RealPersonaChat: A Realistic Persona Chat Corpus with Interlocutors' Own Personalities

Sanae Yamashita[1] **Koji Inoue**[2] **Ao Guo**[1] **Shota Mochizuki**[1]
Tatsuya Kawahara[2] **Ryuichiro Higashinaka**[1]

[1] Graduate School of Informatics, Nagoya University
{yamashita.sanae.w7@s.mail, guo.ao.i6@f.mail,
mochizuki.shota.k8@s.mail, higashinaka@i}.nagoya-u.ac.jp

[2] Graduate School of Informatics, Kyoto University
inoue.koji.3x@kyoto-u.ac.jp kawahara@i.kyoto-u.ac.jp

Abstract

Personality is an important element when designing dialogue systems that sound human-like. The PersonaChat corpus (Zhang et al., 2018), in which interlocutors interact based on given personas, was proposed for this purpose, but since the interactions are not based on the interlocutors' own personas, the dialogue tends to be unnatural, potentially leading to dialogue models making unnatural utterances. In this study, we constructed the RealPersonaChat (RPC) corpus by collecting the actual personality traits and personas of interlocutors and having them freely engage in dialogue. This corpus contains 14K dialogues in Japanese and is currently the world's largest corpus of dialogue data that includes personas and personality traits. We compared our corpus with an existing one to clarify the features of RPC and found that the frequency of persona information in RPC utterances is significantly lower than that in the existing corpus. Moreover, we found that our corpus contains dialogues with high dialogue satisfaction. Additionally, by using RPC, we successfully extracted expressions related to high and low scores of personality traits.

1 Introduction

In recent years, sophisticated dialogue systems based on neural networks (Roller et al., 2021; Shuster et al., 2022), including large language models like GPT (Brown et al., 2020; OpenAI, 2023), have been extensively researched.

When designing an engaging and human-like dialogue system, personality plays an important role (Mairesse and Walker, 2007; Neff et al., 2010; Zheng et al., 2019). To realize human-like systems, the PersonaChat corpus (Zhang et al., 2018), where interlocutors engage in conversations based on given personas, has been created and is widely used. However, this corpus suffers from the limitation that the interlocutors perform dialogue on the basis of given personas, not on their own, potentially leading to unnatural dialogues. When training dialogue models on such data, they may not express truly human-like personality and behavior.

In this study, we constructed the RealPersonaChat (RPC) corpus consisting of 14K Japanese dialogues by collecting actual personas and personality traits from interlocutors and having them engage in free chit-chat conversations. In addition to clarifying the basic statistics of our corpus, we compared the characteristics of the RPC corpus with those of an existing corpus, JPersonaChat (JPC; the Japanese version of PersonaChat), and investigated the differences with regards to persona frequency and dialogue quality. We also mined expressions related to the personality traits of interlocutors. The contributions of this study are as follows.

- We constructed RealPersonaChat (RPC), a large-scale realistic dialogue corpus in Japanese that includes the actual personas and personality traits of the interlocutors.
- We compared RPC with an existing corpus (JPC) and showed that the frequency of persona information in utterances is significantly lower in RPC than in JPC. We also found that our corpus contains dialogues with higher dialogue satisfaction.
- We mined expressions related to the high and low scores of personality traits and successfully obtained expressions related to personality traits.

In Section 2 of this paper, we present related work. In Section 3, we show the procedure for constructing RPC, and in Section 4, we describe the statistics of the RPC dialogues. In Section 5, we compare RPC with an existing corpus. In Section 6, we mine expressions related to personality traits by using RPC. Finally, in Section 7, we summarize the paper and discuss future work.

2 Related Work

To build a data-driven chit-chat dialogue system, a large dialogue corpus is generally necessary. To this end, a number of dialogue corpora have been developed (Lowe et al., 2015; Li et al., 2017; Rashkin et al., 2019; Smith et al., 2020). To enable human-like dialogues, it is necessary to equip systems with personality, so dialogue data with persona have also been collected. For the creation of PersonaChat (Zhang et al., 2018) and its Japanese version, JPersonaChat (JPC) (Sugiyama et al., 2023), profile texts called persona consisting of sentences for each interlocutor were provided, and the dialogues were collected on the basis of these given personas. However, since the personas are not the interlocutors' own, the dialogue has the risk of sounding unnatural. In contrast, the RPC corpus proposed in this paper is based on information about the interlocutor's own persona and personality traits. Note that there is one corpus called PANDORA (Gjurković et al., 2021) that contains interlocutors' own personality information, but the data is based on Reddit threads and thus does not contain the naturalistic conversations that we require.

