Automatic Disambiguation of Discourse Particles

Kerstin Fischer and **Hans Brandt–Pook**
Graduate Program Task-oriented Communication
University of Bielefeld, Germany
Postfach 100131, D-33501 Bielefeld
fischer@nats.informatik.uni-hamburg.de
hbrandt@techfak.uni-bielefeld.de

Abstract

In spite of their important quantitative role, discourse particles have so far been neglected in automatic speech processing for two reasons: Firstly it is not clear what they may contribute to the aims of automatic speech processing, and secondly their functions seem to vary so much that it seems difficult to identify the information relevant to such aims. The approach presented here therefore attempts to provide automatic means to distinguishing the different readings of discourse particles and to filtering out the information which can be useful for speech understanding systems, employing positional information and their role within a dialogue model of the respective domain, two types of information which are especially easy to obtain. First results indicate that discourse particles can indeed be automatically disambiguated on the basis of the model proposed.

1 Introduction

Discourse particles, as for instance German *ja, nein, ach, oh,* and *ähm,* and English *well, yes, oh, ah,* and *uhm* (Schiffrin, 1987), are extremely frequent phenomena of spontaneous spoken language dialogues. For instance, in informal German human-to-human communication, their frequency ranges between 8.8% and 9.8% (Fischer and Johanntokrax, 1995). In human-computer interaction, this prominent quantitative role decreases; however, they may still constitute 6.6% of the 150 most frequent words (Fischer and Johanntokrax, 1995). In spite of their important quantitative role, discourse particles have so far been neglected in automatic speech processing; if they are identified at all, then only in order to eliminate them (O'Shaugnessy, 1993). Reasons may be firstly that it is not clear what they could contribute to the aims of automatic speech processing, and secondly that they may fulfill so many different functions that it seems difficult to identify the information relevant to such aims.

In this study, it will firstly be investigated what these discourse particles contribute to natural human-to-human conversation with the aim to determine which of their functions can be useful for automatic speech processing. In order to make use of the information they provide, however, they need to be disambiguated. Two different strategies will be employed to disambiguate them automatically: On the one hand their position with respect to the turn and utterance in which they occur will be investigated in order to see how much it contributes to the interpretation of a discourse particle occurrence, on the other their role in the dialogue structure, especially regarding a dialogue model, will be analyzed. These two types of information, position and dialogue acts, are particularly easy to obtain during automatic speech processing. The aim of this investigation is thus to determine

- what discourse particles contribute to human-to-human dialogues;

- what they may contribute to automatic speech processing;

- in how far their surface properties, in particular their position regarding the

turn and the utterance, influence their functions;

- what the dialogue structure may contribute to their automatic disambiguation;

- how these aspects interact and how this interaction can be modelled in an actual automatic speech processing system.

It will be shown that the two types of information involved suffice to disambiguate a considerable portion of discourse particle occurrences and that there is consequently no reason to eliminate them from automatic speech processing. Finally, a model for an implementation in a semantic network for the automatic disambiguation of discourse particles will be proposed.

2 The Functions of Discourse Particles

The following examples from a corpus recorded in the appointment scheduling domain (Verbmobil Database, TP13) show several different functions the English discourse particle *yeah* may fulfill:

(1) 13BAR: what about the 18th of December?
14RIC: <P> *yeah, yeah*, that work.

(2) 1UMI: *yeah*, we've got to get together and discuss <P> Stufe A für die Studienordnung.[1]

(3) 124ENG: so that won't work either.
125UMI: *yeah*, that's not good.

(4) 3RIC: I'm Ric M. and I am <P> what do I do? (whispering) software . *yeah*, I'm working for a software account[2]

In example (1), *yeah* is a positive answer to a proposal. In (2), it functions as a signal by means of which the speaker introduces a new topic, here it opens up the conversation, constituting the first utterance in the dialogue; in example (3), it serves to give feedback to the hearer, indicating perception and understanding and also basic agreement with the partner (Schegloff, 1982). Finally, in example (4), *yeah* occurs after a speech management problem, that is, after the speaker does not know how to continue.

Other discourse particles may fulfill similar functions: in example (5), the discourse particle *mmm* functions as a signal of perception and understanding, and so does the first *well* in example (6) while it furthermore relates the current utterance to the previous. The second instance of *well* in this example marks a thematic break, concluding the topic concerning the partner's cold and returning to the previous attempt to find a possible date for a meeting. This discourse structuring function can also be found for the hesitation marker *uh* in example (7).

