Construction of a Japanese Dialog Corpus Annotated with Speakers' Intimacy

Takuto Miura, Kiyoaki Shirai, Hideaki Kanai, Natthawut Kertkeidkachorn

Proceedings of the 38th Pacific Asia Conference on
Language, Information and Computation (PACLIC 38)

Nathaniel Oco, Shirley N. Dita, Ariane Macalinga Borlongan, Jong-Bok Kim (eds.)

2024

Construction of a Japanese Dialog Corpus Annotated with Speakers' Intimacy

Takuto Miura[1], Kiyoaki Shirai[1], Hideaki Kanai[1], Natthawut Kertkeidkachorn[1]
[1] Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan
{s2460005, kshirai, hideaki, natt}@jaist.ac.jp

Abstract

In recent years, several studies have been devoted to the estimation of a speaker's intimacy with his/her partner in a dialog. This is because intimacy is considered to be one of the key factors in the development of a friendly dialog system. To train a model to guess the level of the speaker's intimacy, a labeled dialog corpus is required. Since manual annotation of intimacy labels is very costly, however, the number of such dialog corpora in Japanese is rather limited. This study aims to construct a Japanese dialog corpus annotated with a speaker's level of intimacy as well as other information, i.e., the depth of self-disclosure and the speaker's personality. The corpus compiles transcriptions of approximately 7,000 utterances from 18 dialog sessions. Each dialog session consists of three short dialogs by two speakers, where the labels of the level of the intimacy and the depth of the self-disclosure are attached at the beginning, interval, and end of continuous dialogs. It enables us to observe changes in the level of the intimacy and the depth of the self-disclosure during the course of the dialog. Furthermore, the constructed corpus was utilized to verify the correlation between the speaker's intimacy and self-disclosure/personality. As a result, a significant correlation between the level of the intimacy and the depth of the self-disclosure is found. We also analyze the relationship between the speaker's level of the intimacy and the use of polite and casual speech styles. It is found that speakers tend to utilize a polite style when the level of intimacy is low and a casual style when it is high.

1 Introduction

A dialog system that can carry out a free conversation with a user has received a great deal of attention (Khatri et al., 2018; Higashinaka et al., 2021; Dinan et al., 2020). These systems are expected to build long-term friendship with users by conversing comfortably with them (Ram et al., 2018).

In human conversation, control of a style, which is the human's behavior to change a speech style according to the intimacy with a partner and/or social relationship, is often observed to communicate with others smoothly (Wardhaugh and Fuller, 2021; Hovy, 1987; Silverstein, 2003). The use of polite and casual expressions is an example of the control of a style (Aapakallio, 2021; Liu and Kobayashi, 2022). Casual expressions are often used when a speaker is friendly with his/her partner, while polite expressions are used when a speaker is not intimate with a partner. The styles are also different due to social relationships such as the relationship between a boss and his/her staff and that between a wife and her husband. The control of a style should be considered not only in human-to-human dialogs but also in conversations between a dialog system and a user (Kageyama et al., 2018). Our final goal is to develop a dialog system that can control a speech style appropriately. Although there are various factors to be considered to achieve control of a style, this study focuses on the level of the user's intimacy. Our desired dialog system identifies the user's level of the intimacy with the dialog system during a conversation, then generates responses with polite expressions when the user's intimacy is low and responses with casual expressions when the intimacy is high.

A common method to identify the level of the intimacy for a given content of a dialog is supervised learning, which requires a dialog corpus annotated with intimacy labels. However, such corpora in Japanese have not been well developed. The goal of this paper is to construct a corpus of free conversation between humans annotated with the intimacy they feel toward their partners. The questionnaire is administered not only before the dialog but also in the middle of and after the dialog to annotate the corpus with labels of speaker's intimacy. In addition, we also annotate the corpus with the depth of self-disclosure and personality as information

about the speaker. These are supposed to be related to the speaker's intimacy, so the correlation between intimacy and self-disclosure/personality is empirically investigated in this paper.

Furthermore, we analyze the relationship between the speaker's intimacy and the style. Specifically, we suppose that speakers use a casual style when the intimacy is high and a polite style when the intimacy is low. This assumption is then subjected to verification.

The contributions of this paper are summarized as follows.

- We construct a corpus of free dialog in Japanese annotated with the level of the intimacy. In addition to the intimacy, the corpus also includes the information of the depth of self-disclosure and the personality of the speaker.
- We analyze the correlation between the speaker's intimacy and the other two annotations (the depth of the self-disclosure and the personality) using the constructed corpus.
- We analyze correlation between the speaker's intimacy and the style.