Personality-annotated corpora are useful for the analysis of expressions related to personality traits. Pennebaker and King (1999) utilized the Linguistic Inquiry and Word Count (LIWC) dictionary (Pennebaker et al., 2001) to count word frequencies in essays and revealed their correlation with the writers' Big Five personality traits. For example, people with high extraversion tend to use more first-person pronouns (e.g., '*I*') and socially related words (e.g., '*talk*', '*friend*'). In contrast, according to Mairesse and Walker (2007), people with low extraversion tend to use phrases that avoid articulation, leading to expressions such as '*kind of*' and '*it seems that*'. Jurafsky et al. (2009) investigated the relationship between personality traits and expressions by examining the frequency of LIWC category words in male-female spoken dialogues in order to automatically predict interlocutors' personalities. In our corpus, we also mine expressions related to the personality traits.

3 RealPersonaChat

For the corpus creation, we collected actual personas (profile texts) and personality traits from participants and had them freely engage in chit-chat conversation in order to construct a natural dialogue corpus including personas and personality traits. The data collection process involved recruiting participants, administering pre-questionnaires, creating personas, collecting dialogues, and administering post-dialogue and follow-up questionnaires. In this section, we provide the details for each procedure. This data collection was approved by the ethics committee of Nagoya University.

3.1 Recruiting Participants

We recruited 233 participants via a recruiting agency[1] in Japan. The participants were required to meet two criteria: being a native speaker of Japanese and being familiar with text chat (able to input over 200 characters per minute). Gender and age were balanced as much as possible to avoid bias. All participants agreed, prior to data collection, to not disclose personal information, to waive copyright, and to allow the collected data to be made publicly available.

3.2 Administering Pre-Questionnaires

First, the participants answered a number of pre-questionnaires about their personality traits. To cover a wide range of personality traits of the participants, we used the following five questionnaires, including those used in the prior study by Guo et al. (2021), who examined the impact of personality traits on the dialogue task performance.

- Big Five (Goldberg, 1990; McCrae and John, 1992): A questionnaire to assess human personality consisting of five factors. Participants responded to 60 questions (Wada, 1996) on a 7-point Likert scale.

- Kikuchi's Scale of Social Skills (KiSS-18) (Kikuchi, 2004): A questionnaire to assess social skills consisting of six factors. Participants responded to 18 questions on a 5-point Likert scale.

[1] https://www.lancers.jp/

- Inclusion of Others in the Self (IOS) (Aron et al., 1992): A questionnaire to assess the degree of felt inclusiveness towards others. Participants responded to one question[2] on a 7-point Likert scale.

- Adult Temperament Questionnaire (ATQ) (Evans and Rothbart, 2007): A questionnaire to assess human temperament consisting of 13 factors. Participants responded to 77 questions[3] on a 7-point Likert scale.

- Self-Monitoring Scale (SMS) (Iwabuchi et al., 1982): A questionnaire to assess the ability to appropriately control behavior in various situations. Participants responded to 25 questions on a 5-point Likert scale.

Refer to the notes under Table 1 for the full set of factors (personality traits) measured by each questionnaire.

In addition to the above personality traits, as subsidiary information, we asked for participants' demographic information and experience with using text chat. Demographic information consists of gender, age, education, employment status, and region of residence. The questionnaire for the experience with using text chat includes age of first text-based chat, frequency of using text chat, number of chatting partners, and typical chat content.

3.3 Creating Personas

After answering the pre-questionnaires, the participants wrote their own personas in sentences. The following instructions were given:

- Create ten brief sentences that describe yourself, ensuring that there are no contradictions or duplications in the content. Each sentence should be around 5–30 characters long.

- Make the content of each persona statement descriptive enough for the others to understand you well.

- Do not include personal information, such as real names, addresses, or phone numbers, including those of third parties.

[2]We used the question in http://sparqtools.org/mobility-measure/inclusion-of-other-in-the-self-ios-scale/ after translating it into Japanese.

[3]https://research.bowdoin.edu/rothbart-temperament-questionnaires/instrument-descriptions/the-adult-temperament-questionnaire/

Refer to the top sections of Table 1 for examples of the personas.

3.4 Collecting Dialogues

After creating the personas, the interlocutors were matched up and engaged in text-based conversations in pairs. We instructed them to include a minimum of 30 utterances per dialogue session, with each utterance up to around 50 characters to prevent the dialogue from becoming unnaturally lengthy. To ensure balanced participation, each interlocutor was allowed to participate in up to 200 dialogues with the restriction of up to 40 dialogues with the same partner. The interlocutors were instructed to freely choose topics by themselves and to have enjoyable conversations without specific instructions.

