ISCA Archive
http://www.isca-speech.org/archive

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

DIALOGUE SEMANTICS FOR AN ORAL DIALOGUE SYSTEM§¶

Paul Heisterkamp
Daimler-Benz AG
Inst. for Information Technology
Wilhelm-Runge-Str. 11
D-7900 Ulm, Germany
heisterkamp@dbulm1.uucp

Scott McGlashan
Social and Computer Sciences
Research Group
University of Surrey
Guildford, U.K.
scott@soc.surrey.ac.uk

Nick Youd
Logica Cambridge Ltd.
Betjeman House
104 Hills Rd.
Cambridge, U.K.
nick@logcam.uk

10.21437/ICSLP.1992-213

ABSTRACT

This paper describes the component in the SUNDIAL spoken dialogue system responsible for semantic interpretation of user input. This component, the 'belief module', constructs interpretations using several sources of knowledge, including static knowledge definitions and a dynamic model of dialogue context, and addresses the problem of the hypothetical nature of the recognition and parsing results as well as the general problems of underspecification and ambiguity.

1 INTRODUCTION

We begin by elucidating our notion of dialogue semantics so as to motivate the need for a distinct semantic interpretation component in a spoken dialogue system. We then give an overview of the SUNDIAL system, focusing on interactions between components relevant to the interpretation process. The interpretation process itself is described and illustrated. Finally, we suggest that this interpretation process can be seen as an abductive process.

2 DIALOGUE SEMANTICS

Semantic interpretation in spoken dialogue systems must deal with three sources of uncertainty.

The first is the ambiguity of language. Not only are there words and sentences that have many different meanings (cf. [11]), but every utterance, whether spoken or written, is understood using the knowledge of the interpreting agent. Thus the agent needs to use the appropriate knowledge in interpreting the utterances of other agents.

A second source of uncertainty is that the interpretation of an utterance may depend upon the agent's knowledge of the preceding dialogue context – interpretation requires more than a knowledge of the current utterance. Phenomena like supra-sentential anaphora and ellipsis can only be interpreted referring to the dialogue context to which they belong (cf. [4]). Other dialogue phenomena, such as deictic expressions, require knowledge about the situational setting of the dialogue. In addition, depending upon the purpose of the dialogue, an utterance may require a non-literal interpretation; e.g. in information providing dialogues, answering only 'yes' or 'no' to the question 'Is there a flight to London tomorrow?' suggests that the question has been given an inappropriate interpretation for this type of dialogue.

Dealing with these uncertainties suggests that semantic interpretation of user input to a dialogue system must be constructed using knowledge about the whole dialogue. Semantic interpretation which takes account of the dialogue context can be termed *dialogue semantics*.

The third source of uncertainty is specific to computer dialogue systems. Human agents are normally capable of dealing with corrupted signals or incorrect processing during interpretation (misreadings, misspellings, mishearings and mispronunciation). In spoken dialogue systems however, the recognition process is already a source of severe problems (cf. [6]). Suffice it to say here that problems of segmentation, word recognition and parsing lead to the fact that the resulting syntactico-semantical representation is a hypothetical one[1]: i.e., the representation is the one that has been assigned the highest probability.[2]

§The research reported here is partly funded by the CEC, DG XIII, in the ESPRIT programme as project 2218.

¶The authors would like to thank their colleagues in the SUNDIAL project, who have contributed to this research through critical discussion of the work presented here: E. Bilange, A. Cozannet, M. Danieli, W. Eckert, N. Fraser, N. Gilbert, M. Guyomard, M. Kharoune, J.-Y. Magadur, D. Sadek, J. Siroux and R. Wooffitt.

[1] Or, in cases where there is an ordered set of strings or parses etc, a set of hypotheses.

[2] Probability here refers to the probability of the representation relative to that of other possible representations of the

In sum, a spoken dialogue system needs to be responsive to these uncertainties when constructing a semantic interpretation of user utterances. To achieve this, it is necessary to constuct interpretations using as much knowledge as possible, and for the interpretation process to be sensitive to the hypothetical nature of the input representations. In the SUNDIAL system, dialogue management – interpreting user utterances and deciding what the system says – is crucially dependent upon a belief module which constructs semantic representations in this manner.

3 THE SUNDIAL SYSTEM

The SUNDIAL (Speech Understanding and Dialogue) project is concerned with handling information dialogues over the telephone (cf. [8]). Versions of the system are being built for English, French, German and Italian, the application domains being flight and train timetable enquiries and reservations.

