Speaker Identification of Quotes in Japanese Novels based on Gender Classification Model by BERT

Yuki Zenimoto, Takehito Utsuro
Degree Programs in Systems and Information Engineering,
Graduate School of Science and Technology, University of Tsukuba,
1-1-1, Tennodai, Tsukuba, Ibaraki, 305-8573, Japan
s2220753_@_s.tsukuba.ac.jp, utsuro_@_.tsukuba.ac.jp

Abstract

In the Japanese language, particles and auxiliary verbs in utterances tend to vary depending on the speaker's gender. Thus, the linguistic expressions within the utterance are an important hint in identifying the speaker. This research proposes a method for identifying the speaker in novels using linguistic expressions within the utterances that reflect the gender. We constructed a dataset of utterances with gender-specific linguistic expressions by automatically collecting utterances from novels that contained the first-person pronouns "俺 (ore)" or "私 (watashi)", considering that "俺 (ore)" is primarily used by males and "私 (watashi)" is primarily used by females. We fine-tuned a BERT (Devlin et al., 2019)-based gender-specific language model that classifies the gender of the speaker of a given utterance using this dataset as the training data. The fine-tuned gender classification model achieved an accuracy of 87.9% when evaluating the utterances of two main characters in a Japanese romance novel, demonstrating that this gender classification model is effective in speaker identification.

1 Introduction

Novels are valuable resources for enriching dialogue systems. Various characters converse with each other in novels, and their behaviors are described in the narratives. If such information is appropriately extracted, it can be used in dialogue systems and various research such as personality analysis. Dialogue systems are required to respond appropriately to the users' utterances, and it is also expected to have a specific personality. To train dialogue models, it is necessary to collect dialogue data from web services such as Twitter and Reddit (Serban et al., 2015; Mazaré et al., 2018). However, these dialogue data are insufficient for training dialogue models capable of acting as if they have specific personalities. This problem is simply because these dialogue data consist of a mixture of utterances of numerous personalities and are difficult to utilize as a source for building dialogue models with specific personalities. Considering the preceding discussion, it is required to prepare dialogue data that are accompanied by profiles (Zhang et al., 2018; Sugiyama et al., 2021). Crowdsourcing is commonly used to collect such dialogue data with profile information. However, crowdsourcing large scale dialogue data with profile information is extremely expensive.

Considering this limitation, this paper introduced an approach for gathering dialogue data from novels, examining the task of speaker identification in novels to generate persona data. Previous work on speaker identification in English novels (O'Keefe et al., 2012; Muzny et al., 2017) relied heavily on narrative elements surrounding the target utterance and the utterances before and after it. However, the target utterance itself was not effectively used. Here, it is important to note that, in the Japanese language, particles and auxiliary verbs in utterances differ depending on the speaker's gender (Miyazaki et al., 2015; Murai, 2018). Thus, the linguistic expressions within the utterance are also an essential clue in speaker identification.

This paper proposes a method to identify the

First-Person Pronouns	Example Sentences	Significant Words
"俺" (ore) (typical for males)	俺の番だな ore no ban da **na** (It is my turn)	na, ze
	俺は家に戻るぜ ore wa ie ni modoru **ze** (I am going home)	
"私" (watashi) (typical for females)	私の番だね watashi no ban da **ne** (It is my turn)	ne, wa
	私は家に戻るわ watashi wa ie ni modoru **wa** (I am going home)	

Table 1: Significant Words for First-Person Pronouns "俺 (ore) " and "私 (watashi)"

speaker in novels, which uses the linguistic expressions within the utterance that reflect the gender. We gathered utterances within novels containing gender-specific linguistic expressions on a large scale, concentrating on the first-person pronouns within sentences as signals. In the Japanese language, there are numerous varieties of personal pronouns. Personal pronouns in Japanese can approximately indicate various attributes, such as gender, age, temperament, and characters' social status. Significantly, the first-person pronouns "俺 (ore)" and "私 (watashi)" are remarkably contrasting, with "俺 (ore)" being primarily used by males and "私 (watashi)" being primarily used by females. As shown in Table 1, terms commonly used by males frequently occur in utterances containing " 俺 (ore)", whereas terms commonly used by females frequently appear in utterances containing "私 (watashi)". Thus, we constructed a dataset of utterances containing gender-specific language expressions by automatically collecting utterances from novels that contained the first-person pronouns "俺 (ore)" or "私 (watashi)". We fine-tuned a BERT (Devlin et al., 2019)-based gender-specific language model that classifies the speaker's gender of a given utterance using linguistic expressions within the utterance as clues[1]. In this research, we refer to this gender-specific language model as the *gender classification model*.

