AN ANALYSIS OF DIALOGUES WITH OUR DIALOGUE SYSTEM THROUGH A WWW PAGE

Tadahiko Kumamoto[1] *Akira Ito*[2]

[1]Communications Research Laboratory, MPT, Kobe, Hyogo, Japan
[2]Faculty of Engineering, Yamagata University, Yonezawa, Yamagata, Japan

ABSTRACT

Many researchers have been developing natural language dialogue systems as a human-friendly man-machine interface. The human factors in a man-machine dialogue, however, are not obvious enough to understand with regard to how people talk with a dialogue system. 141 dialogues which our dialogue system had in DiaLeague '97 were analyzed at the utterance and dialogue levels, where DiaLeague '97 was the second dialogue contest in which a dialogue system engaged in a dialogue with a human in order to solve a specific problem. For the analyses at the utterance level, we investigated the users' speaking styles, the richness of the users' utterances in a variety of surface patterns, and the influence of the system's utterance pattern on the users' utterance. For the analyses at the dialogue level, we investigated the instances of confusion observed in the 141 dialogues and we also show how the users behaved when the confusion occurred.

1. INTRODUCTION

Natural language dialogue systems have been developed as a human-friendly man-machine interface [1, 2]. The human factors in a man-machine dialogue, however, are not obvious enough to understand with regard to how people talk with a dialogue system.

A dialogue is a collaboration between dialogue participants. Therefore, identification of a conversational partner influences not only the form and contents of an utterance but also the context and expansion of a dialogue. Dialogue data for analyzing such human factors have been collected by using the following two methods. One is based on the Wizard of Oz setting [3], and the other is based on using a speech understanding system [4]. In the Wizard of Oz method, a person poses as a dialogue system and talks with a subject. The subject is made to believe that the other participant is not a human but a machine. This setting enabled us to collect the high level dialogues that people have. We consider, however, that these dialogues do not have the various phenomena that may occur in a dialogue with an actual working dialogue system. Therefore, the dialogues collected by the Wizard of Oz method are not effective for investigating the influence of such phenomena on a dialogue. According to Ref. [3], in the dialogue collection experiment, what the subjects actually said was not expected in a real situation with a dialogue system since the subjects did not have enough motivation to talk with the dialogue system. This implies that a small number of subjects may not have the same behavior patterns with the people who will talk with a dialogue system in daily life. In the meantime, the dialogues collected using the speech understanding method are not rich enough in variety for analyzing a user's behavior since a low accuracy of speech understanding requires a simple task, but it is useful for analyzing voice-dependent characteristics of user utterance such as hesitations and repairs.

We analyzed 141 dialogues which our dialogue system had in DiaLeague '97 [5]. DiaLeague '97 was the second dialogue contest in which a dialogue system engaged in a dialogue with a human in order to solve a specific problem. It was proposed as a method for objectively and synthetically evaluating the overall performance of a dialogue system. Each of the dialogue systems that participated in the contest obtained a score according to the amount of dialogue it had with a contest participant. One of the main features of this contest is that a human and a dialogue system talked with each other through the world-wide web (www). Such use of www enabled us to collect real man-machine dialogues with many people. Note that in this contest the dialogue system input (user utterance) was a Japanese kanji-kana mixed sentence input by keyboard and the mouse's cut and paste function, while output from the dialogue system (system utterance) was a Japanese kanji-kana mixed sentence displayed on the user's window.

In this article, the 141 dialogues were analyzed at the utterance and dialogue levels. For the analyses at the utterance level, we investigated the users' speaking styles, the richness of the users' utterances in a variety of surface patterns, and the influence of the system's utterance pattern on the users' utterance. For the analyses at the dialogue level, we investigated the instances of confusion observed in the 141 dialogues and we also show how the users behaved when the confusion occurred.

2. DIALEAGUE '97

DiaLeague '97 was a contest for synthetic evaluation of dialogue systems. It was held on the www pages for 168 hours [5]. This contest was announced on NetNews and information was sent to people registered on mailing lists. The five dialogue systems that registered in advance had

728 dialogues with the Internet users in this contest.