(5) 16UMI: yeah, just one of the students about San Soto anyway.
17ENG: *mmm*

(6) 6UMI: I've got <P>
7ENG: *well*, you've got a very heavy cold, but *well* <P> wh/ wh/ why is this week out of the question?

(7) 114UMI: yeah, this weekend's k k k completely chaotic. *uh*. let's see.

Thus, the functions discourse particles may fulfill are as follows: they mark thematic breaks, thus making the macrostructural organisation of the dialogue transparent for the hearer; they signal the relatedness of the current utterance to the previous, they indicate whether the information transfer is successful, and whether the channel is still available. Finally they support the formulation process in case there are speech management problems. Thus, they may fulfill a large but not an unlimited number of functions.

Furthermore, several of these functions have been found to co-occur in stable feature bundles. For instance, in example (6), the first instance of *well* displays perception and understanding of what the partner has said and furthermore indicates that the speaker is going to add something to the same topic. The same information is conveyed by *yeah* in example (3). This implies that there are not only co-occurrence relations between different functional features but also that a general inventory of pragmatic interpretation of discourse particles can be proposed, such that the same pragmatic function can be fulfilled by different discourse particles.

The linguistic model presented here consequently assumes that each discourse particle can be assigned a context-specific interpretation (Stenström, 1994) which belongs to a limited inventory and which can be seen as a bundle of functional features which may be fulfilled by several different discourse particles. Additional information is carried by the lexemes themselves (Fischer, 1996), for instance, in the case of *yes* agreement, and rejection in the case of *no*.

The descriptive inventory, which was arrived at by means of several hypothesis-test-cycles on the basis of four large corpora (Fischer and Johanntokrax, 1995; Schmitz and Fischer, 1995), consists of the following items:

- **take-up**: signals perception and understanding regarding the previous utterance and takes up the current topic;

- **backchannel**: gives feedback and supports the other's turn;

- **frame**: marks a break in the thematic structure;

- **repair marker**: signals problems in the formulation process;

- **answer**: signals agreement on, or rejection of, a proposition;

- **check**: signals the hearer that the speaker would like to get positive feedback;

- **modal**: refers to the common ground or the information 'on hand;' establishes a relation of coherence between the utterance and a pragmatic pretext.

In how far can these functions now be used for automatic speech processing? On the one hand, the most trivial point is the ambiguity of certain discourse particles concerning propositional and other information, for example German *ja, nein, doch* or *nicht*, or English *yes* or *no*. Information on this ambiguity is essential for automatic speech processing systems since if it is not resolved, contradictory or just wrong information will be inferred. Consider the following proposal for an appointment in example (8).

(8)　　geht es *nicht um 13 Uhr?*
　　　　*(is*n't it possible at 1 p.m.?)

In this example the discourse particle *nicht* does not contribute to the propositional information; however, it does so in the following example:

(9)　　es geht *nicht um 13 Uhr.*
　　　　*(it is*n't possible at 1 p.m..)

The examples demonstrate the need for the disambiguation of discourse particle occurrences since the recognition of keywords such as *nicht* would lead to wrong conclusions regarding the first example; however, disregarding it would obscure the proposition in the second example.

On the other hand, if the different functions of discourse particles can be automatically determined, they can also support further semantic and pragmatic analyses concerning the information structure and the macro-structural organisation of a dialogue, segmentation, and the recognition of repair sequences in the same way as they are taken as signals by human speech processors. Automatic disambiguation of discourse particles is therefore essential to their further employment in automatic speech processing systems.

3 Automatic Disambiguation of Discourse Particles

To account for the multifunctionality of discourse particles, two strategies are combined: the analysis of the syntactic position of a discourse particle and the role of the pragmatic environment the discourse particle occurs in.

3.1 The Influence of the Position in Turn and Utterance

Positions in turn and utterance seem to be good candidates for determining the interpretation of German discourse particles; for instance, if *ja* occurs utterance finally, it is likely to be used in a checking function, for example:

(10)   elf Uhr sagten Sie, *ja*? (FRS001)
        (you said 11 a.m., right?)

