```

The system has apparently fulfilled the user's request.

```
Q3: What is the average salary paid
    to Jones's employees?
R3: $0.
Q4: List Jones's employees.
R4: NIL
```

From these responses, the user realizes that something has gone wrong.

```
Q5: List the employees and their
    managers.
R5: EMP     MGR
    Adams   Fisher
    White   Baker
    Brown   Baker
    Smith   Baker
    Pullum  Baker
```

The user sees that the system has made two unanticipated changes – changing Smith's and Pullum's managers – in addition to the one that was requested.

From the user's point of view, his request is meaningful and unambiguous. He sees a set of values, and asks to change one of them. (He might not even know that employees and managers are linked via their departments.)

The problem lies in the fact that his update request can be performed in two ways:

a) by making the manager of the Sales department be Baker.
b) by moving Brown from the Sales department to the Marketing department;

Both of these literally fulfill the request. The system, lacking any means for deciding between these, has apparently chosen (a), making Baker the manager of the Sales department, with the unanticipated effect that two other employees have had their managers changed.

3. A More Formal Characterization

This problem can be explained more formally. Given a data base structure, define the user's *view function* **F** as the transformation that is applied to the data base to yield the conceptualization with which the user works. For instance, in the example in section 2, the view function, as defined by Q1, is a transformation consisting of a join and a projection, which is applied to the original two files to yield a single new file with only two attributes. Define the user's view as the result of applying the view function to a given state of the data base; in the example, this produces a file with five entries, as shown in R1.

A user's update request (call it **u**) is a request to update the view. In the example, the request is stated in Q2. Since the view is only 'virtual' (derived from the data), we cannot modify it directly, but must make changes to the underlying data base. Call the result of translating the update request to the data base level **T(u)**. The object is to find the change to the underlying data base that comes closest to having the desired effect on the user's view. That is, we want the translation **T(u)** that produces a revised data base such that, when the view function is applied to that data base, the result is the view requested by the user.

In graphical terms:

D represents the initial state of the data base, **D'** the state that results after applying the translated update **T(u)**;

$$\begin{tikzcd} \mathbf{F(D)} \arrow[r, "\mathbf{u}"] & \mathbf{u(F(D))} \overset{?}{=} \mathbf{F(D')} \\ \mathbf{D} \arrow[u, "\mathbf{F}"] \arrow[r, "\mathbf{T(u)}"'] & \mathbf{D'} = \mathbf{T(u)(D)} \arrow[u, "\mathbf{F}"'] \end{tikzcd}$$

In mathematical terms, the mapping **F** from the underlying data base **D** to the user's view **F(D)** induces a homomorphism. Loosely defined, a homomorphism is a function that preserves the structure of its arguments under given operations. In this case, the operations are changes to the underlying data base, and corresponding changes to the user's view. The difficulties with updates expressed on the view, rather than the underlying data base, arise from the characteristics of the inverse of this homomorphism: elements in the user's view (states of the "conceptual" data base) map under $\mathbf{F}^1$ into a set of states of the underlying data base. This set may be empty (if the view update cannot be accomplished in any way), or have many elements (in the case of a request that is ambiguous with respect to **D**). If the mapping **F** is invertible, i.e. $\mathbf{F}^1$ is also a function, then an isomorphism is induced. In this case, each requested update will have a single, unambiguous interpretation in the underlying data base, and the difficulties addressed here do not arise. However, in general this is not the case.

The ideal update translation will produce a state of the data base that, when transformed by the user's view function, exactly yields the revised state that he requested. In actuality, our implementation will consider changes to the data base that literally fulfill the user's request but may not yield precisely the intended view **u(F(D))**. In the example, there were two translations of the user's request; update (b) yielded the exact view, update (a) a different one.

4. Description of the PIQUE System

We have implemented a prototype system (*PIQUE*) that addresses this problem by processing update requests in four phases.

(1) Decide what the user's current view of the data base is.

The system maintains an ongoing model of the user's conception of the data base, derived from the dialogue.

(2) Use the view to generate a set of *candidate updates* **T(u)**, which perform the update.

When an update comes in, it is assumed to be an update to the user's view. That is, the user requests changes with respect to his conceptualization of the data base. The candidate translations are updates to the data base, each of which literally accomplishes the user's request.

(3) Use a set of ordering heuristics to rank these candidates, in terms of how accurately they fulfill the user's request.

These candidates are evaluated according to the ordering heuristics, to measure how much impact they have on the user's view. For example, a candidate that causes side effects (unrequested changes to the user's view) is ranked lower than one that does not cause such side effects. "Pragmatic" information contained in the data base schema is also used in making the decision.

