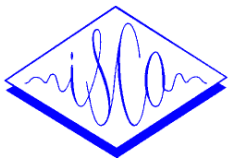

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

COPING WITH ABOUTNESS COMPLEXITY IN INFORMATION EXTRACTION FROM SPOKEN DIALOGUES

Megumi Kameyama, Isao Arima***

* Artificial Intelligence Center, SRI International,
333 Ravenswood Ave., Menlo Park, California 94025, U.S.A.
** Laboratory for Information Technology, NTT DATA Communication Systems Corp.,
Kowa Kawasaki Nishi-guchi Bldg.,
66-2 Horikawa-cho, Saiwai-ku, Kawasaki-shi, Kanagawa 210, JAPAN
megumi@ai.sri.com, arima@rd.nttdata.jp

ABSTRACT

We report on the strategies in automatic summarization of spontaneous spoken dialogues. Our dialogue summarization system, MIMI, recognizes key linguisitc patterns and merges information to construct a summary of conference room scheduling dialogues. Dialogues about single reservations can be accurately summarized with a simple merging scheme, but we need several extensions and changes to cope with the *aboutness complexity* in unrestricted dialogues — knowing exactly *how many* reservations are being discussed and *which* reservation values must be updated by subsequent utterances. A side effect of these extensions is a new problem of overrecognition caused by spoken language disfluencies.

Keywords: Spontaneous Dialogues, Summarization, Information Extraction

1. INTRODUCTION

A full understanding of spoken dialogues would need a full discourse understanding plus robustness with disfluencies. Such a system would require a lot of linguistic and world knowledge, and, if at all possible, would take a long time to develop and to run. We approach the problem with an opposite question — How *little* knowledge is needed to extract simple summaries of task-oriented spoken dialogues?

We have collected and transcribed 150 spontaneous Japanese dialogues between a client and clerk doing conference room scheduling. Their simple summaries are records of who reserved what room(s) for what date(s) and time(s). We are developing an efficient and robust dialogue summarization system extending the technology of FASTUS™ [3, 1, 2], developed for information extraction from written texts. The system is called MIMI (for "ears" in Japanese).

The work reported here is the result of directly processing the written transcripts. This is an interim phase towards a real-time dialogue summarizer with a speech recognition component. We have recently made a transition from summarization of dialogues about single reservations to *unrestricted* summarization. Here, we discuss the extensions and changes that MIMI underwent in this transition. We believe that many of our design decisions are directly applicable to similar summarization tasks across different domains.

2. MIMI'S SUMMARIZATION METHOD

MIMI is an automatic spoken dialogue summarizer with the architecture of a cascaded finite-state transducer. A dialogue is treated as a sequence of *utterances* that can be flexibly defined. Each utterance is processed through four phases that recognize *words*, *basic phrases*, *complex phrases*, and *domain patterns*, respectively. MIMI's domain patterns are patterns of who says what — for instance, the client saying who he or she is, the clerk saying which room is available. MIMI robustly ignores any words or patterns of phrases that it does not know.

As a dialogue is processed, a summary is gradually built up. When a domain pattern is recognized, the summary of the given dialogue is updated. Each new value updates the existing reservation object(s) in one of three possibilities — `New`, `Merge`, or `Override`:

`New` When there is no information of the same type in the current summary, the new information is simply added.

`Merge` Two values of the same type that are mutually consistent are merged into a single value. This includes the cases of two identical values.

`Override` A new value replaces the most recent value of the same type when they are mutually inconsistent.

The Appendix contains a short dialogue example with its summary output.

3. COMPLEXITY OF SPOKEN DIALOGUE SUMMARIZATION

The major factor that determines the complexity of the summarization process is the number of reservations being made in the dialogue. Those dialogues that make or cancel single reservation instances (*single-reservation dialogues*) are the easiest (if we know that they are about only one reservation). Those that make or cancel multiple reservations (*multiple-reservation dialogues*) are more difficult, representing the *aboutness complexity*. The first difficulty is to know exactly *how many* reservations are being discussed, and the second difficulty is to know *which* reservation values must be updated by subsequent utterances.

The first version of MIMI was designed only for single-reservation dialogues [4]. We used utterance units segmented by pauses and speaker turns and a simple information-updating strategy. The most successful strategy was to *merge* redundant values and to let new values *override* older ones when inconsistent. It achieved a recall of 82.5% and precision of 90% for a blind test set of 45 dialogues in October 1993.[1] The second version of MIMI, whose results to date are reported here, is designed for unrestricted dialogues, and we have made a number of extensions to cope with the difficulty of summarizing multiple-reservation dialogues.