In this contest, a route task was employed. A user and a dialogue system each had a railway route map. The two route maps were the same in number and arrangement of stations, but had different patterns regarding passable routes. The goal was to find a route from a starting station to a destination station, where the route was the same for both route maps, was passable in either map, and was the only route. The starting station was located at the upper-right hand corner of the route map and the destination station was located at the bottom-left hand corner of the route map. Some of the stations in a route map did not have a name, and were denoted by a question mark followed by a numeral. If a name of a station was unknown in one route map, it was known in the other route map. The dialogue system was required to know all of the names of the stations on the common route they found.

The Internet users accessed the www page (http://www.etl.go.jp/etl/nl/dialeague/webdialogue/) which worked as a dialogue server, and they registered an entry name. If the registration was completed, the dialogue server randomly selected one out of the five dialogue systems registered in advance, determined which (the user or the dialogue system) first passed a message to the other, and determined the route maps that were presented to both. The dialogue server then constructed the www page on which the user had a dialogue with the selected dialogue system. The www page has three frames. The top frame shows a route map of the user and the middle frame shows user's and system's utterances during the dialogue. The bottom frame has a text input field which allows the user to pass a message to the dialogue system. In this contest our dialogue system obtained the highest average score out of the five dialogue systems.

3. ANALYSES AT THE UTTERANCE LEVEL

3.1. Extraction of utterance patterns

The 141 dialogues that we analyzed had 2,930 user utterances and 2,898 system utterances. Since an individual concept and a deviation in transcription were observed in these utterances, we abbreviated the utterances in the following ways.

- Abbreviated the name of a station as *A*.
- Abbreviated two or more consecutive station names as *A* ... *A*.
- Abbreviated words representing direction as *direction*.
- Abbreviated words representing distance as *distance*.
- Removed interjections.

By these operations, for example, the utterance "Dazaifu wa Kasukabe no migi desuka (Is Dazaifu at the right of Kasukabe?)" is abbreviated as "*A* wa *A* no (direction) desuka (Is *A* at the *direction* of *A*?)," and "E-to, Fujimi, Neyagawa, Higashimurayama, Narashino to ikemasuka (Well, shall I take the Fujimi, Neyagawa, Higashimurayama, and Narashino route?)" as "*A* ... *A* to ikemasuka (Shall I take the *A* ... *A* route?)."

Table 1: End-of-sentence expressions of user utterances

Types	Examples	Freq.
Yes/No	Hai (yes, right), Iie (no)	1,095
"masu"	suffix representing politeness	985
"desu"	polite form of the verb "to be"	705
Fragment	nominals and postpositional words representing case and subject	54
"suru"	basic form of a verb	21
Imperative	imperative form of verb	9
Others		61
Total		2,930

Table 2: Sentence types of user and system utterances

	Decl.	YN-Q	WH-Q
system utt.	559 (19.3%)	1,517 (52.3%)	470 (16.2%)
user utt.	1,061 (36.2%)	444 (15.2%)	210 (7.2%)

3.2. End-of-sentence expressions and sentence types

People are expected to talk to a dialogue system carefully so that the dialogue system could understand their utterances. We analyzed the end-of-sentence expressions and the sentence types of 2,930 user utterances in order to investigate the users' speaking styles.

In Table 1, the end-of-sentence expressions were classified. It shows that 1,690 user utterances were classified into the "desu" and "masu" types which represent politeness. In our prospect before the contest, the users were considered to talk bluntly like "*A* ni ikeruka (Can I go to *A*?)," and to talk fragmentarily like "*A* no (direction) wa (At the *direction* of *A*?)." A short sentence fragment may especially give the user a higher score. However, most of actual user utterances were polite.

The sentence types of the user and the system utterances were compared. The results are listed in Table 2. It shows that the users preferred making a declarative sentence rather than making a yes-no question or a "wh" question. To clarify why they did so, we investigated how the users knew the name and position of a station when they wanted to know it. The results are listed in Table 3. We know that the simplest way to know a name and a position is to directly ask the dialogue system in the form of a "wh" question. However, as Table 3 shows, declarative sentences like "*A* no (distance)(direction) e ikimasu (I go *distance* to the *direction*.)" and yes-no questions like "*A* wa *A* no (direction) desuka (Is *A* at the *direction* of *A*?)" were often observed. This means that people use indirect expressions even in a dialogue with a machine.

