Gefördert vom

bmb+f

Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

Verbmobil

Verbundvorhaben

Collaboration in Automatic Dialogue Interpreting

Birte Schmitz

Technische Universität Berlin

Vm

Report 212
August 1997

August 1997

Birte Schmitz

Kontaktadresse:
Technische Universität Berlin
Fachbereich Informatik (13)
Projektgruppe KIT
Sekr. FR 6-10
Franklinstraße 28/29
D-10587 Berlin

Tel.: (030) 314 - 24928
Fax: (030) 314 - 24929
e-mail: `birte@cs.tu-berlin.de`

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 701 P 9 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

ISSN 0949-6084

In Proceedings of the Workshop "Collaboration, Cooperation and Conflict in Dialogue Systems", IJCAI-97, Nagoya, Japan, 1997, 79–88

Collaboration in Automatic Dialogue Interpreting

Birte Schmitz*
Technische Universität Berlin, FR 5-12
Franklinstr.28/29
D-10587 Berlin, Germany
e-mail: birte@cs.tu-berlin.de

November 25, 1997

Abstract

This paper starts from the assumption that collaboration is a fundamental prerequisite of dialogues. This view is applied to a special instance of human-human dialogues that are mediated by an interpreter. I demonstrate how this situation differs from a normal human-human-dialogue situation: The responsibility for mutual understanding and the control over the dialogue lies with the interpreter. The paper then focuses on the question how a human-human dialogue mediated (i.e. interpreted) by a machine can be adequately modeled. Based on the assumption that the machine cannot bear the responsibility for understanding, I propose certain strategies to compensate for the machine's incapability to take the control.

1 Introduction

Acting collaboratively is a fundamental prerequisite for dialogues. Obviously, participating in a dialogue implies interpreting utterances, and interactants can only determine the intended interpretation if they act collaboratively: Interpretations are not predetermined entities but they rather have to be co-constructed by the participants.

I start from this view and apply it to a special dialogue situation, namely dialogue interpreting. In this situation a human-human dialogue is mediated by an interpreter. I

*This work was funded by the German Federal Ministry of Education, Science, Research and Technology (BMBF) in the framework of the Verbmobil Project under Grant 01 IV 101 Q 8. The responsibility for the contents of this study lies with the author.

investigate the impacts of this situation on collaboration: In a dialogue interpreting situation the main effort lies with the interpreter. It is her responsibility to ensure mutual understanding.

I then compare this to a situation where the human interpreter is replaced by a machine. As the machine cannot bear the responsibility for mutual understanding, the labour of keeping a dialogue going has to be distributed in a different way. I propose strategies for both the machine and the users to compensate for the machine's lack of real collaborative activities.

2 Human-Human Dialogues

Following approaches in conversational analysis I regard dialogues as *social interactions*. [Stenström 94, p.1] characterizes spoken interaction as a

> here-and-now social activity which is governed by two main principles:
>
> - speakers take turns
> - speakers cooperate

She adds that "When studying transcripts of genuine conversation one is struck by the general atmosphere of cooperativeness and harmony. This does not mean, of course, that the listener always waits for the speaker to finish before taking over. Nor does it mean that speakers never disagree, object, or contradict each other." Elaborating on this, I see the interpretation of dialogue contributions as negotiated and finally co-constructed by the participants. All participants are mutually responsible for the understanding of each utterance. Certain linguistic means are available in order to interactively ensure understanding: *Reformulations* are used to correct, specify, or expand a description; *back-channel responses*, such as 'mhm' or 'yes' are used to give feedback to the dialogue partner; and *clarification queries* are used to signal problems in deriving the originally intended interpretation.

I will illustrate this view by focusing on a special linguistic phenomenon, namely the use of definite references. All examples are taken from the Verbmobil corpus[1]. Verbmobil [Wahlster 93] is a German government-funded long-term project on automatic dialogue interpreting that in its first phase has been confined to dealing with appointment-scheduling dialogues. These dialogues are characterized by the fact that

[1] These data can be accessed via the Bavarian Archive for Speech Signals (http://www.phonetik.uni-muenchen.de/Bas/). To improve readability, I left out certain markers indicating breathing or lip smacks, which seem not to be relevant for this discussion.

most contributions refer to temporal entities. With examples taken from these dialogues I will illustrate the view that the process of referring in human-human dialogues can be regarded as a collaborative action (cf.[Clark, Wilkes-Gibbs 86]).

