

# Linguistic Individuality Transformation for Spoken Language

Masahiro Mizukami, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura

**Abstract** In text and speech, there are various features that express the individuality of the writer or speaker. In this paper, we take a step towards the creation of dialogue systems that consider this individuality by proposing a method for transforming individuality using a technique inspired by statistical machine translation (SMT). However, finding a parallel corpus with identical semantic content but different individuality is difficult, precluding the use of standard SMT techniques. Thus, in this paper we focus on methods for creating a translation model (TM) using techniques from the paraphrasing literature, and a language model (LM) by combining small amounts of individuality-rich data with larger amounts of background text. We perform an automatic and manual evaluation comparing the effectiveness of three types of TM construction techniques, and find that the proposed system using a method focusing on a limited set of function words is most effective, and can transform individuality to a degree that is both noticeable and identifiable.

## 1 Introduction

In language, the words chosen by the speaker or writer transmit not only semantic content but also other information such as aspects of their individuality, personality, or speaking style. While not directly related to the message, these aspects of language are extremely important to achieve rapport between the person creating the message and its intended target. We can assume that this observation will also carry over to human computer interaction [1].

For example, in a situation where a dialogue system is used to represent famous characters in movies or comics, we would like to reproduce the character's well know and unique expressions. It is also natural that a dialogue system can realize smoother communication by talking in a more polite way to adults, and a more friendly and informal way to children [2]. To make these sorts of applications pos-

---

Masahiro Mizukami, Graham Neubig, Sakriani Sakti, Tomoki Toda, Satoshi Nakamura  
NAIST, 8916-5 Takayama-cho, Ikoma, Nara 630-0192, Japan,  
e-mail: {masahiro-mi, neubig, ssakti, tomoki, s-nakamura}@is.naist.jp

sible, it is necessary the ability to express a rich variety of individuality and atmosphere depending on the type of user or scene [3, 4].

In this paper, we define the individuality as the elements which allows us to distinguish unique person from other person. Individuality is closely related to personality, and previous work has modeled personality using measures such as the Big Five Traits [5]. Previous work has also noted that coherence of acoustic and linguistic traits has a strong influence on perceptions of individuality [6].

Handling of individuality of features of the voice (i.e. “acoustic individuality”) is a widely researched topic in speech synthesis and translation [7, 8, 9]. On the other hand, there are few studies that attempt to control the individuality of each speaker as expressed on the lexical level through choice of words or expressions, etc. (i.e. “linguistic individuality”). There are some works that attempt to generate sentences that express a certain personality based on rule-based sentence generation [4], personality infused  $n$ -gram models [3]. However, while controlling personality is certainly a first step in the direction of creating a richer user experience, research in the area overall is sparse, and controlling personality will not allow us to, for example, reproduce the unique expressions of a single speaker.

In this paper, we propose a technique that takes text as input, and converts the text into text that reflects the individuality of a target speaker. This approach has two differences from the previously mentioned work on personality-sensitive natural language generation. The first is that our method handles not *generation*, but *transformation*, taking as input a natural language sentence and converting the individuality of the source speaker into that of the target speaker. This has the advantage that it can be used as a post-processing step either for dialogue systems where generation is used as a black box, or for other applications that do not explicitly use generation, such as machine translation. In addition, by focusing not on *personality*, but *individuality*, we are able to cover applications such as the previously-mentioned dialogue system mimicking a famous character.

We propose the probabilistic framework for transforming individuality like statistical machine translation. This framework is based on previous work [10, 11, 12] that uses machine translation techniques to translate between speaking or writing styles. However, in contrast to these works, which rely on parallel data of the source and target styles, it is difficult to prepare a large quantity of parallel data between source and target speakers for individuality translation.

In this framework, we define a translation model (TM) that has the ability to translate between individualities of speakers, and a language model (LM) that has reflects the individuality of the target speaker. For the LM, we use a small collection of text created by the target speaker and a larger background model. For the TM, as it is difficult to create the parallel data necessary to train standard MT systems, we examine techniques from the paraphrasing literature, acquiring paraphrases using a thesaurus, distributional similarity, and bilingual parallel text.