The fine-tuned gender classification model was evaluated against the utterances of two main characters in a Japanese romantic novel. The results showed that the fine-tuned gender classification model achieved an accuracy of 87.9%, indicating that this model is effective in speaker identification. Additionally, we found that the speaker alternation constraint employed in previous studies (He et al., 2013) and the constraint on personal nouns within utterances improve the speaker identification performance.

2 Related Works

Previous studies on constructing Japanese persona datasets have usually relied on crowdsourcing to gather sentences that reflect a given personality. Sugiyama et al. (2021) used crowdsourcing to create 100 different personalities and collected a total of 61,794 utterances in 5,000 dialogues. They set the unit price at 300 yen (approximately three dollars) per dialogue. Ishii et al. (2021) proposed the use of role play-based question-answering to efficiently gather paired utterances expressing an anime character's personality. Users who were familiar with the anime character provided them with 15,112 of these utterance pairs for free. However, finding relevant users who are familiar with such characters and then collecting utterance pairs using the same cost-free method are difficult.

Two corpora that could be used for speaker identification in Japanese novels are Aozora Bunko[2] and the balanced corpus of contemporary written Japanese (BCCWJ) (Maekawa et al., 2014). Aozora Bunko is a Japanese digital library that contains thousands of out-of-copyright Japanese works. BCCWJ[3] is a Japanese first 100 million words balanced corpus covering 2,663 novels (published between 1976 and 2005) and 270,388 individual utterances,

[1] In both training and testing, the first-person pronouns "俺 (ore)" and "私 (watashi)" of the input sentence were replaced with the [MASK] token.

[2] https://www.aozora.gr.jp/

[3] https://clrd.ninjal.ac.jp/bccwj/

<table>
<tr><th>Example Sentences</th><th>Speaker (gender)</th><th>Personal Nouns (Japanese/Pronunciation/English)</th></tr>
<tr><td>わたしは背が高すぎて、
お姉様の服は着られませんから......
(I am too tall to wear your clothes. ......)</td><td rowspan="2">Marie
(female)</td><td>わたし/“watashi”/I
お姉様/“one-sama”/sister</td></tr>
<tr><td>......キュロス様は、今、どちらに......？
(Where ...... is Mr. Kyuros now ......?)</td><td>キュロス様/“Kyurosu-sama”/Mr. Kyuros</td></tr>
<tr><td>どうもありがとう、使用人のお嬢さん。
(Thank you very much, servant girl.)</td><td rowspan="2">Kyuros
(male)</td><td>お嬢さん/“ojo-san”/girl</td></tr>
<tr><td>おはよう、マリー。
(Good morning, Marie.)</td><td>マリー/“mari”/Marie</td></tr>
</table>

Table 2: Examples of the Annotation of Utterances Within the Novel for Evaluation

each linked to the speaker and some attributes (gender, age group, occupation, and so on) (Yamazaki et al., 2018). Both Aozora Bunko and BCCWJ offer collections of older novels, and many utterances are written in a different style compared with those in contemporary novels. Therefore, these corpora are not appropriate for gathering distinctive characters' utterances, such as those in anime and comics.

Early works on quote attribution in English novels focused on textual indications to determine the mention corresponding to the speaker of a quote (Elson and McKeown, 2010; Muzny et al., 2017). They mainly used patterns like quote-mention-verb and dependency parses to extract mentions and speakers. Only vocatives were extracted from the utterance and were incorporated in the classification of the next speaker or the listener as features within the utterance (He et al., 2013; Yeung and Lee, 2017); however, linguistic styles such as auxiliary verbs were not used in the task for English novels.

In the field of a character's linguistic style analysis, Miyazaki et al. (2016) identified 13 categories of linguistic peculiarities that can be used to identify the linguistic styles of most Japanese fictional characters. Akama et al. (2018) proposed style-sensitive word vectors that capture the stylistic similarity between two words. To measure the intensity of the persona characteristics of an utterance, Miyazaki et al. (2021) proposed a persona speaker probability that distinguishes the persona of the speaker of each utterance. In contrast to their method that classifies utterances into specific characters, our proposed method classifies the speaker of the utterances as male or female.

3 The Novel for Analysis

To evaluate the performance of speaker identification by the gender classification model and other additional constraints, we selected a contemporary novel[4] with a romance theme between two main characters — an aristocratic man ("Kyuros") and an aristocratic woman ("Marie").

3.1 Annotation

The first author of the paper is the annotator of our dataset. Quotes in Japanese text were represented by "「" at the start and "」" at the end, and we describe such utterance as a quote. Table 2 shows examples of the annotation. The annotator was instructed to associate the following attributions to each utterance.

