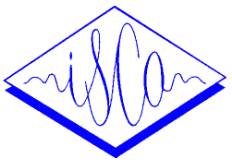

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

WHAT'S NEXT: A CASE STUDY IN THE MULTIDIMENSIONALITY OF A DIALOG SYSTEM

Robert Belvin, Ron Burns and Cheryl Hein
(robin, ron, cheryl)@hrl.com

HRL Laboratories, LLC
3011 Malibu Canyon Road
Malibu, California 90265

ABSTRACT

In this paper we argue that conversational dialog systems must model a multidimensional space defined by syntactico-semantic, circumstantial, and interactional dimensions. We take as a test-case the interpretation of a very simple query type within a conversational navigation system that we have implemented. We outline the decision logic this system follows in interpreting next-turn and numbered turn queries, and discuss some of the interesting results that have come about from user testing of the prototype.

1. INTRODUCTION

One of the challenges in developing any open dialogue system lies in identifying the various dimensions along which the dialogue may vary and then determining which information should be carried along as the dialogue continues, which should be updated when, and which can safely be ignored. But the application type can influence what is appropriate, and one can distinguish two broad categories of applications. There are those wherein the user's circumstances are static (or irrelevant) to the application's internal logic, and there are those in which the user's circumstances change as a result of the interaction, for example, making air-travel reservations. In the latter type of application, not only must the dialogue system keep track of utterance-level and discourse-level semantics, but it must also have some way of detecting and tracking what the user's circumstances are, since interpretation of user queries and generation of appropriate system responses are affected by changes in the user's circumstances.

In this paper we examine certain aspects of a conversational in-vehicle route-guidance system. Such an application presents a particularly dramatic example of a system which must be attentive to user circumstance. We focus attention on a family of query types, those having to do with upcoming turns, and show how there are a number of apparently straightforward queries whose interpretation in fact requires a fairly sophisticated series of implicatures. The information required to determine the interpretation is decidedly heterogeneous (multidimensional). It is partly circumstantial, partly semantic and syntactic, and partly dependent on discourse history. Section two outlines the data, section three considers the interaction of driver circumstance and discourse history, while section four shows a small-scale decision table which calculates items in SQL queries based on this kind of heterogeneous information, and describes an additional dimension which should be included in such a matrix for a truly robust conversational system. Section five is devoted to a summary, and description of future work .

2. THE OBJECT OF STUDY

The data we are concerned with in this paper are transcribed recordings of people interacting with a conversational navigation system prototype.[1] The design and implementation of the prototype are described in a companion paper in this volume [1]. The data are somewhat artificial in that the nine users from whom the data were collected were almost all from the conversational systems project team, and the data were collected in a lab. We have designed a large data-collection project which is scheduled to begin later this year. In that work we will collect conversational data from naï ve users in a vehicle.

In the current study, the data consists of approximately 1000 user-system turns. Within these data, approximately 40% of the user queries are about future turns, and 40% of those are about the immediately approaching turn. But we believe these percentages are artificially low, and that the number of future-turn and next-turn queries will in fact be higher when we carry out our larger data-collection effort. The reason for this is that about one-third of the user utterances in our corpus are devoted to odometer readings, an artifact of this current implementation which will disappear in later versions as GPS is integrated into the system. (See the companion paper [1] for more explanation.) If we ignore these odometer sentences, then the percentage of user queries about future turns is 60%, and about the immediately approaching turn is approximately 25%. That is, users devote about a quarter of their queries to asking about the next turn.

We have gone to the trouble of explicating this characteristic of our route-guidance data to underline the fact that this family of queries (next-turn and numbered turn queries) is well-represented in conversational navigation data. But it is perhaps not surprising that in a conversational route-guidance system, a large percentage of queries would be about the next turn. The purpose of such a system, after all, is to allow the user to get information about their route on a turn-by-turn basis as they proceed down the route, rather than getting the information all at once in a hard-to-remember instruction set.

[1] The initial prototype was developed based on the developers' intuitions about what people would need to ask, in addition to a small amount of data which was collected from a thought-experiment asking participants to tell us the kinds of queries they would ask an in-vehicle route-navigation system.