Thus, different functions can be assigned to certain contextual positions of *ja*; as we have seen, turn-final *ja* usually functions to elicit feedback from the hearer; utterance-medial *ja*, if the utterance is not interrupted and a new utterance is begun, serves as a modal particle, otherwise as a repair marker. The largest group is constituted by turn- and utterance-initial discourse particles; these serve either to signal perception and understanding and that one is going to add something to the same subject matter, or as answer signals. The latter may however also occur as a complete turn or turn-finally, so that other types of information are necessary to identify the function of *ja* as an answer signal carrying propositional content than the position in turn and utterance only. Nevertheless, the syntactic position in which a discourse particle occurs provides a valuable source for the automatic disambiguation of discourse particles, as will be shown by means of the following experiments.

First of all, 150 examples of the especially multifunctional German discourse particle *ja* were analyzed regarding their surface and pragmatic properties. Secondly, to each discourse particle a context-specific interpretation from the inventory proposed in the previous section could be assigned. By means

of an artificial neural network classificator (Scheler and Fischer, 1997), it was determined whether the eight classes described in the previous section can be assigned automatically on the basis of a selection of structural and pragmatic features. In particular, training a feedforward-backpropagation network simulated by SNNS p.d. software with ca. 100 examples of German *ja*, and testing on ca. 50 new examples can show firstly whether the classification proposed is sufficiently contrastive and thus learnable. Furthermore it is determined of which influence on the correct classification of the new examples particular properties are such as the syntactic position or the respective intonation contour.

Good results in the experiments employing the whole feature set indicate that the classes are learnable and that the classification can be replicated by an automatic classifier (Scheler and Fischer, 1997). Concerning the role of particular surface features, ca. 55% (standard deviation 4.38) of the context-specific interpretations can be determined only on the basis of surface properties such as syntactic position, combination with other discourse particles, intonation contour, or the left and right context. The position in turn and utterance alone produces similar results. On the basis of the intonation contours alone, ca. fifty percent can be assigned correctly (however, the standard deviation here is quite high). Position and prosody together reach 57.7% (standard deviation 4.31). The other surface properties therefore only provide redundant information compared to these two types, adding them for training and testing the network does not increase the assignment rate. So it seems promising to regard aspects of the surface realization as a valuable source for the automatic disambiguation of the functions of discourse particles. This information is exploited in the implementation of a speech processing system in the appointment scheduling domain described in section 4 especially regarding the position with respect to the utterance since it is the information most easily to obtain for an automatic
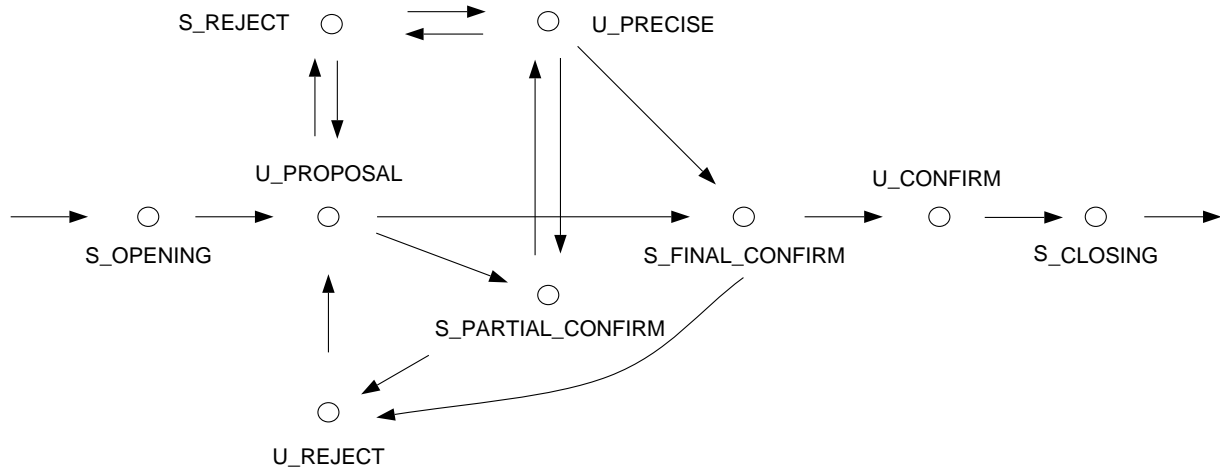
Figure 1: Dialogue model

speech processing system.