(4) Take action, depending on the number of candidates and their ranking.

When the candidates have been ranked, action is taken. This might consist of performing one of the candidates, offering a choice to the user, or explaining why the update cannot be performed at all.

These phases are considered in turn.

4.1 Inferring the user's view

The user of a natural language data base system typically has a conception of the data base that is a subset of the relations, attributes, connections, and records actually present. In order to interpret updates correctly, the system must take into account the user's current conception of the data base. Our approach is to build a user model based on the concepts of which the user has indicated an awareness, those that have occurred in his queries and updates.

This is implemented by making use of the *connection graphs* corresponding to the user's inputs. A system that processes natural language inputs must find paths through the data base, defined by operations such as *joins,* which connect the concepts mentioned in the input. (The LADDER system, for example, provides this service with the help of navigation information stored in a separate *structural schema.*) This set of paths is called the connection graph.

The importance of this work is that the connection graph provides a good model for the structure of the user's view. That is, each query implicitly induces a view of the data base that the user holds, at least until the next input. When an update is received, it can be checked for compatibility with the current view, to see if it could be an attempt to update that view. This compatibility test basically checks to see whether the concepts and relationships mentioned in the update are completely contained in the view. (The actual matching criterion is more complicated than simple inclusion, but this will serve for explanatory purposes.) If the update and view are compatible, the user is assumed to be continuing an interaction with that view.

Consider the example of section 2. The user poses a query that mentions employees and their managers. He then makes an update request of a similar form. Because the update request is compatible with the view induced by the previous query, the user is assumed to be referring to that view and to be asking to change it. Note that although *departments* are needed in the connection graph, they are not mentioned by the user, and therefore do not appear in the view.

Views are *stacked* as the dialogue progresses, and updates can be checked for compatibility with all previous views (most recent first). This enables the system to correctly handle a situation in which a user returns to a previous view for further work.

Note that an update also induces a connection graph, just as a query does. If an update request is not compatible with *any* of the views defined previously, the connection graph for the update itself can be used to define the view. This occurs if the user is making an update unrelated to any of the information that he has examined. In this case, the view must be inferred from the update alone. Thus, to return to the example of section 2, "Change Brown's manager from Jones to Baker" might be meaningful even if the user has not previously asked about these things.

This strategy is conservative, in that the only concepts that will appear in views are those of which the user has indicated at least some awareness. As a result, the system will never assume a view that is more complex than the one actually held by the user, and thus will never mislead him by introducing a new concept during a response or explanation. The errors that occur will consist of underestimating the user's familiarity with the data base; the system will tend to be pedantic, rather than mysterious.

This strategy also provides a notion of *focus:* as the user discusses different parts of the data base, the view changes automatically. This is important, because the notion of *side effect* changes as the user's focus changes. Changes occurring to previous views are less important than changes occurring to the current view.

The concept of user modelling is well known in artificial intelligence (Mann et al. 1977). A common approach is to record an explicit list of the things the user knows (for example, Cohen 1978). Our model, however, is much simpler. Given the role of the view information in the inferencing heuristics, this model is adequate for our purposes. Davidson (1982) discusses the issue of modeling in more detail.

4.2 Generating candidate updates

One of the crucial steps of the algorithm described above is the generation of *candidate updates* that can then be evaluated for plausibility. In most cases, an infinite number of changes to the data base are possible that would literally carry out the request (mainly by creating and inserting "dummy" values and links). However, this process can be simplified by generating only candidate updates that can be directly derived from the user's phrasing of the request. This limitation is justified by observing that most reasonable updates correspond to different readings of expressions in *referentially opaque* contexts.

A referentially opaque context is one in which two expressions that refer to the same real world concept cannot be interchanged in the context without changing the meaning of the utterance (Quine 1971). Natural language data base updates often contain opaque contexts.

