EFFECTS OF SELF-DISCLOSURE AND EMPATHY IN HUMAN-COMPUTER DIALOGUE

Ryuichiro Higashinaka, Kohji Dohsaka, and Hideki Isozaki

NTT Communication Science Laboratories, NTT Corporation
{*rh,dohsaka,isozaki*}@*cslab.kecl.ntt.co.jp*

ABSTRACT

To build trust or cultivate long-term relationships with users, conversational systems need to perform social dialogue. To date, research has primarily focused on the overall effect of social dialogue in human-computer interaction, leading to little work on the effects of individual linguistic phenomena within social dialogue. This paper investigates such individual effects through dialogue experiments. Focusing on self-disclosure and empathic utterances (agreement and disagreement), we empirically calculate their contributions to the dialogue quality. Our analysis shows that (1) empathic utterances by users are strong indicators of increasing closeness and user satisfaction, (2) the system's empathic utterances are effective for inducing empathy from users, and (3) self-disclosure by users increases when users have positive preferences on topics being discussed.

Index Terms— Natural language interfaces, Interactive systems

1. INTRODUCTION

Conversational systems need to build trust or cultivate long-term relationships with users through social dialogue when they want users to engage in possible face-threatening tasks such as real-estate transactions [1] or to disclose personal information for personalized services such as health-care [2]. Social dialogue includes greetings, chatting about general topics like the weather, exchanging personal preferences, and so forth. Since social dialogue comprises various linguistic phenomena, systems must be able to make an appropriate linguistic choice to have the desired effect. However, because most research has been primarily concerned with the effect of social dialogue as a whole in human-computer interaction, little work has sought to qualitatively examine the individual effects of linguistic phenomena within social dialogue.

This paper investigates the individual effects of linguistic phenomena through dialogue experiments. Focusing on self-disclosure and empathy (agreement and disagreement), we collected dialogue data between a system and users to calculate the correlations between the counts of self-disclosures or empathic utterances and the dialogue quality perceived by users. We then analyzed the correlations to reveal the individual effects of self-disclosure and empathy. Figure 1 shows an excerpt of a dialogue in our experiments. In the dialogue, the user and system discuss likes and dislikes about animals (fireflies in this case), disclosing preferences and agreeing with each other. Since dialogue participants can choose to self-disclose or to be empathic at any moment, it is important for a system to know the individual effects of its actions to improve dialogue quality.

Section 2 describes our approach in detail. Section 3 describes our dialogue system and the dialogue data collection. Section 4 shows the results of the analysis and describes our main findings. Section 5 summarizes and mentions future work.

S1	Hello (GREETING)
U1	Hello (GREETING)
S2	Let's discuss likes and dislikes about animals. (OPEN-DIALOGUE)
U2	Do you like insects? (Q-DISC-P)
S3	Do you like insects? (Q-DISC-P)
U3	I like fireflies. (DISC-P)
S4	You like fireflies. (REPEAT)
U4	I like fireflies because they glow. (DISC-P-R)
S5	I also like them very much. (DISC-AGREE-P)
U5	Why do you like fireflies? (Q-DISR-R)
S6	I like them too. (DISC-AGREE-P) They are evanescent. (DISC-R)
U6	I like them because they are beautiful. (DISC-P-R)
S7	I see (ACK), I understand. (EMP) I also think they are beautiful. (DISC-AGREE-R) Can you tell me more about why you like fireflies? (Q-DISC-R-OTHER)
U7	I like them because they glow at night. (DISC-P-R)

Fig. 1. An excerpt of a dialogue in our experiment. S stands for system utterance and U for user utterance in the dialogue. Dialogue acts corresponding to utterances are shown in parentheses (See Section 3.1.2 for their meanings). The dialogue was originally in Japanese and was translated by the authors.

2. APPROACH

As linguistic phenomena in social dialogue, we focus on self-disclosure and empathy, both of which are known to contribute greatly to the closeness between conversational participants [3, 4, 5]. In addition to closeness, we are also concerned with user satisfaction of a dialogue because it is commonly used to evaluate conversational systems. Our objective is to reveal their individual effects on the closeness and user satisfaction in a qualitative manner. Our approach for this is to collect as many dialogues as possible and calculate the correlations between the counts of self-disclosure and empathy and the closeness or user satisfaction evaluated from questionnaires.