The first type of information which the automatic speech processing system is going to rely on for the automatic disambiguation of discourse particles is thus the prototypical syntactic position for each discourse function.

3.2 The Role of a Dialogue Model

As a second major source of information for the automatic disambiguation of discourse particles, a significant correlation between certain dialogue acts, domain-specific speech acts (Schmitz and Quantz, 1995), and certain discourse functions, the context-specific readings of discourse particles, was determined. For instance, in the informal task-oriented human-to-human dialogues, 40.43% (standard deviation 7.9) of the discourse functions of *ja* could be assigned correctly by the artificial neural network classificator only on the basis of the information which the preceding and which the following dialogue act was, although in this corpus the number of dialogue acts assigned is quite high, and the relationship between them is highly variable. Consequently, the reduction of complexity caused by the specialization on one domain in task-oriented settings can be exploited here by postulating a limited inventory of speech acts occurring in the corpus under consideration. A dialogue model which represents the possible relations be-

tween dialogue acts in a flow chart can thus make predictions about the following dialogue act and also about a probable function of the respective discourse particle. Thus, information for each discourse particle about its occurrence in certain dialogue acts may support its automatic disambiguation. In human-computer interaction, for example, this sequential relationship between dialogue acts may be even more straightforward, and the domain may be furher restricted.

For analyses of the appointment scheduling domain, a small corpus of 30 dialogues with altogether 145 turns was gathered in a human-to-machine scenario to arrive at statistical correlations between functional information concerning discourse particles and the dialogue acts in which they occur. On the basis of these instances of spoken interaction, a dialogue model was constituted (see Fig. 1). The descriptive categories were based on those proposed for the Verbmobil corpus (Alexandersson et al., 1997). By means of these dialogue acts, it can be partly determined what the function of a respective discourse particle may be in the particular setting; for instance, it is much more likely that *ja* is an answer particle if it occurs in an utterance after the system has asked for the confirmation of a date (as in example 11) rather than an uptaking signal when it is used at the beginning of a dialogue between system S and user U (as in example

12).

(11)   S: Bitte bestätigen Sie diesen Termin.
(Please confirm this appointment.)
U: *ja*, dieser Termin ist richtig. (10_2_02)
(yes, this is the right date.)

(12)   U: *ja*, guten Tag, ich hätte gern einen Termin. (6_1_00)
(yes, hello, I'd like to schedule an appointment.)

In the linguistic model, information on the correlation between dialogue acts and certain functions of discourse particles as well as syntactic information are combined. The following section describes the implementation of these types of information.

4 Implementation

The model described above was implemented and tested in a system for appointment scheduling (Brandt-Pook et al., 1996). For the representation of linguistic knowledge, this system uses the semantic network language ERNEST (Kummert et al., 1993). The linguistic knowledge of this system is modelled hierarchically on five levels of abstraction: The word hypothesis level establishes the interface to the speech recognition system; the syntactic, semantic and the pragmatic level declaratively represent the interaction of lexical and other knowledge; and the dialogue level manages dialogue strategy and dialogue memory. The concepts at the different levels are connected by means of *abstraction links*. Within a level, *has-part* and *is-a* links are defined.

Each occurrence of a discourse particle is interpreted bottom-up. Thus, a word-hypothesis is created so as to identify the relevant word form; connected to this is a concept PARTICLE, for which the utterance position for the respective discourse particle is identified. On the dialogue level the interpretation of a discourse particle occurrence and its functional role is completed. The attribute *att_disc_particle* realizes rules

that combine the different types of information which participate in the model for the pragmatic interpretation of discourse particles. Thus, on the one hand, the position in turn and utterance is treated as a condition on the interpretation of a discourse particle, on the other, due to the fact that the previous system-output influences the next user utterance, the dialogue act of the following utterance and consequently the function of a respective discourse particle can be predicted. For instance, a system request for an explicit confirmation predicts an answer signal by the user, as implemented in the rule concerning the word form *ja* (see Fig. 2). With this information, an utterance

```
IF (word == ja)
  IF (position == INITIAL)
    IF (last_sys-output==S_P_CONFIRM ||
        last_sys-output==S_F_CONFIRM )
      discourse_function := ANSWER
    ELSE
      discourse_function := TAKE-UP
  ELSE IF (position == MEDIAL)
    discourse_function := MODAL
  ELSE    /* position == FINAL */
    discourse_function := CHECK
```

Figure 2: Rule concerning the word form *ja*

of *ja* can be correctly classified as either an answer signal or as a discourse particle with certain pragmatic functions.