Based on the results of the analysis, we find that in the system proposed in this paper, conversion of function words allows for detectable and identifiable increases in the individuality of the target sentence. On the other hand, conversion of content words is less successful, leaving important challenges for future work.

## 2 A Probabilistic Framework for Transforming Individuality

In this section, we describe our proposed method for translation of speaker individuality. To create a method capable of this conversion, we build upon previous work that has studied conversion of writing or speaking style [10, 11, 12].

Specifically, we build upon the work of Neubig et al. [12], which was originally conceived for translation from spoken to written text, or for translation of text from one style to another. Given a string of input words  $V$  (representing a spoken language sentence) and a string of words  $W$  (representing a written language sentence), we transform  $V$  to  $W$  using the noisy channel model. In consideration of the quantity of available corpora, the posterior probability  $P(W|V)$  is decomposed into TM probability  $P(V|W)$ , which must be estimated from a corpus of parallel sentences, which is more difficult to find, and LM probability  $P(W)$ , which can be estimated from a corpus of only output side text which we can secure in large quantities:

$$P(W|V) = \frac{P(V|W)P(W)}{P(V)}. \quad (1)$$

Given this probabilistic model, the output is found by searching for the output sentence  $\hat{W}$  that maximizes  $P(W|V)$ .  $P(V)$  is not affected by choice of  $W$ , so this maximization is expressed as follows:

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(V|W)P(W). \quad (2)$$

In addition, because the LM probability  $P(W)$  tends to prefer shorter sentences, we also follow standard practice in machine translation [13] in introducing a word penalty proportional to sentence length  $|W|$ . We combine these three elements in a log-linear model, with parameters  $\lambda_{tm}$ ,  $\lambda_{lm}$ , and  $\lambda_{wp}$  as follows:

$$\hat{W} = \underset{W}{\operatorname{argmax}} \lambda_{tm} \log P(V|W) + \lambda_{lm} \log P(W) + \lambda_{wp} |W| \quad (3)$$

Following this framework, we consider a setting in which we translate from utterance  $V$  that expresses the individuality of the source speaker to utterance  $W$  that expresses the individuality of target speaker. However, compared to the previously mentioned style transformation or standard SMT, we are faced with a drastic lack of data. The amount of target side data  $W$  is limited, and we will often have no parallel data with identical semantic content expressed with the individuality of the target and source speakers. In fact, when we had one author of the paper attempt to make this data in preliminary experiments, we found that even when an annotator is available, creation of the data is quite difficult and time-consuming. If the annotator attempted to follow the semantic content of the input faithfully, it was difficult to express a rich variety of individuality, and when the annotator attempted to edit more freely, the individuality was expressed abundantly, but in many cases the semantic content changed too much to be used reliably training or testing data for the system.

In the next two sections, we describe how we build a system even in situations where no parallel data is available to train that TM probability  $P(V|W)$ .

### 3 Language Model

For transforming individuality, it is necessary to build an LM that express the individuality of the target speaker.

#### 3.1 Language Model Training

For transforming individuality, it is necessary to build an LM that expresses the individuality of the target speaker. In order to do so, we need to collect data that expresses the target speaker’s speaking style. In addition, it is better if the data used to train the LM matches the content of the data to be converted. Thus, an initial attempt to create an LM that expresses the speaking style of the target will start with gathering data from the speaker, and training an  $n$ -gram LM on this data.