3.5 Administering Post-dialogue and Follow-up Questionnaire

After each dialogue, the interlocutors gauged the dialogue quality on six items: informativeness of the partner's utterance, comprehension of the partner's utterance, sense of familiarity with the partner's utterance, interest in the topic, proactiveness of oneself, and satisfaction with the dialogue on a 5-point Likert scale. In addition, after all the dialogues were completed, the interlocutors provided free-text opinions about their dialogue experience.

3.6 Dialogue Example

Table 1 shows a dialogue example of the RealPersonaChat corpus (RPC). The first row shows the personas of interlocutors A and B. The second row shows personality traits. The personality traits are expressed by binary values, 'high' and 'low', meaning that they are above or below the median of all interlocutors' personality trait scores. The third row shows the dialogue (an excerpt of a longer dialogue). In this dialogue, interlocutors A and B began with talking about the weather, then moved on to talking about camping, and later, about their families. As indicated by the underlined parts, interlocutor A utilized information related to camping and interlocutor B utilized information related to the place of residence, camping, and a fourth-grade child, from their respective personas. In this dialogue, both interlocutors gave the maximum scores to all items in the post-dialogue questionnaire, indicating that it is a satisfactory dialogue.

Persona A	Persona B
• I enjoy checking out information about new shops. • I'm not a big fan of outdoor activities. • I prefer glamping over camping. • I practice yoga. • I don't like cucumbers, and I can immediately notice them in any dish. • I'd like to go on a cruise someday. • I have a strong preference for things I like or dislike. • I'm from Seoul, South Korea. • I can play the piano and saxophone. • I enjoy attending live performances.	• I like playing Taiko drums, and I also teach drumming. • I like dishes made with flour or noodles. • Cooking is my favorite household chore. • I was born and raised in Hyogo Prefecture, and I have continued to live here even after getting married. • I start my day by washing my face to feel refreshed. • I enjoy cycling and often commute by bicycle for about three stations. • Recently, my main leisure activity has been camping. • I have a child in the fourth grade. • I frequently wear black clothing. • I drink about five cups of coffee a day.

Personality traits A	Personality traits B
Big Five* O=High, C=High, E=High, A=Low, N=Low **KiSS-18**** BS=Low, AS=High, EMS=High, OMS=High, SMS=High, PS=High **IOS** IOS=Low **ATQ**† Fe=Low, Fr=Low, Sa=High, D=High, AcC=High, AtC=High, IC=High, So=High, HIP=Low, PA=Low, NPS=High, APS=High, AsS=High **SMS**‡ Ex=Low, OD=Low, Act=High	**Big Five*** O=Low, C=High, E=High, A=Low, N=High **KiSS-18**** BS=High, AS=Low, EMS=Low, OMS=High, SMS=High, PS=High **IOS** IOS=High **ATQ**† Fe=High, Fr=High, Sa=High, D=High, AcC=High, AtC=High, IC=High, So=High, HIP=Low, PA=High, NPS=Low, APS=Low, AsS=Low **SMS**‡ Ex=High, OD=High, Act=Low

Dialogue	
A:	Hello! Nice to meet you!
B:	Hello. Nice to meet you too!
A:	It looks like it's going to be hot again today. Did you experience any damage from the typhoon?
B:	I'm in Kansai region, so we didn't have much impact from the typhoon.
A:	I'm in Kansai too! We didn't have much rain either, right?
B:	I was at a campsite in the mountains at that time, so it became a different kind of rain for me.
A:	I see, so it rained in the higher areas. But, you went camping during the Obon holiday, right?
B:	Yes, that's right. It was thunderstorms all the time over there.
A:	Oh no, that must have been tough. Do you go camping often?
B:	Yes, I do. Whenever there's a three-day weekend, I usually go camping.
A:	Wow, that's impressive! I have a desire to go camping, but I haven't been able to give it a try yet.
B:	Is that so? It's fun, you know. Did you go on any trips during the Obon holiday?
A:	I tried it once, but it was tough because of heavy rain, so I haven't done it for a while. I just wandered around nearby during the Obon holiday!
B:	Exploring nearby places can also be a refreshing change and enjoyable, right?
A:	Yes, it is. We had a meal in Umeda and also went to see a ballet performance for children!
B:	I love going out to Umeda too.
A:	Umeda has been developing even more recently, and there are more fun shops, right?
B:	That's right. When I go shopping with my child, she always asks for a lot of things.
A:	So you have children! It can be difficult to respond when children ask for things, right?
B:	I have a daughter in elementary school, and she seems to like dressing up recently.

