A Human-Human Train Timetable Dialogue Corpus

Filip Jurčíček[1], *Jiří Zahradil*[2], *and Libor Jelínek*[2]

[1]Center of Applied Cybernetics
[2]Department of Cybernetics
University of West Bohemia in Pilsen
Pilsen, 306 14, Czech Republic
{filip,jzahrad,jelinekl}@kky.zcu.cz

Abstract

This paper describes progress in a development of the human-human dialogue corpus. The corpus contains transcribed user's phone calls to a train timetable information center. The phone calls consist of inquiries regarding their train traveler's plans. The corpus is based on dialogues's transcription of user's inquiries that were previously collected for a train timetable information center. We enriched this transcription by dialogue act tags. The dialogue act tags comprehend abstract semantic annotation. The corpus comprises a recorded speech of both operators and users, orthographic transcription, normalized transcription, normalized transcription with named entities, and dialogue act tags with abstract semantic annotation. A combination of a dialogue act tagset and a abstract semantic annotation is proposed. A technique of dialogue act tagging and abstract semantic annotation is described and used.

1. Introduction

Statistical approach to spoken language understanding (SLU) and dialogue management (DM) gains on popularity. Whereas traditional methods use hand-crafted rules, statistical techniques learn from data. As a result, we need a suitable spoken dialogue corpus to train a statistical model of a spoken dialogue system (SDS).

Although Stuttle in [1] argues that for successful incorporation of techniques in to SDS is needed a specially collected human-human (HH) corpus, we believe that regular HH corpus that is reasonably large can suppress all disadvantages mentioned in [1], e.g. almost errorless communication in terms of ASR. Therefore, we proceeded to extend our current dialogue corpus [2] in terms of coverage of sentences and modify semantic annotation to lower time needed to semantically annotate whole dialogue. Furthermore, we decided to add dialogue act tags to our corpus in order to study divers user's dialogue strategies.

We acquired the HH dialogue corpus because it does not suffer from fixed policy of a human-computer (HC) dialogue system. A dialogue policy is a mapping from dialogue states (situations) to system actions. From HC dialogue corpus with a fixed policy, the statistical methods are not able to learn a robust model. Consequently, we would not be able to successfully operate dialogues. [1]

In addition, general HH dialogues are structurally similar to mixed-initiative (user-driven) dialogues. Intuitively, in HH dialogues there is no restriction on participant roles, e.g. who is moving a conversation ahead at a given point or selecting a new topic for conversation. Likewise, in mixed-initiative dialogues users have weaker constraints about what can be said. For these resons, DM is much more complicated in mixed-initiative SDS than in system-initiative (system-driven) SDS. Our objective is to prepare a spoken dialogue corpus that can be used for developing more sophisticated dialogue management techniques.

To cover needs of both SLU and DM, we need to provide not only dialogue management related information (dialogue acts) but also need a structured meaning of these moves. More specifically, we need semantic information related to given application domain. We use an abstract semantic annotation according to [3]; this annotation is necessary for all utterances. Important question is how to obtain this annotation. The annotation has to be simple and unambiguous for sake of quality and consistency.

The rest of the paper is organized as follows. Section 2 describes a train timetable dialogue corpus. Section 3 details our dialogue act tagging scheme and the three dimension of our dialogue act tagset. Section 4 elaborates on the annotation process. Finally, section 5 concludes the paper.

2. Train timetable dialogue corpus

The corpus was collected in a train timetable information center. We had been recording the corpus since April, 2000 to September, 2000. We collected 6584 calls from which we transcribed 6353 calls (dialogues). Callers were mainly Czechs. A small portions of callers were from abroad (0.7%); their Czech language was heavily influenced by their native language. Therefore, we labeled this type of recordings as unsuitable for further processing.

The whole corpus contains 106 hours of speech. The signal is single channel, sampled at 8kHz with A-Law compression. The average length of a dialogue is 60 seconds. The audio files were divided into utterances. Effective user's and operator's speech (without noise and silence) is 32% and 39%. The corpus consists of 81543 turns. Each turn starts with a speaker change. The size of the vocabulary of the whole corpus is about 12k words, and there is almost 600k tokens in it. The operator's vocabulary (5839 words) is smaller than user's vocabulary (9485 words). While a dialogue has 6 user's turns on average, the first user's turn contains 35% of user's tokens in the dialogue on average.

