Criteria for the Segmentation of Spoken Input into Individual Utterances

Marion Mast[1]
Elisabeth Maier[2]
Birte Schmitz[3]

[1]FAU,[2]DFKI,[3]TUB

Report 97
Dezember 1995

Dezember 1995

Marion Mast[1]
Elisabeth Maier[2]
Birte Schmitz[3]

[1]Lehrstuhl für Mustererkennung (Inf. 5)
Friedrich–Alexander–Universität Erlangen–Nürnberg
Martensstr. 3
D–91058 Erlangen

[2]DFKI GmbH
Stuhlsatzenhausweg 3
D–66123 Saarbrücken

[3]Technische Universität Berlin
Fachbereich Informatik (13)
Projektgruppe KIT-VM 11
Sekr. FR 5-12
Franklinstr. 28/29
D–10587 Berlin

Tel.: (09131) 85 - 7799
e-mail: `mast@informatik.uni-erlangen.de`

Gehört zum Antragsabschnitt: TP 10 Übersetzungsorientierte Dialogverarbeitung und TP 11.1 Lokale Interpretationshypothesen Deutsch

Das diesem Bericht zugrundeliegende Forschungsvorhaben wurde mit Mitteln des Bundesministers für Bildung, Wissenschaft, Forschung und Technologie unter den Förderkennzeichen 01 IV 102 H/0, 01 IV 101 K/1 und 01 IV 101 Q gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei den Autoren.

Introduction

This report describes how spoken language turns are segmented into utterances in the framework of the VERBMOBIL project.

The problem of segmenting turns is directly related to the task of annotating a discourse with dialogue act information: an utterance can be characterized as a stretch of dialogue that is attributed one dialogue act. Unfortunately, this rule in many cases is insufficient and many doubtful cases remain. We tried to at least reduce the number of unclear cases by providing a number of hands-on rules for segmenting dialogues. In that sense, we hope that this document is helpful for labeling dialogues with discourse information.

This report has to be seen as an extension of the VERBMOBIL Report No. 65 [2] that includes a detailed discussion of the various dialogue act types employed in VERBMOBIL. All dialogue act types used throughout this report are taken from the set defined there.

VERBMOBIL - A brief Overview

VERBMOBIL combines the key technologies speech processing and understanding and machine translation. The long-term goal of this project is the development of a prototype for the translation of spontaneous spoken dialogues between two (business) partners who want to find a date for a meeting (for more detail on the objectives of VERBMOBIL see [4]). A special characteristic of VERBMOBIL is that both participants are assumed to have at least a passive knowledge of English which is used as intermediate language. Translations are produced *on demand* so that only parts of the dialogue are processed Therefore the system distinguishes between two processing modes:

- *Deep Processing:* when one of the dialogue participants doesn't speak English and therefore requests a translation. In this case the input goes through phases of speech recognition, syntactic and semantic analysis, dialogue processing, transfer, re-generation and synthesis, delivering spoken output in the target language.

- *Shallow Processing:* when a dialog participant speaks English which means there is no necessity for translation.

The dialogue model describes the expected actions of the participants in an appointment scheduling dialogue. Basic units in the dialogue model are dialogue acts. A model based on dialogue acts seems to be an appropriate approach also from the point of view of machine translation theory and of transfer in particular: While in written discourse sentences can be considered the basic units of transfer, this assumption is not valid for spoken dialogues. In many cases only sentence fragments are uttered, which often are grammatically incomplete or even incorrect. Therefore for spoken language different descriptive units – in VERBMOBIL dialogue acts – have to be chosen.

Segmentation Criteria

We are concerned with the segmentation of turns into utterances, where each utterance corresponds to a dialogue act. As definition of turn we use the rather technical definition from [3, p.34]: "The turn is everything A says before B takes over and vice versa." Because there is no equivalent definition of the term utterance as for the term sentence in written language, we composed some rules, which take the special forms of spoken language into account.

Since it is our goal to divide turns into single utterances each fragment of a turn belongs to an utterance. The following criteria are used both for manual and automatic segmentation of turns in the framework of the VERBMOBIL project.

In the following English and German examples (from the VERBMOBIL corpus) each utterance is followed by the performed dialogue act (in small capitals). Omitted is the marking of the speaking direction which can be AB which means A speaks to B or vice versa (In VERBMOBIL only dialogues with two participants are recorded).