For example, consider that a particular individual (in a suitable data base) may be referred to as "Dr. Smith", "the instructor of CS100", "the youngest assistant professor", or "the occupant of Room 424". While each of these expressions may identify the same data base record (that is, they have the same *extension*), they suggest different methods for locating that record (their *intensions* differ). In the context of a data base query, where the goal is to unambiguously specify the response set (extension), the method by which they are accessed (the intension) does not normally affect the response. Updates, on the other hand, are often sensitive to the substitution of extensionally equivalent referring expressions. "Change the instructor of CS100 to Dr. Jones." may not be equivalent to "Change the youngest assistant professor to Dr. Jones." or "Change Dr. Smith to Dr. Jones." Each of these may imply different updates to the underlying data base.

For operating with an expression in an opaque context, therefore, we must consider the sense of the expression, in addition to its *referent* (Frege 1952). In a data base system, this sense is embodied in the procedure used to evaluate the referring expression; the referent is the entity obtained via this evaluation. A request for a *change* to a referring expression is thus not specifically a request to perform a substitution on the referent of the expression, but rather a request to change the data base so that the sense of the expression now has a new referent. That is, after the update, evaluating the same procedure should yield the new (requested) result.

For example, consider a data base of ships, ports, and docks, where ships are associated with docks, and docks with ports. Assume that there is currently a ship named Totor in dock 12 in Naples (and no other ship in Naples), and consider the following updates:

Change *Totor* to Pequod.
Change *the ship in dock 12* to Pequod.
Change *the ship in Naples* to Pequod.

The referring expressions (italicized) have the same referent in all three cases, but the senses differ. The expression "Totor" is resolved by means of a lookup in the *ships* relation; "the ship in dock 12" requires a join between the *ships* and *docks* relations; "the ship in Naples" requires a join between all three relations.

Consider the ways of performing each request, as indicated by the sense of the referring expression. The first version can be implemented only by making a direct substitution on the *ships* relation, corresponding to renaming the ship. The second admits this possibility, but also the possibility of moving a new ship into the dock (if there is already a ship named Pequod). The third allows the first two, plus the possibility of moving a different dock into Naples (if there is a dock somewhere else with the Pequod in it). (This will later be ruled out for other reasons, as explained in the next section, but cannot be excluded on purely linguistic grounds.)

Thus, the particular referring expression selected by the user motivates a set of possible actions that may be appropriately taken, but does not directly indicate which is intended or preferred.

This characteristic of natural language updates suggests that the generation of candidate updates can be performed as a *language driven inference* (Kaplan 1978) without severely limiting the class of updates to be examined. Language driven inference is a style of natural language processing in which the inferencing process is driven (and hence limited) by the phrasing of the user's request.

In this instance, the candidate updates are generated by examining the referring expression presented in the update request. The procedure implied by this expression follows an "access path" through the data base structure. The candidate updates computed by the program consist of changing links or pointers along that path, or substituting values in the final record(s) identified.

For example, consider the structure of the "ships" data base:

$$\begin{tikzcd} \boxed{\text{PORTS}} \arrow[r, "(a)"] & \boxed{\text{DOCKS}} \arrow[r, "(b)"] & \boxed{\text{SHIPS}} \end{tikzcd}$$

The candidate translations for the third request (changing "the ship in Naples") correspond to the following changes to the data base:

(1) making a change to the Ships file (i.e., renaming the ship);
(2) changing link (b) (moving a new ship into the dock);
(3) changing link (a) (moving a new dock into the port).

If the expression "the ship in dock 12" were used, only options 1 and 2 would be generated; similarly, if "Totor" were used, only option 1 would be generated.

4.3 The selection of appropriate updates

At first examination, it would seem to be necessary to incorporate a semantic model of the domain to select an appropriate update from among the candidate updates. While this approach would surely be effective, the overhead required to encode, store, and process this knowledge for each individual data base may be prohibitive in practical applications. In general, the required information might not be available. What is needed is a general set of heuristics that will select an appropriate update in a reasonable majority of cases, without specific knowledge of the domain.

The heuristics that are applied to rank the candidate updates are based on the idea that the most appropriate one is likely to cause the minimum disruption to the user's conception of the data base. This concept is developed formally in the work of Lewis, presented in his 1973 book, *Counterfactuals*. In this work, Lewis examines the meaning and formal representation of such statements as "If kangaroos had no tails, they would topple over." (p. 8) He argues that to evaluate the correctness of this statement (and similar counterfactual conditionals) it is necessary to construct in one's mind the possible world minimally different from the real world that could potentially contain the conditional (the "nearest" consistent world). He points out that this hypothetical world does not differ only in that kangaroos don't have tails, but also reflects other changes required to make that world plausible. Thus he rejects the idea that in the hypothetical world kangaroos might use crutches (as not being minimally different), or that they might leave the same tracks is the sand (as being inconsistent).

