

# Segmentation of Spoken Language for NLP

Jan W. Amtrup and Susanne J. Jekat

University of Hamburg, Computer Science Dept.  
email: {amtrup,jekat}@informatik.uni-hamburg.de

The segmentation of spoken language into adequate units for NLP is a prominent theoretical and empirical problem. The workshop aims at the presentation of different aspects of segmentation and the discussion of unsolved problems.

Segmentation of spoken texts refers to the separation of a text into a set of non-overlapping units that on the one hand reveals the inherent structure of the whole (macro-structure) and on the other hand consists of elements with a homogeneous internal structure (micro-structure). Segmentation below word boundaries does not concern us here; rather, we focus on problems of the relation between segmentation and informational structure and study phenomena above the word level<sup>1</sup>.

Within this task, spontaneous dialogues represent a central research field, because they contain most phenomena of spoken language, as for example hesitations, repairs, breaking offs and others. Up to now, many of these phenomena cannot be adequately processed by NLP systems, e.g. the complete utilization of prosodic cues. Thus, the starting point for research should be a limited domain where some but not all of the problems arise. In our work, we concentrate on cooperative negotiation dialogues (e.g. appointment schedulings). To a large extent those dialogues show a clear composition, regarding both the dialogue structure and the turn structure<sup>2</sup>. Below the level of dialogue modelling and above the level of word recognition, a special task of NLP is to recognize and/or summarize the function(s) of the turn in question. The specific purpose of a given NLP system, as well as its architecture and formalisms, further constrain the kind of analysis and the set of phenomena to be tackled.

Written language is often segmented by way of conventionalized methods, e.g. splitting into sentences, phrases or — based on structural analysis — constituents. But these methods only cover phenomena of wellformed, standard language. As to spoken language, there is a richer repertoire of possibilities to structure texts and their informational content. This repertoire among others includes

- prosody and intonation,
- lexical items used to organize a turn,
- performance phenomena (e.g. pauses, hesitations, repairs, reduplications, breaking offs),

---

<sup>1</sup> Which traditionally are called *supra-segmental* phenomena. . .

<sup>2</sup> We use the term *text* as a generic term for all linguistic manifestations (oral or written), whereas *turn* is understood as a speaker's complete contribution within a dialogue.

- structural constraints on the organization of informational content (e.g. topicalization) and the
- establishment of rhetorical relations on the basis of preceding dialogue elements.

The major part of this workshop is dedicated to the presentation and discussion of some of these aspects in combination with methods for their computational processing. The choice of problems to be discussed is guided by one special purpose of NLP, namely machine interpreting. We select this area of application as a central topic of discussion because it is to be expected that current work in the joint research project *Verbmobil*<sup>3</sup> will yield innovative approaches in the field. Provided that analysis is done for a whole turn of a negotiation dialogue to be interpreted, we consider the following questions to be of importance:

- 1 How to segment the turn into smaller parts that can be handled more efficiently (see example 1a)?
- 2 How to cope with contradictive information provided within the same turn (see example 2a)?
- 3 How to establish a hierarchy among the topics and/or speech acts introduced by the turn (see example 3a)?

Related to concrete examples<sup>4</sup>, the questions which have been listed above might result in the following problems for machine interpreting, which show some possible topics for discussion:

- 1a The turn contains important information and redundant performance phenomena, which have to be separated, because redundant phenomena need not to be translated: "But uh ehm what's see. The Monday or Tuesday and Wednesday on the seventh or eighth of June I have free."
- 2a The turn contains contradictory information, in this case acceptance of a proposed time but in the following a counterproposal, so information has to be weighted or modified: "June is okay but I would prefer May."
- 3a The turn contains two proposals that first are introduced as equivalent, but at the end of the turn one of the proposals is presented as the preferred one. For adequate translation the preferred proposal may have to be identified: "I could go any time between Wednesday and Friday eh so it could be either Wednesday and Thursday, that's May eighteenth and nineteenth, or eh Thursday and Friday, that's nineteenth and the twentieth. But I must say, I would prefer Wednesday and Thursday, because it's not directly before the holidays."

---

<sup>3</sup> *Verbmobil* aims at the construction of an interpreting assistant for cooperative negotiation dialogues and shall cover English, German and Japanese.

<sup>4</sup> Examples are taken from the *Verbmobil Tp13-Corpus*, containing monolingual and interpreted bilingual dialogues within the domains **appointment scheduling** and **travel planning**.