#### 3.2 Language Model Adaptation

When we collect the utterance of only one target speaker and build an LM, it is difficult to collect a large number of utterances from any one speaker. Thus the contents covered by the LM are restricted. Therefore, an LM made with only data from the target speaker cannot estimate the LM probability  $P(W)$  accurately. To remedy this problem, in this paper we build a target LM that interpolates a small LM  $P_t(W)$  that is trained as explained in the previous section and an LM  $P_g(W)$  that is trained from a large-scale corpus. Using an interpolation coefficient  $\lambda$ , we combine these two models using linear interpolation

$$P(W) = \lambda P_t(W) + (1 - \lambda) P_g(W). \quad (4)$$

We calculate  $\lambda$  to generate LM  $P(W)$ , such that we achieve the maximum LM probability on a held out development set also created using data from the target speaker. Note that this framework is flexible, so we could also add an additional LM considering the personality of the speaker [3], but in this paper for simplicity we only use two models: the general domain, and with the target speaker’s individuality.

### 4 Translation Model

Now that we have modeled individuality in the LM, we must next create a translation model  $P(V|W)$  that expresses the possible transformations changing the style, but not the semantic content, of the utterance. However, as mentioned in Section 2, it is non-trivial to collect a corpus of sentences spoken by the source and target speaker while having the same meaning, so we will have to create this model without relying on a parallel corpus.

In this paper, we solve this problem by building the TM using techniques from paraphrasing. In this work, we focus on methods for paraphrasing using a thesaurus,  $n$ -gram-based distributional similarity, and bilingual parallel text, with each of the three resources playing a different role.

#### 4.1 Translation Model Using Thesauri

Thesauri are language resources specifying groups of synonyms and thus are a good resource for reliably finding semantically plausible transformations. The most

widely used thesaurus in the NLP community is Wordnet [14], and its counterpart in Japanese, our target language, is Japanese Wordnet [15]. The TM built using a THESAURUS is used to find replacement candidates based on synonyms for nouns and verbs, similarly to previous works on paraphrasing using thesauri [16].

Using this thesaurus, we build the TM according to the following procedure.

1. For each word in the input, search the WordNet with the word as the query.
2. When the word is found, acquire all synonyms from WordNet using the synset.
3. Calculate the TM probability (Section 4.3) for all words, and store them in the TM.

We show an example of the TM acquired by this method in Table 1.

#### ***4.2 Translation Model Using Distributional Similarity***

Thesauri have the advantage of providing broad coverage, but they also consist mainly of synonyms for nouns and verbs, and don't have data regarding synonymy of fillers, exclamations, particles and other function words. However, these elements are very important in expressing a number of aspects of language [17]. Especially in Japanese, particles at the end of the sentence and auxiliary verb particle have been noted as playing an important role in expressing individuality [18].

The TM is built according to the following procedure.

1. Prepare a list of function words by performing POS tagging on the training corpus and extracting all non-content words.
2. Count all 3-grams in the target speaker's utterances.
3. Find groups of  $n$ -grams that have a function word in the second position and the same first and third words, and add them to the set of potential synonyms.  
e.g.) that's so great, that's really great
4. Calculate the TM probability for all words, and store them in the TM.

We show an example of a TM acquired by this method in Table 2. We can extract non-content word and particle paraphrases. In this method, we don't consider meaning of words, and we sometime get wrong paraphrases of the meaning, for example, "it for you" and "it from you". We check this problem by evaluating transforming word error rate.

#### ***4.3 Calculation of Translation Model Probability***

While the two previous methods can find potential candidates for translation, it gives us no mechanism to determine how reliable these candidates are. However, we also found in preliminary experiments that simply assigning a uniform probability to all transformations in the previous sections was not sufficient to accurately decide when words are interchangeable. To solve this problem, we calculate TM probabilities using  $n$ -gram similarity.

We base our method on techniques to acquire synonyms from non-parallel corpora [19, 20]. In the previous works, similarity of the word itself is calculated from a non-parallel corpus according to the contextual similarity of the words.

