Combining Multiple Information Layers for the Automatic Generation of Indicative Meeting Abstracts

Thomas Kleinbauer and **Stephanie Becker** and **Tilman Becker**
German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany
`<firstname.lastname>@dfki.de`

Abstract

We describe a new application for NLG technology: the generation of indicative, abstractive summaries of multi-party meetings. Based on the freely available AMI corpus of 100 hours of recorded meetings, we are developing a summarizer that uses the rich annotations in the AMI corpus.

1 Introduction

The automatic summarization of documents has been a research topic for half a century now. Most prominently, the automatic creation of document *extracts* has been studied extensively. However, when applying such approaches to natural dialogs, such as meetings, the resulting texts may differ vastly from hand-written summaries: instead of concise and coherent prose, the expected output consists of a concatenation of speaker contributions taken from the original dialog. Yet, these utterances were made from each speaker's own perspective and thus are likely to contain first-person wording, inept for a comprehensible summary. Additionally, speech disfluencies or–in automated settings–speech recognition errors might further decrease the readability of the text. Finally, the extracted utterances are reproduced out of context which can be problematic in numerous ways, including the acceptance of the service by the meeting participants.

In this paper, we present our ongoing research on the generation of meeting abstracts that aims at overcoming the outlined shortcomings. So far, we have concentrated on *indicative* summaries that allow the reader to quickly assess whether the underlying meeting is relevant for her current information need.

2 Related Work

Extractive summarization of documents has been studied extensively over the last decades (s. Mani and Maybury (1999) for an overview), but faces additional challenges when applied to natural language dialogs. Unlike carefully authored articles, spontaneous utterances are often ungrammatical and contain speech disfluencies (Liu et al., 2006). Moreover, free discussions are naturally less well structured, e. g., when speakers switch topics or digress. For an automated system, additional difficulties arise from the limitations of current ASR systems, introducing recognition errors into all subsequent processing steps. Zechner (2001) and Murray et al. (2005) show ways to cope with such issues.

Generative approaches, on the other hand, are based on an internal representation of summary contents verbalized through NLG techniques (e. g. Kan et al. (2001)). Such approaches have been applied to natural discourse domains before, for instance, Alexandersson (2003) generates summaries of machine–translated phone conversations. However, we are not aware of any prior work attempting to generate full abstracts of multi-party interaction.

3 Annotated AMI Meetings

In the AMI project[1], circa 100 hours of meetings have been recorded, annotated and stored in a freely available multimodal corpus (McCowan et al., 2005). The meetings are semi-staged, in the sense that they are based on the predefined scenario of a virtual company in which a project team works on the task of designing a new innovative remote control. The roles of the four project team members are played by

[1] http://www.amiproject.org

subjects which act as a project manager, a user interface designer, a production designer, and a marketing expert. However, the discussions of the meeting participants are free and not prescripted. Meetings typically last about 30-40 minutes.

In addition to multiple video and audio streams, a number of annotations are included in the corpus, such as speech transcription, syntactic chunks, named entities, dialog acts, addressing, argumentative structure, hot spots, decision points and topics.

The goal of the AMI project is to develop automatic recognition systems for all of these annotation layers. In section 4 we show which layers are already used by our summarizer, see figure 5 for an example from the current system. However, all of these annotations are potentially useful, very rich resources for further extensions of our system.

3.1 Propositional Content

Additionally, we have annotated a small subset of the AMI corpus with categories from a domain ontology to represent the propositional content of speaker utterances. The AMIMATTER ontology that we created for this purpose models the remote control design scenario in a formal ontology based on Dolce-Lite-Plus (Masolo et al., 2003). Embedded in a comprehensive theory of representing situations and descriptions, it provides a taxonomy of relevant terms, ordered by an `IS-A` relation that expresses subsumption, or specialization. For instance, it contains information such as (remote_control IS-A technical_device) which expresses that the category remote_control is a sub-category of the category technical_device. Hence, a reasoner can infer that all remote controls (which technically would be considered *instances* of the category remote_control) are technical devices.

The AMIMATTER ontology covers over 20 different subdomains, with a total of 53,319 categories. 52,072 of those are extracted from WordNet (Fellbaum, 1998), the remaining 1,247 cover scenario-specific concepts and the Dolce-Lite-Plus upper model. Three subdomains—physical objects, meeting-related categories and project-related categories—were used to annotate the discourse transcription. The current system relies only on the annotation of relevant categories, ignoring relations within or beyond the dialog act segment boundaries[2]. Fig. 1 shows an example of such an annotation: three instances from the physical object subdomain were created (shown as boxes) and linked to the respective words in the source utterance above.