- **Speaker Name of the Utterance**
  The speaker of the utterance was identified by a unique name. Utterances by characters whose names were not mentioned in the novel, such as store clerks and crowds, were labeled as "Others".

- **Personal Nouns appearing within the Utterance**
  Personal nouns, such as first- and second-person pronouns, were extracted from the utterance. Personal nouns such as "会社員 (kaishain)" (company employee) and "使用人 (shiyonin)" (servant) were not tagged in this context because they were mentioned in the same way by any speaker. However,

[4] `https://ncode.syosetu.com/n1860fv/`

<table>
<tr><th colspan="2">Type</th><th>Number of dialogues</th><th>Number of utterances by Marie (female) and Kyuros (male)</th><th>Number of utterances by characters other than Marie (female) and Kyuros (male)</th></tr>
<tr><td rowspan="2">Dialogues consisting of utterances only by Marie (female) and Kyuros (male)</td><td>Two speakers take turns alternately (speaker alternation constraint satisfied)</td><td>298</td><td>900</td><td>0</td></tr>
<tr><td>One speaker continues more than one utterance (speaker alternation constraint unsatisfied)</td><td>(1) (consisting of utterances by one speaker only)</td><td>2</td><td>0</td></tr>
<tr><td colspan="2">Total</td><td>298 (+1)</td><td>902</td><td>0</td></tr>
<tr><td colspan="2">Dialogues including characters other than Marie (female) or Kyuros (male)</td><td>504</td><td>821</td><td>851</td></tr>
<tr><td colspan="2">Isolated single utterance immediately preceded/followed by narratives</td><td>0</td><td>485</td><td>426</td></tr>
<tr><td colspan="2">Total</td><td>504</td><td>1,306</td><td>1,277</td></tr>
</table>

Table 3: Statistics of the Dataset [Focusing on utterances by “Marie” (female) or “Kyuros” (male), where we exclusively evaluate utterances by “Marie” (female) or “Kyuros” (male) in this research. A total of 902 utterances of the upper half of the table were evaluated with the speaker alternation constraint (Figure 2), whereas a total of 1,306 utterances of the lower half of the table were NOT evaluated with the speaker alternation constraint (Figure 3).]

when suffixes (for example, representing politeness) were added to those personal nouns, such as “会社員さん (kaishain-san)” (company employee + politeness) and “使用人くん (shiyonin-kun)” (servant + politeness), those personal nouns were also extracted because suffixes often help readers identify the speaker.

3.2 Dataset

We obtained 3,993 utterances as a consequence of the annotation, in which the speaker and personal nouns within the utterance were linked. The main characters were “Marie” (female), and “Kyuros” (male). Of the 3,993 utterances, 1,178 were by “Marie” (female) and 1,030 by “Kyuros” (male), and the remaining 1,785 utterances were by minor characters [5]. In this research, the performance of the speaker identification was exclusively evaluated on the 2,208 utterances by “Marie” (female) and “Kyuros” (male). Due to the gender classification model’s constraints on input token size, only 32 tokens from the beginning of the utterance were used for classification.

Table 3 shows the dataset statistics, focusing on utterances by “Marie” (female) or “Kyuros” (male). The upper half of the table shows the number of dialogues and utterances consisting only of utterances by “Marie” (female) or “Kyuros” (male), whereas the lower half shows that of dialogues including characters other than “Marie” (female) or “Kyuros” (male) and the number of isolated single utterances immediately preceded/followed by narratives. The upper half of Table 3 shows that the speaker alternation constraint (described in section 5) was satisfied for almost all the dialogues consisting only of utterances by “Marie” (female) and “Kyuros” (male). Those consecutive utterances to which the speaker alternation constraint could be applied were those immediately preceding or immediately following another utterance, where isolated single utterances immediately preceded/followed by narratives were excluded.

Table 4 shows the statistics for the utterances subject to the constraint on personal nouns within quotes (described in section 6). Utterances that meet the constraint on personal nouns within quotes

[5] Of the 1,785 utterances, 1,277 of the dialogues were with “Marie” (female) or “Kyuros” (male) and the remaining 508 utterances were from other dialogues. (shown in the lower half of Table 3)

Character	Proportion
"Kyuros" (male)	49.2% (507/1,030)
"Marie" (female)	39.8% (469/1,178)
Total	44.2% (976/2,208)

Table 4: Applicability of the Constraint on Personal Nouns Within Quotes

were those that contained at least one personal noun within the quote.