3. CIRCUMSTANCE AND DIALOG HISTORY

3.1 Time-dependent discourse coherence

As noted, it is not that surprising that queries about the next turn comprise a significant percentage of the utterances in a conversational navigation system log. What *is* surprising, and in our opinion interesting, about this kind of query is that it is not always transparent to interpret. Queries such as "what's next?" are in fact ambiguous. Determining the correct interpretation requires consideration of the discourse history and the user's circumstances. For example, in the following dialog sequence (drawn from our data), there are at least two possible interpretations for "what is next?" in the third turn:[2]

```
   ---------------
   U:  what is my next turn
   S:  From Malibu Canyon Road turn left
       onto Pacific Coast Highway.
   ---------------
   U:  and after that
   S:  From Pacific Coast Highway go
       straight onto East I-10 freeway
   ---------------
→ U:  what is next
```

Figure 1. Sample dialog containing ambiguous "What is next". (U:user, S:system)

The system response has deliberately been left out so the reader can observe the ambiguity more clearly. This query could be requesting information about the next turn from the driver's current position (i.e. the immediately approaching turn), or it could be requesting information about the *third* turn from the driver's current position, that is, the next turn from the most recently referred to turn. We will henceforth refer to these two interpretations as *next-from-here* and *next-after-that*, respectively.

The factor which appears to have the most influence on which interpretation is given to this utterance originates neither in the utterance itself nor in the preceding dialog, but is almost purely circumstantial, namely, how much time has passed since the last utterance. The more time that has passed since the last system instruction, the more likely "what is next" must be given the *next-from-here* interpretation.

The reason for this is clearly related to the time-dependent nature of coherent discourses. While "what is next" is still within the scope of the preceding discourse context, it may (most likely will) be given the *next-after-that* interpretation. But after a certain length of time has elapsed, "what is next" cannot be interpreted as referring to some previously uttered instruction, but only as referring to the driver's current position. If we think of this in terms of the user's frame of reference for talking about their real or imagined location (we'll refer to this as the FROM value), then we could characterize this phenomenon as the value of FROM defaulting to HERE in the absence of immediate discourse context.

It is worth mentioning that the foregoing is not an isolated example. There are many other cases which demonstrate the same kind of ambiguity, with perhaps even more interesting variability of interpretation:

```
   ---------------
   U:  what's my next turn
   S:  From Pacific Coast Highway go
       straight onto East I-10 freeway
   ---------------
→ U:  how much further is it
```

Figure 2. Sample dialog containing ambiguous "How much further is it?"

Although in this sequence it seems quite obvious that "it" in the second user utterance refers to the transition from Coast Highway onto Interstate 10, imagine the case that there was a ten minute delay between the first exchange and the second query. In this case (especially if the system knew via GPS that the driver had already passed the PCH/I-10 transition), the best interpretation for "it" becomes the final destination, rather than the most recently mentioned turn. "It" in this case is similar to the "there" in "Are we there yet?" in that in both cases, the locative proform gets a referential value from the most cognitively salient destination in a journey, namely the final destination. Note the similarity of this phenomenon to the "default to here" pattern we discussed in connection with the prior example.

As of the writing of this paper, we are setting up an experiment to determine if there is some threshold beyond which values in human discourse reference frames (at least in this application domain) get reset to defaults such as FROM=HERE or TO=FINAL DESTINATION.

This time-dependent interpretation can be seen as a special case of a more general discourse rule which "resets" what may be interpreted as given and what must be interpreted as new. Relating this phenomenon to discourse theories would take us beyond the scope of this short paper, though we note the potential contributions these theories may make to the design of this kind of system [2,3,6].

3.2 Turn References at the Start of the Trip

Interpretations of numbered turn references can vary depending on another purely circumstantial factor, namely whether the driver is querying the system while preparing to begin the trip, or after she has begun driving. Some drivers will want to preview trip information before beginning to drive, and in this situation, interpretation of certain query types may differ from interpretation done during the trip. When the driver is querying the system before beginning to drive, she is more likely to conceive of and speak of the route in an "absolute" sense (cf. [5]). That is, the driver may conceive of the route as a fixed plan,

[2] There is at least one further possible interpretation to "what is next?" here, at least if the proper prosodic features are present. If heavy emphasis is placed on "what," the query has a quasi *echo-question* interpretation, indicating either that the user did not hear, or else is surprised at the prior instruction and is asking for clarification or repetition.

wherein each turn and segment have a unique and constant order in a sequence. When conceiving of the route in this way, one may refer to turns by number in the route, rather than by number relative to current position. Consider the first part of the route we were using in the prototype, shown below:

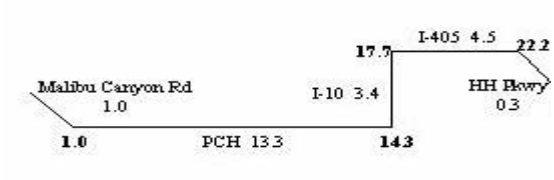

Figure 3. Schematic of part of the route from HRL to LAX

In a query like "How far is the second turn?", given the "absolute" perspective on the route, the second turn is from PCH onto I-10. Although we have yet to gather real user data bearing on this question, our intuition is that once the trip is underway, especially once any significant distance has been traveled, users will be much more likely to use numbered turn references relative to their current position.