4. EXTENSIONS TO HANDLE MULTIPLE–RESERVATION DIALOGUES

To know exactly how many reservations are being discussed, we now recognize a set of *disjoint reservation instances*. To know which reservation values must be updated in the subsequent dialogue, we now interpret *anaphoric reference* to reservation instances.

Disjoint Reservations. Disjoint reservation instances are created by recognizing conjoined or quantified phrases. The merging scheme is defined so that disjoint reservations never merge with one another, and new reservation values mentioned in the subsequent dialogue merge or override one or more of these disjoint reservation objects.

The following are the phrase patterns that indicate multiple disjoint reservations:

1. Conjunction: *konnsyuu no suiyou to raisyuu no kayoubi* 'Wednesday this week and Tuesday next week'
2. Iteration: *kayou, suiyou, mokuyou, kinnyou* 'Tuesday, Wednesday, Thursday, Friday'
3. Quantification: *konnsyuu kara maisyuu mokuyoubi* 'Thursday every week starting from this week'
4. Date Range: *asita kara konnsyuu ippai* 'Starting from tomorrow through this week'
5. Enumeration: *mazu konnsyuu no suiyoubi ... sono tugi raisyuu no kayoubi* 'First Wednesday this week ... Next Tuesday next week'

The exact set of dates is determined from the *dialogue date* in the context and the restriction that reservations can be made only up to four weeks in a row.

MIMI's earlier *breath-groups-based* utterances ending with pauses and speaker turns turned out to be too small because these complex phrases are often broken up by pauses and the listener's confirmations. Two such examples are given here (A=client, B=clerk):

1. A: *getuyou to* 'Monday and' B: *hai* 'Yes' A: *kayou* 'Tuesday'
2. A: *raisyuu no getuyou* 'Monday next week A: *suiyou* 'Wednesday' A: *kinnyou* 'Friday' B: *hai* 'Yes' A: *totte kudasai* 'Please reserve'

Examples like these motivated a larger utterance unit ignoring pauses and short confirmations. The result is a *turn-based* utterance segmentation, which packs the above small units as follows:

1. A: *getuyou to kayou* 'Monday and Tuesday'
2. A: *raisyuu no getuyou suiyou kinnyou totte kudasai* 'Please reserve Monday Wednesday Friday next week.'

Anaphoric Reference to Reservation Instances. Once a set of disjoint reservation objects is in the summary, a particular mention of room or time may be about one or all of these reservations. The linguistic signals for the range of distribution come in various forms, and they are often implicit. For instance, an overhead projector need usually applies to all of the reservations in each dialogue unless something to the contrary is explicitly said. Times and rooms, however, widely vary. Some multiple-reservation dialogues continue to talk about the whole set of reservations at once, reserving the same room for the same time range. Others discuss each reservation in turn, concluding on different rooms and times. When different reservations are discussed in turn, we need to keep track of the reservation being discussed. This is a *topic recognition problem*. Our strategies include resolving partial date reference to reservations and recognizing topic shifts.[2]

Since dates distinguish different reservation instances in this domain, once the dates of multiple reservations are mutually understood, the speakers refer to them using partial dates as *anaphoric expressions*. As is generally the case with anaphoric expressions, the distinguishing properties among the particular set of reservations pick out instances. Here are some examples:

[1] *Recall* is the number of answers the system got right divided by the number of possible right answers. It measures how comprehensive the system is. *Precision* is the number of answers the system got right divided by the number of answers the system gave. It measures the system's accuracy.

[2] The topic recognition problem has not entirely been solved in MIMI. When some values are still unfilled at the end of the dialogue, existing values are randomly distributed.

- Reservations on different days of the same week are referred to with days of week, e.g., *kayou* 'Tuesday'.
- Reservations on the same day of week in different weeks are referred to with weeks, e.g., *saraisyuu* 'the week after next'.
- Reservations on different days across different weeks are referred to with the days of month, e.g., *mikka* 'the third'.

When a partial date is used in the context of multiple reservations, MIMI interprets it as a reference to one of them and updates only the relevant reservation instance with the new information.

When a multiple-reservation dialogue discusses one reservation at a time, the result is a sequence of dialogue segments about single reservation instances. In one such multiple-segment example, four reservations on different dates (8/28, 9/4, 9/11, and 9/18) are discussed one by one, starting the segments with the following utterances:

B4: *itibann saisyo no kinnyoubi hati gatu ni zyuu hati niti wa desu ne.* 'As for the first Friday, August 28th.'

