Coupled pragmatic and semantic automata in spoken dialogue management

Jolanta Bachan

Adam Mickiewicz University in Poznań
ul. Wieniawskiego 1, 61-712 Poznań, Poland
jolabachan@gmail.com
http://www.bachan.speechlabs.pl

Abstract. Dialogue managers are often based explicitly on finite state automata, but the present approach couples this type of dialogue manager with a semantic model (a city map) whose traversal is also formalised with a finite state automaton. The two automata are coupled in a scenario-specific fashion within an emergency rescue dialogue between an accident observer and an ambulance station, i.e. a stress scenario which is essentially different from traditional information negotiation scenarios.
The purpose of this use of coupled automata is to develop a prototype dialogue system for investigating semantic alignment and non-alignment in a dialogue. The research on alignment of interlocutors is to improve human-computer communication in a Polish adaptive dialogue system, focusing on the stress scenario. The investigation was performed on two dialogue corpora and resulted in creating a working text-in-speech-out (TISO) dialogue system based on the two linked finite-state automata, evaluated with about 130 human users.

Keywords. Dialogue systems, finite-state automata, dialogue modeling, dialogue corpus analysis

1 Aims and theoretical background

The main goal of the present investigation is to define a strategy for providing explicit models of alignment and accommodation in human-computer and human-human communication. Alignment means the adaptation of users to each other in speech style, vocabulary, pronunciation, gestures and body movements (e.g. [10], [13]). Acceptable human-computer interaction is the subject of much research, but the literature on these topics does not consider alignment of synthetic speech with a human interlocutor. The present research focuses specifically on stressed (not necessarily emotional) speech in crisis situations. The models should account for speech style alignment in these situations. Speech style is a well-studied parameter, in contrast to emotion.

The research was carried in two steps: first, a preliminary study was performed on a small subset of dialogue corpus and tools for dialogue processing were created

along with first dialogue automata, second, a dialogue corpus was recorded in an emergency scenario which analysis was the ground for creating a prototype dialogue system. The goal is not to provide a product or a comprehensive dialogue system but to give a proof-of-concept implementation in a new, previously unexplored semantic alignment domain in crisis scenarios and demonstrate the methodology with a Text-In-Speech-Out (TISO) approach.

2 First steps in realistic automata creation: preliminary study

2.1 Selected material for analysis

For examining the details of alignment and dialogue act theory, a corpus linguistic study on a small sample performed. For the study, the map-task dialogues from the PoInt corpus [12] were used. Two dialogues (18min of speech) were annotated on the dialogue act level using the selected dialogue act categories from Bunt DIT++ categories [7]. In this pilot study the following were looked at:

1. dialogue annotation at the dialogue act level
2. annotation of turns – semantic dialogue flow
3. finite state automata of dialogue act sequences
4. most frequent dialogue acts sequences

2.2 Dialogue act annotation

For the preliminary analysis, two dialogues (18 min) between two females and a male and a female were annotated on the dialogue act level. The dialogue act categories for the annotations were selected from Bunt's main categories of the Dynamic Interpretation Theory, DIT [6].

More than one dialogue act category was assigned to a speaking turn, because one utterance can have more than one communicative function. In principle, multiple categories require feature-based finite state treatment [5], but the combinations were treated as atomic symbols. Abbreviations of 12 dialogue act functions chosen from Bunt's categories and used in the annotation of the selected dialogues are allo: allo-feedback, auto: auto-feedback, cnt: contact management, dir: directives, infpr: information providing, infsk: information seeking, open: open meeting, own: own communication control, partner: partner communication management, social: social obligations management, time: time management, turn: turn management.

2.3 Processing of annotations for dialogue analysis

To analyse the dialogue for Finite State Automata (FSA) creation the information on the dialogue act annotation and utterance transcription tiers from one map-task dialogue recording was extracted and analysed. The material was prepared using Linux scripts and was automatically divided into 49 *parts* and the beginnings and ends of those parts were used to determine the initial and terminal states of the automata.