Two specific aims are to determine what linguistic behavior most affects closeness and user satisfaction, and to clarify how dialogue systems should make its linguistic choice in order to increase them. The former will help us derive appropriate dialogue strategies, and taken together, these findings will help us achieve our eventual goal of creating systems that can achieve better closeness and user satisfaction. We are also interested in how systems can be made to encourage self-disclosure by users because one of the important goals of dialogue systems is to have them elicit user information for personalization or marketing purposes.

To collect dialogues, we built an automated dialogue system using text input. We did not take a Wizard-of-Oz approach because we are simply interested in the correlation between the observed linguistic phenomena in a human-computer dialogue and the user's perception of that dialogue. Unless automated dialogues end up in complete failure, systems do not need to have human-level dialogue processing capability. In addition, we are also interested in collecting realistic dialogue data with the current level of dialogue technology. Currently, the system works only on text input in order to avoid the influence of speech recognition errors. We aim to deal with speech input in the future.

The selection of the dialogue domain is important because the domain has to be one in which users can easily self-disclose or show empathy. Further, the content to be discussed should not be too complex in order for our system to work sufficiently. For these reasons, we decided to develop our system in the animal domain where conversational participants talk about their likes and dislikes regarding animals. Our reasoning here is that everyone knows at least some animals and have preferences for them.

3. EXPERIMENT

We first implemented a dialogue system in the animal domain and then collected dialogues using human subjects. Dialogue data annotation was performed to find exact counts of self-disclosure and empathy.

3.1. System

The system is a Japanese keyboard-based dialogue system and has a chat-like interface. It has some manually prepared knowledge of animals and preferences and can discuss likes and dislikes about them with users. A dialogue starts with the system's greeting. The system then prompts the user to input an utterance. The user marks the end of an utterance by pressing Enter. Turns are strictly controlled to alternate. A dialogue ends when one of the conversational participants explicitly declares the end of the dialogue.

When a user enters an utterance, it is first parsed by the *utterance understanding component* into a meaning representation called a *dialogue act*. A dialogue act typically comprises (i) a dialogue act type that identifies the main intent of the user's utterance and (ii) its auxiliary information encoded as attribute-value pairs. The dialogue act is then passed onto the *dialogue manager*, which decides the next action to take considering both the dialogue act and the dialogue history. Finally, the *utterance generation component* receives the next action of the dialogue manager to generate the system's utterance. In what follows, we describe how the knowledge of animals and preferences were prepared, the dialogue acts, and how each component works.

3.1.1. Knowledge of Animals and Preferences

The system can recognize 90 animals and has a list of attributes (adjectives) for them. The attributes are used to state reasons for preferences. The animals and their attributes were taken from an Associative Concept Dictionary [6], which lists various nouns with their associative words. We extracted entries of animals in this dictionary together with adjectives in their associative words. For example, *hotaru* (fireflies) has four attributes; *kirei* (beautiful), *chiisai* (small), *utsukushii* (beautiful), *hakanai* (evanescent). The system in Fig. 1 uses this list.

The system's preferences for animals are decided randomly under a probability distribution: positive, 45%; negative, 45%; neutral, 10%. We limited the occurrence of the neutral polarity for fear of the system not being able to say anything due to its indifference. Reasons are randomly selected from the animals' attributes with one exception; some obvious positive attributes (e.g., beautiful) are not selected as reasons to have a negative polarity and likewise for several negative attributes. In the case of fireflies, *kirei* and *utsukushii* can only be reasons for the system to like fireflies. *chiisai* and *hakanai* can be reasons for both polarities. No reason is given to animals for which the system has the neutral polarity.

3.1.2. Dialogue Acts

The system recognizes 22 dialogue acts, which fall into one of six categories: self-disclosure, agreement (empathy), disagreement (opposite of empathy; antipathy), dialogue-control, question, and response. We list the dialogue acts in each category as shown below.

Self-disclosure: We have DISC-P, DISC-R, DISC-P-R, DISC-R-OTHER and RES. DISC-P is used to disclose a proposition P. A proposition is either like(X,Y) or dislike(X,Y), meaning that a conversational participant X likes an animal Y or vise versa. DISC-R discloses a reason R for some aforementioned proposition P. DISC-P-R discloses P and R at the same time, and DISC-R-OTHER discloses R in addition to already mentioned reasons (e.g., "another reason is that…"). We consider RES, which is a response to a yes-no question (i.e., Q-DISC-P), to be self-disclosure.