* O: Openness, C: Conscientiousness, E: Extraversion, A: Agreeableness, N: Neuroticism
** BS: Basic Skill, AS: Advanced Skill, EMS: Emotional Management Skill, OMS: Offence Management Skill, SMS: Stress Management Skill, PS: Planning Skill
† Fe: Fear, Fr: Frustration, Sa: Sadness, D: Discomfort, AcC: Activation Control, AtC: Attentional Control, IC: Inhibitory Control, So: Sociability, HIP: High Intensity Pleasure, PA: Positive Affect, NPS: Neutral Perceptual Sensitivity, APS: Affective Perceptual Sensitivity, AsS: Associative Sensitivity
‡ Ex: Extraversion, OD: Other Directedness, Act: Acting

Table 1: Personas, personality traits, and dialogue excerpt from RealPersonaChat. In the corpus, there is also information about the interlocutors' demography and their experience with text chat. Underlined text indicates persona information included in the dialogue. This dialogue has been translated from the original Japanese to English by the authors.

4 Statistics of Collected Dialogues

In this section, we first describe the basic statistics of RPC and compare it with an existing corpus. Then, we discuss the distribution of interlocutors' demographic information and personality traits, the results of interlocutors' ratings on the quality of the dialogues, and the reflections by the interlocutors at the end of the data collection.

	RealPersonaChat (RPC)	JPersonaChat (JPC)	PersonaChat (PC)
No. of dialogues	14,000	5,000	10,907
Dialogue length	28–49 utts. (Avg. 30.09)	11–26 utts. (Avg. 12.36)	12–50 utts. (Avg. 14.86)
No. of utterances	421,203	61,793	162,064
Utterance length	1–124 chars. (Avg. 22.92)	1–100 chars. (Avg. 40.25)	1–60 words (Avg. 11.71)
Vocabulary	46,591	18,329	20,275
No. of tokens	5,551,830	1,459,322	1,897,757
Type-Token ratio	0.008	0.013	0.011
Herdan's C	0.692	0.692	0.686
No. of participants	233	Unknown	Unknown
No. of personas	233	100	7,027
Persona writer	Interlocutor themselves	Crowdworker	Crowdworker
Persona length	10 sentences 59–607 chars. (Avg. 182.08)	5 sentences 39–90 chars. (Avg. 62.87)	3–5 sentences 10–78 words (Avg. 26.98)
Associated information	Personality traits, demographic information, questionnaire on text-based chat	None	None
Language	Japanese	Japanese	English

Table 2: Statistics of RealPersonaChat, JPersonaChat, and PersonaChat datasets.

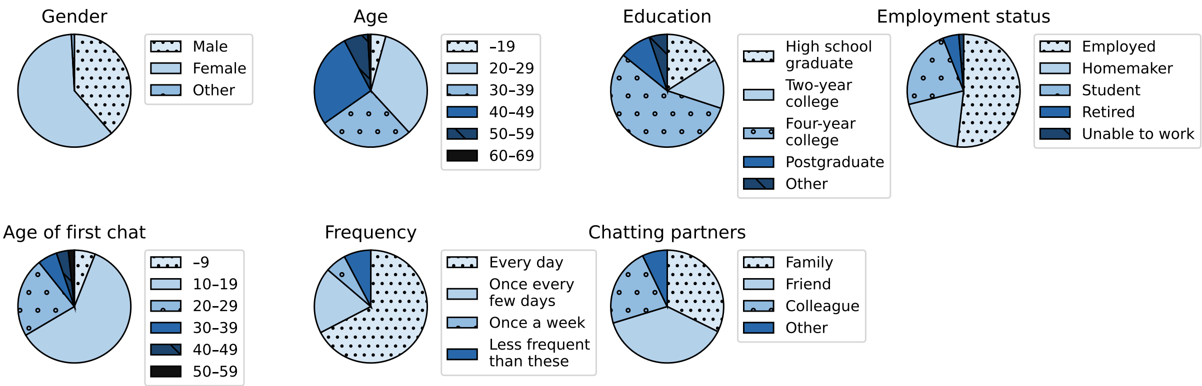

Figure 1: Distribution of the interlocutors' demographic information (upper row) and experiences with using text chat (lower row).

4.1 Basic Statistics

Table 2 lists the basic statistics of RPC. Additionally, for comparison, we present the statistics of an existing Japanese corpus, JPersonaChat (JPC), and an English corpus, PersonaChat (PC). JPC is a corpus in which the personas from PC were translated into Japanese, and dialogues were collected by using the translated personas. RPC consists of 14K dialogues, making it nearly three times the size of JPC. The dialogues in RPC are longer, with an average of 30.09 utterances per dialogue. It has the largest number of utterances at 421,203.

