Using Discourse Structure in a Dialogue System to Search in Databases

Christian Hying
IMS, Universität Stuttgart
Azenbergstr. 12
70174 Stuttgart
Germany
christian.hying@ims.uni-stuttgart.de

Sunna Torge
ASL, Sony International (Europe) GmbH
Hedelfinger Strasse 61
70327 Stuttgart
Germany
torge@sony.de

Introduction We present the functionality of a *discourse processing component* (DPC) for dialogue systems that are applied to the task of browsing a database.[1] The DPC is implemented according to (Grosz and Sidner, 1986). It contains a *focus stack* which keeps information about the *intentions* and the linguistically relevant objects (*discourse objects*) which occur in the course of the dialogue. The intentions control the focus stack. They are computed by employing a simple semantics of utterances: utterances are mapped onto intentions to specify a database query. We identify the underlying *discourse purpose* with the goal of picking a single database item. This is subsumed by the database query specified. Thus, computing the relation between utterance intentions and discourse purposes boils down to comparing database queries. The focus stack is used to build a salience structure which contains *discourse objects*. These *discourse objects* serve as possible antecedents for anaphoric expressions. For each *discourse object* the salience structure holds information about salience, surface form and meaning in order to support an anaphora resolution component.

In order to show the applicability in user directed dialogue we have chosen an ill-structured task, cf. (Bernsen and Dybkjær, 2000), namely picking a song from a music database. There is no natural order in which the attributes *title*, *artist*, and *genre* have to be specified. We assume that the system has a small text display and the ability to produce spoken output. The user can provide input to the system only by way of spoken input.

Interaction The interaction between user and system is predetermined by the following interaction pattern: first, the user specifies a database query, and second, the system offers the user options to refine that database query. Note, that the latter also comprises the offering of single database items. Since our approach heavily relies on discourse processing, a closer look at possible user input shows, that generally speaking there are two possibilities: (i) The user specifies a new database query which does not relate to any previous material. Examples are shown in Figure 1 in utterances (1) and (5b). And (ii), the user can take up one of the options offered by the system by using an anaphoric expression such as a definite description, a name, an abbreviation of a name, or a pronoun. In Figure 1 utterance (3), the abbreviation “Folk” is anaphorical on the option named “Irish Folk”. Similarly (5a) is anaphoric, too.

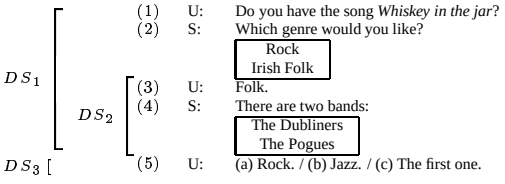

Figure 1: Example dialogue with discourse structure

Computation of Discourse Structure The setting of the task is such that we only need Grosz

[1]The work was done as a diploma thesis for the University of Stuttgart at Sony International Stuttgart. We would like to thank Ulrich Heid and Jan van Kuppevelt for their support.

and Sidner's *dominance* relation. We compute it by establishing a subsumption relation between database queries. A database query is represented by a set of attribute-value pairs where the attribute specifies a field of the database and the value specifies the value of the field.

(1) Let A and B be database queries. A *subsumes* B, iff $A \subset B$.

That means that A subsumes B, if and only if any attribute-value pair that is element of A is also element of B and B contains at least one pair that is not element of A.

The specification of database queries relates to the structure of discourse in the following way: each discourse segment is assigned exactly one database base query which characterises its discourse purpose. A discourse segment starts with the specification of a database query and comprises all successive utterances which do not specify another query that is *not* subsumed by it. A discourse segment embeds another discourse segment if the database query that is associated with it subsumes the query of the other segment. In the example DS_1 is associated with the database query ⟨title, "whisky in the jar"⟩ and DS_2 with {⟨title, "whiskey in the jar"⟩, ⟨genre, "irish folk"⟩}. If the user uttered (5a), DS_3 would be associated with ⟨genre, "jazz"⟩ and accommodated on top level. If he uttered (5b), DS_3 would be associated with the query {⟨title, "whiskey in the jar"⟩, ⟨genre, "rock"⟩} and embedded under DS_1. Finally, (5c) would yield DS_3 being embedded under DS_2.

Discourse Processing The DPC updates the focus stack with every utterance, so that the stack holds the information which is in the focus of attention at each point of the dialogue. The elements of the focus stack are *focus spaces*. In our implementation they are realized as feature structures of the type shown in Figure 2.

A focus space representation contains of three features: the feature PURPOSE holds the discourse purpose of the associated discourse segment in form of a database query. Each of the other two features, i.e. DISC and GRAPH, holds a list of representations of *discourse objects*. The first list (DISC) contains representations of *discourse objects* which have occurred in a natural language utterance, and the second one (GRAPH) contains representations of *discourse objects* which have been presented on the display. Object representations contain a database query as their denotation (SEM feature) and information about their surface realization (SURF feature).

$$\begin{bmatrix} \text{PURPOSE} & : \textit{database-query} \\ \text{DISC} & : \textit{List of} \begin{bmatrix} \text{SURF} & : \begin{bmatrix} \text{SYN} & : \textit{NumberGenderFS} \\ \text{TYPE} & : \textit{DatabaseFieldSymbol} \end{bmatrix} \\ \text{SEM} & : \textit{database-query} \end{bmatrix} \\ \text{GRAPH} & : \textit{List of} \begin{bmatrix} \text{SURF} & : \begin{bmatrix} \text{TYPE} & : \textit{DatabaseFieldSymbol} \end{bmatrix} \\ \text{SEM} & : \textit{database-query} \end{bmatrix} \end{bmatrix}$$

Figure 2: Type of *FocusSpaceFS*

Salience Structure After each update of the focus stack a copy of it is sent to the natural language understanding unit. We call this copy the *Salience Structure*. It provides a structured view on salient *discourse objects* which are possible candidates for antecedents of anaphoric expressions. We claim that it contains important information about *discourse objects* which serve as possible antecedents for anaphoric expressions: (i) salience, (ii) modality (DISC/GRAPH), (iii) the *order* of occurrence, (iv) syntactic properties, and (v) semantic denotation.

For example see the alternative options in Figure 1: (5a) and (5c) are treated anaphorical, because they can be uniquely matched by a *discourse object* in the salience structure. The expression "rock" is matched by the displayed option "rock" introduced in (2). And the expression "the first one" is matched by the option "the dubliners" introduced in (4).

References

N. O. Bernsen and L. Dybkjær. 2000. From single word to natural dialogue. In M. V. Zelkowitz, editor, *Advances in Computers*, volume 52, pages 267–327. Academic Press, London.

B. Grosz and C. Sidner. 1986. Attention, intentions and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Christian Hying. 2002. A context manager exploiting discourse structure. Diploma Thesis, IMS, Universität Stuttgart.