2 Related Work

2.1 Mentally Annotated Dialog Corpus

Several dialog corpora have been created to develop a dialog system that takes the relation between the user and the system into account. Rashkin et al. (2019) constructed a dialog corpus containing many sympathetic utterances by recording dialog in a situation where two speakers tend to show their sympathy to others, aiming to construct a dialog system that can generate sympathetic responses. Following their method, a similar dialog corpus in Japanese was constructed by Sugiyama et al. (2023). Specifically, they translated Rashkin's instructions into Japanese to encourage the participants to show their sympathy. Komatani and Okada (2021) constructed a dialog corpus containing conversations between a human and a dialog system using the Wizard-of-Oz method, where the dialog system was actually impersonated by another human, aiming to construct a dialog system that can control the contents of dialog according to the user's impression of the system. In their corpus, each dialog was annotated with the users' impression, such as "How well can you converse with the dialog system?"

Similar to our study, there have been a few attempts to construct a Japanese dialog corpus annotated with the intimacy of a speaker. Yamazaki et al. (2020) constructed a multimodal corpus of Japanese free conversation. The participant was asked to answer the questionnaire to show how strongly they feel the intimacy with their dialog partner, then the obtained the degree of the intimacy was added to the corpus. In addition, each utterance was labeled with its dialog act. However, this corpus is publicly unavailable.

This paper also constructs a Japanese dialog corpus annotated with the level of the speakers' intimacy. In addition, the depth of self-disclosure and personality, which are considered to be highly related to the intimacy, are added as the information about the speaker.

2.2 Intimacy Estimation

Chiba et al. (2021) trained a multimodal model that identifies the speaker's intimacy using a text (transcriptions of utterances), speech (prosody), and video (Action Units of speakers during utterances) as inputs. However, the task is designed as a binary classification, where the two classes are "high" (speakers are known to each other) and "low" (speakers are strangers), and the classification is limited to this coarse level.

Pei and Jurgens (2020) implemented an intimacy estimation model using a pre-trained language model, and analyzed questions in social media, books, and films using this model. They showed that the pragmatic choices in the questions vary according to the degree of the intimacy, and that the intimacy can be modified by social norms such as gender, social distance, and anonymity. The intimacy label in their dataset is objective, i.e., it is determined by the annotator's estimation of the writer's level of the intimacy. On the other hand, this study focuses on subjective intimacy, where the intimacy label is assigned by the speaker.

2.3 Analysis of Style

The nature of styles as they appear in text has been examined in several studies. Warriner et al. (2013) analyzed the correlation between lexical features of texts and emotions. Chhaya et al. (2018) investigated the correlations between formal attitudes, frustration, and politeness in 960 emails. Dankers et al. (2019) and Mohammad et al. (2016) studied the interaction between figurative expressions and emotions in texts. Brooke and Hirst (2013) con-

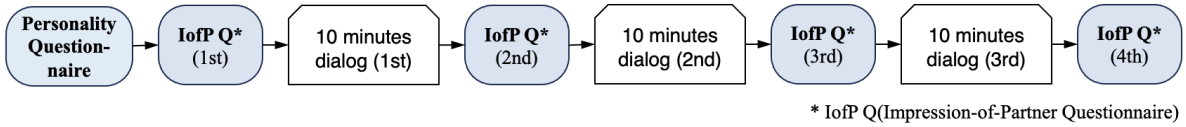

Figure 1: Flow of dialog session

ducted a topic analysis of six perspectives on texts of various genres: literary, abstract, objective, colloquial, concrete, and subjective. Liu and Kobayashi (2022) constructed a corpus of Japanese honorifics and analyzed the characteristics of Japanese honorific sentences.

These studies have not analyzed the interrelationship between the style and speakers' inherent characteristics such as intimacy. In this study, we analyze the relationship between the use of the polite or casual style in Japanese and the speaker's intimacy.

3 Intimacy Annotated Japanese Corpus of Free Conversation

This section describes the construction of the corpus of Japanese free conversation annotated with the level of the speaker's intimacy.

3.1 Dialog Session

We designed schemata for recording and annotating the dialog corpus following Yamazaki et al. (2020). The flow of the recording of the dialogs is shown in Figure 1. Two subjects are asked to freely chat about any topics. The subjects have 10 minutes of dialog, three times. In addition, "Impression-of-Partner Questionnaire" is administered before each of three dialogs and after the last dialog: in it, the subjects are asked about their impressions of the partner. Another questionnaire called "Personality Questionnaire" to reveal the personality of the subjects is also administered before starting the conversation. Hereafter, we call a series of the above procedures "dialog session."

Nineteen Japanese students, 16 males and 3 females, participated in this recording of dialogs. The two people conducting a dialog session were randomly combined from among these subjects. One pair of the subjects performed a dialog session only once, but one subject participated in several sessions with different partners.

