Gefördert vom
bmb+f
Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

Verbmobil
Verbundvorhaben

Eliminating Traces of Spoken Language Production in Automatic Dialogue Interpreting

Birte Schmitz

Technische Universität Berlin

Vm

Report 213
Juli 1997

Juli 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 7th International Conference on Theoretical and Methodological Issues in Machine Translation, TMI-97, Santa Fe, New Mexico, 1997, 184–192

Eliminating Traces of Spoken Language Production in Automatic Dialogue Interpreting

Birte Schmitz*

Technische Universität Berlin, Projekt KIT-VM11, Franklinstr.28/29, D-10587 Berlin, Germany
e-mail: birte@cs.tu-berlin.de

Abstract. This paper focuses on a particular problem of Automatic Dialogue Interpreting. As the input is unplanned spoken language, it contains certain *traces of the effort to formulate the utterances* (such as false starts, filled pauses, self-repairs or repetitions). In an adequate translation, these traces are obviously to be eliminated. I introduce a pragmatic-based approach to Automatic Dialogue Interpreting that handles these traces adequately. This approach is characterized by the fact that *dialogue acts* representing both the propositional content and the communicative intention of an utterance are used as a translation invariant.

1 Introduction

Dialogue Interpreting is characterized by the fact that the input is unplanned spoken language produced in a dialogue. This kind of input differs considerably from planned written texts: In a dialogue situation a speaker holding the turn is obliged to produce her utterances under pressure of time[2] – otherwise she will loose the turn. While she produces the turn, the hearer receives and interpretes it – these processes take place nearly simultaneously. This implies for the speaker that there is no way of correcting the output before the hearer receives it. In addition to repairs, all kinds of *traces of the effort to formulate the utterances* are witnessed by the hearer. These traces include typical disfluencies like false starts, filled pauses, repetitions of syllables, words or phrases, corrections, and other kinds of reformulations as well as markers which the speaker uses to relate to problems in formulating his contribution.

In order to produce an adequate translation of unplanned spoken language these traces are to be eliminated in the translation process. This, at least, is expected from a human interpreter, as [Keith 84, p.309] points out:[3]

> … many writers on practical aspects of interpreting have … pointed to the fact that, as a result of its oral representation, the text the interpreter is faced with often displays structural faults and inadequacies which force him to modify and reorganise it in order for the text to attain optimum coherence and efficiency (in the text-linguistic sense of these terms).

* 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.

[2] "The average speed of spoken English, including pauses, is in the neighborhood of 180 words per minute" [Chafe 82, S.36].

[3] A similar position is taken in [Shlesinger 91, p.150] and [Herbert 52, p.50].

This should be expected from a machine as well, as a cohesive and coherent discourse is a fundamental prerequisite for successful communication. Rendering the traces of spoken language production is problematic: As these traces usually do not provide propositional aspects of information they cannot easily be translated with semantic-based methods. If you decide to translate them anyway, you will gain as a result a discourse that is hardly understandable at all, as example (1), a turn taken from an appointment-scheduling dialogue,[4] demonstrates. In the preceeding discourse the topic was introduced, namely finding a date for a five-day meeting in Hamburg. In the last turn Speaker A gave the information that it is difficult for him to find a suitable date and that January is not possible for him at all. The following turn by Speaker B can be interpreted as an answer to this:

(1)

⟨äh⟩	das wär' auch bei mir sehr schlecht,	⟨äh⟩	ich find' im Februar auch ganz schlecht
um	*that would also be very bad with me*	*um*	*I find February totally bad as well*

also/-	doch,	Februar,	⟨äh⟩	vom zwanzigsten bis vierundzwanzigsten
so	*but*	*February*	*um*	*from the 20th to the 24th*

Februar würd' es gehen.	⟨äh⟩	im Januar	geht 's bei mir auch nicht,
of February it would be possible	*um*	*in January*	*it is not possible for me too*

aber +/am/+ im Februar,	am zwanzigsten bis vierundzwanzigsten Februar.
but in February	*20th to 24th of February*

A much more adequate translation is the following:

(2) January does also not suit me, but what about the 20th to 24th of February?

In this paper I develop a framework which allows the automatic derivation of such a translation. The prerequisite for this is a pragmatic-based approach to Automatic Dialogue Interpreting. The underlying ideas stem for experiences within the project Verbmobil [Wahlster 93].

This paper is structured as follows. First, I distinguish three different types of traces of spoken language production and discuss for each type whether it is more adequate to render or to eliminate it in the translation. I then describe the pragmatic-based approach to Automatic Dialogue Interpreting, which is characterized by the fact that the translation process is based on dialogue acts. Next, I show for the different types of traces how they can be reduced or elmininated within the pragmatic-based approach. Finally, I summarize the main advantages of this approach and refer to further applications of its main ideas.