1. Each beginning of a turn / end of a turn is also the beginning of an utterance / end of an utterance. In some cases the syntactic or the semantic structure of an utterance is incomplete. Typical examples for such fragmentary utterances can be found in cases which can be characterized as FEEDBACK_ACKNOWLEDGEMENT[1].

Example taken from Verbmobil Memo 24, Appendix I		
SUS	027 :	Ah ⟨Atmen⟩, so ⟨ehm⟩ ⟨2⟩ ⟨Schmatzen⟩ maybe, it would be the easiest way that ⟨äh⟩ fetch you at home, ⟨P⟩ if you like? SUGGEST_SUPPORT_LOCATION
DAN	028 :	Yeah. FEEDBACK_ACKNOWLEDGEMENT
SUS	029 :	Yeah? So you should give me your address CLARIFY_STATE, then I ⟨P⟩ will fetch you at ⟨P⟩ seven ⟨P⟩ PM. CONFIRM

2. An important aspect for segmentation is the existence of at least one finite verb within an utterance. The following procedure is carried out for the segmentation of a turn: all material that belongs to the verb frame of a finite verb belongs to the same utterance. A new utterance begins where a word belongs to a new verb frame. That way it is guaranteed that both the obligatory and the optional elements of a verb are included in the same utterance. Of course it has to be taken into account that in spoken language syntactic constraints can be violated, i.e. that obligatory elements of a verb frame are missing; In most cases it is still possible to comprehend such utterances.

Example CDROM 8, 127c.trl		
JBT	004 :	⟨A⟩ two ⟨#Mikrobe⟩ o'clock in the afternoon sounds fine ⟨;period⟩ ⟨;seos⟩ ACCEPT_DATE where would you like to meet ⟨;period⟩ ⟨#Klicken⟩ ⟨#Mikrobe⟩ ⟨;seos⟩ REQUEST_SUGGEST_LOCATION

[1]The transliterations of the dialogues include German labels as used within VERBMOBIL. Mostly they refer to noises like breathing (Atmen), lip smacks (Schmatzen), touches of the microphone (Mikrobe), klicking (Klicken); they also hint at segment ends (seos), at hesitations (ehm), at pauses (P) etc.

Example CDROM 8, r131c.trl		
SRH	001 :	⟨Schmatzen⟩ ⟨A⟩ it will ⟨!2 it'll⟩ be impossible for me to make a meeting this afternoon REJECT_DATE ⟨;period⟩ ⟨;seos⟩ ⟨Schmatzen⟩ ⟨A⟩ ⟨ähm⟩ ⟨A⟩ ⟨Schmatzen⟩ ⟨A⟩ how 'bout toward the end of next week ⟨;period⟩ say ⟨;comma⟩ +/the/+ ⟨;comma⟩ Friday the sixteenth ⟨;quest⟩ ⟨#Klicken⟩ ⟨A⟩ ⟨;seos⟩ SUGGEST_SUPPORT_DATE
JBT	002 :	⟨#Mikrobe⟩ ⟨Schmatzen⟩ Friday the sixteenth looks ⟨;comma⟩ ⟨P⟩ good ⟨;comma⟩ as long as I can get away by four *P*M ⟨;period⟩ ⟨#Rascheln⟩ ⟨#Klicken⟩ ⟨;seos⟩ ACCEPT_DATE

Example CDROM 8, r132c.trl		
RJK	005 :	I will ⟨!2 I'll⟩ be back on the seventeenth ⟨;period⟩ ⟨;seos⟩ SUGGEST_SUPPORT_DATE and ⟨;comma⟩ ⟨ähm⟩ I will ⟨!2 I'll⟩ be free in the morning ⟨;comma⟩ ⟨ähm⟩ until ten AM ⟨;comma⟩ that day ⟨;comma⟩ ⟨;seos⟩ ⟨A⟩ ⟨ähm⟩ SUGGEST_SUPPORT_DATE and the next day I am ⟨!2 I'm⟩ free ⟨;comma⟩ until ⟨;comma⟩ two PM ⟨;comma⟩ ⟨;seos⟩ SUGGEST_SUPPORT_DATE and then I have class ⟨;comma⟩ ⟨ähm⟩ until four ⟨;period⟩ ⟨#Klicken⟩ ⟨A⟩ ⟨;seos⟩ SUGGEST_SUPPORT_DATE

It is important to note that elaborations of concepts mentioned in an utterance are not treated as utterances of their own unless they occur with an additional finite verb. In the VERBMOBIL domain of appointment scheduling dialogues typical cases of elaborations concern the introduction of a time frame, e.g. a week, which is then subsequently elaborated upon by giving more specific time frames, as e.g. a day, an hour, etc (see the examples below).