4 Gender Classification Model by BERT

In this section, we describe how to construct the gender classification model, which is based on a fine-tuned BERT-based gender-specific language model (Devlin et al., 2019). Different linguistic styles in the Japanese language convey the speaker's gender (Murai, 2018). Therefore, the linguistic style of the utterance can be used to determine the gender of the speaker. Our gender classification model takes an utterance as input and outputs the classification probability of the speaker's gender. This model is similar to the persona speaker probability model of Miyazaki et al. (2021) used for filtering out inappropriate utterances with respect to the given persona. The persona speaker probability model was trained by collecting utterances from the set of personas. In contrast to their method, we used utterances from a large number of novels other than the target novel as the training data to train the classification model.

4.1 First-Person Pronouns in the Japanese Language

There are numerous types of personal pronouns used in the Japanese language. Personal pronouns in Japanese can approximately indicate various attributes, such as gender, age, temperament, and characters' social status. Significantly, the first-person pronouns "俺 (ore)" and "私 (watashi)" are remarkably contrasting, with "俺 (ore)" being primarily used by males and "私 (watashi)" being primarily used by females. It is observed that words commonly used by males frequently appear in utterances containing "俺 (ore)", and similarly, words commonly used by females appear in utterances containing "私 (watashi)". Considering this, we constructed a dataset of utterances that included linguistic expressions specific to genders, containing the first-person pronouns "俺 (ore)" or "私 (watashi)" from novels.

4.2 Collecting Quotes for Training the Gender Classification Model

We used a novel posting site called "小説家になろう" (Aim to be a novelist)[6] to gather utterances containing the first-person pronouns "俺 (ore)" and "私 (watashi)". We selected 2,000 novels and gathered their text. We next applied the morphological analysis to the text using Sudachi[7]. Quotes in Japanese text were represented "「" at the start and "」" at the end, as present below:

「おはよう、マリー。」
("Good morning, Marie.")

Therefore, to extract quotes from the Japanese text, we extracted string sequences starting with "「" and ending with "」". Furthermore, we selected quotes that included the first-person pronouns "俺 (ore)" or "私 (watashi)". At this stage, we obtained 38,192 quotes that included "俺 (ore)" and 52,052 quotes that included "私 (watashi)". Then, quotes composed of multiple sentences were decomposed according to the symbols "。","?", and "!" and were regarded as multiple quotes[8]. By accumulating each of those multiple quotes, we obtained 83,571 utterances that had linguistic expressions closely related to "俺 (ore)" (typical for males) and 118,997 utterances that had linguistic expressions closely related to "私 (watashi)" (typical for females).

4.3 Training the Gender Classification Model

In this research, we used a pre-trained BERT (Devlin et al., 2019) model for gender classification. Specifically, we used Tohoku University's Japanese version of BERT-base[9], which is trained on Japanese Wikipedia[10]. This model consists of 12 layers,

[6] https://syosetu.com/

[7] https://github.com/WorksApplications/Sudachi

[8] After decomposing the original single quote into multiple quotes, it can happen that those constituent quotes may not include the first-person pronouns "俺 (ore)" or "私 (watashi)".

[9] https://github.com/cl-tohoku/bert-japanese

[10] We also evaluated a base-sized Japanese RoBERTa model (Liu et al., 2019) of https://huggingface.

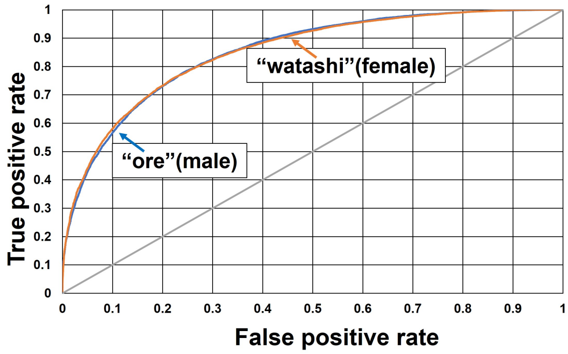

Figure 1: ROC Curves of the Gender Classification Model for the Test Data (%) [The set typical for males denoted as "ore" (male) and the set typical for females denoted as "watashi" (female). The overall test data accuracy is 77.1% and the AUC is 0.851 for both curves]

768 dimensions of hidden states, and 12 attention heads. The first-person pronouns "俺 (ore)" and "私 (watashi)" of the input sentence were replaced with the [MASK] token in both training and testing[11]. For a total of 83,571 utterances that had linguistic expressions closely related to "俺 (ore)" (typical for male) and 118,997 utterances that have linguistic expressions closely related to "私 (watashi)" (typical for female), the ratio of training, validation and test data was 6:2:2. The model with the minimum validation loss was evaluated against the test data.