Those divisions into *parts* indicate the first silence after the last dialogue act in a sequence overlapping with the other speaker's speech, which means that the start/end of the *part* may occur within speaker's turn if the speaker made a pause within his turn. Figure 1 illustrates the division of the dialogue into *parts* starting from the beginning of the dialogue till the 51 second.

2.4 Time structure of the dialogue

A computational analysis of the annotation files shows that the time relations between the utterances are very complex. The relations are not purely sequence relations but involve overlaps in time.

In Figure 1 the temporal division of the dialogue into parts is shown. At the first silence after the last dialogue act in a sequence overlapped with the other speaker's speech, the bars represent dialogue act intervals (chunks of speech), the indices indicate indexed borders and the dash indicates the end of a turn (silence). The diagram is based on the first 51 seconds of the dialogue.

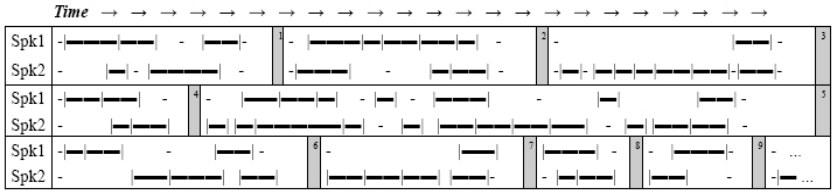


Figure 1: Temporal sequences and overlaps in a dialogue

Ideally the two speakers would be modelled separately, and the parallel automata combined [11] in order to model the temporal relations. However, a simplified approach was adopted. The output dialogue act sequences underwent the following processing for the purpose of automata generalisation:

1. alphabetical sorting of dialogue act sequences,
2. reduction of multi-layered labels of dialogue acts to one-layered labels; only the first dialogue act was preserved, e.g. infpr_dir → infpr,
3. deletion of repetition of the same dialogue act,
4. re-sorting of dialogue act sequences.

The reduction process was carried out, because the fact that to the same utterance more than one communicative function could be assigned made the longer sequences unique and limited the generalisation.

2.5 Loop-free automata creation

The annotation on the dialogue act tier was used to create manually a collection of loop-free automata which modelled each sequence of the dialogue acts for each of the

speakers. To create the loop-free automata a matrix of dialogue acts flow for Speaker 1 and Speaker 2 with time relations was analysed separately for each speaker. Each dialogue act sequence served to define the initial node, the terminal node and the transitions between the nodes for each of the loop-free automata. Such automata were evaluated for correctness using an NDFST interpreter [9] (described below) .

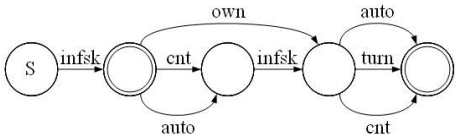


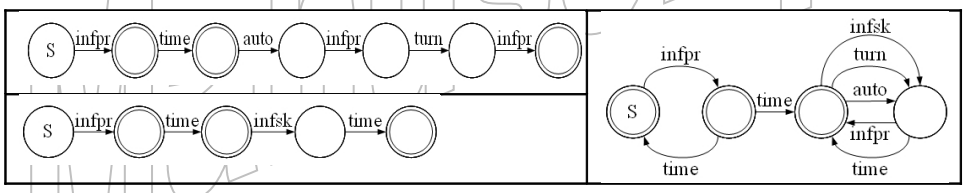
Figure 2: Loop-free automaton for Speaker 1

2.6 Generalisations over non-finite regular languages

Generalisations over non-finite regular languages can be expressed with FSAs with loops, and visualised with directed cyclic graphs. Altogether 22 automata with loops were created, 7 for speaker 1 and 15 for speaker 2.

Analysis of the prefixes and analysis of the loop-free automata allowed the creation of a whole set of automata with loops which model non-finite regular languages. Examples of loop-free automata and their counterparts with loops for speaker 2 are shown in Table 1.