3.2 Recording and Transcription of Dialog

The subjects engaged in a free chat on the online conference system Webex.[1] The video and audio of the dialog were recorded using Webex's recording function. The first author, who is a native Japanese speaker, transcribed the utterances from the recorded audio-visual data. The policy of the transcription is as follows.

- Insert a period at the obvious end of a sentence.
- Insert a question mark "?" instead of a period at the end of an interrogative sentence.
- Put a comma at a pause or breather.
- Errors and self-corrections are included in the script as they are. However, they are omitted when they cannot be heard.

After the transcription, the dialogs were divided into utterances by a period and a question mark. Then, a speaker ID was assigned to each utterance. Figure 2 shows an example of the recorded dialog with English translation in parentheses, where "sub02" and "sub09" are speaker IDs.

3.3 Impression-of-Partner Questionnaire

The Impression-of-Partner Questionnaire was administered four times per dialog session. The subjects answered the same two questions all four times:

Q1 How deeply do you feel intimacy with your partner at this moment?

Q2 How much do you disclose yourself to your partner at this moment?

In Q1, the subjects evaluated the level of the intimacy on a five-point Likert scale (Likert, 1932) based on the following criteria:

- To what extent do you feel your partner is your close friend?

[1] https://www.webex.com/

sub09	何を話しますか? (What shall we talk about?)
sub02	先ほどサークルって言ってましたけど、何サークルに入ってるんですか? (You mentioned a club earlier. What club do you belong to?)
sub09	自分、フットサルサークルに入ってて、同じフットサルのはい、仲間ですね。 (I'm in a futsal club, so we are friends in the sense that we both play futsal.)
sub02	覚えてるかわかんないですけど僕もたまにフットサルの方、行っていて。 (I don't know if you remember, but sometimes I also go to the same futsal club.)
sub09	あっ、だからか、何かどこかで見たことあるような。 (Ah, so I feel like I have met you somewhere.)
sub02	そういう感じですね。 (That's right.)

Figure 2: Example of recorded dialogs

- To what extent do you trust your partner and open your mind?
- To what extent are you frank and comfortable with your partner?

These criteria were proposed by the research that investigated the scale of intimacy in social relationships (Kawano et al., 2017; Sinclair and Dowdy, 2005).

In Q2, the subjects evaluated the depth of their self-disclosure to their dialog partner, which refers to how deeply a person conveyed information about himself/herself to another person. Niwa and Maruno (2010) proposed a scale of the depth of the self-disclosure by four types of information that a person discloses to others: (1) superficial information about oneself, (2) one's past experiences, (3) one's faults and weaknesses that are not crucial, and (4) one's negative characteristics, lack of ability, and crucial weaknesses. We showed these criteria to the subjects and asked them to rank the depth of their self-closure on a five-point Likert scale.

Administering Impression-of-Partner Questionnaire four times in a dialog session enables us to analyze how the level of intimacy and the depth of self-disclosure change through a dialog.

3.4 Personality Questionnaire

The Personality Questionnaire was administered, only once, at the beginning of the dialog session. We measured the strength of the Big 5 factors (extraversion, cooperativeness, diligence, neuroticism, and openness) (Costa and McCrae, 1992). The Japanese version (TIPI-J) of the Ten Item Personality Inventory (TIPI) (Oshio et al., 2012), a Japanese translation of the TIPI (Gosling et al., 2003), was used to measure the Big 5 factors. TIPI-J consists of ten questions; each of the two questions corresponds to one of five factors. The strength of each of the Big 5 factors was measured by asking subjects to answer those questions on a 7-point Likert scale. The final strength of each factor is determined by averaging the answers to two questions, resulting in a value between 1 and 7 with a step of 0.5 (e.g., 3.5).

3.5 Summary of Constructed Dialog Corpus

The dialog corpus consists of multiple dialog sessions. Each dialog session consists of two speaker IDs, transcriptions of three dialogs, eight intimacy labels (two speakers $\times$ four times), eight self-disclosure labels (two speakers $\times$ four times), and ten personality scores (two speakers $\times$ the five factors of the Big 5). Each transcription of a dialog contains segmented utterances with the speaker IDs. We call the constructed dialog corpus the "Japanese Intimacy Dialog Corpus" or the "JID corpus" for short. Table 1 shows its statistics.