Each system is an instance of a common architecture composed of speech recognition, linguistic analysis, dialogue management and generation components. A language-dependent recognition component for continuous speech based on HMM techniques delivers a graph of word hypotheses to a linguistic analysis component. This component parses the input and constructs a syntactic and (context-independent) semantic representation. The realisations of these two components are different for each of the demonstration systems, because they are dependent on the structures of the respective language.[3] However, the semantic representations constructed by different parsers are expressed in a standardised knowledge representation language. The output of the linguistic analysis component is then sent to a dialogue management component which provides a semantic, dialogue and task interpretation of the user input (see below). When it is the system's turn to speak, a dialogue and semantic representation of the system utterance is sent to the generation component. Using this information, the generation component then generates surface forms and synthesizes the system utterance.

The dialogue management component is independent of the application and the language of the dialogue. The application and language customisations are completely declarative knowledge that is used or disregarded according to software flag settings. The dialogue manager architecture relevant to interpretation (the 'belief module') is shown in Figure 1. The semantic representation returned by the parser is treated as input. In order for the dialogue to continue in a coherent and appropriate way, two results of contextual interpretation need to be produced. Firstly, a task-oriented representation of information must be presented to the task-control component, whose purpose is to manage the current task with respect to the given application, in our case a database. As a result, information

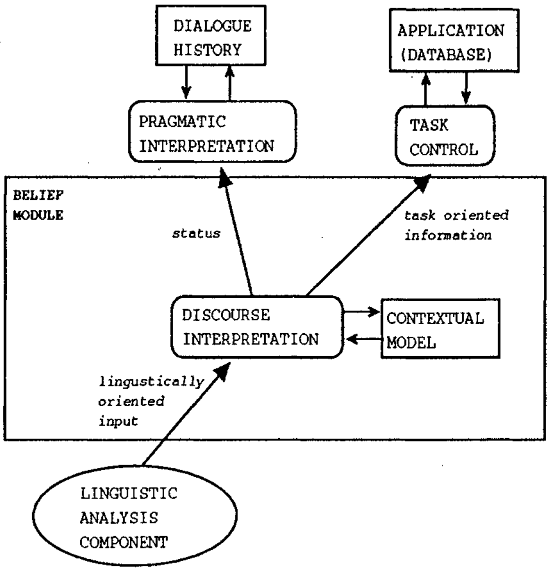

Figure 1: Architecture for interpretation cycle

can be provided for the user, or more information may be sought. The second result contributes towards *pragmatic interpretation*: i.e., determining how the user's input may be fitted onto the ongoing internal representation of the conversation, or *dialogue history*. The flag *status* reflects how the contextual model has changed: the value *repeat*, for example, indicates that nothing new has been added. The pragmatic interpretation component may use this result together with other knowledge, to decide that the utterance is functioning to provide confirmation of information specified earlier.

In a separate generation phase the state of the dialogue history, together with outstanding goals such as the need to request or provide information, is used to motivate the production of system utterances (cf. [12]).

4 THE BELIEF MODULE

The belief module performs the semantic interpretation by using several sources of knowledge. Starting from the semantic description of an utterance, it builds an extension of its contextual model which is coherent both with this input as well as with other knowledge sources. We call a coherent, but hypothetical, semantic interpretation a *belief* state, or *world*. Accordingly, the system's reaction, based on this hypothetical understanding, manifests the system's belief of what its knowledge state should be like.

The contextual model is constructed from two knowledge sources. The first is a hierarchy of surface-oriented concepts, which is shared with the linguistic analysis component. The second source is a hierarchy of task-oriented concepts, known to the task control module, which correspond to relations in the application database. When

utterance, not the probability of the utterance's grammatical correctness.

[3] The same or at least compatible formalisms are used wherever possible.

instances of surface-oriented concepts are created which reference task-oriented concepts, instances of these later concepts are also created in the contextual model. This mapping between surface-oriented and task-oriented concepts is guided by inference rules associated with particular types of concepts. For example, in the user utterance 'The noon flight to Paris', a rule maps the property of the flight (of being 'at noon') onto a value for the departure time of that flight connection the user requested. Thus, the belief module translates surface-oriented semantic representation into task-oriented representations suitable for the application database.