Example CDROM 8, r133c.trl		
NBS	000 :	(...) I ⟨;comma⟩ only seem to have Thursdays and Fridays free ⟨;period⟩ ⟨Schmatzen⟩ ⟨A⟩ ⟨ähm⟩ in the afternoon ⟨;comma⟩ on Thursday but ⟨;comma⟩ any time on Fridays ⟨;period⟩ ⟨;seos⟩ SUGGEST_SUPPORT_DATE how 'bout you ⟨;period⟩ ⟨#Klicken⟩ ⟨Schmatzen⟩ ⟨A⟩ ⟨;seos⟩ REQUEST_SUGGEST_DATE

Example CDROM 8, r136c.trl		
NBS	002 :	⟨#⟩ ⟨Schmatzen⟩ ⟨A⟩ no ⟨;comma⟩ I am ⟨!2 I'm⟩ going to ⟨!2 gonna⟩ be ⟨;comma⟩ out of town on Tuesday ⟨;period⟩ ⟨A⟩ as well as Wednesday and Thursday ⟨;comma⟩ ⟨;seos⟩ ⟨A⟩ SUGGEST_EXCLUDE_DATE but ⟨;comma⟩ ⟨Schmatzen⟩ I can meet ⟨;comma⟩ ⟨Schmatzen⟩ on Friday ⟨;comma⟩ ⟨Schmatzen⟩ ⟨A⟩ from ⟨;comma⟩ ⟨A⟩ ⟨Schmatzen⟩ one PM to three ⟨;comma⟩ or ⟨;comma⟩ four to five ⟨;comma⟩ ⟨#Klicken⟩ ⟨;seos⟩ SUGGEST_SUPPORT_DATE

Example CDROM 3, g143a.trl		
BAC	003 :	⟨;T⟩ann würd' ich noch gerne mit Ihnen einen Besprechungstermin im Oktober abmachen, SUGGEST_SUPPORT_DATE und zwar ⟨A⟩ am⟨Z⟩ ⟨A⟩ Montag, ⟨A⟩ den siebzehnten Oktober, ⟨P⟩ würd' ich gerne mich mit Ihnen treffen, oder ⟨A⟩ wahlweise Montag, vierundzwanzigster Oktober. SUGGEST_SUPPORT_DATE

3. For dependent clauses, the preceding rule is also applicable. Each dependent clause which contains a finite verb is seen as an utterance of its own.

Example CDROM 8, r202c.trl		
YAW	003 :	⟨Schmatzen⟩ well ⟨;comma⟩ Monday is ⟨;Mikrobe⟩ kind of bad REJECT_DATE because I have a seminar from ⟨#⟩ ten *AM* to five *PM* ⟨#⟩ ⟨;seos⟩ GIVE_REASON unless you are ⟨!2 you're⟩ free some time after that ⟨;period⟩ ⟨A⟩ ⟨#Quietschen⟩ ⟨#Klicken⟩ ⟨#Rascheln⟩ ⟨;seos⟩ SUGGEST_SUPPORT_DATE

Example CDROM 8, r136c.trl		
RJK	001 :	⟨(⟩. . .) Monday I am ⟨!2 I'm⟩ ⟨;comma⟩ ⟨Geräusch⟩ quite busy ⟨;comma⟩ ⟨A⟩ ⟨Schmatzen⟩ all day ⟨;comma⟩ ⟨;seos⟩ SUGGEST_EXCLUDE_DATE but Tuesday I am ⟨!2 I'm⟩ free in the afternoon ⟨;comma⟩ ⟨#Klicken⟩ ⟨#Rascheln⟩ ⟨;seos⟩ SUGGEST_SUPPORT_DATE

Example CDROM 1, n003k.trl		
PS	1005 :	⟨#Klicken⟩ ⟨P⟩ dienstags um zehn is' bei mir jetzt ⟨!jetz⟩ wiederum schlecht REJECT_DATE weil ich da noch⟨Z⟩ ⟨äh⟩ ⟨P⟩ trainieren bin. GIVE_REASON ⟨A⟩ ich denke wir sollten das Ganze dann doch auf die nächste Woche verschieben. SUGGEST_SUPPORT_DATE geht 's bei Ihnen da ⟨P⟩ ⟨#Klicken⟩ ⟨P⟩? REQUEST_COMMENT_DATE

Exeptions of this rule are constructions which are mentioned in the following rule.