5 Results and Conclusion

To determine in how far the system relying only on the two types of information described, syntactic position and the previous dialogue act, can automatically disambiguate occurrences of discourse particles, a test was designed on the basis of natural authentic dialogues. These, however, were typed into the system in order to be able to analyze the contribution of the implementation most clearly; the results are therefore not obscured from errors by the speech recognizer.

Good results were achieved regarding the

test-sentences with altogether 75 discourse particles. In comparison with a double-checked hand-tagged version, 83% were automatically assigned the correct discourse function. These results show that automatic speech processing systems can indeed disambiguate discourse particle occurrences on the basis of extremely reduced linguistic knowledge. Consequently it is not the case that automatic speech processing systems cannot identify the relevant information. Thus, discourse particles do not constitute such a chaotic domain after all; already very simple and easily obtainable linguistic information on their occurrences in dialogues can lead to their automatic classification. A system which could make use of prosody, for instance, may get even better results. This opens up the way to employing discourse particles in automatic speech processing systems, for instance, as keywords regarding propositional information, to infer dialogue structure, or in order to control the information flow and to support speech management tasks, which all can support the aims of automatic speech processing systems.

References

Jan Alexandersson, Bianka Buschbeck-Wolf, Tsutomu Fujinami, Elisabeth Maier, Norbert Reithinger, Birte Schmitz, and Melanie Siegel. 1997. Dialogue acts in verbmobil-2. Report 204, Verbmobil.

Hans Brandt-Pook, Gernot A. Fink, Bernd Hildebrandt, Franz Kummert, and Gerhard Sagerer. 1996. A Robust Dialogue System for Making an Appointment. In *ICSLP96: Proc. of the Int. Conf. on Spoken Language Processing*, volume 2, pages 693–696, Philadelphia.

Kerstin Fischer and Michaela Johanntokrax. 1995. Ein linguistisches Merkmalsmodell für die Lexikalisierung von diskurssteuernden Einheiten. Report 18, SFB 360 "Situierte künstliche Kommunikatoren". University of Bielefeld.

Kerstin Fischer. 1996. A construction-based approach to the lexicalization of interjections. In M. Gellerstam, J. Järborg,

S. Malmgren, K. Nören, L. Rogström, and C. Rojder Papmehl, editors, *Euralex '96: PROCEEDINGS*, pages 85–91. University of Gothenburg.

Franz Kummert, Heinrich Niemann, Reinhard Prechtel, and Gerhard Sagerer. 1993. Control and Explanation in a Signal Understanding Environment. *Signal Processing, special issue on 'Intelligent Systems for Signal and Image Understanding'*, 32:111–145.

Douglas O'Shaugnessy. 1993. Locating disfluencies in spontaneous speech: An acoustic analysis. In *Proceedings of the 3rd European Conference on Speech Communication and Technology*, pages 2187–2190.

Emanuel A. Schegloff. 1982. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In Deborah Tannen, editor, *Analysing Discourse. Text and Talk*. Washington: Georgetown University Press.

Gabriele Scheler and Kerstin Fischer. 1997. The many functions of discourse particles: A computational model of pragmatic interpretation. In *Proceedings of Cogsci 1997*.

Deborah Schiffrin. 1987. *Discourse markers*. Number 5 in Studies in Interactional Sociolinguistics. Cambridge University Press.

Birte Schmitz and Kerstin Fischer. 1995. Pragmatisches Bescheibungsinventar für Diskurspartikeln und Routineformeln anhand der Demonstratorwortliste. Memo 75, Verbmobil.

Birte Schmitz and Joachim J. Quantz. 1995. Dialogue acts in automatic dialogue processing. In *Proceedings of the Sixth Conference on Theoretical and Methodological Issues in Machine Translation, TMI-95, Leuven*, pages 33–47.

Anna-Brita Stenström. 1994. *An Introduction to Spoken Interaction*. Learning about Language. London/ New York: Longman.