2 Types of Traces of Spoken Language Production

There are different types of traces of spoken language production, each requiring a different treatment in the translation process. Following [Gülich, Kotschi 95], I distinguish *verbalization procedures*, *qualification procedures*, and *treatment procedures*.

[4] This and all other transliterations are taken from the corpus of dialogues that were recorded and transliterated in the project Verbmobil. 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.

Verbalization procedures are marked by all kinds of "phenomena of performance" such as hesitation phenomena, filled and unfilled pauses, false and fresh starts and phenomena that can be interpreted as self-repairs.

Which of these markers should be eliminated in the translation? It is obviously not adequate to translate *self-repairs* into the target language. According to [Levelt 83, p.44] a self-repair typically consists of three parts: the *original utterance*, which contains the trouble spot, the *editing part*, i.e. a period of hesitation, possibly containing an editing term such as 'uh', 'well', and the *repair*. In order to eliminate a self-repair, the trouble spot in the original utterance has to be replaced by the repair. False and fresh starts are special types of self-repairs; false starts should be removed in the translation. Note that self-repairs can refer to morphological, syntactic or semantic aspects of single words or phrases. All self-repairs marking verbalization procedures are characterized by the fact that they are performed within one utterance.

Filled and unfilled Pauses serve at least two purposes in the discourse: On the one hand they express a hesitation (such as the occurrences of 'äh' in (1)), on the other hand they are used to structure the discourse. Insofar as discourse-structuring pauses are used when segmenting the input, they can be regarded as contributing to the translation. Hesitation-marking pauses, however, should not be rendered in the target language.

Qualification Procedures are signalled by all utterances the speaker produces to relate to certain problems that she encounters while formulating her contribution. Examples are utterances such as 'I don't know', 'this is difficult' or 'just let me see'. [Gülich, Kotschi 95, p.35] characterize these procedures as "explicit manifestations of a speaker's constant cognitive monitoring of his/her discourse production."

What happens with these markers in the translation process? At least in German and in English this type of procedure is typically expressed by *routine formulae* – linguistic patterns which are tied to certain recurrent social situations. An important aspect of routine formulae is presented by the fact that usually their "literal" (i.e. compositionally built up) meaning is not their intended meaning in an utterance. Therefore their translation cannot be produced compositionally, but should instead be constructed on the basis of their function in the discourse. For example, in German appointment-scheduling dialogues many utterances are introduced by phrases like 'ich würde sagen' (*I would say*), 'ich muß sagen' (*I must say*), 'ich wollte fragen' (*I wanted to ask*); these phrases do not contribute to the meaning of the utterance, but are rather used to give the speaker planning time and to make the output appear more fluent. Their literal translations does generally not fulfill the same function.

Treatment procedures According to [Gülich, Kotschi 95, p.34] "it is characteristic of them (these phenomena, B.S.) that they refer to a preceding segment by means of a new utterance, which somehow changes, modifies, reformulates or expands the earlier utterance. In other words, the speaker performs some kind of treatment on an utterance which has already been produced. "

In the Verbmobil corpus, reformulating treatment procedures are mainly expressed by *repetitions*, *corrections* and *specifications*. Whether they should be reduced in the translation or not depends on their discourse function. Only if they are not intended as a support for the hearer in deriving the originally intended interpretation, they should be reduced. In this case they only mark the speaker's effort to produce the utterance – the translation is much more coherent without them. For repetitions this is illustrated in (1).

A correction signalling a treatment procedure differs from a self-repair signalling a verbalization procedure with respect to the repaired unit. Treatment procedures refer to a complete utterance that has already been produced. In the repairing utterance either its propositional content or its communicative intention is corrected. An examples is given in (1), where the speaker first rejects February as a suitable date and then corrects herself by suggesting a certain date in February. As this correction does not provide any information for the hearer, it should be eliminated in the translation.

Are specifications to be reduced too? An example for a specification is given in (1). The speaker first suggests February, and in the following utterance he specifies his suggestion to the 20th to 24th of February. In the translation it suffices to render the most specific date proposed.

In the following section I introduce an approach to Automatic Dialogue Interpreting that makes it possible to eliminate the traces of spoken language production discussed above.