4.4 Evaluating the Gender Classification Model

Figure 1 shows the ROC-curves of our gender classification model for male (denoted as "ore" (male), i.e., utterances that have linguistic expressions closely related to "俺 (ore)") and for female (denoted as "watashi" (female), i.e., utterances that have linguistic expressions closely related to "私 (watashi)") for the test data, where the accuracy of the overall test data is 77.1% and each of the two curves has AUC of 0.851. Table 5 shows examples of probabilities by the gender classification model. The utterances including "だぜ (daze)" are identified as related to the first-person pronoun "俺 (ore)" (male). In contrast, the utterances including "よ (yo)" are identified as related to the first-person pronoun "私 (watashi)" (female), which is consistent with the observation of Murai (2018). In addition, the utterance with "お前 (omae)" which is used primarily by males, and that with "あなた (anata)" which is used primarily by females, are also appropriately identified by our gender classification model.

co/rinna/japanese-roberta-base in the same task, where the performance was almost similar to that of BERT (Devlin et al., 2019).

[11]The performance with [MASK] token replacement was superior to than that without [MASK] token replacement for both training and testing.

5 Speaker Alternation Constraint

In consecutive utterances in a novel, two speakers usually take turns alternately. This observation is represented in this section by the "speaker alternation constraint", where we restricted the turns of the speakers to constantly alternating. We also made a minor modification to the speaker alternation constraint, where we defined a dialogue as a sequence of utterances with no intervening narratives. This modification stems from the observation that when narratives interrupt in the middle of utterances, the same speaker is often consecutive. He et al. (2013) proposed integrated classifier features for speaker identification with the speaker alternation constraint. In contrast to He et al. (2013), we introduce the speaker alternation constraint as the hard constraint in this research, so that the results of speaker identification strictly follow the constraint.

This constraint was satisfied for almost all the dialogues consisting only of utterances by "Marie" (female) and "Kyuros" (male), as shown in the upper half of Table 3. More specifically, as mentioned in section 3.2, as the target speakers in the novel for evaluation, we considered only the two major characters of the romance novel, "Marie" (female) and "Kyuros" (male). In terms of the speaker alternation constraint, we only considered those two major characters.

Let A and B be the two speakers for which we consider the speaker alternation constraint. Given an utterance sequence $(U_1, U_2, \ldots)$ and suppose that $(S_1, S_2 \ldots)$ is the speaker sequence of $(U_1, U_2, \ldots)$, then under the speaker alternation constraint, we considered only the following two types of speaker

Example Sentences	Probability of the Model	
	“俺 (ore)” (typically for male)	“私 (watashi)” (typically for female)
俺のものだぜ ore no mono **da** **ze** (It’s mine)	**0.965**	0.035
私のものよ watashi no mono **yo** (It’s mine)	0.005	**0.995**
お前の名前は? **omae** no namae wa ? (What is your name?)	**0.769**	0.231
あなたの名前は? **anata** no namae wa ? (What is your name?)	0.083	**0.917**

Table 5: Examples of the Probabilities by the Gender Classification Model

Personal Nouns (Japanese/pronunciation/English)	# of times the quotes including the personal noun occurrences judged by the gender classification model		Total
	Marie (female)	Kyuros (male)	
わたし/“watashi”/I (only used by Marie)	215 (100.0%)	0 (0.0%)	215
マリー/“mari”/Marie (mostly used by Kyuros)	19 (8.9%)	194 (91.1%)	213
俺/“ore”/I (only used by Kyuros)	3 (1.7%)	174 (98.3%)	177
君/“kimi”/you (only used by Kyuros)	40 (33.1%)	81 (66.9%)	121
ミオ/“mio”/Mio (used by both Marie and Kyuros)	41 (54.7%)	34 (45.3%)	75

Table 6: Examples of the Personal Nouns and # of Times their Occurrences Judged by the Gender Classification Model

sequences:

$$(S_1, S_2, S_3, \ldots) = (A, B, A, \ldots)$$
$$(S_1, S_2, S_3, \ldots) = (B, A, B, \ldots)$$

Here, the probability of the speaker sequence of the utterance sequence $(U_1, U_2, \ldots, U_n)$ being $(S_1, S_2, \ldots, S_n)$ by the gender classification model P_g is given as below:

$$P_g((U_1, U_2, \ldots), (S_1, S_2, \ldots)) = \prod_{i=1}^{i=n} P_g(U_i, S_i)$$

Therefore, we compared the two probabilities below:

$$P_g((U_1, U_2, \ldots), (A, B, A, \ldots))$$
$$P_g((U_1, U_2, \ldots), (B, A, B, \ldots))$$

As the result of the gender classification model and the speaker alternation constraint, the speaker sequence with the higher probability was chosen.