The application of this work to processing natural language data base updates is to regard each transaction as presenting a "counterfactual" state of the world, and request that the "nearest" reasonable world in which the counterfactual is true be brought about. For example, the request "Change the teacher of CS345 from Smith to Jones." might correspond to the counterfactual "If Jones taught CS345 instead of Smith, how would the data base be different?" along with a speech act requesting that the data base be put in this new state.

To select this nearest world, three sources of information are used:

(a) the side effects entailed by the different candidates;
(b) pragmatic information contained in the data base schema;
(c) semantic constraints attached to the data base schema.

(a) Side effects

As illustrated in the example of section 2, updates may have effects on the user's view and the data base beyond those literally requested. Using the rationale of "minimal disruption", updates that do not have side effects are preferable to those that do. For each candidate, we consider the number and type of side effects caused, and rank the candidates accordingly. In data base management terms, the update with the fewest side effects on the user's data sub-model is selected as the most appropriate.

Considering the example from section 2, note that the two candidates have different effects on the user's view. The one that was actually performed – candidate (a), changing the name of the manager of the Sales department – also changes two other values in the view. The other candidate – (b), moving Brown to the Marketing department – does not have these effects. Therefore, the latter more exactly fulfills the user's request, and would be preferred.

The side effects that actually occur for a particular candidate are in a sense accidental, in that they depend on the particular state of the data base. For example, the number of side effects caused by changing the manager of the Sales department depends upon how many other employees happen to work in that department. To avoid this property of contingency, a more stable approach is to consider what side effects could result from performing the given candidate in *any* state of the data base. This set of *potential* side effects can be determined by analyzing the restrictions in the data base schema concerning the cardinality and dependency of relationships between entities. The significance of this concept is that the constraints on cardinality and dependency may be strong enough to ensure that the set of potential side effects (and hence the set of actual ones) is empty – indicating that the given candidate does not have any side effects in the current state, and more important, could not have side effects in *any* state.

Consider once again the example of section 2. Of the two updates, (a) causes actual side effects, (b) doesn't. A stronger reason for preferring (b) is that it *cannot* cause side effects to the user's view, regardless of the state of the data base. To see this, note that the cardinality of the relationship between employees and departments is typically N:1 – each employee works for only one department. Thus, an employee can have only one manager, and moving the employee to a new department cannot cause any changes to this aspect of the view beyond the one requested. The potential side effects of (a) consist of changes to the managers of employees other than Brown; the two actual side effects are an example of this.

This calculation can be generalized, by considering a graphical representation of the view, in which nodes represent relations, and arcs stand for the joins (relationships) between relations. For relationships that are N:1 as in the example, the arc is labeled to indicate the direction of the functional determination. Thus, the graph for the example would be:

$$\begin{tikzcd}\boxed{\texttt{Employees}} \arrow[r] & \boxed{\texttt{Departments}}\end{tikzcd}$$

The view graph can be used to evaluate the side effects for each translation, with the following rule:

> Consider the value or link being changed by the translation in question, and the relation of which it is a part. If that relation is a root of the view graph, that is, if there exist paths following the arrows, from the relation in question to all the other relations of the graph, then the translation will not have any side effects.

For the example in question, translation (a) consisted of a change to *Departments,* while (b) entailed a change to *Employees* (to move Brown to another department). In the graph, *Employees* is a root, but *Departments* is not – the link from *Departments* to *Employees* runs the "wrong" way. Thus, the translation (b) cannot entail side effects, although (a) may; this is consistent with the previous observation.

In Davidson (1983), this analysis is carried further and developed more formally. We identify a number of different types of side effects and establish graphical conditions for the presence and absence of each. Further, theorems are introduced concerning comparison of side effects for different translations, and the optimality of certain translations is proved.

In the ranking of candidates for appropriateness, only potential side effects are considered. Explanations, when needed, are phrased with respect to actual side effects, if any exist, otherwise to potential ones.

(b) Pragmatic information