The number of dialogues conducted by each interlocutor ranged from 5 to 204, with a mean of 120.17 and a standard deviation of 78.47. There were 1,572 pairs of interlocutors in the corpus, and the number of dialogues conducted by the same pair ranged from one to 40, with a mean of 8.91 and a standard deviation of 6.47. Because of the presence of multiple dialogues conducted by the same pairs, this corpus can be a useful resource for the analysis of long-term conversations, such as the one done with the Multi-Session Chat (Xu et al., 2022) dataset. It also allows for the analysis of dialogues conducted between the interlocutors with various combinations of personality traits.

4.2 Interlocutors' Demographic Information and Personality Traits

The top and bottom rows of Fig. 1 show the distribution of demographic information and the distribution of experience with using text chat, respectively. While not shown here for brevity, the personality trait scores mostly follow a normal distribution, indicating that RPC is a well-balanced corpus without significant bias.

4.3 Results of Interlocutors' Ratings on the Quality of the Dialogues

Table 3 lists the results of the post-dialogue questionnaire. The average ratings for informativeness, comprehension, familiarity, interest, proac-

	Mean	SD
Informativeness	4.51	0.74
Comprehension	4.55	0.69
Familiarity	4.60	0.68
Interest	4.49	0.79
Proactiveness	4.52	0.75
Satisfaction	4.54	0.73

Table 3: Results of interlocutors' ratings on the quality of the dialogues.

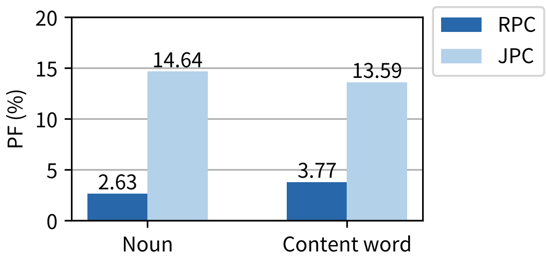

Figure 2: Persona frequency (PF). Content word indicates nouns, verbs, adjectives, and adverbs.

tiveness, and satisfaction were all approximately 4.5 points (1 is the worst and 5 the best). Given that more than 60% of the ratings were given the maximum score of five, this confirms that the RPC contains high-quality dialogues.

4.4 Reflections by the Interlocutors

According to the questionnaire conducted at the end of data collection, the interlocutors particularly liked dialogues related to their hobbies. They also liked the dialogues in which they had the same experience, such as having seen the same movie or being of the same generation. They also liked talking about household chores such as childcare and cooking. Although more analysis is needed, the corpus seems to cover a wide variety of everyday conversations.

5 Comparison with an Existing Corpus

To clarify the characteristics of RPC, we compared it with JPC, the Japanese version of PersonaChat. First, we compared the frequency of persona-related words appearing in the utterances between JPC, where interlocutors were given their personas, and RPC, where no specific instructions were given about their personas. Next, to assess the quality of the dialogues and the perception of personas and personality traits from the dialogues, we performed a subjective evaluation.

5.1 Persona Frequency

To compare the frequency of persona-related words appearing in the utterances, we counted the occurrence of persona information in each utterance and calculated the persona frequency (PF) as

$$\mathrm{PF} = \frac{\text{Number of persona words in an utterance}}{\text{Total number of words in an utterance}}, \quad (1)$$

where 'persona word' refers to a word in the persona of an interlocutor.

Since the persona length (L) differs between RPC and JPC ($L_{\mathrm{RPC}} = 182.08$, $L_{\mathrm{JPC}} = 62.87$), RPC is estimated to have $L_{\mathrm{RPC}}/L_{\mathrm{JPC}}$ times more persona words. Therefore, when calculating the PF in RPC, we adjusted Eq. (1) by multiplying by $L_{\mathrm{JPC}}/L_{\mathrm{RPC}}$. For word segmentation, we used MeCab[4] (version 0.996) as the morphological analyzer and NEologd[5] (Release 20200827-01) as the dictionary.

Figure 2 shows the PF in RPC and JPC. As we can see, RPC has a PF of approximately 2–3%, while JPC has a very high PF of 13–14%. Mann-Whitney U tests revealed that RPC's PF is significantly lower ($p < 0.05$) than that of JPC, regardless of whether nouns or content words were used as counting units. This finding suggests that the dialogues of JPC probably contain unnaturally frequent persona information.