4. There are a number of exceptions from the previous rules:[2] one of them concerns routine formulas that are followed by sentential complements as in the following examples:

   - I mean . . .
   - I think . . .
   - you know . . .
   - I see . . .

[2] The discourse functions of conventionalized / fixed / idiomatic expressions are described in [1].

- I'd say that ...
- I would propose to ...
- I wanted to ask whether ...
- if you think that ...

Example CDROM 8, r137c.trl		
RJK	005 :	⟨Schmatzen⟩ ⟨A⟩ okay ⟨;comma⟩ but I might be a minute or two late ⟨;period⟩ ⟨;seos⟩ ACCEPT_DATE so ⟨;comma⟩ ⟨A⟩ do not ⟨!2 don't⟩ be surprised ⟨;comma⟩ if I am ⟨!2 I'm⟩ there at ten thirty seven ⟨;period⟩ ⟨#Klicken⟩ ⟨A⟩ ⟨;seos⟩ CLARIFY_STATE
NBS	006 :	⟨#⟩ ⟨A⟩ no ⟨;comma⟩ I like promptness ⟨;period⟩ ⟨;seos⟩ GIVE_REASON ⟨;comma⟩ ⟨;seos⟩ I hope that you are ⟨!2 you're⟩ on time FEEDBACK_RESERVATION ⟨;period⟩ ⟨;seos⟩ ⟨A⟩ I mean if you ⟨;comma⟩ if you are ⟨!2 you're⟩ planning to be late ⟨;comma⟩ we could just make it at a later time ⟨;period⟩ ⟨#Klicken⟩ ⟨A⟩ ⟨;seos⟩ REQUEST_SUGGEST_DATE

Example CDROM 1, n001k.trl		
PS	1009 :	⟨P⟩ ich muss sagen mir wär's dann lieber wenn wir die ganze Sache auf Mai verschieben. SUGGEST_SUPPORT_DATE ⟨P⟩ geht es da bei Ihnen auch? ⟨#Klicken⟩ ⟨P⟩ REQUEST_COMMENT_DATE

5. Another exception concerns conventionalized expressions that do not necessarily contain a verb. Examples are phrases for greeting a person, for expressing thanks, etc. Typical phrases are: *hello, good morning, thanks, see you*, etc.

Example Dialogue menm_ffmw_7		
MEN	007 :	(...) Tuesday November twentythird eight to ten AM SUGGEST_SUPPORT_DATE how's that sound to you? REQUEST_COMMENT_DATE
FFM	008 :	that's fine ACCEPT_DATE I'll see you Tuesday November twentythird then CONFIRM thanks THANK

Example TP 13 Gespräch 22		
DAN	000 :	Hello, GREET my name is Daniel Jones, (. . .) INTRODUCE_NAME

Example CDROM 3, m127d.trl		
BLO	025 :	ja, guten Tag, GREET mein Name ist Leyweber, L E Y W E B E R. INTRODUCE_NAME

6. Difficulties arise for particles which can perform several functions. E.g. the problem is to distinguish between *yes* as an uptake and *yes* as an confirming answer. To solve this problem we designed some heuristic default rules for the treatment of discourse particles etc. These rules are used, if no further disambiguating information is available. For discourse particles and interjections like e.g. *alright, I see, wonderful, great, ok, oh, yes, well, ah, no* etc. the following rules apply:

   - if such a word occupies the sentence-initial position it is not segmented as an utterance of its own; this approach is motivated by the fact that for such words it is less likely that they form a prosodic unit – also due to their brevity. Rather, these particles have to be attributed the function to signal the uptake of a turn[3].