In the evaluation, the speaker alternation constraint was applied to the 902 utterances of dialogues only by “Marie” (female) and “Kyuros” (male) in the upper half of Table 3.

6 Constraint on Personal Nouns Within Quotes

Personal nouns, as well as Japanese suffixes representing politeness such as “君 (kun)” and “さん (san)” within utterances, often help readers in identifying the speaker. As shown in Table 6, the first-person pronoun “わたし (watashi)” was only used by “Marie” (female), whereas the first-person pronoun “俺 (ore)” was only used by “Kyuros” (male). Furthermore, the second-person pronoun “君 (kimi)” was only used by “Kyuros” (male). However, as in Table 6, according to the gender classification model, 33% of the utterances including “君 (kimi)” were incorrectly judged as uttered by “Marie” (female). To correct those 33% error cases, we introduced the score $Q_{G+PN}(U, S)$ of the speaker of the utterance U to be S by the gender classification model adjusted by the constraint of the personal nouns within quotes as shown below:

$$Q_{G+PN}(U, S) \;=\; P_g(U, S) + \alpha \sum_{k=1}^{m} r_g(n_k, S) \quad (1)$$

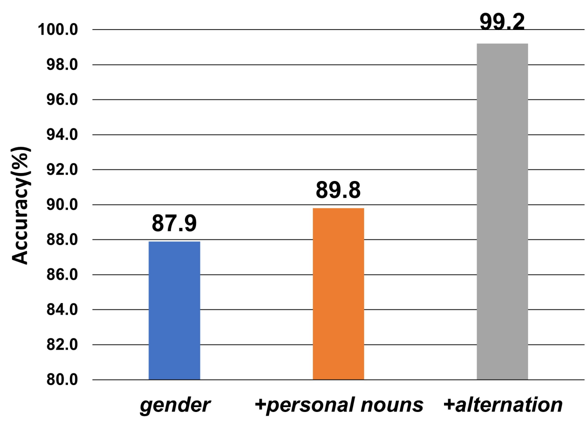

Figure 2: Evaluation Results for 902 Utterances in the Upper Half of Table 3 Consisting of Utterances Only by Marie (female) and Kyuros (male) (*gender*: gender classification model, +*personal nouns*: gender classification model + the constraint on personal nouns within quotes, +*alternation*: gender classification model + the speaker alternation constraint)

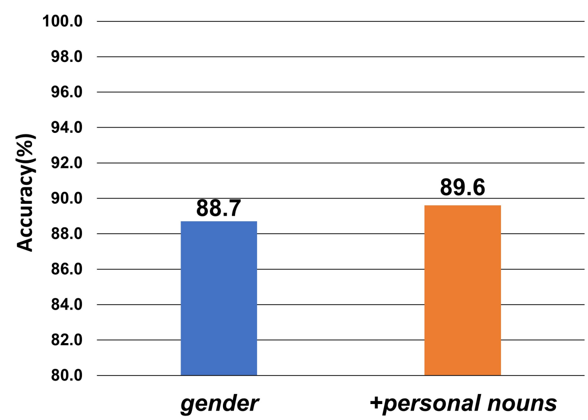

Figure 3: Evaluation Results for 1,306 Utterances [by Marie (female) and Kyuros (male)] in the Lower Half of Table 3 (*gender*: gender classification model, +*personal nouns*: gender classification model + the constraint on personal nouns within quotes)

where P_g is the probability of the gender classification model, $n_1, n_2, \ldots n_m$ represent the m kinds of personal nouns within the utterance U, $r_g(n_k, S)$ represents the proportion of utterances classified by the gender classification model as speaker S out of the total utterances containing the personal noun n_k, and $\alpha = 0.895$ is a hyper-parameter to be optimized with accuracy through 10-fold cross-validation. The intuitive motivation of the formula (1) is to revise the probability $P_g(U, S)$ of the gender classification model by adding the proportion $r_g(n_k, S)$ of utterances classified by the gender classification model for all of the personal nouns $n_1, n_2, \ldots n_m$ within the utterance U.