3 A Pragmatic-Based Approach to Automatic Dialog Interpreting

An approach that handles traces of spoken language production adequately has to fulfill certain requirements with respect to the representation of utterances and turns:

- The representation of utterances should abstract away from traces signalling verbalization procedures;
- The representation of utterances should contain information about discourse functions of particular words, phrases or the whole utterance. This is required in order to adequately translate markers signalling qualification procedures.
- In order to eliminate a repetition, specification or correction of an utterance, these phenomena must be automatically recognized in the turn representation. Therefore the representation of turns should consist of a list of representations of utterances. A specification can be only recognized on the basis of a subsumption hierarchy which makes it possible to compare two concepts with respect to their specificity.

These demands are met by a pragmatic-based approach to Automatic Dialogue Interpreting [Schmitz, Quantz 95] (see also [Morimoto et al. 92],[Levin et al. 95]). In this approach translation is based on *dialogue acts*. Dialogue acts are abstract types of information that represent both the propositional content and the communicative intention (or illocution) of an utterance. In Verbmobil we are currently using a set of 40 dialogue acts [Jekat et al. 95] which proved to be relevant in the scenario of appointment-scheduling dialogues. Typical examples are dialogue acts like **Greet** ('hello'), **Accept_Date** ('Monday suits me'), **Suggest_Support_Date** ('What about Thursday at two o'clock?') or **Give_Reason** ('I'll be on a conference then'). Some dialogue acts, such as greeting or thanking someone, are purely illocutional. Others either contain certain aspects of propositional information such as the information that the topic of a sentence is a date (as in **Accept_Date**) or an instance of a date (as in **Suggest_Support_Date**) or they contain a whole proposition (as in **Give_Reason**).

The basic idea of the pragmatic-based approach is to use dialogue acts as a translation invariant: They function as a target representation of the analysis process and as a source representation for the generation process. Note that this approach is not neccessarly an interlingual one, as the propositional content of the utterance represented in the dialogue act can be rendered into the target language using a traditional transfer approach. In Table 1 the results of a pragmatic-based approach are compared to a more “literal” or semantic-based approach of a turn taken from the Verbmobil data.[5] In this example, with the exception of **Clarify_State**, all the dialogue acts do not provide any propositional information. **Clarify_State** refers to an informing dialogue act; its representation contains a proposition (“the time period we are talking about is from the 21st until the 25th of August.”) The translation produced by an pragmatic-based approach is much easier understood, since the discourse is more coherent. Note that there might be a chance to understand the intended interpretation of the semantic-based translation, if it is read with correct prosodic marking. Produced by a state of the art speech-synthesis system without the intended prosodic marking, it will most probably be not understandable at all.

Certain utterances, which are analysed as being traces of spoken language production, are not rendered into the target language: **Accept_Date** is repeated in the original, but it is translated only once. The second **Feedback_Acknowledgement** occurs within the turn (**not** at the beginning or end of it). In this context it can be left out in the translation. (In a more interactive kind of conversation some kind of **Feedback_Acknowledgement** is typically given by a single turn, as Levinson points out. In [Levinson 83, p.316] he characterizes the closing section of certain conversations as: “… a sequence of *Okays* just prior to a final exchange of *Good-byes*.” In that case a **Feedback_Acknowledgement** clearly has to be translated.)

Pragmatic-based Translation		
SL Utterance	**Dialogue Act**	**TL Utterance**
das ist ja wunderbar, daß wir uns so schnell einigen konnten,	**Feedback_Acknowledgement**	allright,
das geht hervorragend,	**Accept_Date**	
also da hab’ ich auch Zeit.	**Accept_Date**	that is okay with me,
das ist vom einundzwanzigsten August bis zum f”unfundzwanzigsten, das ist ein Freitag.	**Clarify_State**	that is from the 21st of August until Friday the 25th,
dann sehen wir uns doch da,	**Confirm**	see you then,
gut.	**Feedback_Acknowledgement**	
Wiederschauen.	**Bye**	bye
Semantic-based Translation		
this is really wonderful, that we could reach an agreement so quickly, that works very well, so I have time then, that is from the 21st of August until the 25th that is a Friday, then we will see us there, good, bye-bye		

Table 1. Comparison of a pragmatic-based and a semantic-based translation

In the following section I demonstrate how this pragmatic-based approach can handle each type of trace of spoken language production adequately.

[5] Note that both translation are not produced automatically but by the author.

4 Eliminating Traces of Spoken Language Production

4.1 Markers Signalling Verbalization Procedures

It is an advantage of the pragmatic-based approach that it uses dialogue acts as a representation of utterances, and that these dialogue acts abstract away form all kinds of markers signalling verbalization procedures: Since translation is based on the dialogue-act representation of utterances, the elimination of markers signalling verbalization procedures in the translation is guaranteed in this approach.