Table 1: Loop-free automata (left) and its counterpart with loops for Speaker 2 (right)



2.7 Coupled turn automata

Coupled turn automata are based on real dialogue act sequences for both speakers. The turn automata were made by combining dialogue act automata for speaker 1 (spk1) and speaker 2 (spk2). Each speaker has his/her own automaton, and they are coupled in/by negotiations. An example of the coupled turn automaton is presented in Figure 3. S is the initial node. The node more to the left shows which speaker starts the sequence. The dotted arrows show the transition of turns between the speakers.

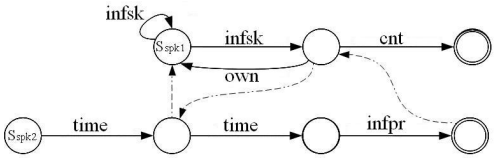


Figure 3: Coupled turn automaton. Spk 2 starts the sequence, Spk 1 follows.

2.8 Evaluation of dialogue act automata

The automata were evaluated for *coherence* (whether the automata are syntactically correct and actually work when operational), *completeness* (whether the automata describe all the phenomena they are intended to describe, not necessarily only restricted to a particular corpus, but including generalisations, and possibly also judged by native speaker intuitions), *soundness* (whether the automata describe *only* the phenomena they are intended to describe not necessarily restricted to a particular corpus, but including generalisations, and possibly also judged by native speaker intuitions), *consistency* (whether the modelling is done in the same way for similar observations of utterances).

The Nondeterministic Finite State Transducer (NDFST) online tool [9] was used in the present study to evaluate the dialogue acts automata. This NDFST tool allows to evaluate the correctness of the FST transitions and it generates output specified by those transitions. The NDFST takes the information about the initial states and terminal states as input, together with the transition quadruples:

```
<currentstate, inputsymbol, outputsymbol, nextstate>
```

All the automata underwent testing and were positively evaluated in the NDFST.

3 Corpus linguistic study

Two types of dialogue were recorded under laboratory conditions: a map task dialogue and a picture description dialogue, so-called diapix task [4] in stressed and neutral scenarios. For creating a dialogue model for the prototype-dialogue system, only the map task dialogues in stress conditions were analysed. The map of the task is presented in Figure 4 (left) – the black circles and street names were not marked. 24 subjects were recorded in a public setting [3] in emergency scenario. Additionally, 3 pairs of close friends were recorded as a control group. For the project, 15 males and 15 females were chosen and recorded in pairs: male – male, male – female, female – female. The corpus contains 4h 12min of recordings, out of which 38min 45sec are map-task emergency dialogues. See [1] for more details.

4 Finite-State Transducer model of the map

The emergency map can be represented as a finite state transducer (FST) where each junction corresponds to the transition node (Figure 4, right). Not all the streets are open and some junctions cannot be reached by dialogue system. There is a traffic jam on the way or roadworks, and at one place the street has been blocked because of a school race. Such blockages are not taken into account when designing the FST. In the FST, q0 is the start node and q13 is the end node. Latin letters are used for transition labels.

5 Dialogue system implementation

The prototype dialogue system is based on two FSAs: Figure 4 (right), for map traversal, and an additional FSA for the dialogue manager (Figure 5, left). The instruction to the human caller is to direct an ambulance from the hospital to the person with a heart attack along the streets. The human user inputs 'chat' text into the system in writing. The dialogue system communicates with the caller via audio output producing synthetic speech. The caller also has a street map on a computer screen. Either formal or informal speech style is selected by an experimenter for the dialogue. The dialogue manager schema is presented in Figure 5 (right).

The dialogue system is a TISO configuration [2]. Written input from the user is entered on the command line and the system produces synthetic speech output via loudspeakers.