For example, suppose the utterance U includes the second-person pronoun “君 (kimi)”. Then, the score $Q_{G+PN}(U, S)$ for S =“Marie” (female) and S =“Kyuros” (male) can be represented as follows:

$$Q_{G+PN}(U, S = \text{“Marie” (female)}) = P_g(U, S) \\ +\alpha r_g(\text{“君 (kimi)”}, S) \;\; (r_g = 0.331)$$

$$Q_{G+PN}(U, S = \text{“Kyuros” (male)}) = P_g(U, S) \\ +\alpha r_g(\text{“君 (kimi)”}, S) \;\; (r_g = 0.669)$$

Thus, even if in some cases, $P_g(U, S = \text{“Marie” (female)}) > P_g(U, S = \text{“Kyuros” (male)})$, by adding $\alpha r_g(\text{“君 (kimi)”}, S = \text{“Kyuros” (male)})$ ($= 0.895 \times 0.669$) to $P_g(U, S = \text{“Kyuros” (male)})$, finally $Q_{G+PN}(U, S = \text{“Kyuros” (male)})$ becomes greater than $Q_{G+PN}(U, S = \text{“Marie” (female)})$, then the classification error by the gender classification model can be recovered.

7 Evaluation

We used the dataset of utterances from two major characters described in section 3 for the evaluation experiment. When the gender classification model categorized an utterance as male, it was labeled as “Kyuros” (male), and when it classified an utterance as female, it was labeled as “Marie” (female).

Figure 2 shows the accuracy of the gender classification model (denoted as *gender*), the gender classification model + the constraint on personal nouns within quotes (denoted as +*personal nouns*), and the gender classification model + the speaker alternation constraint (denoted as +*alternation*) for 902 utterances in the upper half of Table 3 consisting of utterances only by “Marie” (female) and “Kyuros” (male). The gender classification model that was trained using an automatically collected dataset achieved an accuracy of 87.9% for the classification of the two main characters of the target novel. Both the speaker alternation constraint and the constraint on personal nouns within quotes were found to improve the classification performance. The speaker alternation constraint was satisfied for

Additional Constraints	Applicable Utterances	Classification Error before Application	After Application	
			Improved	Damaged
Speaker Alternation Constraint	902	109 (12.1%)	106 (11.8%)	4 (0.4%)
Constraint on Personal Nouns within Quotes	976	82 (8.4%)	31 (3.2%)	5 (0.5%)

Table 7: Detailed Analysis of the Performance of Each Constraint

Example Sentences	Speaker	Probability of the Model		corrected by constraints
		Marie (female)	Kyuros (male)	
どういたしまして。これは余り do, itashi mashite. kore wa amari (You are welcome. This is a surplus.)	Kyuros	**0.812**	0.188	corrected by the speaker alternation constraint
さあ、マリー saa Marie (Here, Marie)	Kyuros	**0.512**	0.487	corrected by the constraint on personal nouns within quotes

Table 8: Examples of Classification Error by the Gender Classification Model and Corrected by the Speaker Alternation Constraint / the Constraint on Personal Nouns Within Quotes

almost all cases, as shown in the upper half of Table 3. Because of this high rate, the speaker alternation constraint achieved an almost perfect accuracy of 99.2%. However, the constraint on personal nouns within quotes slightly improved the accuracy of the gender classification model by just 1.9 points.

The evaluation results for 1,306 utterances by "Marie" (female) and "Kyuros" (male) in the lower half of Table 3 are also shown in Figure 3, where we did not apply the speaker alternation constraint because those dialogues included characters other than "Marie" (female) and "Kyuros" (male). Their evaluation results are mostly similar to those in Figure 2, where the constraint on personal nouns within quotes slightly improved the accuracy of the gender classification model.

Table 7 shows a detailed analysis of each constraint's performance. As shown in Tables 3 and 4, the proportion of utterances to which each constraint is applicable was not insignificant. However, the performance of the gender classification model was relatively high. Table 7 shows that the proportion of classification error before the application of each constraint was not particularly high. Therefore, the improvement that can be achieved by each constraint was also relatively insignificant.

Table 8 shows examples of classification errors by the gender classification model, where those errors were corrected by the speaker alternation constraint or the constraint on personal nouns within quotes. As in the first example, utterances containing honorifics that are used regardless of gender, such as sentence-final auxiliary verbs "です (desu)" and "ます (masu)", have a higher probability of being classified as utterances by a female. The first example contains a conjugation form "まして (mashite)" of "ます (masu)" and causes a classification error. This example is corrected by the speaker alternation constraint. Next, the second example consists of only a few words or lacks linguistic features that indicate gender and is incorrectly classified. This example, however, contains the personal noun "Marie" within the quote, which enables it to be correctly classified as an utterance by "Kyuros" (male).