In example (1) the speaker proposes a date by first referring to days of the week and then specifying these days by information about the particular dates. By adding a more specific description the speaker indicates that he anticipated a possible difficulty for his dialogue partner to identify the intended referent ('Friday or Saturday next week').

(1) how about on say Friday or Saturday +/w=/+ ⟨ähm⟩ next week *that'll be I guess October ⟨ähm⟩ first and second*

In other cases it is left to the hearer to request a further specification of a proposed date:

(2) **GRA037:** What time do you have to be back on Tuesday?
BOB038: ⟨ähm⟩ I would like to be back by seven o'clock.
DON039: In the morning?
BOB040: At night.

Usually the hearer can understand the utterance on the basis of a context-dependent preference for a particular interpretation. But, as the following example shows, the preferred interpretation can still be modified or changed in the ongoing discourse.

(3) **PS1002:** ja am Dienstag den sechsten April hätt' ich noch einen Termin frei allerdings nur nachmittags. geht es da bei Ihnen auch ?
(*On Tuesday, the sixth of April I'm free, but only in the afternoon. Is that possible for you?*)

BS1003: oh das is' schlecht, da habe ich um vierzehn Uhr dreißig einen Termin beim Zahnarzt. aber Donnerstag Vormittag so um neun wär' mir recht.
(*that is bad, I have an appointment with the dentist at two thirty. But Thursday in the morning would be okay.*)

PS1005: ist das der achte April ?
(*Is that the eighth of April?*)

BS1006: oh ich dachte eigentlich an den fünfzehnten April. eine Woche später.
(*In fact I meant the fifteenth of April, one week later*)

The point is that the intended interpretation of an utterance is not only highly context-dependent, but very often – even in a given context – not specified enough for the hearer to infer it. In such a case, if the identification of a referent is relevant for him, he must request a clarification, a specification, or a correction from his dialogue partner.

The examples above demonstrate that both speaker and hearer are provided with certain means to work collaboratively on an interpretation. As I pointed out above the responsibility for mutual understanding lies with all participants. In the following section I investigate how this situation changes for human-human dialogues that are mediated by an interpreter.

3 Dialogue Interpreting

I restrict my investigation to a special type of interpreting – *liaison interpreting* – and follow a definition from [Zähner 90]:

> What we will understand here by liaison interpreting is the immediate translation of each speaker's contribution in a conversation, discussion or similar exchange between two or more individuals speaking different languages. [Zähner 90, S.297]

How can the task of the interpreter be characterized? In accordance to a relevance-theoretic approach to translation [Gutt 91], it is the task of the translator or interpreter to produce a target-language utterance that enables the hearer in the target language to infer the originally intended interpretation. In this view the interpreter is responsible for the successful communication. In liaison interpreting the interpreter is co-present in the dialogue situation. According to [Keith 84] liaison interpreting includes three different dialogues:

- one macro-dialogue between the primary interactants. This dialogue is mediated by the interpreter.
- two micro-dialogues, one between speaker A and the interpreter and one between speaker B and the interpreter respectively.

What does this mean for the interpreter? She clearly has to fulfill different roles: On the one hand she has to mediate between the primary interactants. She hears the source-language utterance, determines the intended interpretation and formulates a target-language utterance that enables the hearer in the target language to infer this intended interpretation.

On the other hand she has to fulfill the role of an interactant in the conversations she has with the primary interactants. These dialogues are subordinated to the aim of the main dialogue, namely to make sure that the interpreter in fact derives the interpretation that the source-language speaker intended and to make sure that the hearer succeeds in deriving this intended interpretation.

The change of roles is usually manifested by a code change (marked by paralinguistic phenomena such as loudness or the change from first to third person).