**Table 1** A sample of the TM using thesauri.

Source	Target	TM prob.
カメラ	カメラ (camera)	0.95
(camera)	カメラ (kamera)	0.01
	ビデオカメラ (video camera)	0.01
	写真機 (photo machine)	0.01
	and other 2 words	
良い	良い (good)	0.4
(good)	いい (nice)	0.4
	よろしい (fine)	0.01
	見事 (excellent)	0.01
	and other 42 words	

**Table 2** A sample of the TM using  $n$ -grams.

Source	Target	TM prob.
です (is)	です (is)	0.7
	だ (is: informal)	0.3
けど (but)	けど (but)	0.8
	よ (yes)	0.2
も (also)	も (also)	0.6
	で (at)	0.4
が (SUBJ)	が (SUBJ)	0.6
	は (SUBJ)	0.4

In order to calculate this contextual similarity, we prepare a bigram LM with vocabulary  $L$ , and decide the similarity  $\text{Sim}(w, v)$  for two words  $w$  and  $v$  as follows:

$$\text{Sim}(w, v) = 1 - \frac{1}{2|L|} \left( \sum_{l \in L} |P(w|l) - P(v|l)| + \sum_{l \in L} |P(l|w) - P(l|v)| \right). \quad (5)$$

Similarity  $\text{Sim}(w, v)$  is decided by the similarity of  $n$ -gram distributions, based on the distributional hypothesis that words that appear in similar contexts have a similar role. For the calculated similarity  $\text{Sim}(w, v)$ , we normalize over values of  $\text{Sim}(w, v)$  for all words, so that the probabilities sum to one

$$P(w|v) = \frac{\text{Sim}(w, v)}{\sum_{l \in L} \text{Sim}(l, v)}. \quad (6)$$

Thus, we can approximate TM probability of words  $w$  and  $v$  without using a parallel corpus.

#### 4.4 Translation model Using Bilingual Text

The final method we examine for creating the TM is based on [21]’s method for using bilingual text to train a paraphrasing model. Paraphrases acquired by this method have the advantage of providing broad coverage (theoretically it is possible to cover both content and function words), and allowing for the acquiring of multi-word transformations.

Assume we have two phrases  $v$  and  $w$  in the language under consideration (in our case, Japanese), and also have a phrase-based TM indicating the translation probabilities to and from a phrase  $e$  in a different language (in our case, English).

**Table 3** The details of the phrase table.

Corpus	BILINGUAL corpus including Wikipedia, lecture, newspaper, magazine and dialogue
Words	24.2M (en) 29.6M (ja)
Phrases	67.1M
Max length	7 words
Alignment	Nile [23]
Parsing	Kytea [24]

**Table 4** A sample of paraphrase acquired from bilingual data for “翻訳された (translated)”.

Translation	TM prob
翻訳された (translated)	0.083
に翻訳された (translated to)	0.034
翻訳 (translate)	0.012
共訳 (joint translation)	0.011
訳される (was translated)	0.011
と訳された (was translated to)	0.002
and 20 other phrases	

We decide the paraphrase probability  $P(w|v)$  using translation probabilities  $P(w|e)$  and  $P(e|v)$  by using the English phrase  $e$  as a pivot as follows:

$$P(w|v) = \sum_e P(w|e)P(e|v). \quad (7)$$

The TM probabilities can be computed using standard methods from SMT [22]. The details of the phrase table that we used in the construction of paraphrases for this work is shown in Table 3.<sup>1</sup>

We show an example of a TM acquired by this method in Table 4.

## 5 Evaluation Measures for Individuality Transformation

In previous work, they evaluate the relationship between some automatic evaluation metrics and various human judgments. Automatic metrics based on LMs have better correlation with human judgments than existing metrics in the context of previous work. We evaluate our proposed method under the same conditions as previous work’s automatic evaluation metrics as LM.

In manual evaluation, they evaluate based on human judgments of semantic adequacy, lexical dissimilarity and stylistic similarity, because they clarify style and the relations with individual elements. We consider it, propose several evaluation measures for transforming of individuality that focus on *individuality* of the target speaker, *accuracy* of conversion, and *breadth* of possible conversion.

### 5.1 Automatic Evaluation

In automatic evaluation, we use the two following measures.