In the pragmatic-based approach these markers have to be eliminated in the analysis process, when for each utterance a dialogue-act representation is built up. Most of the hesitation phenomena (i.e filled and unfilled pauses[6]) can already be reduced in the process of speech recognition. There also exist approaches dealing with self-repairs. In [Heeman, Allen 94] the authors present a method to automatically recognize and correct certain self-repairs (false starts are not dealt with), which does not rely on any syntactic or semantic pre-proccesing besides a part-of-speech tagger. Their algorithm is based on the recognition of certain repair patterns which consist of words, substitutions of words, repair markers, and word fragments. Their results are rather promising: they can recognize and correct 80% of the repairs they are dealing with.

4.2 Markers Signalling Qualification Procedures

As I argued above, formulae indicating qualification procedures should be translated according to their function in the discourse. One class of these formulae marks a qualification procedure that in Verbmobil is referred to as **Deliberation**. In performing this dialogue act the speaker shows that she needs some time to plan her further contribution but endeavours not to loose the turn. Typical examples from German Verbmobil dialogues are ‘ich sehe gerade’ (*I just see*), ‘ich schau’ mal nach’ (*I have a look*), or ‘wenn ich hier auf meinen Terminplaner schaue’ (*if I have a look at my diary here*). Obviously these expressions cannot be translated compositionally into English. If at all, they should be rendered by English phrases expressing a similar discourse function such as ‘let’s see’, which frequently occurs in English Verbmobil dialogues, as the following part from a turn (taken from Verbmobil dialogue r342c) demonstrates.

(3) I’m booked ⟨ähm⟩ Monday November the twenty second because I’ve got a class nine to twelve a meeting one thirty to three and a lab from three to five but <u>let’s see</u> on Tuesday the twenty third I’m free from three to five PM and <u>let’s see</u> ⟨ähm⟩ on Thursday the twenty fifth I’m free from twelve to two

In the pragmatic-based approach to automatic dialogue interpreting these phrases are translated adequately: Their translation is based on their function in the discourse, which is expressed by the respective dialogue act.

Other examples for phrases marking qualification procedures are ‘I’m afraid’ or ‘I guess’. These phrases do not perform a dialogue act themselves but are part of an utterance performing a dialogue act, as the following examples demonstrate:

[6] [Batliner et al. 95] present the results of an automatic classification of unfilled pauses as either hesitation or repair markers.

(4) well it's not real good I'm afraid I have a meeting from two to four on that day (r394c)
(5) okay so I guess we should really get together sometime (r365c)

'I'm afraid' can be interpreted as a marker for the dialogue act **Reject_Date**. In the pragmatic-based approach it is not translated compositionally but expressed in the target language by a phrase expressing the same dialogue act. 'I guess' is used in utterances expressing a variety of dialogue acts. Its main function consists of giving the utterance an appearance of greater fluency. This function is expressed by different means in different languages. In the pragmatic-based approach it is the task of the generation component to create fluent target language discourse.

4.3 Markers Signalling Treatment Procedures

As argued above, repetitions express redundant information which does not contribute to the intended interpretation of the turn and are therefore to be eliminated in the translation. How this can be achieved within the pragmatic-based approach is demonstrated with the example given in Table 1, where two successive source language utterances express an **Accept_Date**. The turn is represented as a list of dialogue acts. If this list contains two successive dialogue acts representing the same information, only one is handed over to the generation component.

In the pragmatic-based approach specifications can be reduced, too. In the example in Table 2 (taken from Verbmobil dialogue m455d), the speaker makes two proposals in two successive utterances, with the second proposal being a specification of the first one. This can automatically be recognized by comparing the proposed dates. In this example, the dialogue act **Give_Reason** is not translated because it is regarded as non-relevant to the intended interpretation of the turn.

Pragmatic-based Translation		
SL Utterance	**Dialogue Act**	**TL Utterance**
vielleicht noch Ende Mai, +/Anfang Juli/+ Anfang Juni,	**Suggest_Support_Date**: date_spec: end of May, beginning of June	
da hätt' ich Zeit.	**Give_Reason**	
neunundzwanzigster Mai bis zweiter Juni?	**Suggest_Support_Date**: date_spec: 29th of May to 2nd of June	Are you free from the 29th of May to 2nd of June?
Semantic-based Translation		
Maybe still at the end of May, the beginning of June, I would have time then, 29th of May until the 2nd of June?		

Table 2. Translating the most specified utterance

That the pragmatic-based approach handles corrections in the same fashion, is demonstrated in Table 3 (taken from Verbmobil dialogue j544a).