5.1 Evaluation

The dialogue system was evaluated using EAGLES standards [8]. After successful diagnostic evaluation, the dialogue system faced functional testing with the human users. 52 people took part in the evaluation. In the evaluation, mainly young students took part around between 19 and 23 years old, with some people in their late 20's and one 52-year-old man. The dialogue model, the scenario and the success of communication were tested by means of user judgments.

After each dialogue the test participant was asked to assess different domains of the system on a 5-point rating scale (1 – lowest, 5 – highest). The four system components were evaluated: speech style selection module, speech synthesis, dialogue manager and system design. For simplification, the test participants were asked to evaluate seven categories: friendliness, speech quality, speech intelligibility, dialogue, dialogue naturalness, system attractiveness and ease of usage.

Last but not least, the system was tested in field conditions at the Researchers' Night 2011 in Poznań, Poland. Around 80 people took part in the evaluation and they were informed that the time of their dialogue was being measured.

5.2 Results

In the test 14 females and 12 males took part to evaluate each of the two scenarios: formal and informal. Altogether 52 people took part in the evaluation. The duration time of all the dialogues in formal and informal scenarios lasted about 75min 34sec and 76min 48sec respectively. The number of inputs inserted during one dialogue is almost the same and equals 20.54 inputs for the formal and 20.26 inputs for informal scenarios. All subjects accomplished the communication with the computer successfully, meaning that semantic alignment was successful despite obstacles. Misalignments happened, but the dialogue system effectively recovered from misalignments.

The results of the judgement testing of the system were good: 4.11 for the formal dialogue scenario and 4.30 for the informal scenario, where 5 was the highest grade.

When it comes to the field testing at the Researchers' Night, all the people carried out the task successfully. The age of the youngest child was 5 years old. The best time was 55s, the longest time was 4min 30sec.

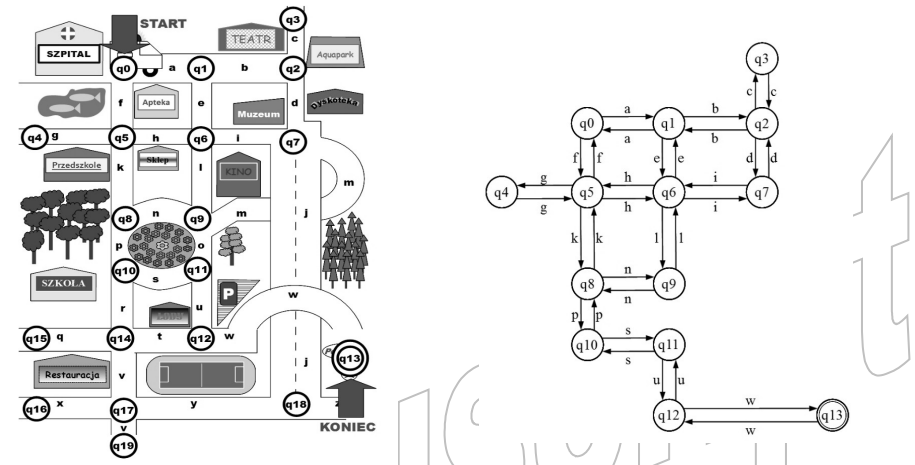


Figure 4: Emergency map with junctions marked (left); Finite State Automaton representing the reachable junctions (right)

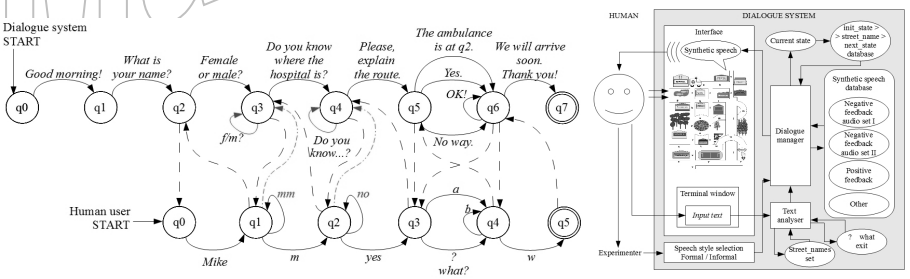


Figure 5: Dialogue automaton with speciment utterance labels (left), dialogue manager schema (right)