The semantic representation of user utterances can be underspecified. This can be the case with deictic (temporal, e.g. 'the last flight to Paris', and spatial, e.g. 'a flight from there to Berlin'), anaphoric (e.g. 'does it call at Duesseldorf') and elliptic (e.g. 'to Rome') expressions, as they are particularly common in spoken language in general as well as in our application domains.

In the course of instantiation the belief module tries to anchor these partially specified inputs to a concept in its contextual model. In deciding where to anchor a concept, the belief model uses as knowledge sources the order of recent concepts marked in the current world as being accessible, and the concepts which were in the immediate context of the last system utterance. From this set of concepts, it identifies candidates for anchoring the description. In the case of ellipsis, the candidate that matches the semantic role of the ellipsed concept is selected. In the case of anaphora, the belief module makes use of a linguistic history – indexed syntactic and semantic representations of all utterances in the dialogue – to find a candidate concept with the appropriate grammatical properties to be the antecedent of an input anaphoric expression.[4]

Where the referent of a temporal deictic expression is a concept which can only be known via a query to the application database, the belief module draws all the inferences it can, and then passes the unresolvable part of the semantic description onto the task control module. This module then maps this underspecified part onto a query to the application database, and returns the result to the belief module, thereby either enabling it to instantiate the referent looked for or to identify the cause for a failure of the instantiation.

A user utterance which is consistent with the current world leads to a monotonic extension of the current world. A new world is constructed when the current user input is inconsistent with the current world, but this new world inherits all the instances of concepts from the current world which are consistent with the current utterance, plus the new diverging instantiations due to that utterance.

5 AN EXAMPLE

This example illustrates both the contextual interpretation of fragmentary input – typical of speech dialogues – as well as the use of worlds for storing and keeping track of non-monotonic changes in belief.

Consider the dialogue fragment (1), handled by the belief module.

(1)

C1:	Are there any flights from Luton to Paris?
S1:	From London to Paris.
S2:	You want to travel on what date?
C2:	Not London, Luton.

The semantic component of the input $C1$ (not shown here) is structured similarly to the natural language utterance. The inferred task-oriented representation is:

(2)

$$\begin{bmatrix} id : dbflight1 \\ type : dbflight \\ goalcity : paris \\ sourcecity : london \end{bmatrix}$$

The *status* message is used to indicate that the user is beginning a new task. When $C2$ is uttered, the belief module is expecting firstly a date; it allows however for the possibility that the utterance may refer back to earlier concepts. The input, in linguistically oriented form, is represented as in (3).

(3)

$$\left\langle \begin{array}{l} \begin{bmatrix} id : A \\ type : location \\ thecity : \begin{bmatrix} id : B \\ type : city \\ value : london \end{bmatrix} \\ modus : \begin{bmatrix} number : sg \\ pol : neg \end{bmatrix} \end{bmatrix} \\ \begin{bmatrix} id : C \\ type : location \\ thecity : \begin{bmatrix} id : D \\ type : city \\ value : luton \end{bmatrix} \\ modus : \begin{bmatrix} number : sg \end{bmatrix} \end{bmatrix} \end{array} \right\rangle$$

Figure 2 shows part of the belief module state as a result of interpreting this input. The previous state, at world W_1 is overlaid with a new state (W_2) inheriting all the information in W_1 except that pertaining to the source city, which is redefined. The interpretation returned (4) incorporates status features which indicate how information has been changed.

(4)

$$\left\langle \begin{array}{l} \begin{bmatrix} id : dbflight1 \\ type : dbflight \\ sourcecity : london \\ status : negated \end{bmatrix} \\ \begin{bmatrix} id : dbflight1 \\ type : dbflight \\ sourcecity : luton \\ status : modified\_by\_user \end{bmatrix} \end{array} \right\rangle$$

Taken together, these amount to a re-specification of the task.

[4] This procedure can be seen as a simplified and straightforward version of the anaphora resolution mechanism described in [10]. It is not yet fully implemented.