Pragmatic-based Translation		
SL Utterance	**Dialogue Act**	**TL Utterance**
im Dezember geht es bei mir,	**Suggest_Support_Date**; date_spec: December	
⟨ähm⟩ und zwar erst ab dem siebten Dezember.	**Suggest_Support_Date**; date_spec: from the 7th of December on	
nein, oh, Entschuldigung.	**Repair_Marking**	
da hab' ich einen Termin übersehen	**Give_Reason**	
ab dem zehnten ⟨äh⟩ Dezember könnte ich	**Suggest_Support_Date**; date_spec: from the 10th of December on	I'm free from the 10th of December on
Semantic-based Translation		
December is okay, namely from the 7th of December on, no, oh, sorry, then I have failed to see an appointment, I have time from the 10th of December on		

Table 3. Eliminating corrections

5 Conclusion

I argued here that spoken language cannot be handled by traditional Machine Translation approaches, since it is not adequate to translate all kinds of traces of spoken language production. I therefore proposed an extension to the semantic-based transfer approach, which is characterized by the fact that translation is not only based on the propositional contents of an utterance but also on its communicative intention or illocution. A further extension is given by the fact that the translation takes contextual information into account: In order to translate an utterance, its role in the turn is investigated. Only if it is regarded as relevant with respect to the intended interpretation of the turn, the utterance is translated. I showed that this is the way to construct an adequate translation, as such a translation enables the target-language hearer to derive the originally intended interpretation. Certain aspects of this approach are already implemented in the Verbmobil system: The dialogue act is provided to the transfer component and can be used to abstract away from traces of spoken language production that occur within an utterance. Currently we are working on the integration of the other, turn-based elimination procedures.

Note that the basic idea of this approach can also be useful for other applications: In an Automatic Dialogue Interpreting situation one of the dialogue partners might ask for a summary of the preceding dialogue or a protocol of the whole dialogue. For this task the elimination of traces of spoken language production is a fundamental prerequisite.

References

[Batliner et al. 95] A. Batliner, A. Kiessling, S. Burger, E. Noeth, *Filled Pauses in Spontaneous Speech*, Verbmobil-Report 88, Universität Erlangen-Nürnberg, 1995

[Chafe 82] W. L. Chafe, "Integration and Involvement in Speaking, Writing, and Oral Literature", in D. Tannen (ed.), *Spoken and Written Language*, Ablex Publishing Corporation, Norwood, New Jersey, 1982, 35–53

[Gülich, Kotschi 95] Elisabeth Gülich und Thomas Kotschi, "Discourse Production in Oral Communication" in U. Quasthoff (ed.), *Aspects of Oral Communication*, de Gruyter, Berlin, New York, 1995, 30–66

[Heeman, Allen 94] P. Heeman, J. Allen, "Detecting and Correcting Speech Repairs" in "32nd Annual Meeting of the Association of Computational Linguistics", Las Cruces, New Mexico, 1994, 295–302

[Herbert 52] J. Herbert, *Manual de l'interprète*, Georg, Genève

[Jekat et al. 95] S. Jekat, A. Klein, E. Maier, I. Maleck, M. Mast, J.J. Quantz, *Dialogue Acts in VERBMOBIL*, Verbmobil Report 65, Universität Hamburg, 1995

[Keith 84] H.A. Keith, "Liason 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

[Levelt 83] W.J.M. Levelt, "Monitoring and Self-Repair in Speech", *Cognition*, **14**, 1983, 41–104

[Levin et al. 95] L. Levin, O. Glickman, Y. Qu, D. Gates, A. Lavie, C.P. Rosé, C. Van Ess-Dykema, A. Waibel, "Using Context in Machine Translation of Spoken Language" in "Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation", TMI-95, Leuven, 1995, 173–187

[Levinson 83] S.C. Levinson, *Pragmatics*, Cambridge University Press, 1983

[Morimoto et al. 92] T. Morimoto, M. Suzuki, T. Takeazawa, G. Kikui, M. Nagata, M. Tomokiyo, "A Spoken Language Translation System: SL-TRANS2", in "Actes du quinzième colloque international en linguistique informatique COLING-92", Nantes, 1992, 1048–1052

[Schmitz, Quantz 95] B. Schmitz and J.J. Quantz, "Dialogue-Act Type Recognition in Automatic Dialogue Interpreting", in "Proceedings of the The Sixth International Conference on Theoretical and Methodological Issues in Machine Translation", TMI-95, Leuven, 1995, 33–47

[Shlesinger 91] M. Shlesinger, "Interpreter Latitude vs. Due Process. Simultaneous and Consecutive Interpretation in Multiligual Trial", in S. Tirkkonen-Condit (ed.), *Empirical Research in Translation and Intercultural Studies*, Gunter Narr, Tübingen, 1991, 147–155

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