Queries of this type are, for practical purposes, only ambiguous once the user has begun the trip, but prior to the absolute numbered turn. Drivers are very unlikely to be asking about the second turn in the route once they have passed the second turn. Moreover, since people will generally only keep track of turn numbers in the range of 1-3, (give or take 1), numbered turn references will only be ambiguous prior to the third or fourth turn in the route (nobody is likely to be asking "what is the eighth turn in the route"). What is more, if the user asks a numbered turn query before beginning the trip, the system response will be the same, since the relative and absolute turn numbers will at that point coincide. Thus, the only time a true ambiguity must be handled by the system is the time after the trip is underway, and before the fourth turn.

Although numbered turn requests comprise the main source of absolute/relative ambiguities in our limited corpus, there is another source as well, a subset of the queries about distance and time. Thus, questions such as "how far is it?" and "how long will it take?" may be uttered either at the beginning or during the drive, and therefore may be interpreted as asking for the trip distance or time from the very beginning, or as asking for the distance or time remaining. Of course, the ambiguity discussed in the previous subsection will also exist if there is some recent utterance referencing a turn to which "it" may be referring. However when queries of the type "how long will it take" or "how far is it" are uttered in contexts *other* than that discussed in the previous section, then for practical purposes they may always be treated as asking from the current position, since it is immaterial whether that position is at the start or in-route (though see footnote 3). One presumes the user is not interested in distance or time of the entire route once they are underway, but is instead only interested in remaining distance or time.

If one looks at the overall query interpretation problem as entailing a determination of whether the user is asking a question relative to their current position, or some other position, then the absolute/relative distinction is just a special case that.

In addition to the foregoing, there is another important driver circumstance which should be included in calculating interpretation of user queries, and this is whether the user is on-route or off-route. We will have to treat this issue as falling outside of the scope of this short paper, though we note that we will be including a study of off-route navigation conversations in our upcoming data-collection effort.

4. MULTIDIMENSIONALITY

As we have indicated, there are a variety of information types one needs to bring to bear on the problem of interpreting the query type of interest. In addition to knowing the elapsed time since the last system utterance, whether the trip is underway and if so, whether the driver has passed the fourth turn, (and ultimately also whether the driver is on- or off-route), one of course must identify relevant linguistic features of the utterance in order to give it a correct interpretation. Since we are restricting ourselves to queries about next and numbered turns, we are taking as a given that the basic semantic content denoting this has been identified in the utterance. There are, then, many linguistic indicators which allow us to distinguish whether a driver wants information based on current or other position. Most obvious of these are explicit adverbial phrases which serve precisely this disambiguating function, as in "How far is the next turn *from here*?" , "What do I do *after that*?" or "What's the second turn *in the route*?". The adverbial phrases in these sentences tell us to interpret the query as next-from-here, next-after-that, and absolute, respectively.

In the following table we show the basic decision logic that we follow in determining how next-turn and numbered turn sentences should be interpreted. The table is to be read column-by-column. Thus, the first column tells us that if we have a query with a numbered turn reference and a phrase which is semantically equivalent to "from here" (which is also the default), then the instruction number which will be requested (via SQL query) from the database is **current+number**. (See our companion paper for more detail on the implementation of the prototype.)

In-route[3]	--	--	√	--	--	--	--
Reset threshold reached	--	--	--	--	--		√
Next turn		√			√	√	√
Numbered turn	n		n	n			
"From here"	√	√		x	x		
"After that"				√	√		
SQL Turn number	c+n	c+1	n	r+n	r+1	r+1	c+1

Ø = set **c** = current position

blank = not set **r** = most recently mentioned turn

n = number value **--** = irrelevant

Figure 4. Decision matrix showing some determining factors for interpreting "next" and numbered turn queries.