5.2 Subjective Evaluation

We compared the quality of dialogues and the perceived personas and personality traits in RPC and JPC. We first randomly extracted 30 dialogues (no overlapping interlocutors across the dialogues) each from RPC and JPC. Then, 60 workers recruited through a crowdsourcing service[6] evaluated them. Specifically, 30 evaluators rated RPC and the remaining 30 rated JPC, with each evaluator assessing three dialogues from either RPC or JPC. Here, we made sure that each dialogue was rated by three evaluators.

This subjective evaluation utilized three metrics: dialogue quality, persona accuracy, and correlation coefficient of personality trait scores, described as follows.

- Dialogue quality: The evaluators subjectively rated the dialogues on coherence, informativeness, and satisfaction on a 5-point Likert

[4] https://taku910.github.io/mecab
[5] https://github.com/neologd/mecab-ipadic-neologd
[6] https://crowdworks.jp

scale. These criteria were taken from (Mehri et al., 2022). Note that for RPC, there are interlocutors' own evaluations of the dialogues (see Section 4.3), but since such evaluations do not exist for JPC, to ensure a fair comparison, we conducted a separate subjective evaluation by third parties.

- Persona accuracy: Following (Zhang et al., 2018), we prepared the correct persona (of one of the interlocutors; let this interlocutor be A) and an incorrect persona (randomly chosen from other interlocutors) and asked the evaluators to choose the persona of interlocutor A from the two presented personas. The baseline performance of random guessing is theoretically 50%. This metric indicates how related the dialogue is to the personas.

- Correlation coefficient of personality trait scores: Evaluators read a dialogue and answered the Big Five personality questionnaire (TIPI-J) (Oshio et al., 2012) with regards to one of the interlocutors (let this interlocutor be A). This procedure is the same as that in (Jiang et al., 2020) to obtain personality trait scores from third parties. Then, the scores were compared with those of the actual personality traits of interlocutor A to derive correlation coefficients. This metric indicates how well the personality traits are perceivable from the dialogue. Note that this is only done for RPC, since JPC does not have personality trait scores for the interlocutors.

Table 4 lists the results. Regarding dialogue quality, the satisfaction score for RPC surpassed that of JPC, while coherence and informativeness were comparable to those of JPC. The results of Mann-Whitney U tests indicate that RPC's satisfaction was significantly higher than that of JPC ($p < 0.05$).

Regarding persona accuracy, JPC achieved a high accuracy of 82%, whereas RPC fell behind with 61% accuracy ($p < 0.05$, Mann-Whitney U test). This is expected, as JPC is based on the given personas. In contrast, in RPC with no specific instructions about personas, the accuracy was 61%, further suggesting the possibility that JPC may be biased in the use of personas.

Table 5 lists the correlation coefficient of personality trait scores. The Pearson's correlation coefficient for conscientiousness was the highest with 0.17, and the Spearman's correlation coefficients for conscientiousness and agreeableness were 0.12, indicating low correlations. This suggests that in naturally collected RPC, interlocutors' personalities are not explicitly discernible.

	RPC	JPC
Coherence	**4.39**	4.34
Informativeness	4.09	**4.10**
Satisfaction	**4.29***	4.03
Persona accuracy	61.11%	**82.22%***

Table 4: Results of subjective evaluation for dialogue quality and persona accuracy. **Bold** indicates the maximum value for each column. An asterisk (*) indicates a statistically significant difference at a significance level of 5%.

	Pearson	Spearman
Openness	−0.09	−0.05
Conscientiousness	**0.17**	**0.12**
Extraversion	−0.04	−0.04
Agreeableness	0.14	**0.12**
Neuroticism	0.04	0.06

Table 5: Correlation coefficient (Pearson and Spearman) of personality trait scores. **Bold** indicates the maximum value for each column.

6 Analysis of the Relationship between Personality Traits and Expressions

To examine whether our corpus can be used to mine expressions related to particular personality traits, we extracted expressions that are more likely to be used by interlocutors with high/low scores in each personality trait.

6.1 Procedure

First, considering 4-grams as meaningful sequence of words, we extracted 4-grams from all utterances in the corpus. Here, we used MeCab as the morphological analyzer and NEologd as the dictionary. To avoid noise, we only retained the 4-grams used by 50 or more interlocutors and that appeared at least 100 times, resulting in 1,628 4-grams used in this analysis.

Next, for each of the extracted 4-grams, we classified all utterances with regard to whether they contain the 4-gram or not and whether the interlocutor's personality trait of that utterance is high or low, thereby creating a 2×2 matrix to perform Fisher's exact probability test in order to test whether the 4-gram is significantly related to the high/low of a certain personality trait.