B12: *ku gatu yokka no kinnyoubi na no desu kedomo.* 'As for Friday September 4th.'

B16: *tugi no syuu wa ku gatu zyuu iti niti na no desu keredomo.* 'The next week is September 11th.'

B19: *zyuu hati niti mo dai go, dai roku, aite imasu node.* 'Also on the 18th, the 5th and 6th are available.'

MIMI recognizes an initial set of segment-opening phrases such as *tugi no* 'the next' and *saigo no* 'the last'. Topic shifting, however, is not always explicitly signaled, as in the B19 utterance above. Segment-based information updating thus does not always lead to correct results.

5. OVERRECOGNITION PROBLEM

After MIMI underwent the above set of extensions, we had to face another problem that the first version of MIMI never had to worry about — MIMI started finding multiple reservations in some single-reservation dialogues. In the training set of 48 dialogues, MIMI still finds too many reservations in 8 dialogues, and too few in 5.

One major cause of this type of error is *repairs* — parts of utterances where the speaker corrects himself. Some mentions of dates and times with repairs look like iterations, and thus match MIMI's key patterns. Utterances with problematic repair patterns (all in uppercase) are

- *RAISYUU NO GETUYOU, KONNSYUU NO MOKUYOU KA RAISYUU NO GETUYOU no gogo.* 'the afternoon of NEXT MONDAY, THIS THURSDAY OR NEXT MONDAY.'
- *TOUKA TO ZYUU SITI NITI, NI ZYUU YOKKA, ZYA NEE YA MIKKA, TOUKA, ZYUU SITI NITI, NI ZYUU YOKKA.* 'the 10th and 17th, 24th, no the 3rd, 10th, 17th, 24th.'

	Training Set			Blind Test Set		
	Total	Single	Multi	Total	Single	Multi
MIMI1 R	—	88.4	—	—	82.5	—
MIMI1 P	—	91.0	—	—	90.0	—
dialogues		(30)			(45)	
MIMI2 R	79.1	96.8	70.0	56.4	71.0	48.8
MIMI2 P	68.9	66.7	70.6	62.6	55.9	69.2
dialogues	(48)	(30)	(18)	(100)	(64)	(36)

Table 1: MIMI1 and MIMI2 Performance Comparison (R=Recall(%),P=Precision(%))

An obvious strategy is to make sure that a conjoined list of dates and times is not redundant. Redundancy checking removed some of the superfluous reservations, but other complex repair patterns still mislead MIMI.

There are 40 instances of repairs and false starts in the training set of 48 dialogues — about one instance per dialogue. Of these 40, 22 occur in MIMI's key patterns, and the remaining 18 are either in informationally insignificant utterances or the disfluency does not affect MIMI's key pattern. Of the 22 disfluencies in key patterns, 8 occur in utterances about dates and 11 are in utterances about hours. These 19 utterances are the prime suspect for the cause of overrecognition — we found that 7 of them account for *most* (6 out of 8) of MIMI's overrecognition.

6. PERFORMANCE RESULTS

We have achieved a recall of 79.1% and precision of 68.9% for a training set of 48 dialogues, 30 of which are single-reservation and 18 are multiple-reservation dialogues. For a blind test set of 100 dialogues, 64 of which are single-reservation and 36 are multiple-reservation, the same system achieves a 56.4% recall and 62.6% precision. Table 1 shows the recall and precision of the two versions of MIMI — MIMI1 (10-20-93) summarizes only single-reservation dialogues. MIMI2 (5-18-94) is unrestricted.

The general drop of performance in the blind test results from a sharp difference between the scores for single-reservation and multiple-reservation sets. Both the single-reservation recall and multiple-reservation precision are around 70% — no significant drop of precision from the training set result. The problem areas are the *multiple-reservation recall* (48.8%) and *single-reservation precision* (55.9%). The low multiple-reservation recall is due to the wide range of linguistic patterns that signal the reserver's intentions, and the low single-reservation precision is due to the misleading disfluencies in the key linguistic patterns. We believe that this general pattern — a low recall in multiple-topic dialogues and a low precision in single-topic dialogues with disfluencies — can be expected in summarizing dialogues in other domains.

7. CONCLUSION

Our minimal information extraction approach to spoken dialogue summarization has been extended to cope with the difficulty of summarizing multiple-topic dialogues. The extensions include the introduction of disjoint referents into the information–merging scheme and a limited reference resolution to achieve minimal topic recognition. The cause of the new overrecognition problem was found to be the disfluencies in spoken language.