[3] Previously we argued that one could ignore ambiguities created by the at-start/in-route dichotomy. However, since we are here reporting on the actual prototype, we have included this distinction since it factored into the real system implementation.

Notice that there are two distinct types of information in this matrix, syntactico-semantic, and circumstantial. There is a third dimension, in fact, which should really be included in interpreting user queries (and also in generating system responses), what we may call the *interactional* dimension. The interactional dimension would help to ensure that the intent of the user query was correctly identified, and that the form of the system response satisfied the user query in a natural way. To take a very simple response generation example from our system, if one asks a question like "Is my next turn a right or a left?", the system will respond with a full instruction, like "From Howard Hughes Parkway turn left onto Sepulveda boulevard." Although this kind of response is comprehensible, the usability factor would greatly increase were the system to respond with "your next turn is a left," or even just "a left". Users are accustomed to this kind of efficiency in normal conversation, and we believe systems which can achieve this will be more effective and require less effort on the part of the user. In order to achieve this kind of tight fit between user query and system response, the interactional dimension must be considered, where the salient values along this dimension can be viewed as a highly enriched speech-act inventory.

The space that these dimensions define can be thought of as constituting an application domain. Interpreting a user query in a conversational dialog system, then, can be thought of as identifying points in a multidimensional space with at least three axes. In figure (5) we have outlined some of the salient points along the axes that would be found in a navigation system, though in a truly robust system there would be many more.[4]

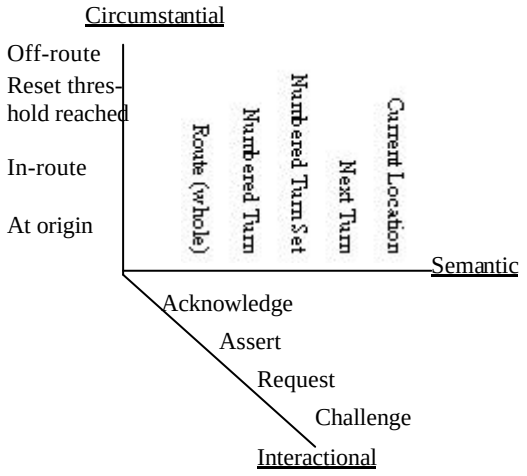

Figure 5. Multi-dimensional space for modeling some of the conversational features of a navigation domain

Defining an application domain can, in this view, be thought of as identifying the salient points along these axes for that domain.

[4] In a more generalized and robust model, we would also probably characterize semantics and syntax as two different dimensions.

5. SUMMARY AND FUTURE PLANS

In this paper we have argued that for a conversational dialog system to interpret even relatively simple-looking queries correctly, its internal logic must include consideration of information of a variety types from a variety of sources. It must not only correctly identify the semantic content of any given user utterance, but it must keep track of how this utterance is related to prior utterances, and it must have some way of determining or at least estimating aspects of the user's circumstances. Ultimately it must also be able to consider the interactional devices that are characteristic of natural conversation.

We are beginning to investigate how best to model this multidimensional space which defines the application domain for a robust conversational dialog system. This will comprise an important part of our future research. Included in this work will be an investigation of the degree to which consideration of discourse markers of the type discussed in [4] may impact performance of both analysis and generation components. Also included in our research plans will be ways of dealing with the problem of proper name recognition. It may well be that proper name recognition, which is a problem in dialogue systems, will benefit from an analogous approach to achieving disambiguation based on the same kind of multidimensional factors discussed in this paper. See our companion paper for some discussion of this issue. Overall, we anticipate that our data-collection plans, which include both human-to-human and human-system components, will help to keep this research rooted in abundant and natural conversational data.

Acknowledgments. This work was supported in part by a research contract from General Motors. Thanks to Tim Clausner, Angela Kessell and Matt Shomphe for comments on an early draft.

6. REFERENCES

1. R. Belvin, R. Burns & C. Hein "Spoken Language Navigation Systems for Drivers," *These Proceedings*, Beijing, China, Oct. 2000.
2. B. Grosz & C. Sidner *The Structure of Discourse Structure*, CSLI Technical Report no. 39, Stanford, 1985.
3. H. Kamp & U. Reyle *From Discourse to Logic*, Kluwer, 1993.
4. D. Schiffrin, *Discourse Markers*, Cambridge, 1987.
5. L. Suchman, *Plans and Situated Actions*, Cambridge, 1987.
6. M. Walker, A. Joshi & E. Prince, eds., *Centering Theory in Discourse*, Oxford, 1998.