Transcription was done using the Transcriber 1.4.1 speech editing tool (http://www.ldc.upenn.edu). In addition to orthographic transcription, the following non-speech sounds were marked: tongue click, lip smack, laughter, breath, background noise, silence, and unintelligible. In addition, we divided each utterance in segments that allow us to assign one dialogue act to

each segment. We chose orthographic transcription because it is more suitable for transcription of Czech spontaneous speech [4]. Spontaneous Czech contains words and usages not found either in standard written or in formal spoken Czech.

In [2] is described previous work on the presented dialogue corpus. Named entities were labeled. Normalized transcription was generated automatically with the aid of a dictionary containing only formal Czech words. Only context independent normalization was performed. Collecting of semantic annotation for the first user's inquiry was started.

3. A dialogue act tagging scheme

Defining a new tagset is a challenging task that is not easy and not always desirable. Therefore, we wanted to reuse an existing tagset that would fulfill following requirements:

1. Robustness: reliability of human annotation of typical data (high interannotator agreement, at least potentially).
2. Evaluation: relation to one or more existing systems so that we can compare already early obtained experimental results with new results.
3. Reusability: avaiable mapping of existing tagsets to the selected tagset at least partially so that useful insights are preserved.
4. Separation: separation of dialogue control related information from task related information.
5. Universality: general and expandable semantic representation maintaining easy annotation.

First of all, we were interested in DAMSL (Dialogue Act Markup in Several Layer) [5]; however, it was too sophisticated for our purposes, covering many aspects of dialogue structure that were not necessarily relevant for our task. Clark et. all [6] states that dialogue act tagging on Switchboard data that initially started with full DAMSL scheme (about 4 million possible combination of DAMSL tags) resulted in a clustered 42 mutually exclusive dialogue act tags. The given reason was that only minimal portion of all possible DAMSL tag's combinations had been seen. In addition, the interannotator agreement levels reported for this scheme were quite low. As a result, we rejected the DAMSL. Nonetheless, we decided to reuse some ideas.

Secondly, we investigated DATE (Dialogue Act Tagging for Evaluation) scheme [7]. Each DATE's dialogue act consists of three dimensions: (1) DOMAIN, (2) SPEECH-ACT, (3) TASK-SUBTASK (sometimes referred as a FRAME domain). We found its limitations. It was originally designed for evaluation and comparison of spoken dialogue systems. It mainly focuses on HC dialogs; in [7] only operator (system) part of COMMUNICATOR data [8] was annotated. The TASK-SUBTASK dimension do not contain enough information about a particular utterance for successful control of dialogue.

However neither of both tagsets was suitable on its own, we adopted a combination of these two tagsets that suppresses their disadvantages and boosts their advantages. We started to build on a DATE for its simplicity and because it had been designed for task-oriented dialogues. We removed the TASK-SUBTASK dimension and replaced it with a new dimension SEMANTICS. The new dimension covers full semantic annotation. Additionally, we found necessity of new tags for annotating user's utterances and complicated (human) operator's answers. Therefore, we extended tagset by a few elements borrowed from DAMSL tagset (agreement, politeness - thanking, speech-repair); it was necessary to fully cover HH interaction. The selected tags were sufficiently clear and simple that we believed we would be able to tag the data reliably. Because DAMSL is differently structured, some of these tags were incorporated into the SEMANTIC dimension (e.g. concepts ACCEPT, REJECT) and rest to the SPEECH-ACT dimension.

We describe the DOMAIN, the SPEECH-ACT, and the SEMANTIC dimensions in three following sections.

3.1. The DOMAIN dimension

The DOMAIN dimension assigns every utterance to three areas of conversational action. We adopted this approach from DATE. The first area of DOMAIN is the task domain, which is train timetable inquiry answering. The second area is managing communication channel. Finally, the third area is a situation frame, which refers to an apology or an instruction contained in a sentence.

Task - the act for an utterance that deals with a task completion. For example, a task utterance asks about departure time or presents departure time of a questioned train. See Table 1.