There may be information in the data base schema to help the selection among candidate updates. For example, certain attributes and links in the schema may be designated at design time as *static,* indicating that they rarely change, or *dynamic,* indicating that they frequently change. This information is used during implementation to select methods for accessing the information. It may also be of use when ranking candidate updates.

Considering the last example from section 4.2, we note that one of the candidates changes the ship by moving a new dock into Naples. This is consistent within the data base and fulfills the update request; but, the data base schema would indicate that such a change is unlikely (because the location of a dock is a *static* attribute), and this candidate's desirability would be downgraded Similarly, there may be general rules that *names* change less often than other attributes.

Note that this information is merely heuristic; if the only candidate is one that involves such a change, it will be performed.

(c) Semantic constraints

The schema will often contain *semantic constraints* that restrict the allowable states of the data base. Examples of these are *functional dependencies* (for example, "Two employees cannot have the same employee number."), *range constraints* ("No employee can make more than $45,000."), and *existence constraints* ("If an employee works in a particular department, there must be a record for that department in the *departments* relation.").

These figure in the process of update interpretation, in the elimination of candidates that are otherwise acceptable. In the example of section 4.2, if there is already a ship named Pequod in the data base, the renaming change could cause a name conflict, resulting in the rejection of this candidate.

Whereas the pragmatic information discussed above was *heuristic,* the semantic constraints are *absolute.* Candidates that violate semantic constraints will never be performed. However, it is still advantageous to generate and consider these candidates, since it is often possible to formulate a meaningful *explanation* for the user about the nonfulfillment of the request.

Our current ordering heuristics incorporate these sources of information. In increasing order of preference, they are:

- updates that violate semantic constraints associated with the deta base;
- updates that violate pragmatic guidelines;
- updates with side effects on the user's current view;
- updates with no side effects.

While this approach can certainly fail in cases where complex domain semantics rule out the "simplest" change, in most cases it is sufficient to select a reasonable update from among the various possibilities.

Consider again Lewis' "Counterfactual" framework. We see that the method of generating candidates discussed in section 4.2 defines the *accessibility* of different states of the world (data base); the semantic constraints define *consistency;* pragmatic constraints and side effect information are measures of *distance* between states of the data base.

4.4 Action taken

If one candidate is better than the others, it is performed. If there are a number of candidates that cannot be distinguished by the heuristic ranking, the user is told about them and offered a choice. If no candidate is admissible (because, for instance, all candidates violate *semantic constraints* on the data base), the user is so informed.

In a number of cases, circumstances must be *explained* to the user. For instance, if the candidate actually performed has side effects, the user must be notified. If a semantic constraint is violated, the user must be told how.

Our approach to explanation assumes that the user is familiar only with his own view of the data base, and so all explanations must be phrased with respect to this understanding (following McKeown 1979). Therefore, options are presented in terms of their effects on the user's view (rather than the actual changes proposed), and violations of semantic constraints are discussed with respect to attributes that the user has already seen. In this way, we ensure that explanations are always comprehensible.

5. Examples of the System in Operation

PIQUE runs in INTERLISP (Teitelman 1978) on the DEC System-20 at SRI International as part of the *KBMS* system (Wiederhold et al. 1981). The natural language parser is written in *LIFER,* a semantic grammar system designed by Gary Hendrix (1977). The data base access uses *SODA,* a LISP-compatible variant of the relational calculus developed by Bob Moore (1979); the SODA interpreter used was written by Bil Lewis, and has been modified and extended by Jim Davidson to handle updates.

Note that some of the information printed by the current system is presented merely for pedagogical purposes, to show the intermediate stages of the computation. In the course of a real run, such information (shown indented in the transcripts below) would be suppressed. The user's input is preceded by >.

Assume a sample data base containing the following information:

Individual employees, with salary, department, and employee number (employee number and name are assumed to be unique):

ESD:

EMP	SAL	DEPT
Adams	30	Invntry
White	35	Mkting
Brown	25	Sales
Smith	30	Sales
Pullum	25	Sales

EE:

EMP	EMPNO
Adams	103
White	431
Brown	554
Smith	222
Pullum	181

Departments have managers and location, and are grouped into divisions (department names are assumed to be unique):

DMLD:

DEPT	MGR	LOC	DIV
Sales	Jones	SF	I
Mkting	Baker	LA	II
Invntry	Fisher	SD	I
Advert	Larkin	NY	II

Divisions are assigned vice presidents:

DV:

DIV	VP
I	Lasker
II	Kline
II	Hannan
III	Burgess

1. Example of an update performed using side effect heuristics.

Consider once again the example shown in section 2, this time with the real system:

```
>(LIST THE EMPLOYEES AND THEIR MAN-
AGERS)

EMP       MGR
ADAMS     FISHER
WHITE     BAKER
BROWN     JONES
SMITH     JONES
PULLUM    JONES
```

Next, the user enters a request to update his view.

```
Enter next command:
>(CHANGE BROWN'S MANAGER FROM JONES
TO BAKER)
```

The program now generates the candidate updates. As indicated, there are two of these.

```
The possible ways of performing
the update:

1. In the relation DMLD change the
MGR attr of the tuple
DEPT    MGR      LOC    DIV
SALES   JONES    SF     I
to the value BAKER

2. In the relation ESD change the
DEPT attr of the tuple
EMP     SAL   DEPT
BROWN   25    SALES
to the value MKTING
```

Note that the update of changing the DEPT attribute of the tuple (MKTING BAKER) to Sales would make Baker one of the managers of Brown. But, it would also leave Jones as a manager of Brown, and thus does not fulfill the user's request, even literally. For this reason, it has not even been considered by the system.

These candidates are then evaluated, in terms of the accuracy with which they fulfill the user's request. In the absence of any strong preference for either one, the decision is made on the basis of potential side effects to the user's view.

```
These translations have the fol-
lowing effects:
1. Effects are:
In the view: potentially changing
the MGR of other EMPS.
2. Effects are:
None.
```

The program concludes that update (2) is superior to (1), since (1) has the possibility of changing the manager of other employees. (In actuality, (1) would change the managers of both Smith and Pullum.)

```
Desired translation is: 2.
Revised view is:
EMP       MGR
ADAMS     FISHER
WHITE     BAKER
BROWN     BAKER
SMITH     JONES
PULLUM    JONES
```

The result accords with the user's wishes; only the requested change has been made to his view.

2. Example of an update performed using pragmatic information

Consider the same data base, with a slightly different dialogue:

```
>(WHAT ARE THE LOCATIONS OF THE EM-
PLOYEES)
```

```
EMP     LOC
ADAMS   SD
WHITE   LA
BROWN   SF
SMITH   SF
PULLUM  SF
Enter next command:
>(MOVE ADAMS FROM SD TO LA)
```

The system interprets this as a request to change Adams' location.

```
The possible ways of performing
the update:
1. In the relation DMLD change the
LOC attr of the tuple
DEPT       MGR      LOC   DIV
INVNTRY    FISHER   SD    I
to the value LA
2. In the relation ESD change the
DEPT attr of the tuple
EMP     SAL     DEPT
ADAMS   30      INVNTRY
to the value MKTING
```

Two candidates are identified, corresponding to (1) physically moving the department to a different location, or (2) reassigning the employee.

Now, the candidates are evaluated.

```
These translations have the fol-
lowing side effects on the view:
1. Effects are:
Violation of pragmatic const-
raints.
2. Effects are:
None.
```

The "location" attribute of the DMDL relation, representing the location of the department, is marked in the data base schema as "static", indicating that it rarely changes. Thus, update (1) is unlikely. The system detects this. Note that update (1) also has potential side effects on the user's view, but the violation of the pragmatic constraint is a stronger reason for rejection.

```
Desired translation is: 2.
Revised view is:
EMP     LOC
ADAMS   LA
WHITE   LA
BROWN   SF
SMITH   SF
PULLUM  SF
```

Once again, the user's request has been fulfilled exactly.

3. Example of a request blocked by a semantic constraint:

Assume a *semantic constraint* – a functional dependency stating that each employee number may be assigned to only one employee.

```
>(LIST THE NAMES AND EMPLOYEE NUM-
BERS FOR ALL EMPLOYEES IN THE SALES
DEPARTMENT)
EMP     EMPNO
BROWN   554
SMITH   222
PULLUM  181
>(CHANGE SMITH'S EMPLOYEE NUMBER TO
103)
```

The program generates the ways of performing the update. There is only one of these.

```
The possible ways of performing
the update:
1. In the relation EN change the
EMPNO attr of the tuple
EMP     EMPNO
SMITH   222
to the value 103.
```

[The effects engendered by this candidate are now listed; the candidate would violate a semantic data base constraint.]

```
These translations have the fol-
lowing effects:
1. Effects are:
Violation of semantic constraints.
```

The system now tells the user what has happened, explaining why the update couldn't be performed, and how the semantic constraint would be violated.

```
This update could not be performed,
because of semantic constraints:
The EMPNO value of 103 has already
been assigned to the tuple
EMP     EMPNO
ADAMS   103
which has the DEPT value of INVNTRY.
This update would violate the func-
tional dependency EMPNO --> EMP.
```

Note that, without the DEPT value printed out, the user may not realize why he cannot see the (ADAMS 103) tuple. The explanation is thus phrased with respect to the user view.

4. Example of a genuinely ambiguous update.

Now, a dialogue concerning a different part of the data base:

```
>(WHICH VP'S ARE IN CHARGE OF WHICH
DEPARTMENTS?)
```

```
VP       DEPT
LASKER   SALES
LASKER   INVNTRY
KLINE    MKTING
KLINE    ADVERT
HANNAN   MKTING
HANNAN   ADVERT
Enter next command:
>(REPLACE LASKER WITH KLINE AS VP IN
CHARGE OF THE SALES DEPT)
```

The system now generates the list of candidate updates. There are two of these.

```
The possible ways of performing
the update:
1. In the relation DV change the
VP attr of the tuple
DIV  VP
I    LASKER
to the value KLINE
2. In the relation DMLD change the
DIV attr of the tuple
DEPT    MGR     LOC    DIV
SALES   JONES   SF     I
to the value II
```

Again, the effects of each on the user's view are computed.