Table 1: Statistics of Japanese Intimacy Dialog Corpus

Subject	Dialog Session	Dialog	Utterance
19	18	54	6,984

Table 2: Distribution of intimacy labels

Intimacy Label	1	2	3	4	5
Numbers	24	18	31	24	11

Table 3: Distribution of self-disclosure labels

Self-disclosure Label	1	2	3	4	5
Numbers	19	30	28	24	7

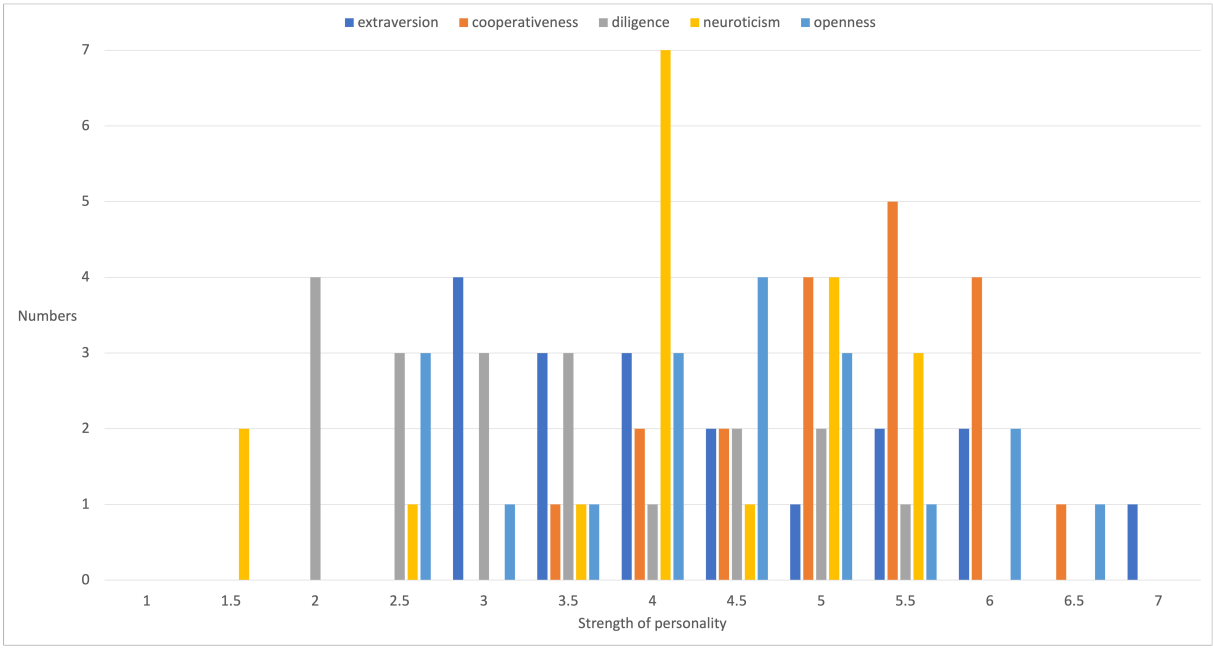

Figure 3: Distribution of strength of personality

Table 4: Results of Intimacy Estimation

window size	Number of data		Precision		Recall		F1-score	
	training	test	BERT	MFC	BERT	MFC	BERT	MFC
2	5,336	1,540	0.29	0.10	0.22	0.20	0.22	0.15
4	5,168	1,492	0.28	0.10	0.24	0.20	0.24	0.15
6	5,000	1,444	0.30	0.09	0.26	0.20	0.26	0.14
8	4,832	1,396	0.29	0.09	0.23	0.20	0.22	0.14
10	4,664	1,348	0.35	0.09	0.27	0.20	0.28	0.14
12	4,496	1,300	0.28	0.09	0.26	0.20	0.24	0.14

Table 2 shows the distribution of the intimacy labels. Each label has more than ten samples, thus the dataset is relatively balanced. Table 3 shows the distribution of the self-disclosure labels. Figure 3 shows the distribution of the strength of the Big 5 factors.

3.6 Intimacy Estimation

The constructed JID corpus is used to train a baseline model for identifying the level of the intimacy that a speaker has with his/her partner. Here, the level of the speaker's intimacy is estimated for a given sequence of utterances of that speaker. A data sample for this task is a set of n consecutive utterances of the same speaker (called "window"). Its ground-truth label is the speaker's answer in the Impression-of-Partner Questionnaire before the dialog. Multiple samples are extracted by repeatedly sliding the window forward by one utterance in a dialog. The parameter n is set to 2, 4, 6, 8, 10 or 12.

We fine-tune the pre-trained Japanese BERT[2] (Devlin et al., 2019). As for the hyperparameters, a batch size is set to 4, the number of epochs to 10, and a learning rate to $5e^{-6}$. The Adam optimizer is used for training.

Table 4 shows the number of samples of training and test data used in the experiment, as well as the macro-averages of the precision, the recall, and the F1-score for the intimacy estimation. MFC (Most Frequent Class) represents the method that classifies all of the data into the most frequent intimacy class. The model demonstrated the best performance in terms of three criteria when $n = 10$. However, a definite correlation between the number of utterances n and the performance of the intimacy estimation was not found. In addition, the F1-score was not particularly high, indicating that

[2]https://huggingface.co/tohoku-nlp/bert-base-japanese-v2

the intimacy estimation is a challenging task.