Communication - the act for an utterance that manages the verbal channel and provides evidence what has been understood. For example, in the early annotation stage we have found implicit confirmation very common in the corpus. A typical example is repetition of departure time.

Frame - the act for an utterance that describes a state of a dialogue e.g. instruction, apology. As is stated in [7], these parts of dialogue are very common in a HC dialogue (instructions), but rare in a HH dialogue. Nonetheless, we kept a frame value in our tagging scheme mainly to maintain compatibility with DATE. Because FRAME acts are related to task-oriented dialogue flow, the SEMANTICS dimension (task-related information) is empty for them.

3.2. The SPEECH-ACT dimension

The SPEECH-ACT dimension refers to an utterance's communicative goal, independently on an utterance form. This dimension differentiates utterances that have the same value of the SEMANTICS dimension. For instance, the SPEECH-ACT dimension values REQUEST-INFO and PRESENT-INFO can refer to the same value in the SEMANTICS dimension, e.g. DEPARTURE(TIME, FROM(STATION)). See details in 3.3

In our annotation scheme, we use all speech acts from DATE. They are namely: request-info, present-info, offer, acknowledgment, status-report, explicit-confirmation, implicit-confirmation, instruction, apology, opening, and closing.

In addition, we have used a speech act for thanking (and a reply for thanking) and speech act speech-repair because they occur very often in HH dialogues. [6]

The instruction speech act is related to the frame act of the DOMAIN dimension and mainly to HC dialogues. Beacause users were familiar with calling to the train timetable information center, we did not find any instance of instructions in the train timetable dialogue corpus. However, we did not exclude the instruction speech act from our annotation scheme. We wanted to maintain compatibility with DATE tagged corpuses for evaluation purposes.

For acknowledgment speech act, we further extended its meaning maintaining DATE tagset. We use this tag for agreement as is described in DAMSL. The second annotation layer (accept, reject, maybe) is contained in the SEMANTIC dimension, see Table 1.

3.3. The SEMANTIC dimension

The SEMANTIC dimension captures task relevant information from each utterance. Our domain task is train timetable inquiry answering; therefore, the goal of communication is to determine information needed to answer an inquiry, e.g. a departure train station or time of desired departure. In the sentence "Is there any train to Pilsen at eight am", the semantics is REQUEST=departure, TO=Pilsen, and TIME=eight am.

Semantic annotation should be simple and easy to obtain. He [9] described a semantic annotation that preserved the hierarchical structure of an utterance, but it still prevailed simplicity. For instance, the semantic annotation of the previous sentence would be DEPARTURE(TO(STATION), TIME). Moreover, the same semantic annotation could belong to the utterance "does any train go to Prague around four pm" This generalization is very precious. The He's annotation is moderate to acquire. We do not need fully annotated treebank data. Dialogue transcribers have to define the semantics that represents each training utterance, but they need not provide a word/concept parse tree. Because abstract semantic annotation is fairly simple, transcribers do not have to have a prior linguist knowledge. Sentence modality do not influence abstract semantic annotation. For example, we do not distinguish between a question or an answer in abstract semantic annotation; thus, we have to assign either request-info or present-info into the SPEECH-ACT dimension.

We defined four main semantic concepts. DEPARTURE is a concept for an utterance that represents question about departure of a particular train (answer is usually exact time when the train leaves a particular train station). DEPARTURE_CONF is a concept for an utterance that contain confirmation requests about departure of a particular train. Similarly, we defined concepts for an arrival (ARRIVAL, ARRIVAL_CONF).

Each of previous semantic concepts is allowed to have following non-terminal leaves (concepts): FROM, TO, THROUGH, IN_DIRECTION, TRAIN_TYPE, TIME. All previous concepts except TRAIN_TYPE and TIME have to be linked with the concept STATION. You can find variations of these concepts in [9].

In addition, we defined four supporting semantic concepts. VERIFY is a concept for an utterance questioning a special property of a train connection, e.g. train type. The concepts ACCEPT, REJECT, and MAYBE represent possible responses to the concept VERIFY. All preceding concepts can be linked with concepts STATION, TRAIN_TYPE, and TIME.

In Table 1 are abstract semantic annotations for a sample dialogue. The word level transcripts (English utterances) are literal translations of original Czech utterances.