```
These translations have the fol-
lowing effects:
1. Effects are:
In the view: potentially changing
the VP of other DEPTs.
2. Effects are:
In the view: potentially inserting
or deleting other VPs for this
DEPT.
```

Thus, *both* candidates have side effects on the view. Since the system cannot decide a priori that one of these is superior to the other, no decision can be made here. The only solution is to ask the user. Note that, since the user is presumed to know nothing about the structure of the underlying data base, the only meaningful way to distinguish between the updates is to describe them in terms of their (actual) side effects on his view. This is another example of explanation phrased with respect to a view.

```
There are 2 methods of performing
this update.
Update (1) will have the side effect
of
replacing the tuple (LASKER INVNTRY)
with (KLINE INVNTRY)
Update (2) will have the side effect
of
inserting the tuple ((HANNAN SALES))
Which would you prefer?
>
```

If the user cannot make a choice, the update is abandoned.

[Note that the actual side effects are in fact examples of the classes described by the potential ones.]

6. Discussion and Evaluation

We discuss the effectiveness of PIQUE as a mechanism for performing natural language updates. Four aspects are considered: coverage, portability, efficiency, and correctness.

Coverage concerns the range of inputs accepted by the system. We distinguish *linguistic* coverage – the range of linguistic phenomena handled by the system – from *logical coverage* – the range of domain capabilities that can be performed using the natural language front end.

Linguistic capabilities have not been stressed in PIQUE, and linguistic coverage is therefore quite limited. Many phenomena – ellipsis, relative clauses, conjunctions, passive voice – are handled only in simple cases, if at all. Extension of the system to fully handle these could be accomplished through expansion of the grammar.

Logical coverage concerns the classes of requests that can be expressed by means of the interface language. Data base interfaces, unlike many artificial intelligence applications, have a task space that is well-defined – specifically, by the capabilities of the underlying data manipulation language. The class of queries accepted by PIQUE is a natural (and common) subset of the relational calculus expressions: *chain-structured, conjunctive* queries. These are queries where the set of predicates to be satisfied is a simple conjunction, and where the set of joins defines a chain; this corresponds to a form of select-project-join expression. The updates are deletions and replacements, again with conjunctive qualifications.

Portability deals with the ease of adaptation for new data bases or domains. The philosophy adopted in the design of PIQUE is somewhat different from that of typical AI systems. Rather than try to capture, represent, and encode the domain- and world-knowledge required to perform a thorough semantic analysis of the problem, we attempt to exploit whatever knowledge is already implicitly or explicitly present in the application (in this case, the content and structure of the data base and the user's phrasing of the update request). Consequently, the implementation is simplified and the techniques are more easily transported to new domains.

The system uses five major modules: parser and grammar; user modeler; candidate generator; ranking heuristics; and data base schema. For transport to a new application, only the first and last (natural language grammar, and data base schema) would require modification. These are both modules that would need revision or replacement anyway, independent of the update capability.

Efficiency concerns the time and space requirements of the system, and how these increase as the data base becomes larger. The algorithms of PIQUE are designed to avoid potentially expensive data base references, by using the data base schema where possible. Thus, the candidate generation and ranking are performed on an intensional (structural) basis, and the set of heuristics used do not refer to the data base state. As a result, the system is relatively insensitive to increases in the size of the data base.

Correctness is used here to indicate the degree to which the system's actions match the intended results; any "intelligent" system has the possibility of error. We consider now the correctness of the PIQUE approach. Note first that one class of errors is unavoidable – these are the "pathological" cases in which the user's intent is beyond the discovery of the program. For example, the user's real meaning in "Change Brown's manager from Jones to Baker." may have been to have Brown fired, and a new employee named Brown hired for Baker. In such situations, it is unlikely that any program will behave correctly; the best result achievable is probably to provide enough feedback to enable the user to discover the error.

Many of the remaining difficulties are associated with the ranking of candidate translations. The emphasis here has been placed on consistency of behavior. The consideration of side effects, and the emphasis on exactness of translations, are designed to ensure that the system's actions will be visible to the user, and unexpected effects will not occur. Side effect considerations have also been favored because of applicability (they are useful in a broad class of situations), perspicuity (they can be understood by the user), and richness (they operate at multiple levels, and provide results more complex than a simple yes-no response).