4 Analysis of Correlation between Intimacy and Personal characteristics

This section outlines two types of correlation analysis, which are employed to investigate the relationship between intimacy and personal characteristics that have been annotated to the JID corpus. One type of analysis examines the relationship between the intimacy and the depth of self-disclosure (subsection 4.1), the other examines the relationship between intimacy and personality (subsection 4.2).

4.1 Intimacy and Self-disclosure

It is known that there is a strong positive correlation between intimacy and self-disclosure (Laurenceau et al., 1998; Agustin and Ilyas, 2019; Hasbiyah et al., 2023; Muñoz, 2022). That is, the greater the level of the intimacy experienced by a speaker, the more personal information is conveyed to the partner.

To verify this assumption, we measured the Pearson correlation coefficient between the intimacy label and the self-disclosure label annotated in our constructed dialog corpus. Recall that both labels are an integer on a 5-point Likert scale. The Pearson correlation coefficient was 0.725, which was considerably high. Its p-value was 2.55e^{-22}, indicating that the correlation is statistically significant. Therefore, it can be concluded that there is a positive interrelationship between the level of the intimacy and the depth of the self-disclosure.

4.2 Intimacy and Personality

Several studies in the field of psychology have reported that the personality of the speaker and/or that of the partner can influence the case with which people establish intimate relationships and the strength of their feelings of intimacy with their partner (Sprecher and Cate, 2004; Karney and Bradbury, 1995; Collins and Read, 1990). Based on this background, we verified the hypothesis that the personality could be one of the clues to predict a change in the level of the intimacy through dialog. In this study, the correlation between the personality and the change in the level of the intimacy in conversation between two strangers was investigated. To this end, we analyzed only subjects who answered the lowest level of the intimacy (score of 1) in Impression-of-Partner Questionnaire prior to the dialog session.

The change in the level of the intimacy was measured by the difference between the intimacy labels obtained by the first Impression-of-Partner and that obtained by the last questionnaire in a dialog session (we call it "intimacy change" hereafter). Besides, the personality of a subject was represented by the personality scores in our corpus, which were the ratings of the Big 5 factors obtained by the Personality Questionnaire. Two kinds of correlation analyses were performed. The first analysis aimed to investigate the correlation between the intimacy change of a speaker and his/her own personality. This was achieved by measuring the Pearson correlation coefficient between the intimacy change of a speaker and his/her own personality score. The second aimed to verify the correlation between the intimacy change of a speaker and his/her partner's personality. This was accomplished by measuring the Pearson correlation coefficient between the intimacy change of the speaker A (or B) and the personality score of the speaker B (or A). These correlation analyses were performed for each of the Big 5 factors.

Table 5 shows the results of the first and second correlation analyses, respectively. The results indicate that the personality of either oneself or one's partner does not significantly correlate with the change in the intimacy level. One exception is that "cooperativeness" factor of the partner exhibits a weak positive correlation with the intimacy change. The cooperative person tends to display thoughtfulness and dedication to others, which may lead to an increase in intimacy with such a partner through conversation. This result indicates the potential for developing a dialog system in which a user can experience friendliness and intimacy by generating responses cooperatively.

5 Analysis of Correlation between Intimacy and Style

This section examines the relationship between the speaker's intimacy and the style. Two styles are considered: polite and casual. We build up a hypothesis that speakers use the polite style when their level of the intimacy is low and use the casual style when they are intimate with a partner. The proportion of utterances in the polite style, P_{po}, and the proportion of utterances in the casual style, P_{ca}, are calculated for each set of utterances annotated with the intimacy label i ($i = 1, 2, 3, 4, 5$) using the JID corpus. The hypothesis is then tested by

Table 5: Pearson correlation coefficient and p-value between intimacy change and personality. ($^{+}p < 0.1$)

Big 5	Self-personality		Partner's personality	
	coefficient	p-value	coefficient	p-value
extraversion	−0.178	0.440	+0.304	0.180
cooperativeness	−0.076	0.742	+0.419	0.059^{+}
diligence	−0.060	0.795	+0.208	0.367
neuroticism	+0.080	0.729	−0.334	0.139
openness	+0.104	0.654	−0.274	0.230

verifying whether P_{po} is low and P_{ca} is high when the level of the intimacy is high, and vice versa.

To obtain P_{po} and P_{ca}, it is necessary to identify the style of each utterance. Two distinct methods are utilized to achieve this objective. The first employs style-specific words, while the second is based on a style classifier. The succeeding subsections describe the analyses based on these two methods.