This view of the role of the interpreter as both a *translator* and a *coordinator* is also taken in [Apfelbaum, Wadensjoe 97].

The interpreter's role as a coordinator is demonstrated with the following example taken from a Verbmobil dialogue that was a recorded and transliterated role play with two primary interactants and a human interpreter. In the preceding dialogue the dialogue partners agreed on meeting on Wednesday, the third of November, in the morning. This is confirmed in the contribution of speaker A. The interpreter translates this contribution and than realizes that the primary interactants have not agreed on the time of day they want to meet. In order to make sure that the aim of the dialogue, namely finding a date which suits both primary interactants, is reached, the interpreter takes control over the dialogue and brings the question of clock time to the primary interactants attention.

(4) **Speaker A:** Dann äh sehen wir uns dann äh am Mittwoch, dem dritte November vormittags
(*Then we will meet on Wednesday, the third of November, in the morning*)

Interpreter: Mister Mayer's looking forward to meeting you on November the third which would be a Wednesday
⟨ with a low voice ⟩ Wann sollte das genau sein an diesem Mittwoch?
(*When exactly was this supposed to be on this Wednesday?*)

This example demonstrates that although both the primary participants and the interpreter collaboratively try to reach the goal of the dialogue, the full responsibility for a successful communication lies with the interpreter.

In the following section I investigate how the situation changes if the interpreter is not a human but a machine.

4 Automatic Dialogue Interpreting

In automatic dialogue interpreting the human-human dialogue is mediated by a machine. Ideally, the machine should fulfill the same task as a human interpreter. As stated in the previous section, the human interpreter functions as a *translator* **and** as a *coordinator*. For each of these roles I first describe the problems that occur if a machine is used as an interpreter. I then propose certain strategies to compensate for the machine's incapabilities.

As a *translator*, the machine has to derive the intended interpretation of the original utterance and produce a target language utterance that enables the hearer in the target language to recover the originally intended interpretation. In principle, deriving the intended interpretation implies different subtasks:

First, a semantic representation has to be derived. This is usually done by compositionally applying linguistic rules. Generally the semantic representation is an incomplete logical form. Sperber and Wilson define a logical form as "a well-formed formula, a structured set of constituents, which undergoes formal logical operations determined by its structure" [Sperber, Wilson 86]. Its incompleteness is reflected by the fact that it contains unbound variables, ambiguous expressions and elliptical material. Only after applying contextual processes that resolve these indeterminacies a complete logical form (or *explicature*) is derived. There clearly are cases where an adequate translation of an utterance requires the derivation of its explicature. An example for this is given in (5). Depending on the preceding discourse, (5) can either be a proposal of a meeting place or a request to give a comment to a proposed date. The respective translations are given in (6) and (7).

(5) Ginge es da bei Ihnen?

(6) Shall we meet at your place?

(7) Is that alright with you?

Generally the intended interpretation of an utterance does not consist solely of an explicature but also of *implicatures*. The implicatures of an utterance are not systematically related to its linguistic properties, but are arrived at by further inferences. The hearer is only lead to derive implicatures if the explicature is not sufficiently relevant for her in the respective context. In the following part of a discourse the intended interpretation of (9) is (10) and (11).

(8) Can we meet next week?

(9) I will already be on a holiday then.

(10) We cannot meet when your are on a holiday.

(11) We cannot meet next week.

According to [Sperber, Wilson 86] (10) is the implicated premise, (11) the implicated conclusion. The decision to express an utterance in such a way that its intended interpretation is an implicature relies on assumptions about the background knowledge of the hearer. The central question whether to express information explicitly or implicitly in the translation therefore depends on contextual knowledge, namely about assumptions on the hearers knowledge. The following turn from an appointment-scheduling dialogue demonstrates that sometimes implicatures depending on cultural context cannot be maintained in the translation without risking a failure of communication.