Example CDROM 8, r201c.trl		
TL	002 :	⟨#Klicken⟩ ⟨#⟩ ⟨Schmatzen⟩ ⟨A⟩ how does the twentieth or the twenty first look ⟨;period⟩ ⟨A⟩ ⟨#Klicken⟩ ⟨;seos⟩ SUGGEST_SUPPORT_DATE
SRH	003 :	⟨A⟩ well we could do the twentieth in the morning ⟨;period⟩ ⟨;seos⟩ (. . .) SUGGEST_SUPPORT_DATE

Example CDROM 3, g144a.trl		
BAC	005 :	ja, ⟨A⟩ dieser ⟨#Mikrowind⟩ würd' ich ja gerne ⟨A⟩ danach haben, und zwar⟨Z⟩ am Sonntag, den einunddreissigsten Juli vierundneunzig ⟨A⟩. SUGGEST_SUPPORT_DATE

   - also expressions that precede phrases that are used when addressing a dialogue partner are not considered an utterance of their own[4].

[3]The discourse functions of such particles in the scenario of VERBMOBIL are described in [1].

[4]This case so far only occurs in a subset of the German data.

Example CDROM 3, m127d.trl

BLO	035 :	ja, Herr Nadolmy, ⟨A⟩ wir müssen noch einen Termin für unsere Reisevorbereitung besprechen. INIT_DATE

- two or more particles together, instead, are considered an utterance. The dialogue act of such an utterance in most cases is of type FEEDBACK.

Example CDROM 8, r214c.trl

JAB	004 :	⟨Schmatzen⟩ ⟨A⟩ ⟨ähm⟩ yeah ⟨;comma⟩ two o'clock sounds good to me ⟨;comma⟩ ⟨;seos⟩ ⟨A⟩ ⟨ähm⟩ ⟨A⟩ ⟨Schmatzen⟩ ⟨A⟩ ⟨#Klopfen⟩ ACCEPT_DATE yeah ⟨;comma⟩ that is ⟨!2 that's⟩ great ⟨;period⟩ ⟨;seos⟩ ACCEPT_DATE I will ⟨!2 I'll⟩ just write that down on my other calendar ⟨;period⟩ ⟨A⟩ ⟨Schmatzen⟩ ⟨#Klicken⟩ ⟨A⟩ ⟨;seos⟩ ACCEPT_DATE
ISN	005 :	⟨#Klicken⟩ ⟨Schmatzen⟩ ⟨A⟩ okay good ⟨;comma⟩ ⟨;seos⟩ FEEDBACK_ACKNOWLEDGEMENT then I will ⟨!2 I'll⟩ ⟨;comma⟩ ⟨äh⟩ see you then ⟨;period⟩ ⟨#Klicken⟩ BYE

Example CDROM 3, g147a.trl

BAC	005 :	⟨;T⟩ja, sehr schön FEEDBACK_ACKNOWLEDGEMENT und dann vielleicht wieder⟨Z⟩ ein ⟨A⟩ bisschen späteres Abendtreffen, ⟨A⟩ übermorgen ⟨P⟩ ab achtzehn Uhr wieder? SUGGEST_SUPPORT_DATE würde es Ihnen das gut passen ⟨#Klicken⟩? REQUEST_COMMENT_DATE

- particles at the end of an utterance are segmented separately.

Example CDROM1, n001k.trl

BS	1020 :	⟨A⟩ ⟨#Klicken⟩ also am dritten Mai um viertel vor drei kommen Sie zu mir ins Büro. CONFIRM ⟨P⟩ ⟨#Schmatzen⟩ alles klar ⟨P⟩ ⟨⟨#Klicken⟩⟩ ⟨P⟩. FEEDBACK_ACKNOWLEDGEMENT

References

[1] Kerstin Fischer, Birte Schmitz. Pragmatisches Beschreibungsinventar für Diskurspartikeln und Routineformeln anhand des Demonstratorlexikons, Verbmobil-Memo, 1995. (in German)

[2] Susanne Jekat, Alexandra Klein, Elisabeth Maier, Ilona Maleck, Marion Mast and J. Joachim Quantz. Dialogue Acts in Verbmobil, Verbmobil-Report No. 65, April 1995.

[3] Anna-Brita Stenström, *An Introduction to Spoken Interaction*, Longman, London and New York, 1994

[4] W. Wahlster. Verbmobil — Translation of Face–To–Face Dialogs. In *Proc. European Conf. on Speech Communication and Technology*, volume "Opening and Plenary Sessions", pages 29–38, Berlin, September 1993.