5.1 Analysis by Style-specific Words

Here, the term "style-specific word" is defined as a word that is frequently used in the polite or casual style. The style-specific words are obtained by the following procedure. Let C_{po} and C_{ca} be corpora that consist of sentences written in a polite and casual style, respectively. The KeiCO corpus (Liu and Kobayashi, 2022) is used as C_{po}, while the set of dialogs between acquaintances in the BTSJ corpus (Usami, 2021) is used as C_{ca}. The number of sentences in C_{po} and C_{ca} are 10,007 and 13,351, respectively. Next, the sets of words specific to the polite and causal styles, denoted as W_{po} and W_{ca}, are extracted as follows.

$$W_{po} = \{w \mid w \in C_{po} \wedge w \notin C_{ca} \wedge R_{\text{TF-IDF}}(w) \leq 50\} \quad (1)$$

$$W_{ca} = \{w \mid w \notin C_{po} \wedge w \in C_{ca} \wedge R_{\text{TF-IDF}}(w) \leq 50\} \quad (2)$$

That is, the set of the top 50 words with the highest TF-IDF, which appear only in C_{po} (or C_{ca}), is defined as W_{po} (or W_{ca}). It should be noted that $R_{\text{TF-IDF}}$ is the rank of the TF-IDF of the word w, assuming that the entire C_{po} and C_{ca} are two virtual documents.

When a word in W_{po} or W_{ca} appears in an utterance, the utterance is assumed to be in a polite style or casual style.[3] Then, P_{po} and P_{ca} are calculated for each subset of utterances that have been annotated with different levels of the intimacy. The results are shown in Table 6.

Table 6: Results of Analysis by Style-specific Words (* $p < 0.05$, ** $p < 0.01$)

Intimacy	P_{po}	p	P_{ca}	p
1	0.066	—	0.367	$2e^{-5}$ **
2	0.057	0.312	0.422	0.001 **
3	0.044	0.046 *	0.445	0.039 *
4	0.039	0.001 **	0.470	0.099
5	0.032	0.005 **	0.550	—

When the level of the intimacy is high, P_{po} tends to be small and P_{ca} tends to be large. Thus, it can be argued that the speaker selects a casual style when he/she perceives a sense of closeness with the partner, and a more polite style when the intimacy level is lower. The Welch's t-test is used to verify whether there is a statistically significant difference in P_{po} between the lowest intimacy level (1) and the other levels, and P_{ca} between the highest intimacy level (5) and the others. The p-values are shown in the "p" column of Table 6. The notable differences are found in both P_{po} and P_{ca}.

5.2 Analysis by Style Classifier

First, the classifier that determines whether the style of utterance is polite or casual is trained using C_{po} and C_{ca} as training data. GPT-2 (Radford et al., 2019) is chosen as the classification model. The GPT-2 model[4], which has been pre-trained on a Japanese dialog dataset, is then fine-tuned using 9,957 polite utterances in C_{po} and 13,301 casual utterances in C_{ca} (23,248 in total). The batch size is set to 4, the training epoch to 20, and the learning rate to $5e^{-6}$. The Adam optimizer is used for the fine-tuning of GPT-2. The performance of the trained model is evaluated using test data consisting of 50 utterances in C_{po} and 50 utterances in

[3] When both words in W_{po} and W_{ca} occur in an utterance, its style is classified as "unknown".

[4] https://huggingface.co/rinna/japanese-gpt2-medium

C_{ca}, which are mutually exclusive from the training data. The accuracy of the style classification is 64%.

The style of each utterance in the JID corpus is identified by the trained style classifier, then P_{po} and P_{ca} are calculated for each group of utterances with different levels of the intimacy. Table 7 shows P_{po} and P_{ca} as well as p-values of Welch's t-test.

Table 7: Results of Analysis by Style Classifier

Intimacy	P_{po}	p	P_{ca}	p
1	0.930	—	0.070	0.818
2	0.924	0.60	0.078	0.755
3	0.941	0.30	0.059	0.319
4	0.932	0.88	0.068	0.704
5	0.926	0.75	0.073	—

No clear correlation is observed between the level of the intimacy and the style of utterances. Furthermore, no significant difference is identified by Welch's t-test. One possible reason may be that utterances are not precisely labeled with the style tags due to the relatively low performance (64% accuracy) of the style classifier.

An additional analysis is carried out by using only reliable utterances. Specifically, the style of an utterance is determined only when the prediction probability of the model is 0.7 or higher. The performance of the style classifier is sufficiently high for these reliable cases. The accuracy is 89%, and the precision for the polite and casual classes is 100% and 67%, respectively. However, the style of only 7% (478/6984) of all utterances can be identified. Since the number of utterances available for analysis is small, we introduce three coarse-grained intimacy class: Not-intimate (intimacy label of 1), Low-intimacy (2 or 3), and High-intimacy (4 or 5). Then P_{po} and P_{ca} are measured for each of the three intimacy classes. The results are shown in Table 8.