Agreement: We have DISC-AGREE-P, DISC-AGREE-R, EMP, and REPEAT. DISC-AGREE-P and DISC-AGREE-R show agreement to the propositions or reasons mentioned by the partner. EMP denotes an explicit empathic action (e.g., "I understand"), and REPEAT means the repetition of the partner's previous self-disclosure to show understanding.

Disagreement: We have DISC-DISAGREE-P and DISC-DISAGREE-R. They show disagreement to the propositions or reasons mentioned by the partner; e.g., saying "I don't like cats" to the partner who has already disclosed that he/she likes cats.

Dialogue-control: We have GREETING, GOODBYE, OPEN-DIALOGUE, and Q-OPEN-DIALOGUE, CLOSE-DIALOGUE as dialogue-initiating/ending acts. We also have SHIFT-TOPIC, which introduces a new topic (animal) into the dialogue.

Question: We have four questioning acts, Q-DISC-P, Q-DISC-P-OPEN (an open question such as "how about cats?"), Q-DISC-R, and Q-DISC-R-OTHER, that ask for propositions or reasons of the partner.

Acknowledgment: We have ACK, which acknowledges the partner's utterance using back-channels.

3.1.3. Utterance Understanding

User utterances are first separated into word tokens using a Japanese morphological analyzer (n.b., there is no explicit word boundary in Japanese) and are then converted into dialogue acts using an understanding grammar realized as a weighted finite state transducer (WFST) in a manner similar to [7]. We defined sequences of words that form dialogue acts and from them compiled a WFST that maps a sequence of words into a scored list of dialogue acts augmented with attribute-value pairs. For example, "I like fireflies" would be parsed into (DISC-P polarity='+' animal='fireflies') and "I like fireflies because they are evanescent" into (DISC-R polarity='+' animal='fireflies' reason='evanescent'). In all, our grammar has a vocabulary of 2,276 words, including 1,005 adjectives taken from the evaluative expression dictionary [8].

Table 1. Questionnaire results averaged over all dialogues (900 dialogues). Values in parentheses show the mean of the standard deviations within one participant.

	Questionnaire item	Avg.	SD.
Q1	Quality of system utterance	3.73	1.19 (0.65)
Q2	System understanding quality	2.71	1.24 (0.89)
Q3	Smoothness of dialogue	2.70	1.25 (0.83)
Q4	Closeness perceived by user	2.58	1.21 (0.83)
Q5	Closeness shown by system	2.67	1.16 (0.79)
Q6	User satisfaction	2.52	1.21 (0.79)
Q7	Willingness for future use	2.46	1.25 (0.72)

3.1.4. Dialogue Manager

The system uses four flags to keep track of the dialogue: SELF-DISC-P, SELF-DISC-R, PARTNER-DISC-P, and PARTNER-DISC-R. They take either *true* or *false* and keep a record of whether the system has disclosed a proposition about the animal in question, whether the system has disclosed a reason, and likewise for the conversational partner. In addition to these flags, the system maintains the name of the animal being discussed and the lists of animals introduced so far in a dialogue. Reasons expressed by the user are also stored as a list so that they can be referred to later to make agreement/disagreement. Hand-crafted rules process the user's dialogue acts to set the flags, and the dialogue manager decides the next action based on them. For each combination of flags, the system has a predefined list of action sequences and randomly selects one as its next action. This random selection procedure is affected by three 0–1 probabilistic parameters that increase/decrease self-disclosure, agreement, and disagreement, respectively.

3.1.5. Utterance Generation

We prepared possible surface realizations for each dialogue act. The system makes its utterance by randomly selecting one of the realizations. For example, we have "I [don't] like OBJ because it is EXP" as a realization for DISC-P-R. At runtime, OBJ is replaced with the animal name in question, [don't] is added when the system's polarity for the animal is negative, and EXP is replaced with one of the reasons.

