Processing dialogue-based data in the UIMA framework

Milan Gnjatović, Manuela Kunze, Dietmar Rösner
Otto-von-Guericke-Universität Magdeburg
Institut für Wissens- und Sprachverarbeitung
P.O.box 4120, 39016 Magdeburg
gnjatovic|makunze|roesner@iws.cs.uni-magdeburg.de

Abstract

Processing dialogue-based data requires handling of different forms of data, such as video, audio, text, etc. This paper presents our experiences in using the UIMA[1] framework to process and analyze dialogues from the NIMITEK[2] corpus.

1. Introduction

The aim of this paper is to provide an overview of benefits from and problems in using the UIMA framework to analyze and process dialogue-based data. For the purpose of our contribution, we focus on two principles that are widely accepted in investigation of language: (1) language is characterized by combined modes of meaning; and (2) discourse is hierarchically organized [2,3,4]. Together, these principles imply that different semantic, hierarchical organizations are to be combined in order to represent a given instance of a discourse segment (e.g. a clause, a dialogue turn, etc.). Extensible Markup Language (XML) appears to be an appropriate tool for hierarchical representation due to its underlying tree paradigm. However, boundaries between hierarchical units from different semantic organizations may overlap in such a way that no unit dominates other (cf. annotations 3 and 4 in Figure 1 below). Although it is possible to come up with a single, valid XML structure that will bring together different semantic organizations, it would significantly complicate the process of annotation. In contrast, separated annotations of these organizations in different XML structures complicate detection and analysis of their interdependencies. In order to overcome this discrepancy, we use the UIMA framework to map these separated annotations on to one another.

2. Processing Dialogues

Our research is primarily based on the NIMITEK[2] corpus of human-machine dialogues. This corpus was gathered in the framework of WOZ simulations and comprises recordings of affected speech and accompanying mimic gestures from interactions between subjects and a speech based system with a human operator playing the role of the system [1]. We employed the UIMA framework to analyze interdependencies between linguistic cues in commands produced by subjects and the focusing structure of recorded dialogues. Four different annotations, illustrated in Figure 1, were assigned to the transcription of the observed dialogue:

(1) annotation of semantic classes of utterances (commands, comments and questions);
(2) annotation of syntactic cues (anaphoric references and ellipsis-substitutions uttered by the subject);
(3) annotation of functional elements related to the focus of attention in the dialogue (e.g. entities and actions);
(4) annotation of prosodic cues (functional elements that carry tonic prominences, i.e. local maxima on the pitch contour).

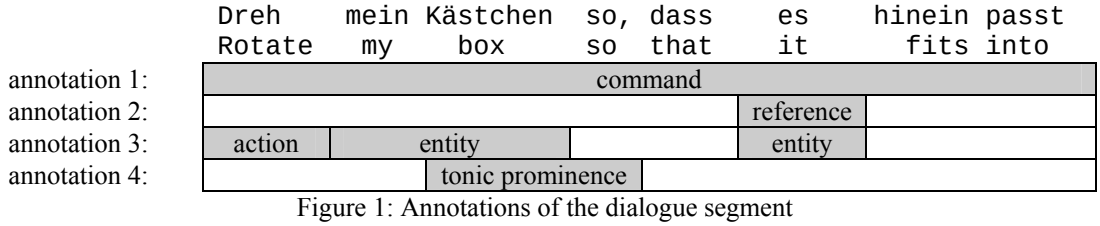

Figure 1: Annotations of the dialogue segment

[1] http://www.alphaworks.ibm.com/tech/uima
[2] http://wwwai.cs.uni-magdeburg.de/nimitek

3. The UIMA Application

Analyzing transcriptions by our UIMA application can be divided in two parts. The first part contains steps for merging several annotations and for importing annotations into UIMA. The second part is concentrated on the analysis of the imported data.

In the first part, we adapted the *FileCollectionReader* and the *Consumer* interfaces. An instance of the *FileCollectionReader* interface reads all annotations of the given transcription and generates UIMA annotations. Document type definitions (DTDs) assigned to each XML based annotations were used to generate UIMA based annotations and to import the transcribed data. In the example below, an excerpt from the transcription with annotations of semantic classes of utterances is given.

```
... <woz> <command>Einen Moment bitte.</command><comment>Bereit.</comment></woz>
    <sub> <command>Den zweitkleinsten Ring auf die Zwei.</command></sub>  ...
```

Furthermore, an instance of the *Consumer* summarizes annotations and generates an XML based representation of the observed transcription comprising all annotations. It facilitated the verification of annotations, i.e. the detection of irregularities, which was a necessary step due to problems that we encountered during the merging process:

- The nodes in XML files contain the original transcription and annotations as embedded nodes. To transform a given annotation into the internal form of UIMA annotations, the position of the annotation node in the document is calculated (document offset). Since the annotations of the transcription are created manually by different users, it is possible that they inadvertently add additional characters (e.g. space) to the original transcription. It changes the calculation of positions of annotation nodes and consequently disturbs the process of merging different annotation.
- In the scope of our research, there is also a need to annotate behavioral manifestations that do not necessarily contain speech: non-verbal sounds produced by the speakers (e.g. cough, laughter, etc), other non-articulated sounds (e.g. clicking, microphone related noise, etc.), subject's emotional expressions with respect to her mimic gesticulation, and non-verbal dialogue acts produced by the system (e.g. performing an action instructed by a subject, etc.). The problem with such annotations is related to the fact that they have the same starting and ending position in UIMA annotation. When using the Document Annotation Viewer, these kinds of annotations are not visible for a user. Instead of that, they should be annotated along the time scale.

In the second part, these XML files are imported and analyzed. The analyses are performed by different analysis engines, resulting in e.g. statistics with respect to occurrences of specific annotations, their interdependencies, etc. Other analysis engines are focused on the processing of natural language, e.g. the structure of noun phrases which are frequently used by the subjects.

4. Conclusion

The UIMA framework provides a set of interfaces for processing and managing a number of different annotations. We used this framework in different applications. In this paper, we described our ongoing work of analyzing dialogue-based annotations. The UIMA framework is used here to merge and to analyze different linguistic annotations of the observed transcription. The resulting statistics with respect to occurrences of specific annotations and their combinations is used to learn more about the role of linguistic and behavioral cues in human-machine interaction. Our future work will include the combination of our textual transcriptions with corresponding video and audio recordings. Therefore, it is necessary to extend annotations of these transcriptions with time related information.

References:

[1] Gnjatović, M., Rösner, D. 2006. Gathering Corpora of Affected Speech in Human-Machine Interaction: Refinement of the Wizard-of-Oz Technique. *Proceedings of the International Symposium on Linguistic Patterns in Spontaneous Speech (LPSS 2006)*. Academia Sinica, Taipei, Taiwan, p. 55-66.

[2] Grosz, B. J., Sidner, C. L. 1986. Attention, Intentions, and the Structure of Discourse, *Computational Linguistics*, 12(3), pp. 175-204.

[3] Halliday, M.A.K. 1994. *An Introduction to Functional Grammar, Second Edition*, London, Edward Arnold.

[4] Roulet, E. 1992. On the Structure of Conversation as Negotiation. *(On) Searle on Conversation*. John R. Searle, et al.; compiled and introduced by Herman Parret and Jef Verschueren, John Benjamins Publishing Company, p.91-99.