Table 8: Results of Analysis Using Reliable Utterances (* $p < 0.05$)

Intimacy	P_{po}	p	P_{ca}	p
Not	0.657	—	0.343	0.069
Low	0.580	0.342	0.388	0.217
High	0.520	0.028 *	0.467	—

A similar tendency is found in Table 6 and 8, i.e., the polite style is used more often when the level of the intimacy is low and the casual style is more preferred when the level of the intimacy is high. This supports our hypothesis. As for Welch's t-test, only the difference of P_{po} between the Not-intimate and High-intimacy is statistically significant.

6 Conclusion

In this study, we constructed the Japanese dialog corpus that compiled the transcription of about 7,000 utterances from 54 dialogs. The corpus was annotated with some information about the speakers: the level of the intimacy with the partner, the depth of the self-disclosure, and the personality. The intimacy and self-disclosure labels were given four times per dialog session, which enabled us to observe their change over the course of a dialog. Furthermore, using the constructed corpus, we examined the correlation between the speaker's intimacy and self-disclosure/personality and found the significant correlation between the level of the intimacy and the depth of the self-disclosure. We also investigated the relationship between the speaker's intimacy and the use of polite and casual styles. The results indicated that speakers tended to use the polite style when the level of the intimacy was low and the casual style when it was high.

In the future, we will develop a response generation model that can adapt the polite and casual style according to the user's level of intimacy with the dialog system. This will allow a dialogue system to achieve human-like control of a style. In the development of such a response generation model, it is essential that the performance of the intimacy estimation is sufficiently high. The findings of this study indicate that there is a significant potential for enhancing the accuracy of intimacy estimation. Therefore, we will investigate methods to improve the performance of the intimacy estimation model.

References

Noora Aapakallio. 2021. *Understanding Through Politeness – Translations of Japanese Honorific Speech to Finnish and English*. University of Eastern Finland.

Anggia Wahyu Agustin and Asmidir Ilyas. 2019. Relationship intimacy and self disclosure young married couple. *Jurnal Neo Konseling*.

Julian Brooke and Graeme Hirst. 2013. A multi-dimensional Bayesian approach to lexical style. In *Proceedings of the 2013 Conference of the North*

American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 673–679, Atlanta, Georgia. Association for Computational Linguistics.

Niyati Chhaya, Kushal Chawla, Tanya Goyal, Projjal Chanda, and Jaya Singh. 2018. Frustrated, polite, or formal: Quantifying feelings and tone in email. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 76–86, New Orleans, Louisiana, USA. Association for Computational Linguistics.

Yuya Chiba, Yoshihiro Yamazaki, and Akinori Ito. 2021. Speaker intimacy in chat-talks: Analysis and recognition based on verbal and non-verbal information. In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue*, pages 1–10, Potsdam, Germany. SEMDIAL.

Nancy L Collins and Stephen J Read. 1990. Adult attachment, working models, and relationship quality in dating couples. *Journal of personality and social psychology*, 58(4):644.

Paul T Costa and Robert R McCrae. 1992. *Neo personality inventory-revised (NEO PI-R)*. Psychological Assessment Resources Odessa, FL.

Verna Dankers, Marek Rei, Martha Lewis, and Ekaterina Shutova. 2019. Modelling the interplay of metaphor and emotion through multitask learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2218–2229, Hong Kong, China. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Emily Dinan, Varvara Logacheva, Valentin Malykh, Alexander Miller, Kurt Shuster, Jack Urbanek, Douwe Kiela, Arthur Szlam, Iulian Serban, Ryan Lowe, et al. 2020. The second conversational intelligence challenge (convai2). In *The NeurIPS'18 Competition: From Machine Learning to Intelligent Conversations*, pages 187–208. Springer.

Samuel Gosling, Peter Rentfrow, and William Swann. 2003. A Very Brief Measure of the Big-Five Personality Domains. *Journal of Research in Personality*, 37:504–528.

Desi Hasbiyah, Mirza Ronda, and Fahruddin Faiz. 2023. Intimate relationship of elderly hajj pilgrimages and clotter officers in the aspect of religiosity through the process of self disclosure during the hajj. *International Journal of Environmental, Sustainability, and Social Science*.

Ryuichiro Higashinaka, Kotaro Funakoshi, Michimasa Inaba, Yuiko Tsunomori, Tetsuro Takahashi, and Reina Akama. 2021. *Dialogue System Live Competition: Identifying Problems with Dialogue Systems Through Live Event*, pages 185–199. Springer Singapore.