The SEMANTIC dimension of our tagging scheme provides training data for a semantic decoder. The goal of the semantic decoder is to label both user's and operator's utterances. From labeled utterances, a simple algorithm should easily extract attribute-value pairs that we are interested in. Therefore, an output of semantic decoder could be following DEPARTURE(does any train go TO(to STATION(Prague)), TIME(around four pm)). The resulting parse tree is shown on Figure 1. The task-related details, e.g. a destination station, time, and train type are the most important information. He [9] designed and tested Hidden Vector State (HVS) parser that was possible to train with abstract semantic annotation.

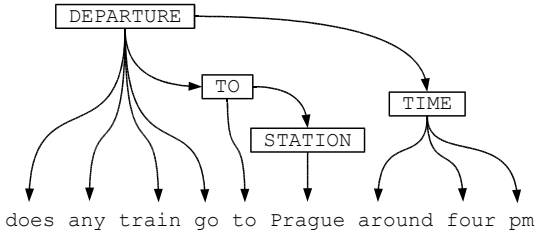

Figure 1: *An example of a semantic parse tree.*

4. Annotation process

To speed-up annotation process, we developed an annotation tool that simplifies whole process. We had a few requirements: configurable task-related tagset, annotation robustness, validation of every possible input, and functions for dialogue viewing. We required multi-platform design and easy maintenance.

We have been concerned about achieving high annotation robustness. Therefore, we implemented an overview mode in our annotation tool in which an annotator can view just dialogue acts without utterances. Standalone dialogue acts, without the utterance, should make sense of story, and they have to be enough to give to the annotator an overview of a dialogue. In the overview mode, we do not show even name entities values because misunderstanding to a dialogue from dialogue act tags is a signal to revise, probably wrong, dialogue act annotation.

Annotation process proceeds in following way. If an utterance contains more than one dialogue act, an annotator chops it into fragments which corespond to one dialogue act each. All utterance fragments are annotated by dialogue acts. The SEMANTIC dimension of each dialogue act is validated by a software validator. The validator can discover typos and mistakenly structured semantics. After annotating whole dialogue, the annotator overviews a dialogue. In the overview mode, the annotator see a simple list of succeeding dialogue acts. The annotator can verify dialogue semantics without distraction from utterances' transcripts.

5. Conclusions

This paper has presented the combination of DATE, DAMSL, and abstract semantic annotation designed for the area of spoken dialogue systems. We have found that the human-computer DATE can be extended so that it covers HH dialogues. Although the DATE scheme was designed primarily for English, we did not encounter any fundamental difficulties with using extended DATE for Czech Train Timetable Dialogue Corpus. Moreover, there was no need to structurally modify DATE except substitution of the TASK-SUBTASK domain by the SEMANTIC domain. Our motivation was to include more task specific information into dialogue act. We have been mainly interested in extraction concept/word pairs of TIME, TRAIN_TYPE and STATION semantic concepts. The extended DATE tends to be reasonably simple to annotate, and it keeps all necessary details for further development of spoken dialogue systems in train timetable inquiry answering.

The benefit of our work is in combination of several established techniques. We can profit from previous work, and in the same time we can compare our results. In the near future, we want to evaluate an interannotator agreement and compare it with dialogue tagging schemes based on DATE or DAMSL. We

Table 1: *A sample of dialogue act tagging including abstract semantic annotation.*

Speaker	Conversational domain	Speech act	Semantics	Literal English translation
operator	communication	opening	NIL	the information please
user	communcation	opening	NIL	hello
	task	request-info	DEPARTURE(TIME, TRAIN_TYPE, TO(STATION))	I have a question how can I go today by regional train to <*station*> staryho plzence </*station*>
operator	frame	status-report	NIL	well we do not have many connections here
	task	present-info	TIME, TIME	now one goes at eight sixteen if you catch it after that only at eleven ten
user	communication	implicit-confirm	TIME	at eleven ten
	task	acknowledgment	ACCEPT(TIME, FROM(STATION))	it is not so bad at eleven ten from <*station*> hlavniho </*station*> yeah
	task	request-info	VERIFY(TRAIN_TYPE)	is it possible to take a stroller with me
operator	task	acknowledgment	ACCEPT	of course madam be sure that you can
user	task	request-info	VERIFY(TRAIN_TYPE)	so there are the new cars
operator	frame	apology	NIL	wait what do you mean with the new cars
user	task	request-info	VERIFY(TRAIN_TYPE)	do I have to take it as baggage
	task	request-info	VERIFY(TRAIN_TYPE)	or can I take it for mothers with children
operator	task	acknowledgment	ACCEPT(TRAIN_TYPE)	for mothers with children
user	communication	closing	NIL	yeah well OK

have annotated 84 dialogues by our dialogue act tagging scheme since January, 2005.