	Distinct 4-grams in high interlocutors	**Distinct 4-grams in low interlocutors**
Big Five: Extraversion	んですね、(*I see*), ですか?? (*is it ...??*), そうですね、(*right*), なんですね (*I see*), ですね、私 (*right, I*), そうなんです (*right*), ね、私は (*right, I*), ですよね、(*right*), お願いいたします。(*please*), んですよ、(*it is ...*)	お話ありがとうございまし (*thanks for chatting*), かもしれないです(*it might be*), ですよね。(*right*), のですね！(*I see!*), 良いですね！(*great!*), お願いします。(*please*), よろしくおねがいします (*please*), のですが、(*but*), ですよね！(*right!*), ございました！(*thanks!*)
KiSS-18: Basic Skill	ですよね、(*right*), んですね、(*I see*), そうですね、(*right*), ですね、私 (*right, I*), なんですよ (*right*), たんですよ (*it is ...*), ね、私は (*right, I*), なんですねー (*I see*), んですよ！(*right!*), そうなんです (*right*)	お願いします。(*please*), ありますか？(*is it ...?*), のですね！(*I see*), はありますか (*have you ever ...?*), ましたか？(*is it ...?*), しますね。(*do ...*), でしょうか？(*is it ...?*), ですか??(*is it ...?*), こんにちは！よろしくお願いします。(*hello! thanks for being here to chat*), んですけど、(*but*)
IOS	んですね、(*I see*), ですよね、(*right*), ています。(*do ...*), そうですね、(*right*), ですね、私 (*right, I*), そうなんです (*right*), いいですね、(*great*), なんですね (*I see*), しています (*do ...*), んですよ、(*right*)	のですね！(*I see*), お話ありがとうございまし(*thanks for chatting*), しますね。(*do ...*), んですが、(*but*), おはようございます！(*good morning!*), なのですね(*I see*), でしょうか？(*is it ...?*), ございました！(*thanks!*), ね。でも、(*but*), というか。(*rather*)
ATQ: Attentional Control	ですよね、(*right*), んですね、(*I see*), なんですねー (*I see*), ですね、私 (*right, I*), ね、私は (*right, I*), そうですね、(*right*), んですよ、(*right*), お願いいたします。(*please*), こんにちは、よろしくお願いします。(*hello, thanks for being here to chat*),、いいですね (*great*)	ですかね。(*think that*), 良いですね！(*great!*), ですか??(*is it ...?*), のですね！(*I see*), でしょうか？(*is it ...?*), 良いですね。(*great!*), んですか？(*is it ...?*), そうなのです (*right*), と思います。(*think that*), んですね。(*I see*)

Table 6: Interlocutors' personality traits and distinct 4-grams in utterances. Items between parentheses are English translations by the authors. As with the convention in Japanese, spaces are omitted between words. Underlined expressions indicate those mentioned in Section 6.2.

6.2 Results

We successfully obtained expressions for all high and low personality traits. In Table 6, we present the top ten distinctive expressions for some of the personality traits with regards to p-values (lower means more distinctive).

As we can see, interlocutors with high extraversion in Big Five tended to talk about themselves using first-person pronouns. In contrast, those with low extraversion exhibited a preference for expressions that avoid making definitive statements. This result aligns well with the findings by Pennebaker and King (1999). For other personality traits, interlocutors with high scores in the basic skill (KiSS-18), IOS, and attentional control (ATQ) also tended to use first-person pronouns. In contrast, interlocutors with low scores in these personality traits showed a preference for interrogative expressions.

As shown in this brief analysis, we can confirm the usefulness of RPC with regards to the mining of expressions related to personality traits. We thus consider RPC to be a valuable resource for training dialogue models that can differentiate the use of various expressions to exhibit a wide variety of personality traits.

7 Conclusion and Future Work

In this study, we constructed the RealPersonaChat (RPC) corpus consisting of natural dialogues by collecting actual personas and personality traits from interlocutors and having them engage in free-form conversations. RPC consists of 14K dialogues and is the world's largest corpus containing personas and personality traits. Through the statistics and comparison with an existing corpus, we confirmed the naturalness and the quality of RPC.

There is much future work to be done. First, by using RPC, we want to realize a dialogue system that can naturally reflect personas and personality traits. We plan to use RPC to train response generation models and verify their dialogue performance. Second, using the corpus, we want to train a personality estimation model, since personality estimation is useful for dialogue systems to generate responses tailored to the user's personality. Here, it is worth noting that it is essential to consider ethical issues related to interlocutors not willing to have their personality predicted. There is also an issue that the estimation results may reinforce stereotypes for people with specific attributes (Tatman, 2020). Third, a more detailed analysis of the corpus will be needed. For example, in addition to mining linguistic expressions, it is important to obtain insight into how personality affects the content and interaction. Last but not least, the analysis of the corpus reported here is still preliminary; after further experiments and scrutiny of the data, we plan to release the corpus to the public.