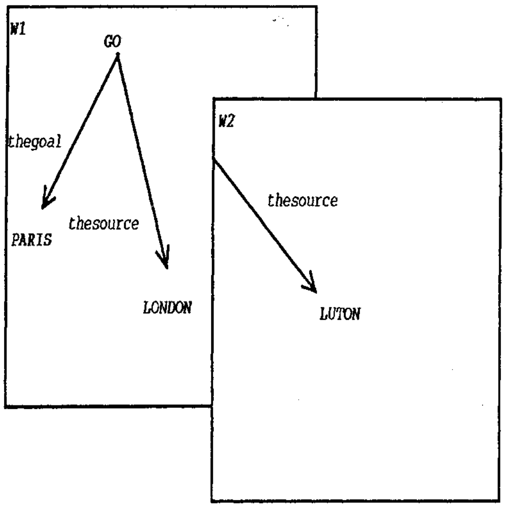

Figure 2: BM state after the input: *Not London, Luton.*

6 ABDUCTION

The interpretation described above is an abductive process, where abduction can be characterised as the search for the appropriate rule or model of which some given instance may be the result (cf. [9]). Starting from the assumption that the input at every stage is an instance of the knowledge that is modeled in the system, the goal of interpretation is to find the part of the task model that the particular input is an instance of. The model is already given. This type of abduction is called *overcoded* abduction (cf. [3]). This is why we do not need to model abduction explicitily as an inferential scheme (cf. [5, 7]).

The worlds mechanism avoids the need for cancellation interactions, i.e whenever there are incompatible relationships, a new world is created. The first compatible hypothetical explanation (i.e. interpretation) can be taken as the best one, because (*i.*) other interpretations can not lead to qualitatively better explanations (as there is no quality measure except consistency), and (*ii.*) every interpretation is subject to implicit or explicit confirmation by the dialogue partner, i.e. the user of the system. So, the type of abduction used here is of the kind shown to be tractable in [1].

Abduction is not a safe mode of inference. Rather, the results of abductive reasoning can only be hypothetical.[5] As our system for spoken language understanding must be able to deal with hypotheses anyway, and as a dialogue system must constantly check its interpretations of the user utterances, it can realize interpretation as abduction without computational problems.

[5] Unless there is exhaustive search across all possible combinations of explanations, which leads to a computationally intractable explosion of the search space (cf. [1]).

7 CONCLUSION

We have described the belief module in the SUNDIAL spoken dialogue system which addresses the problem of uncertainty in semantic interpretation of user input. In constructing an interpretation, the belief module uses surface-oriented and task-oriented knowledge representations together with inference rules which map between the two. The belief module anchors these representations into a dynamic, contextual model of the dialogue. In cases where the representation is inconsistent with the current world in the contextual model, a new world is created which only encodes knowledge inconsistent with the current world. Finally, this interpretation process is treated as an abductive process in that the inference rules attempt to infer instances of concepts pertaining to the task model.

References

[1] Bylander, T., D. Allemang, M. C. Tanner and J. R. Josephson (1991): The computational complexity of abduction. In *Artificial Intelligence 49 (1991).*

[2] Dale, R., E. Hovy, D. Rösner, O. Stock (Eds.) (1992): Aspects of automated natural language generation. Proc. of the 6^{th} International workshop on natural language generation, Trento, Italy, April 1992. Berlin: Springer.

[3] Eco, U. (1984): Semiotics and the philosophy of language. London: Macmillan.

[4] Grosz, B. J. (1981): Focusing and description in natural language dialogues. In Joshi, A. K., B. L. Webber and I. A. Sag: *Elements of discourse understanding*, Cambridge: Cambridge UP 1981.

[5] Hobbs, J. R., M. Stickel, P. Martin, and D. Edwards (1988): Interpretation as abduction. In *Proc. of the 26^{th} Annual Meeting ACL*, Buffalo, N.Y., June 1988.

[6] Lee, K.-F. (1989): Automatic Speech Recognition. The developement of the SPHINX system. Boston, Dordrecht, London: Kluwer.

[7] Lascarides, A., J. Oberlander (1992): Abducing temporal discourse. In *[2].*

[8] Peckham, J. (1991): Speech Understanding and Dialogue over the telephone: an overview of the ESPRIT SUNDIAL project. In *Proc. DARPA Speech and Natural Language Workshop*, Pacific Grove, Ca., Feb. 1991.

[9] Peirce, C.S. (1931-58): Collected Papers. Cambridge, Mass.: Harvard UP.

[10] Rich, E., S. LuperFoy (1988): An Architecture for Anaphora Resolution. In *2^{nd} Conference on Applied Natural Language Processing. ACL.*, Austin, Texas, Feb. 1988.

[11] Wichter, S. (1988): Signifikantgleiche Zeichen. Tübingen: Narr.

[12] Youd, N. J., S. McGlashan (1992): Generating Utterances in Dialogue Systems. In *[2].*