8. FUTURE PROSPECTS

The initial focus should be on improving the precision in summarizing single-reservation dialogues and the recall in summarizing multiple-reservation dialogues. That is, we will need to suppress overrecognition and expand the coverage of multiple-reservation key patterns. One possibility is to add a repair pattern to MIMI so that parts of information can be locally retracted. The second focus should be on improving an overall precision. MIMI currently relies too heavily on the general-purpose merging scheme to update the summary. More precise updating instructions sensitive to the speaker's communicative intentions (e.g., assertions, requests, offers, refusals, and acceptances) would improve the precision because we would know *how* to update the parts of the summary. We also plan to integrate MIMI and a speech recognizer for real-time dialogue summarization.

ACKNOWLEDGMENT

We would like to thank Douglas Appelt, Ken'ya Murakami, Raymond Perrault, and Otoya Shirotuka for their support and encouragement.

References

[1] Appelt, Douglas, Jerry Hobbs, John Bear, David Israel, and Mabry Tyson. FASTUS: a finite-state processor for information extraction ¿from real-world text. In the *Proceedings of the International Joint Conference on Artificial Intelligence*, 1993a.

[2] Appelt, Douglas, Jerry Hobbs, John Bear, David Israel, Megumi Kameyama, and Mabry Tyson. SRI: description of the JV-FASTUS system used for MUC-5. In Sundheim, Beth, ed., the *Proceedings of the 5th Message Understanding Conference*, ARPA, 1993b.

[3] Hobbs, Jerry, Douglas Appelt, John Bear, David Israel, and Mabry Tyson. FASTUS: a system for extracting information from natural-language text. Technical Note No. 519. SRI International Artificial Intelligence Center, 1992.

[4] Kameyama, Megumi and Isao Arima. 1993. A minimalist approach to information extraction from spoken dialogues. In the *Proceedings of the International Symposium on Spoken Dialogue (ISSD-93)*, Waseda University, Tokyo, 137-140.

Appendix. Dialogue Example

This dialogue took place at 13:30 on 24 August 1992. The Client is A, and the Clerk is B. Key patterns appear in boldface, followed by the relevant merging or other contextual operations that take place with the associated values.

B1: *hai, kotira kaigisitu yoyaku gakari desu.* "This is the CRR clerk."

A1: **kotira dai ni ninnsiki nyuuryoku no kita desu keredomo,**[New:name&group] **kaigisitu no yoyaku wo sitai no desu ga.** [New:purpose] "I'm Kita of the 2nd Recognition Input Group. I'd like to reserve a conference room."

B2: *hai hai.* "OK."

A2: *ano,* **hi wa ku gatu hutuka no suiyoubi de.** [New:date] "Let me see. The date is Wednesday September 2nd."

B3: *hai.* "OK."

A3: **zikann wa ni zi kara yozi,**[New:time] **ninnzuu roku ninn desu.** [New:size] "The time is from 2 to 4 o'clock. The number of people is six."

B4: *ee,* **ouettipii nannka wa tukaimasu ka.** [question(overhead)] "Uh will you use something like an overhead projector?"

A4: **ouettipii wa tukaimasenn.** [New:overhead] "We won't use an overhead projector."

B6: *hai, wakarimasita. syousyou o mati kudasai. sou suru to desu ne, kaigisitu,* **dai ni kaigisitu ga aite imasu ne.** [New:room] "OK. I see. Please wait a moment. In that case, conference room, the 2nd conference room is available."

A6: *aa, sou desu ka.* "Oh I see."

B7: *kotira ee teiinn ga zyuu ni mei natte orimasu node daizyoubu da to omoimasu.* "Since this one can take 12 people, I think it will do."

A7: *hai, doumo yorosiku onegai simasu.* "OK. Please do that."

B8: *hai,* **sore dewa yoyaku simasu.** [New:action] "OK. I'll reserve it then."

Summary Output:

```
<RESERVATION-25>
TYPE OF ACTION: RESERVE
ACTION TAKEN: ACCEPTED
RESERVER NAME: KITA
RESERVER GROUP: DAI-NI-NINNSIKI
ROOM: DAI-NI-KAIGISITU
RESERVED DATE: (POINT (92 9 2) NIL)
RESERVED TIME: (RANGE (14 0 0) (16 0 0))
NUMBER OF PEOPLE: 6
OHP REQUIRED: NO
```