Of course, the performance of the system suffers when limited information is available. In part because of its generality, there is a definite risk that the system will take inappropriate actions or fail to notice preferable options. A more knowledge-based approach would likely yield more accurate and sophisticated results. The process of responding appropriately to updates could be improved by taking advantage of domain specific knowledge external to the data base, using partial case-structure semantics, or tracking dialogue focus, to name a few possible extensions. To mitigate these shortcomings, the system is engineered to fail "softly", by presenting options to the user or requesting clarifications by re-phrasing the request. As data bases encode richer semantic knowledge, as suggested by Wiederhold and El-Masri (1979) and Hammer and McLeod (1978), the ranking heuristics can be easily extended to take advantage of these additional knowledge sources.

7. Conclusion

We have studied the subjective problem of interpreting natural language updates to a data base system. In particular, we have examined the problems associated with vague requests, which lack sufficient detail to enable a unique interpretation.

Drawing on work in artificial intelligence, the philosophy of language, and data base theory, we have developed and implemented a domain-transparent approach to this problem. This method is characterized by the maintenance of a form of "user model" for interpreting requests, and the use of a collection of heuristics to rank alternative translations. Particular attention has been paid to the requirements of *efficiency* and *portability.*

8. Bibliography

Cohen, Philip. 1978 On Knowing What to Say: Planning Speech Acts. Technical Report #118, Computer Science Department, University of Toronto.

Davidson, James. 1982 Natural Language Access to Databases: User Modeling and Focus. *Proceedings of the Fourth Biennial Conference of the Canadian Society for Computational Studies of Intelligence.* Saskatoon, Saskatchewan, pp. 204-211.

Davidson, James. 1983 Interpreting Natural Language Database Updates. Ph.D. thesis, Computer Science Department, Stanford University, Stanford, California.

Dayal, Umeshwar. 1979 Schema-Mapping Problems in Database Systems. Technical Report TR-11-79, Center for Research in Computing Technology, Harvard University, Cambridge, Massachusetts.

Frege, Gottlob. 1952 On Sense and Reference. In Geach, P. and Black, M., Eds. (Black, M. (Tr.)), *Translations from the Philosophical Writings of Gottlob Frege.* Blackwell, Oxford, England.

Hammer, Michael and McLeod, Dennis. 1978 The Semantic Data Model: A Modelling Mechanism for Data Base Applications. *ACM SIGMOD Conference Proceedings:* 26-36.

Hendrix, Gary. 1977 Human Engineering for Applied Natural Language Processing. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence.* Cambridge, Massachusetts: 183-191.

Hendrix, Gary et al. 1978 Developing a Natural Language Interface to a Complex System. *ACM Transactions on Database Systems* 3: 105-147.

Henisz-Dostert, Bozena and Thompson, Frederick. 1974 The REL System and REL English. In Zampolli, A. and Calzolari, Eds., *Computational and Mathematical English.* Casa Editrice Olschki, Firenze, Italy.

Kaplan, S. Jerrold and Davidson, James. 1981 Interpreting Natural Language Database Updates. In *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics.* Stanford, California: 139-141.

Kaplan, S. Jerrold. 1979 Cooperative Responses from a Portable Natural Language Data Base Query System. Technical Report HPP-79-19, Computer Science Department, Stanford University, Stanford, California.

Kaplan, S. Jerrold. 1978 Indirect Responses to Loaded Questions. *Proceedings of the Second Workshop on Theoretical Issues in Natural Language Processing.* Urbana-Champaign, Illinois.

Keller, Arthur M. 1982 Updates to Relational Databases Through Views Involving Joins. In Scheuermann, Peter, Ed., *Improving Database Usability and Responsiveness.* Academic, New York: 363-384.

Lewis, D. 1973 *Counterfactuals.* Harvard University Press, Cambridge, Massachusetts.

Mann, William; Moore, James; and Levin, James. 1977 A Comprehension Model for Human Dialogue. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence.* Cambridge, Massachusetts: 77-87.

McKeown, Kathleen. 1979 Paraphrasing Using Given and New Information in a Question-Answer System. *Proceedings of the 17th Annual Meeting of the Association for Computational Linguistics.* LaJolla, California: 67-72.

Moore, Robert. 1979 Handling Complex Queries in a Distributed Data Base. Technical Note 170 (October), Artificial Intelligence Center, SRI International, Menlo Park, California.

Quine, Willard. 1971 Reference and Modality. In Linsky, Leonard, Ed., *Reference and Modality.* Oxford University Press, Oxford, England.

Teitelman, Warren. 1978 Interlisp Reference Manual. Xerox PARC, Palo Alto, California.

Wiederhold, Gio and El-Masri, Ramez. 1979 The Structural Model for Database Design. *Proceedings of the International Conference on Entity-Relationship Approach to Systems Analysis and Design.* North Holand Press, Amsterdam, Holland: 247-267.

Wiederhold, Gio; Kaplan, S. Jerrold; and Sagalowicz, Daniel. 1981 Research in Knowledge Base Management Systems. *ACM SIGMOD Record* 11(3).