Table 3: Sentence types for asking the name and position of a station

	Decl.	YN-Q	WH-Q
Position of Station	5 (3.0%)	41 (24.6%)	121 (72.5%)
Name of Station	28 (17.3%)	64 (39.5%)	70 (43.2%)

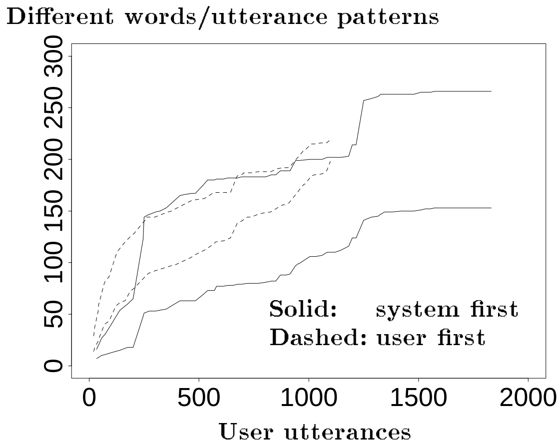

Figure 1: The upper solid and dashed lines show a relation between the numbers of utterances and different words, while the lower ones show a relation between the numbers of utterances and utterance patterns. "User first" means that the user passed a message in their dialogue first.

3.3. Variety of user utterance patterns

302 utterance patterns were extracted from 2,930 user utterances, and 48 from 2,898 system utterances. That is, the number of user utterance patterns is 6.3 times the number of system utterance patterns.

For the 2,930 user utterances, we investigated how the numbers of different words and utterance patterns changed as the number of utterances increased. The results are shown in Figure 1, where the number of utterances was accumulated in every dialogue based on the time when the dialogue began. Figure 1 shows that the number of different words increased similarly either in user first or system first, and the utterance patterns observed in a dialogue were richer in variety on an average when the dialogue began with a user question and request.

3.4. Influence of a system utterance pattern on a user utterance

We investigated whether the users were influenced by the system utterance patterns with regard to making a sentence.

The system and user utterances have only 22 common utterance patterns. Nevertheless, 82.5% (2,392/2,898) of the system utterances and 61.8% (1,811/2,930) of the user utterances correspond to the 22 utterance patterns. This may imply that the system and user utterances have a strong correlation in utterance patterns. However, the 22 utterance patterns are only 7.3% (22/302) of the user utterance patterns, and distributions of the system and user utterances in the 22 utterance patterns also are very different. We focused on two of the 22 utterance patterns, and investigated when the users used the two. One is "*A* wa dokodesuka (Where is *A*?)" for asking the position of a station, and the other "*A* no (direction) no eki wa nandesuka (What is the station at the *direction* of *A*?)" for asking the name of a station. The results are as follows. Although 33 of the system utterances and 108 of the user utterances correspond to the former utterance pattern, both user and dialogue system used it together in only nine dialogues. Moreover, in three of the nine dialogues, the user used it before the dialogue system did. In the meantime, the latter utterance pattern was used by the dialogue system 215 times. However, the users only used it four times. The above findings form a conclusion that the users were not influenced by the system utterance patterns.

4. ANALYSES AT THE DIALOGUE LEVEL

User utterances to a dialogue system are rich in variety but are not always grammatical. In addition, every person has his/her own word concept and grammar, and it is not easy to describe them exactly in a machine readable form. Consequently, even a well-designed dialogue system inevitably causes dialogue confusion. Although such confusion can be observed in a dialogue between people, people often recover from the confusion by the dialogue itself. We analyzed the 141 dialogues between the users and our dialogue system, and investigated how the users behaved when dialogue confusion occurred.

4.1. User's reaction against an unexpected system utterance

We investigated how the users behaved when the dialogue system said an unexpected utterance out of context. The results are listed in Table 4, where the four utterance patterns have the highest frequencies among all the unexpected system utterances. Table 4 indicates that when the unexpected system utterances were questions, the users behaved in a friendly manner in many cases. That is, they did not ignore the unexpected system question. When the dialogue system said "Wakarimasen (I don't understand.)," the users preferred to change the contents of their utterance rather than to change an expression and wording of the utterance. In a dialogue between people, one is expected to modify his/her utterance so that the other could understand it. Because the conversational partner was a dialogue system, the users did not feel a need to modify their utterance.