4 Summary Content Representation

We currently concentrate on three of the above annotation layers, topic labels, dialog acts and propositional content. For the pre-existing topic annotation, the recordings were split into larger segments and labeled with one of 24 topics matching typical activities in the remote control design scenario, e. g., "discussion" or "presentation of prototype(s)". These segments are used by our system as the basic structuring unit for the summaries. In most cases, the label can be used to verbalize the general subject of the topic segment, with the exception of the "other" label which is used for unknown topics.

In a similar practice, all participants' utterances in the manual transcript of the meeting discourse were segmented and labeled with dialog acts such as "inform", "suggest", etc. according to a scheme consisting of 15 distinguished dialog acts. However, our system currently discards the labels themselves, but uses the segments as a common unit for the propositional content annotation outlined in section 3.1. We perform a frequency analysis of all annotated ontology instances and select the three items that occur most often. We found this a useful heuristic, although it sometimes produces unexpected results (s. fig. 5: the term "beep" stems from an ontology category of the same name that was used to annotate a discussion about audio signals in the corpus).

5 Text Generation

The actual generation of the abstracts is done in a three-step pipeline:

1. Analysis of meeting annotation layers
2. Sentence planning
3. Surface realization

In the first step, information drawn from the annotation layers (s. section 3) is transformed

[2]More precisely, annotators were asked to identify those terms in a speaker utterance that belong to one of the three subdomains, identify the appropriate AMIMATTER category and create an instance of it, and connect the instance with the original word.

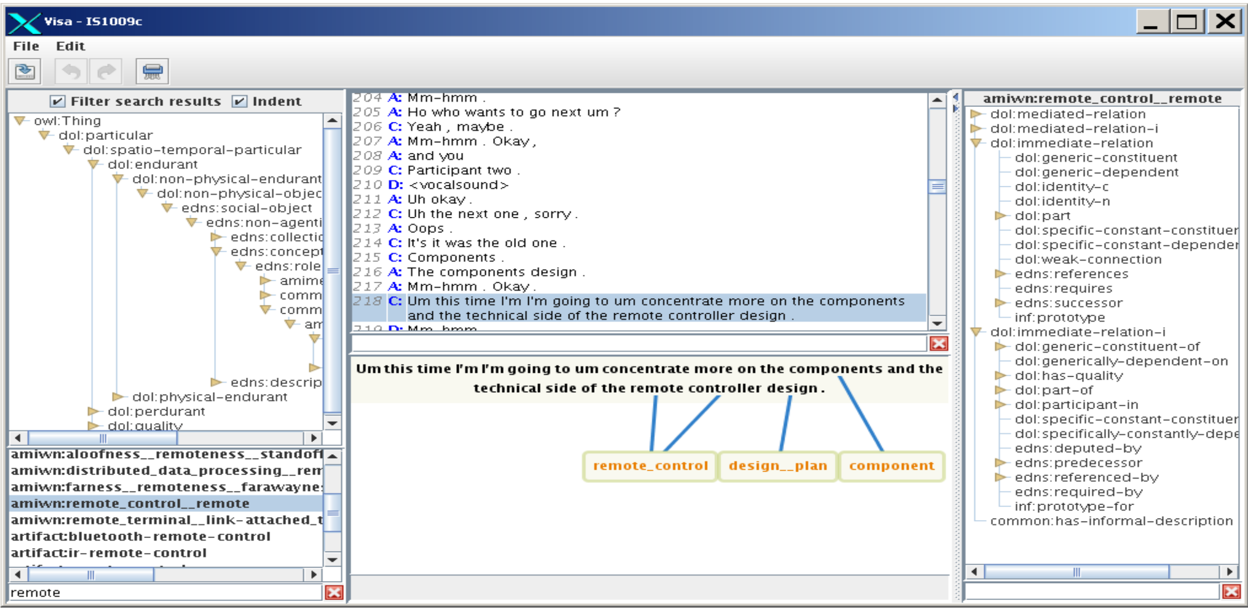

Figure 1: Example annotation of an utterance in meeting IS1009c in the AMI corpus. The outer sides display categories and relations of the AMIMATTER ontology in tree views, the center part contains the meeting transcript (top) and the annotation area (bottom).

into expressions in a propositional logic-like formalism (figure 2). These assertions are used

```
(topic "t0")
(about "t0" "opening")
(content "t0" "introduction")
(content "t0" "project manager")
(after "t0" "t1")  ...
```

Figure 2: The input for the sentence planner: topic t0 which is the opening of the meeting occurs before topic t1 and contains the content items “introduction” and “project manager”.