8 Concluding Remarks

This research described a method for constructing a novel dataset by automatically collecting sentences having gender characteristics using first-person pronouns as a cue and a gender classification model trained on the dataset. This gender classification model shows high classification performance in the male and female speaker classification, indicating that gender-specific linguistic features contribute to the speaker identification task in Japanese novels. Additionally, we found incorporating personal nouns within the utterance and the preceding and following utterances increased the classification performance in Japanese novels, as in the case of English novels reported in previous studies (O'Keefe et al., 2012; Muzny et al., 2017). Although the performance of the gender classification model was relatively high, our task setting of speaker identification was limited to the two major characters of the romance novel. Our future work includes developing a speaker identification model that links all utterances to appropriate speakers using the classification probability by the gender classification model as one feature. It is also necessary to broaden the dataset to a more diverse set of contemporary novels.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 21H00901.

References

Reina Akama, Kento Watanabe, Sho Yokoi, Sosuke Kobayashi, and Kentaro Inui. 2018. Unsupervised learning of style-sensitive word vectors. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 572–578, Melbourne, Australia, July. 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, June. Association for Computational Linguistics.

David Elson and Kathleen McKeown. 2010. Automatic attribution of quoted speech in literary narrative. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pages 1013–1019.

Hua He, Denilson Barbosa, and Grzegorz Kondrak. 2013. Identification of speakers in novels. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1312–1320, Sofia, Bulgaria, August. Association for Computational Linguistics.

Ryo Ishii, Ryuichiro Higashinaka, Koh Mitsuda, Taichi Katayama, Masahiro Mizukami, Junji Tomita, Hidetoshi Kawabata, Emi Yamaguchi, Noritake Adachi, and Yushi Aono. 2021. Methods for efficiently constructing text-dialogue-agent system using existing anime characters. *Journal of Information Processing*, 29:30–44.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. *Computing Research Repository, arXiv:1907.11692*.

Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. 2014. Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, 48(2):345–371.

Pierre-Emmanuel Mazaré, Samuel Humeau, Martin Raison, and Antoine Bordes. 2018. Training millions of personalized dialogue agents. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2775–2779, Brussels, Belgium, October-November. Association for Computational Linguistics.

Chiaki Miyazaki, Toru Hirano, Ryuichiro Higashinaka, Toshiro Makino, and Yoshihiro Matsuo. 2015. Automatic conversion of sentence-end expressions for utterance characterization of dialogue systems. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, pages 307–314, Shanghai, China, October.

Chiaki Miyazaki, Toru Hirano, Ryuichiro Higashinaka, and Yoshihiro Matsuo. 2016. Towards an entertaining natural language generation system: Linguistic peculiarities of Japanese fictional characters. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 319–328, Los Angeles, September. Association for Computational Linguistics.

Chiaki Miyazaki, Saya Kanno, Makoto Yoda, Junya Ono, and Hiromi Wakaki. 2021. Fundamental exploration of evaluation metrics for persona characteristics of text utterances. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 178–189, Singapore and Online, July. Association for Computational Linguistics.

Hajime Murai. 2018. Factor analysis of utterances in Japanese fiction-writing based on BCCWJ speaker information corpus. *Advances in Human-Computer Interaction*, 2018:1–9, 11.

Grace Muzny, Michael Fang, Angel Chang, and Dan Jurafsky. 2017. A two-stage sieve approach for quote attribution. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 460–470, Valencia, Spain, April. Association for Computational Linguistics.

Timothy O'Keefe, Silvia Pareti, James R. Curran, Irena Koprinska, and Matthew Honnibal. 2012. A sequence labelling approach to quote attribution. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 790–799, Jeju Island, Korea, July. Association for Computational Linguistics.

Iulian Vlad Serban, Ryan Lowe, Peter Henderson, Laurent Charlin, and Joelle Pineau. 2015. A survey of available corpora for building data-driven dialogue systems. *Computing Research Repository, arXiv:1512.05742*.

Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, Hiromi Narimatsu, Yuya Chiba, Hideharu Nakajima, and Toyomi Meguro. 2021. Empirical analysis of training strategies of transformer-based japanese chit-chat systems. *Computing Research Repository, arXiv:2109.05217*.

Makoto Yamazaki, Yumi Miyazaki, and Wakako Kashino. 2018. Annotation and quantitative analysis of speaker information in novel conversation sentences in Japanese. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Chak Yan Yeung and John Lee. 2017. Identifying speakers and listeners of quoted speech in literary works. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 325–329, Taipei, Taiwan, November. Asian Federation of Natural Language Processing.

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, Melbourne, Australia, July. Association for Computational Linguistics.