3.2. Data Collection

We recruited 50 Japanese adults (25 males and 25 females) for data collection. They were paid for their participation. The duration of each dialogue was limited to four minutes. The participants were notified to end the dialogue when the limit approached. We prepared 18 parameter patterns to make variations in the system's self-disclosure, agreement, and disagreement. Each participant talked to the system with a randomly selected parameter pattern, resulting in 18 dialogues. Including two test dialogues in the beginning, we collected 20 dialogues from each participant for a total of 1,000 (50×20) dialogues.

After each dialogue, the participants filled out a questionnaire (five-point Likert scale) that asked for their subjective evaluation of the dialogue. Table 1 summarizes the results. From the averaged scores, relatively low system performance can be seen. However, what is important is the standard deviation for Q4 (closeness) and Q6 (user satisfaction). The system succeeded in collecting dialogues of varying closeness and user satisfaction, which is necessary for our analysis.

3.3. Data Annotation

We annotated user utterances with correct dialogue acts. A single annotator, who was not one of the authors, annotated each dialogue. Since users often made utterances that were not defined in our dialogue act set, we newly introduced DISC-OTHER and OTHER, which were annotated for self-disclosures not related to the animal domain and for out-of-domain utterances, respectively. The system's dialogue act type recognition accuracy (excluding DISC-OTHER and OTHER) was 50.03%.

4. ANALYSIS

We used 900 dialogues (excluding test dialogues) for our analysis. For each dialogue, we counted the number of dialogue acts in order to derive 19 counts. Based on the dialogue act category, we have the (1) user disclosure count, (2) user agree count, (3) user disagree count, (4) user question count, (5) user acknowledge count. We also have the (6) user other count for counting OTHER. DISC-OTHER is included in (1). We ignore dialogue-controlling acts. We derived (7)–(11) as counterparts of (1)–(5) for system dialogue acts.

Users are likely to prefer systems that show similarity to users [9]. In addition, users may simply like dialogues in which animals that they like appear. To investigate this issue, we have the (12) preference agree count and (13) preference disagree count, corresponding to the number of animals on which the user and the system agreed on preferences. We also have the (14) user preference positive count, (15) user preference negative count, and (16) user preference neutral count, representing the number of animals for which users showed positive, negative, and neutral preferences, respectively. The counterparts for these for the system were also counted (17–19).

4.1. Closeness and User Satisfaction

We calculated the correlation of the 19 counts with closeness (Q4) and user satisfaction (Q6). Tables 2 and 3 show the top-five counts correlated with closeness and user satisfaction, respectively. Here, we used Spearman's rank correlation coefficient for correlations.

It can be seen that the user agree count is relatively more correlated with both closeness and user satisfaction than other counts, leading us to believe that we should focus on increasing this count for better dialogue. An analysis of variance (ANOVA) also revealed that dialogues with zero, one, or more than one user agree count are significantly different in closeness ($F(2,897)=20.1$, $p<0.0001$) and user satisfaction ($F(2,897)=22.0$, $p<0.0001$), and a non-parametric multiple comparison test (Steel-Dwass test) showed that dialogues with more than one user agree count are significantly better in closeness ($p<0.0002$) and user satisfaction ($p<0.002$) than those with zero or just one user agree count.

It is surprising that the user's self-disclosure shows little correlation with either closeness ($R=0.029$, $p=0.39$) or user satisfaction ($R=0.065$, $p=0.052$). This means that the amount of user's self-disclosure does not necessarily mean they are willing to talk; perhaps self-disclosure is costly for users.

4.2. Increasing User Agree Count

We need to increase the user agree count for better dialogue. Although the user agree count can be increased when the system knows what the user is likely to agree, it is difficult to achieve unless we have some prior knowledge of user preferences. Therefore, we investigate system actions that may increase the user agree count. We

Table 2. Top-five counts correlated with closeness. All counts show significant correlations ($p<0.0001$).

		Correlation coefficient
1	User agree count	0.196
2	System agree count	0.146
3	User preference positive count	0.144
4	User other count	-0.127
5	User preference negative count	-0.113

Table 3. Top-five counts correlated with user satisfaction. All counts show significant correlations ($p<0.0001$).

		Correlation coefficient
1	User agree count	0.202
2	System disclosure count	0.153
3	System agree count	0.152
4	User preference positive count	0.121
5	User preference neutral count	0.115

performed a multiple linear regression analysis with the user agree count as the objective variable and the counts related to system actions as explaining variables. By analyzing the regression model, we can obtain the relative importance of system actions in relation to the user agree count. We do not simply list correlation coefficients here because we specifically want to find out the relative importance of system actions.