(12) Also ich bin gebürtiger Rheinländer, der Februar ist relativ stark besetzt, ich würde den Januar vorziehen.
(*I am born in the Rhine area, February is rather full, I would prefer January*)

The relation between the fact of being born in the Rhine area and the fact that February is not a suitable time for a meeting can only be understood by taking into account that during February people in the Rhine area are busy celebrating carnival. An adequate translation has to either leave out this account or provide the hearer with the necessary information.[2]

It should have become clear by now that the derivation of the intended interpretation heavily depends on contextual knowledge.

Unfortunately only a limited amount of this contextual knowledge can be made available for the interpreting system. Therefore it is in many cases impossible for a system to derive the intended interpretation. In German (13) 'immer' is lexically ambiguous, resulting in two readings and translations (14) and (15).

(13) können Sie immer Dienstag ?

(14) Are you free on Tuesdays?

(15) Are you free the whole day on Tuesdays?

This ambiguity is not easily resolved by an automatic system, as the decision for one of these readings relies on contextual knowledge about the current focus of the negotiation (either day of the week or time of day), which can not in all cases be derived from the context representation.

Another example is given in (16). German 'nächste' is ambiguous, it can be translated as 'next' or as 'following', depending on the perspective of the speaker: in the "next-case" he takes the moment of speaking as point of reference, whereas in the "following" case an already introduced other temporal referent is the point of reference.

(16) gut, ähm in der nächsten Woche is' ja ein Feiertag am Donnerstag, aber sonst hab' ich dort nichts vor. wie sieht 's bei Ihnen aus?
(*okay, next (or 'in the following') week there is a holiday on Thursday, apart from that I'm free. What about you?*)

[2] Identifying an utterance as culture-dependent obviously is very difficult. I am not aware of any Machine-Translation system providing a solution for this.

If the interpreting machine cannot resolve such an ambiguity, what strategies should it pursue in order to fulfill its translation task?

I propose to implement the following three strategies, that should be applied in the following order:

Preserving ambiguities: If possible, the system should preserve ambiguities. As a general principle it is better to leave a translation underspecified than to choose a wrong interpretation.

Prompting a clarification dialogue: As we have seen, a human interpreter in this situation has the possibility of starting a subdialogue with the source-language speaker in order to clarify the ambiguity. This possibility should be given to the system, too.

There are certain problems with system-initiated clarification dialogues that are aimed at eliciting additional information needed in the translation process:

- If there are two many of these subdialogues, the main dialogue between the primary interactants cannot be maintained. This clearly applies to both human- and machine-mediated dialogues. But, as the system will be more often in a situation where it cannot decide between several readings than a human interpreter, in a machine-mediated situation the set of ambiguities eliciting a clarification dialogue must be reasonably restricted.
- There are cases of ambiguities where a clarification dialogue makes no sense to the user. The particle 'ja' is frequently occuring in spontaneously spoken German language. It can at least be interpreted as having the function to mark an uptake (17),[3] as an answer to a question (18) or as an acceptance of a proposal (19).

(17) Ja, dann lassen Sie uns doch einen Termin ausmachen.
((*no translation of 'ja') Let's make an appointment.*)

(18) A: Is the third a Monday?
B: Ja. (yes)

(19) A: Can you make it on Friday?
B: Ja. (yes)

It does not make any sense to prompt a clarification dialogue in this case by asking a question such as: "It what sense do you mean *ja*?".

[3]Uptake particles enable the hearer to adopt to acoustic properties of the speaker's utterances without losing propositional content. They signal a turn taking at the beginning of a turn and a turn holding within a turn. They cannot be translated directly. (Examples: 'ja', 'also').

- As one cannot expect a user to be familiar with linguistic terminology, the clarification queries must be stated in comprehensible terms. If this is not possible, the subdialogue will probably not lead to the intended interpretation.

Using default readings: If an ambiguity cannot be preserved, and a clarification dialogue is inadequate (either because the ambiguities cannot be explained in naive terms or because the ambiguity cannot reasonably resolved by a clarification dialogue), the system should have default rules that describe domain-dependent prefered readings.