We investigated how the users behaved when the dialogue system ignored their question and/or request. The results are listed in Table 5 and show that the users behaved in a friendly manner. In only four of the 33 cases, the users ignored the system question.

Table 4: User's reaction against an unexpected system utterance

S	Wareware wa ima dokoni irundesuka (Where are we now?)	10
U	Responded to the syst. utterance	9
	Ignored the syst. utterance	1
S	Dokokara *A* ni ikundesuka (Where do you go to *A* from?)	22
U	Asked a question about the syst. utterance	4
	Responded to the syst. utterance	14
	Ignored the syst. utterance	2
	Failed to type a sentence	2
S	*A* no (direction) no eki wa nandesuka (What is the station at the *direction* of *A*?)	30
U	Asked a question about the syst. utterance	7
	Responded to the syst. utterance	22
	Ignored the syst. utterance	1
S	Wakarimasen (I don't understand.)	15
U	Modified the self utterance	2
	Changed contents of the self utterance	10
	Kanji code error, etc	3

∗ "S" and "U" denote system and user utterances, respectively.

∗ Numerals in table show frequencies of occurrence.

Table 5: User's reaction when our dialogue system ignored the user's question and/or request

User's reaction	Freq.
Asked a question about the system question	4
Responded to the system question	23
Ignored the system question	4
Input an empty line	1
Quit the dialogue	1
Total	33

4.2. User's reaction against repetition of a system question

Our dialogue system repeated a system question when it failed the spoken language understanding and dialogue processing of a user's reply. We investigated how the users responded to a repetition of a system question. The results are listed in Table 6 and show that the users preferred to change the contents as well as their behavior for "Wakarimasen (I don't understand.)."

For all of the 100 cases regarding repetition, we investigated whether repetition of a system question contributed to recovery from dialogue confusion. The results are as follows. Repetition of a system question contributed to the recovery of an original dialogue context in 67 cases, and contributed to making up a new dialogue context in 13 cases. Consequently, we say that repetition of a system question was effective in recovering from the dialogue confusion caused by the failure of spoken language understanding and dialogue processing. A fatal dialogue situation occurred in 14 cases. It occurred when a user's wrong belief was registered in the dialogue system as a fact, and when a user's reply was wrongly interpreted.

Table 6: User's reaction against repetition of a system question

User's reaction	Freq.
Modified the user reply	22
Repeated the user reply	8
Changed the contents of the user reply	66
Kanji code error, etc	4
Total	100

5. CONCLUSION

We analyzed 141 dialogues between our dialogue system and Internet users through www, and we investigated how the users behaved when they had dialogues with a machine. We found the following results. (1) Users talked to a dialogue system in a polite manner. (2) Users did not always make a sentence using the direct speech. (3) Utterance patterns observed in a dialogue were richer in variety when the dialogue began with a user question and/or request. (4) Users were not influenced by system utterance patterns in making a sentence. (5) Users did not ignore an unexpected system utterance. (6) Users preferred to change the contents of their utterance rather than to express it differently when the dialogue system said "Wakarimasen (I don't understand.)," and repeated a system question. (7) Dialogue confusion which was caused by the failure of spoken language understanding and dialogue processing was often recovered by repeating a system question. (8) A fatal dialogue situation often occurred when a user's wrong belief was registered in the dialogue system as a fact.

We will develop a more robust dialogue system based on the findings described in the present article.

6. REFERENCES

1. Y. Hiramoto et al. A speech dialogue management system for human interface employing visual anthropomorphous agent. In *Proc. 3rd IEEE International Workshop on Robot and Human Communication*, pages 277–282, Nagoya, 1994.
2. K. Yamada et al. Model of utterance and its use in cooperative response generation. In *Proc. East-West International Conference on Human-Computer Interaction*, pages 252–264, 1993.
3. S. Nakagawa et al. Relationship between the structure of spoken dialog systems and user utterances. *Trans. IEICE of Japan*, J79-D-II(12):2139–2145, 1996.
4. K. Itou et al. System design, data collection and evaluation of a speech dialogue system. *IEICE Trans. Inf.& Syst.*, E76-D(1):121–127, 1993.
5. K. Hashida et al. Dialeague: Synthetic evaluation of natural language processing systems. *Journal of Japanese Society of Artificial Intelligence*, 12(3):390–399, 1997.