Eduard Hovy. 1987. Generating natural language under pragmatic constraints. *Journal of Pragmatics*, 11(6):689–719.

Yukiko Kageyama, Yuya Chiba, Takashi Nose, and Akinori Ito. 2018. Improving User Impression in Spoken Dialog System with Gradual Speech Form Control. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 235–240, Melbourne, Australia. Association for Computational Linguistics.

Benjamin Karney and Thomas Bradbury. 1995. The Longitudinal Course of Marital Quality and Stability: A Review of Theory, Method, and Research. *Psychological Bulletin*, 118:3.

Minoru Kawano, Ikuya Murata, Shigeki Ahama, and Motohiro Hasegawa. 2017. Development of Scale of Intimacy in Social Network (in Japanese). *JSiSE (Japanese Society for Information and Systems in Education) Research Report*, 31:159–166.

Chandra Khatri, Anu Venkatesh, Behnam Hedayatnia, Raefer Gabriel, Ashwin Ram, and Rohit Prasad. 2018. Alexa Prize — State of the Art in Conversational AI. *AI Magazine*, 39(3):40–55.

Kazunori Komatani and Shogo Okada. 2021. Multimodal Human-Agent Dialogue Corpus with Annotations at Utterance and Dialogue Levels. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8.

Jean-Philippe Laurenceau, Lisa Barrett, and Paula Pietromonaco. 1998. Intimacy as an Interpersonal Process: the Importance of Self-Disclosure, Partner Disclosure, and Perceived Partner Responsiveness in Interpersonal Exchanges. *Journal of personality and social psychology*, 74:1238–1251.

Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of psychology*, 140(22).

Muxuan Liu and Ichiro Kobayashi. 2022. Construction and validation of a Japanese honorific corpus based on systemic functional linguistics. In *Proceedings of the Workshop on Dataset Creation for Lower-Resourced Languages within the 13th Language Resources and Evaluation Conference*, pages 19–26, Marseille, France. European Language Resources Association.

Saif Mohammad, Ekaterina Shutova, and Peter Turney. 2016. Metaphor as a medium for emotion: An empirical study. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 23–33, Berlin, Germany. Association for Computational Linguistics.

Kyrie Eleison Muñoz. 2022. Predicting travel intentions using self-disclosure, trust and intimacy: the case of tinder users during covid-19. *Journal of Tourism Futures*.

Sora Niwa and Shun'ichi Maruno. 2010. Development of a Scale to Assess the Depth of Self-disclosure (in Japanese). *The Japanese Journal of Personality*, 18(3):196–209.

Atsushi Oshio, Shingo Abe, and Pino Cutrone. 2012. Development, Reliability, and Validity of the Japanese Version of Ten Item Personality Inventory (TIPI-J) (in Japanese). *The Japanese Journal of Personality*, 21(1):40–52.

Jiaxin Pei and David Jurgens. 2020. Quantifying intimacy in language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5307–5326, Online. Association for Computational Linguistics.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Ashwin Ram, Rohit Prasad, Chandra Khatri, Anu Venkatesh, Raefer Gabriel, Qing Liu, Jeff Nunn, Behnam Hedayatnia, Ming Cheng, Ashish Nagar, et al. 2018. Conversational AI: The science behind the alexa prize. *arXiv preprint arXiv:1801.03604*.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy. Association for Computational Linguistics.

Michael Silverstein. 2003. Indexical order and the dialectics of social life. *Language & Communication*, 23:193–229.

Vaughn Sinclair and Sharon Dowdy. 2005. Development and Validation of the Emotional Intimacy Scale. *Journal of Nursing Measurement*, 13:193–206.

S. Sprecher and R. M. Cate. 2004. Intimacy and love in close relationships. In Handbook of closeness and intimacy. *In Handbook of closeness and intimacy*, pages 163–188.

Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, Hiromi Narimatsu, Yuya Chiba, Hideharu Nakajima, and Toyomi Meguro. 2023. Empirical analysis of training strategies of transformer-based Japanese chit-chat systems. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, pages 685–691. IEEE.

Mayumi Usami, editor. 2021. *BTSJ-Japanese Natural Conversation Corpus with Transcripts and Recordings (March 2021)*. National Institute for Japanese Language and Linguistics, Japan.

Ronald Wardhaugh and Janet M Fuller. 2021. *An introduction to sociolinguistics*. John Wiley & Sons.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *BEHAVIOR RESEARCH METHODS*, 45(4):1191–1207.

Yoshihiro Yamazaki, Yuya Chiba, Takashi Nose, and Akinori Ito. 2020. Construction and Analysis of a Multimodal Chat-talk Corpus for Dialog Systems Considering Interpersonal Closeness. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 443–448, Marseille, France. European Language Resources Association.