This train timetable dialogue corpus will be a basis for further development of a mixed-initiative spoken dialogue system. We will use this corpus for training a semantic decoder based on HVS parser. Furthermore, we will build a dialogue act decoder using dynamic Bayes networks. Our plan is to compose both models into a finite state transducer. We aim to use standart techniques and tools such as the Graphical Models Toolkit (GMTK) [10] and AT&T FSM Library™[11]. Because our tagging scheme originate in DATE, we expect to build an automatic dialogue evaluator based on the PARADISE framework [12] for our intended spoken dialogue system.

6. Acknowledgments

This work was supported by the Ministry of Education of the Czech Republic under project 1M6840770004 and MSM235200004.

7. References

[1] Stuttle, M., Williams, J., Young, S., "A framework for dialogue data collection with a simulated ASR channel", INTERSPEECH-2004, ICSLP, 241-244, 2004.

[2] Jelínek, L. and Šmídl, L., "Spontaneous Speech Understanding in Train Timetable Inquiry Processing Based on N-gram Language Models and Finite State Transducers", ORLANDO, FL, USA, 2004.

[3] Young, S., "Talking to Machines (Statistically Speaking)", Proceedings of the International Conference on Spoken Language Processing, Denver, USA, 2002.

[4] Psutka, J., Ircing P., Hajic, J., Radova, V., Psutka, J.V., Byrne, W., and Gustman, S., "Issues in annotation of the Czech spontaneous speech corpus in the MALACH project", Proceedings of the International Conference on Language Resources and Evaluation, LREC, 2004.

[5] Allen, J. and Core, M., "DAMSL: Dialog Act Markup in Several Layer", http://www.cs.rochester.edu/research /cisd/resources/damsl, 1997.

[6] Clark, A. and Popescu-Belis, A., "Multi-level Dialogue Act Tags", Proceedings of 5th SIGdial Workshop on Discourse and Dialogue, Boston MA, USA, 2004.

[7] Walker, M., and Passonneau, R., "DATE: A Dialogue Act Tagging Scheme for Evaluation of Spoken Dialogue Systems", IEEE Trans. Speech and Audio Proc., 7(6):697–708, 1999.

[8] Walker, M., Aberdeen, J., Boland, J., Bratt, E., Garofolo, J. , Hirschman, L., Le, A., Lee, S., Narayanan, S., Papineni, K., Pellom, B., Polifroni, J., Potamianos, A., Prabhu, P., Rudnicky, A., Sanders, G., Seneff, S., Stallard, D., and Whittaker, S., "DARPA Communicator Dialog Travel Planning Systems: The June 2000 Data Collection", EUROSPEECH: European Conference on Speech Processing , 2001.

[9] He, Y. and Young, S, "Semantic Processing using the Hidden Vector State Model", Computer Speech and Language, Computer Speech and Language, 2003.

[10] Zweig, G., Bilmes, J., Richardson, T., Filali, K., Livescu, K., Xu, P., Jackson, K., Brandman, Y., Sandness, E., Holtz, E., Torre, J., and Byrne, B., "Structurally Discriminative Graphical Models For Automatic Speech Recognition", The 2001 Johns Hopkins Summer Workshop, Baltimore, MD, USA, 2001.

[11] Mohri, M., Pereira, F., and Riley, M, "Weighted Finite-State Transducers in Speech Recognition", Computer Speech and Language, 16(1):69-88, 2002.

[12] Walker, M., Litman, D., Kamm, C., and Abella, A., "PARADISE: A Framework for Evaluating Spoken Dialogue Agents", Proceedings of the 35th Annual Meeting of the Association of Computational Linguistics, ACL 97, 1997.