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 Mast¹
Elisabeth Maier²
Birte Schmitz³

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

²DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken

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

Tel.: (030) 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.

<table>
<thead>
<tr>
<th>Example taken from Verbmobil Memo 24, Appendix I</th>
</tr>
</thead>
<tbody>
<tr>
<td>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</td>
</tr>
<tr>
<td>DAN 028 : Yeah. FEEDBACK_ACKNOWLEDGEMENT</td>
</tr>
<tr>
<td>SUS 029 : Yeah? So you should give me your address CLARIFY_STATE, then I (P) will fetch you at (P) seven (P) PM. CONFIRM</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Example CDROM 8, 127c.trl</th>
</tr>
</thead>
<tbody>
<tr>
<td>JBT 004 : (A) two (#Mikro) o’clock in the afternoon sounds fine (period) (;seos) ACCEPT_DATE where would you like to meet (;period) (#Klicken) (#Mikro) (#Mikro) (;seos) REQUEST_SUGGEST_LOCATION</td>
</tr>
</tbody>
</table>

---

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 (Mikro), clicking (Klicken); they also hint at segment ends (seos), at hesitations (ehm), at pauses (P) etc.
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 VERBMOBIIL 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).
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.
4. There are a number of exceptions from the previous rules: 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, r137.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 ⟨period⟩ ⟨comma⟩ if I am ⟨if⟩ 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 ⟨if⟩ you’re on time FEEDBACK RESERVATION ⟨period⟩ ⟨seos⟩ ⟨A⟩ I mean if you ⟨comma⟩ if you are ⟨if⟩ 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
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 seg-
  mented 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\(^3\).

\(^3\)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.
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⟩ ⟨;eos⟩ ⟨A⟩ ⟨ähm⟩ ⟨A⟩ ⟨Schmatzen⟩ ⟨A⟩ ⟨#Klopfen⟩ ACCEPT_DATE yeah ⟨;comma⟩ ⟨;eos⟩ ⟨A⟩ ⟨Schmatzen⟩ ⟨A⟩ ⟨#Klopfen⟩ ACCEPT_DATE I will ⟨!2 I’ll⟩ just write that down on my other calendar ⟨;period⟩ ⟨A⟩ ⟨Schmatzen⟩ ⟨A⟩ ⟨#Kloffen⟩ ⟨;eos⟩ ACCEPT_DATE

ISN 005: ⟨#Klicken⟩ ⟨Schmatzen⟩ ⟨A⟩ okay good ⟨;comma⟩ ⟨;eos⟩ FEEDBACK_ACKNOWLEDGEMENT then I will ⟨!2 I’ll⟩ ⟨;comma⟩ ⟨ähh⟩ 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

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