6 Conclusions

Dialogue modelling strategies were discussed, and semantic alignment was modelled: Analysis of corpora of dialogue recordings in a map task emergency scenario was used to create a finite-state automaton modelling a dialogue in a prototype dialogue system. Finally, a prototype dialogue system was developed and evaluated with human users. The prototype dialogue system combined text input with speech output and its core was based on two linked finite state automata: one for the dialogue man-

ager and one for map traversal. The laboratory setting of the evaluation task demonstrated alignment of the semantic representation of the map, as all the human users finished the task successfully. Additionally, the alignment of the dialogue system was based on speech style selections, formal and informal, not emotions. The methodological goal of modelling semantic alignment in dialogue was attained, using a new methodology with coupled automata, with a proof-of-concept TISO implementation using a Polish speech synthesiser voice developed for the purpose.

Acknowledgments. *This work was partly funded by the research supervisor project grant No. N N104 119838. The author is currently supported by grant "Collecting and processing of the verbal information in military systems for crime and terrorism prevention and control." (OR 00017012).*

References

1. Bachan, J. (forthcoming). Developing and evaluating an emergency scenario dialogue corpus. In: *Proceedings of LREC 2012.* 23-24-25 May 2012, Istanbul, Turkey
2. Bachan, J. (2011). Modelling semantic alignment in emergency dialogue. In: Proceedings of 5th LTC 2011. 25-27 November 2011, Poznań, Poland, pp. 324-328
3. Batliner, A., Steidl, S., Hacker, Ch., Nöth, E. (2008). Private emotions versus social interaction: a data-driven approach towards analysing emotion in speech. In: User Modelling and User-Adapted Interaction - The Journal of Personalization Research 18, pp. 175-206
4. Bradlow, A. R., Baker, R. E., Choi, A., Kim, M. and van Engen, K. J. (2007). The Wildcat Corpus of Native- and Foreign-Accented English. In: Journal of the Acoustical Society of America, 121(5), Pt.2, p. 3072
5. Berndsen, J. (1998). Time Map Phonology: Finite State Models and Event Logics in Speech Recognition. Dordrecht Kluwer Academic Publishers
6. Bunt, H. (2000). Dialogue pragmatics and context specification. In: H. Bunt & W. Black, (Eds.) Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics. John Benjamins, Amsterdam, pp. 81–150
7. Bunt, H. DIT++ Taxonomy of Dialogue Acts. (2008). (Release 3, version 2, February 8, 2008) <https://let.uvt.nl/general/people/bunt/docs/dit-schema3-2.html>, accessed on 2009-10-15
8. Gibbon, D., Mertins, I. & Moore, R. (2000). Handbook of Multimodal and Spoken Dialogue Systems: Terminology, Resources and Product Evaluation. New York: Kluwer Academic Publishers
9. Gibbon, D. (2008). Nondeterministic Finite State Transducer. Version 2008-08-12. <http://wwwhomes.uni-bielefeld.de/gibbon/Forms/Python/FSM/generator.html> , accessed on 2011-10-12
10. Giles, H., Coupland, N., & Coupland, J. (1992). Accomodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), Contexts of accommodation (pp. 1-68). Cambridge: Cambridge University Press
11. Kaplan, R & Kay, M. (1994). Regular Models of Phonological Rule Systems. In: *Computational Linguistics 20(3)*, pp. 331-378
12. Karpiński, M. (2002). The Corpus of the Polish Intonational Database (PoInt). In: *Investigationes Linguisticae, Vol. 8*, pp. 24-25, <http://www.staff.amu.edu.pl/~inveling/pdf/maciej_karpinski_inve8.pdf>, 2010-05-20
13. Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. In: Behavioral and Brain Sciences, 27, pp. 169-225