What about the second role, the interpreting system functioning as a *coordinator*? As described above, a human interpreter is responsible for the success of the communication process. She has to make sure that the primary interactants ultimately reach the goal(s) of the dialogue. Clearly, this responsibility cannot be given to the machine. Instead, both the user and the machine have to be provided with certain strategies in order to manage successful communication. *For the machine* these strategies are restricted to either initiating a clarification dialogue or using certain heuristics leading to a default choice.

The fact that the responsibility for mutual understanding lies not with the system, clearly implies further effort *of the users*. They need to be aware of the machine's lack of any responsibility for (mis)understandings and have to find strategies to compensate for this: They should adapt to this situation by trying to collaborate with each other via the machine. This can only be achieved in a situation where all users can control the performance of the interpreting machine. There are several ways of user control in automatic dialogue interpreting:

- The system can present a paraphrase of the original utterance to the user. From these paraphrases the user can decide whether the system analysed his input utterance correctly. The performance of the translation component can be controlled by re-translating the target-language output back into the source language.

- Another way of control can be achieved by using a *mediating language*. In Verbmobil phase I we used English as the mediating language: The primary interactants (or users) were supposed to have at least passive knowledge of English. They talk to each other either in English or, if their English is not good enough, in their mother tongue (either German of Japanese). Only in the latter case Verbmobil is used as an interpreting device: It translates German or Japanese input into English. This approach has an obvious advantage: The users can control the situation, they can share the responsibility of ensuring understanding and they can act collaboratively.

5 Conclusion

This paper addresses the question how human-machine-human dialogues differ from human-human dialogues with respect to collaborative behavior. I argue that – as a machine cannot be responsible for mutual understanding of the human interactants – certain strategies have to be implemented:

In order to fullfil its *translator role*, the machine should be able to preserve ambiguities, to initiate a clarification dialogue in certain situations or to use domain-dependent default choices if no other contextual knowledge is available.

As the *coordinator role* cannot be fulfilled by the machine, the interpreting system should provide certain means of user control. I proposed either the system's paraphrasing of both input and output in source and target language, or the use of a mediating language that all users at least passively know. Future work, e.g. with Wizard-of-Oz situations, has to show whether the implementation of these proposals suffices to ensure successful communication in automatic dialogue interpreting.

Acknowledgments

This paper has benefited from discussions with Uwe Küssner, Manfred Stede, Joachim Quantz and Elisabeth Maier. I would also like to thank the anonymous reviewers for their detailed comments, which resulted in additional clarification.

References

[Apfelbaum, Wadensjoe 97] B. Apfelbaum, C. Wadensjö, "How does a VERBMOBIL affect conversation? Discourse analysis and machine-supported interaction", in Ch. Hauenschild, S. Heizmann (eds.), *Machine Translation and Translation Theory. Perspectives of Co-operation*, Mouton de Gruyter, Berlin, New York, to appear 1997

[Clark, Wilkes-Gibbs 86] H.H. Clark, D. Wilkes-Gibbs, "Referring as a Collaborative Process", *Cognition*, **22**, 1986, 1–39

[Gutt 91] E.-A. Gutt, *Translation and Relevance*, Basil Blackwell, Oxford, 1991

[Keith 84] H.A. Keith, "Liaison Interpreting – An Exercise In Linguistic Interaction", in W. Wilss, G. Thome (ed.), *Die Theorie des Übersetzens und ihr Aufschlußwert für die Übersetzungs- und Dolmetschkritik*, Gunter Narr Verlag, Tübingen, 1984, 308–317

[Sperber, Wilson 86] D. Sperber und D. Wilson, *Relevance*, Blackwell, Oxford, 1986

[Stenström 94] A.-B. Stenström, *An Introduction to Spoken Interaction*, Longman, London and Newy York, 1994

[Wahlster 93] W. Wahlster, *Verbmobil: "Ubersetzung von Verhandlungsdialogen*, Verbmobil-Report 1, DFKI Saarbrücken, 1993

[Zähner 90] Ch. Zähner, "The Role of the Memory in Teaching Liaison Interpreting", in *Multilingua* 9-3, 1990, 297–311