Acknowledgments

This work was supported by JST Moonshot R&D Grant number JPMJMS2011.

References

Arthur Aron, Elaine N Aron, and Danny Smollan. 1992. Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4):596.

Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

David E Evans and Mary K Rothbart. 2007. Developing a model for adult temperament. *Journal of Research in Personality*, 41(4):868–888.

Matej Gjurković, Mladen Karan, Iva Vukojević, Mihaela Bošnjak, and Jan Snajder. 2021. PANDORA talks: Personality and demographics on Reddit. In *Proceedings of the 9th International Workshop on Natural Language Processing for Social Media*, pages 138–152.

Lewis R Goldberg. 1990. An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59(6):1216–1229.

Ao Guo, Atsumoto Ohashi, Ryu Hirai, Yuya Chiba, Yuiko Tsunomori, and Ryuichiro Higashinaka. 2021. Influence of user personality on dialogue task performance: A case study using a rule-based dialogue system. In *Proceedings of the 3rd Workshop on Natural Language Processing for Conversational AI*, pages 263–270.

Chiaki Iwabuchi, Kunio Tanaka, and Hiroaki Nakazato. 1982. A study of the self-monitoring scale. *The Japanese Psychological Research*, 53(1):54–57. (in Japanese).

Hang Jiang, Xianzhe Zhang, and Jinho D Choi. 2020. Automatic text-based personality recognition on monologues and multiparty dialogues using attentive networks and contextual embeddings (student abstract). In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, volume 34, pages 13821–13822.

Dan Jurafsky, Rajesh Ranganath, and Dan McFarland. 2009. Extracting social meaning: Identifying interactional style in spoken conversation. In *Proceedings of the 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 638–646.

Akio Kikuchi. 2004. Notes on the researches using KiSS-18. *Bulletin of the Faculty of Social Welfare, Iwate Prefectural University*, 6(2):41–51. (in Japanese).

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the 8th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 986–995.

Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau. 2015. The Ubuntu Dialogue Corpus: A large dataset for research in unstructured multi-turn dialogue systems. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 285–294.

François Mairesse and Marilyn Walker. 2007. PERSONAGE: Personality generation for dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 496–503.

Robert R McCrae and Oliver P John. 1992. An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):175–215.

Shikib Mehri, Jinho Choi, Luis Fernando D'Haro, Jan Deriu, Maxine Eskenazi, Milica Gasic, Kallirroi Georgila, Dilek Hakkani-Tur, Zekang Li, Verena Rieser, Samira Shaikh, David Traum, Yi-Ting Yeh, Zhou Yu, Yizhe Zhang, and Chen Zhang. 2022. Report from the NSF future directions workshop on automatic evaluation of dialog: Research directions and challenges. *arXiv preprint arXiv:2203.10012*.

Michael Neff, Yingying Wang, Rob Abbott, and Marilyn Walker. 2010. Evaluating the effect of gesture and language on personality perception in conversational agents. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents*, pages 222–235.

OpenAI. 2023. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*.

Atsushi Oshio, Abe Shingo, and Pino Cutrone. 2012. Development, reliability, and validity of the Japanese version of Ten Item Personality Inventory (TIPI-J). *Japanese Journal of Personality*, 21(1).

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC2001.

James W Pennebaker and Laura A King. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6):1296.

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.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325.

Kurt Shuster, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, et al. 2022. BlenderBot 3: A deployed conversational agent that continually learns to responsibly engage. *arXiv preprint arXiv:2208.03188*.

Eric Michael Smith, Mary Williamson, Kurt Shuster, Jason Weston, and Y-Lan Boureau. 2020. Can you put it all together: Evaluating conversational agents' ability to blend skills. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2021–2030.

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 *Proceedings of 2022 IEEE Spoken Language Technology Workshop*, pages 685–691.

Rachael Tatman. 2020. What I won't build. Keynote at the Fourth Widening Natural Language Processing Workshop at Annual Meeting of the Association for Computational Linguistics.

Sayuri Wada. 1996. Construction of the big five scales of personality trait terms and concurrent validity with npi. *The Japanese Journal of Psychology*, 67(1):61–67. (in Japanese).

Jing Xu, Arthur Szlam, and Jason Weston. 2022. Beyond goldfish memory: Long-term open-domain conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5180–5197.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213.

Yinhe Zheng, Guanyi Chen, Minlie Huang, Song Liu, and Xuan Zhu. 2019. Personalized dialogue generation with diversified traits. *arXiv preprint arXiv:1901.09672*.