Table 4 shows the results of the multiple linear regression analysis. Explaining variables that caused multi-collinearity were removed in the analysis. From the table, we can see an obvious result that the system disclosure count has a high standard partial regression coefficient; users cannot agree unless the system discloses something. What we think is most interesting is that the system agree count has a relatively high regression coefficient. As the system was built so that it would not be affected by whether users agree or disagree with it, we can safely say that users tend to agree with systems that agree with users. It seems to be a good strategy for the system to agree with users as much as possible.

4.3. Increasing User Disclosure Count

There are cases where we want to increase the user disclosure count, such as when we want to elicit information from the user. Although it may be possible to perform a multiple linear regression analysis to find system actions that drives users to disclose, we cannot do it because the user's self-disclosure affects the system's choice of actions considerably, which makes the analysis of the causal relationship difficult. Therefore, we performed a multiple linear regression analysis with the user disclosure count as the objective variable and the counts related to preferences as explaining variables.

Table 5 shows the results, with only two explaining variables because others were removed due to multi-collinearity. From the table, we can say that users tend to self-disclose when they talk about topics they like. They also self-disclose a fair amount on things they do not like. It is surprising that the preference agree count does not appear in the results. Considering that the user preference negative count has a negative correlation with closeness (Table 2), the best strategy to increase the user's self-disclosure is to bring up topics of interest to the user as much as possible.

5. SUMMARY

This paper investigated the individual effects of self-disclosure and empathic utterances in human-computer social dialogue. Our analysis of the collected dialogue data revealed that (1) increasing user agreement is the key to achieving better closeness and user satisfaction, (2) the system's agreement is effective for inducing agreement from users, and (3) self-disclosure by users increases when users have positive preferences on topics being discussed. These results provide useful insight into how dialogue systems should behave in social dialogue. As future work, we plan to verify our findings in a more controlled dialogue experiment. Sequences of dialogue acts rather than their number need to be considered for further analysis. We also plan to improve the system's understanding and generation ability as well as its coverage of topics and domains.

Table 4. Results of the multiple linear regression analysis with the user agree count as the objective variable and the counts related to system actions as explaining variables (Adjusted R^2=0.245).

	Regression coefficient
System disclosure count	0.393
System agree count	0.245
System acknowledge count	0.189
system question count	-0.172

Table 5. Results of the multiple linear regression analysis with the user disclosure count as the objective variable and the counts related to user/system preferences as explaining variables (Adjusted R^2=0.253).

	Regression coefficient
User preference positive count	0.496
User preference negative count	0.358

6. REFERENCES

[1] Timothy W. Bickmore and Justine Cassell, "Relational agents: a model and implementation of building user trust," in *Proc. CHI*, 2001, pp. 396–403.

[2] Timothy W. Bickmore and Rosalind W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 12, no. 2, pp. 293–327, 2005.

[3] Irwin Altman and Dalmas A. Taylor, *Social penetration: The development of interpersonal relationships*, Holt, Rinehart & Winston, 1973.

[4] Harry T. Reis and Phillip Shaver, "Intimacy as an interpersonal process," in *Handbook of personal relationships*, S. Duck, Ed., pp. 367–398. John Wiley & Sons Ltd., 1998.

[5] Youngme Moon, "Intimate exchanges: Using computers to elicit self-disclosure from consumers," *The Journal of Consumer Research*, vol. 26, no. 4, pp. 323–339, 2000.

[6] Jun Okamoto and Shun Ishizaki, "Associative concept dictionary construction and its comparison with electronic concept dictionary," in *Proc. PACLING*, 2001, pp. 214–220.

[7] Alexandros Potamianos and Hong-Kwang J. Kuo, "Statistical recursive finite state machine parsing for speech understanding," in *Proc. ICSLP*, 2000, vol. 3, pp. 510–513.

[8] Hiroya Takamura, Takashi Inui, and Manabu Okumura, "Extracting semantic orientations of words using spin model," in *Proc. 43rd ACL*, 2005, pp. 133–140.

[9] Robert B. Cialdini, *Influence: Science and Practice*, Allyn & Bacon, 2000.