as a knowledge base by the sentence planner PREPLAN, a hierarchical, goal-driven planner (André, 1995). In addition to the assertions, PREPLAN is provided with a library of plan operators, each of which encodes strategies how to reach a given goal. Figure 3 shows an example of such an operator which describes how to reach the goal “ShowSummary” as the result of solving three subgoals, one of which is an iteration over all topics. Here, the “with”-condition is matched against the knowledge base that was generated before.

```
strategy: (ShowSummary)
subgoals: (WriteXMLHeader)
          (for-each ?t with (topic ?t)
             (ShowTopic ?t))
          (WriteXMLFooter)
```

Figure 3: A complex plan operator in PREPLAN

PREPLAN successively finds matching plan-operators until all goals and subgoals are resolved. The outcome of this process is an XML-encoded description of instructions in a logical form which is passed to the surface realizer, NIPSGEN (Engel, 2006), a template-based generator. NIPSGEN converts the semantic input into typed feature structures which are then transformed into a natural language utterance. A derivation tree for the XTAG-grammar (XTAG Research Group, 2001) is created using transformation rules which are applied to the input structure (see figure 4 for a sample rule). The actual syntax tree is constructed using the derivation tree. The generation of the correct morphological inflections is achieved by percolating the morphological features through the XTAG tree and looking up the correct inflections for all lexical leaves in the XTAG lexicon for English. Traversing the lexical leaves from left to right produces the natural language ut-

```
$VP=VP(o:Introduction(has-topic:$T,
           has-agent:$A), not(lex:))
-> $VP(lex:introduce, sub:NP(o:$A),
       obj:NP(o:$T))
```

Figure 4: A NipsGen rule: the semantic concept 'Introduction()' is lexicalized with the verb 'introduce'. The values of the features 'has-topic' and 'has-agent' are realized as NP's in object and subject position, respectively.

terance.

> *"The meeting was opened and the meeting group talked about the user interface, the remote control and the design. They debated the costs, the company and the project while discussing the project budget. The signal, the remote control and the beep were mentioned afterwards. They talked about meeting before closing the meeting."*

Figure 5: Example of a meeting summary.

6 Current and Future Work

We are currently developing the summarization system further by adding more annotation layers to the processing pipeline.

Work has also begun on the evaluation of meeting summaries. To this end, we will use a task based evaluation scheme where summaries are used by subjects to better understand previous meetings in order to join the team, replacing a previous member. The quality of summaries will degrade when we move from hand-annotated layers to automatically generated annotations. Extrinsic evaluations as described above will be a realistic measure for the level of degradation.

Given the richness of the data in the AMI corpus, we have also started work on multimedia summaries that will combine text with pictures from the video signals and links into the meetings[3]. We are experimenting with result-based summaries, presented in a newspaper style and timeline-based summaries, presented in a comic-strip style.

In general, summarization of multi-party meetings poses further challenges, like summaries from a personal perspective, and moving to related domains like instant messaging and IRC interactions.

[3]These links are timestamps that are used by a meeting player to show relevant segments.

References

Jan Alexandersson. 2003. *Hybrid Discourse Modelling and Summarization for a Speech-to-Speech Translation System.* Ph.D. thesis, Universtität des Saarlandes, Germany.

Elisabeth André. 1995. *Ein planbasierter Ansatz zur Generierung multimedialer Präsentationen.* Infix, St. Augustin.

Ralf Engel. 2006. SPIN: A semantic parser for spoken dialog systems. In *Proceedings of the Fifth Slovenian And First International Language Technology Conference (IS-LTC 2006).*

Christiane Fellbaum, editor. 1998. *WordNet–An Electronic Lexical Database.* MIT Press.

Min-Yen Kan, Kathleen R. McKeown, and Judith L. Klavans. 2001. Applying natural language generation to indicative summarization. In *Proceedings of 8th European Workshop on Natural Language Generation*, pages 92–100, Toulouse, France, July.

Y. Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper. 2006. Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Transactions on Audio, Speech and Language Processing*, 14(5):1526–1540.

Inderjeet Mani and Mark T. Maybury, editors. 1999. *Advances in automatic text summarization.* MIT Press.

C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. 2003. Wonderweb deliverable D18 ontology library (final), December.

I. McCowan, J. Carletta, W. Kraaij, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, W. Post, D. Reidsma, and P. Wellner. 2005. The Ami meeting corpus. In *Proceedings of the Measuring Behavior 2005 symposium on Annotating and Measuring Meeting Behavior*, Wageningen, NL, September.

G. Murray, S. Renals, and J. Carletta. 2005. Extractive summarization of meeting recordings. In *Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, September.

XTAG Research Group. 2001. A lexicalized tree adjoining grammar for english. Technical Report IRCS-01-03, IRCS, University of Pennsylvania.

Klaus Zechner. 2001. *Automatic Summarization of Spoken Dialogues in Unrestricted Domains.* Ph.D. thesis, Carnegie Mellon University.