**LM Ratio** Xu et al, [10] proposed a method for evaluating the style of a converted sentence using the ratio of language model probabilities, where  $P_t$  is the probability of a model trained on target domain data, and  $P_s$  is the probability of

<sup>1</sup> This Japanese paraphrase model will be made available upon acceptance of the paper.

a source domain language model:

$$P(\text{style} = \text{target}|\text{sentence}) = \frac{P_t(\text{sentence})}{P_s(\text{sentence}) + P_t(\text{sentence})}. \quad (8)$$

**Coverage** We define coverage as the ratio of words for which there is a conversion candidate in the TM. A TM that can convert various vocabulary will have a higher coverage, and thus coverage can be used to evaluate the breadth of the conversion.

## 5.2 Manual Evaluation

While automatic evaluation is useful for the rapid development of systems, it is difficult to evaluate small differences in nuance. Thus, we also perform manual evaluation to evaluate correctness and individuality of the output. Specifically, we evaluate two following factors.

**Individuality** In order to evaluate individuality, we first have a subject read the training data to learn the individuality of target speaker. The subject is then shown the system output and asked “does this sentence reflect the individuality of person who wrote the training data?” The subject then assigns a score of 1 (do not agree) to 5 (do agree).

**Word Error Rate; WER** This is the ratio of words converted by our method that are syntactically or semantically incorrect in the post-conversion sentence. This is calculated by having the subject look at the sentence before and after conversion and point out conversion mistakes.

## 6 Experimental Evaluation

In order to evaluate the proposed method, we performed an evaluation focused on how well the proposed model can reproduce the individuality of a particular speaker.

### 6.1 Experiment Conditions

As data for our research, we use a camera sales dialogue corpus [25] that consists of one-on-one sales dialogues between three salesclerks and 19 customers. We split the corpus of three salesclerks into one corpus for every speaker each and further divide each of these corpora into training, development, and evaluation data. The details of the data for each of the salespeople is shown in Table 5. All conversations were performed in Japanese by native or highly fluent Japanese speakers. As mentioned in Section 3.2, in order to create an LM that is both sufficiently accurate and expresses the personality of the speaker, we use multiple LMs created using data from the target speaker and a larger background corpus. As our target speaker data, we use the training data from the previously described camera sales corpus. As our large background corpus, we use data from the BTEC [26], and the REI-JIRO<sup>2</sup> dictionary example sentence corpus. The size of these background corpora is also shown in Table 6. We calculate the linear interpolation parameter to maximize

<sup>2</sup> <http://eihiro.jp>

**Table 5** Number of utterances and words in the camera sales dialogue corpus.

	Clerk	Ut.	Word
Train	A	238	11,758
	B	240	12,495
	C	228	9,039
Dev.	A	65	3,016
	B	43	2,271
	C	37	1,462
Test	A	9	173
	B	9	134
	C	9	148

**Table 6** Number of sentences and words in BTEC, and REIJIRO.

Corpus	Sent.	Word
BTEC	465k	4.11M
REIJIRO	424k	8.90M
SUM	889k	13.01M

likelihood on the development data. As a result, the linear interpolation parameter  $\lambda$  became 0.88. For the log-linear model in Equation (3), we set  $\lambda_{tm} = \lambda_{lm} = 1$ , and adjust the word penalty so that the length of sentences before and after transformation is approximately equal.<sup>3</sup>

We perform an evaluation over 3 combinations of TMs for conversion of individuality. We compare the three methods for constructing the TM using the thesaurus (THESAURUS),  $n$ -gram similarity (SIMILARITY), and parallel corpus (BILINGUAL). We also compare with a baseline method that does not perform any conversion at all (SOURCE).

In the experimental evaluation, we first have subjects read the training data of the target speaker. Next, we prepare an input sentence that is selected randomly from other salesclerks. Based on this input sentence, we use the three methods described in the previous paragraph to convert it into the target speaker’s individuality. The subject reads these three results. The subject estimates WER and individuality for each of these four conversion results according to the measures described in Section 5.2, and we also automatically calculate LM measure and coverage according to the measures described in Section 5.1.

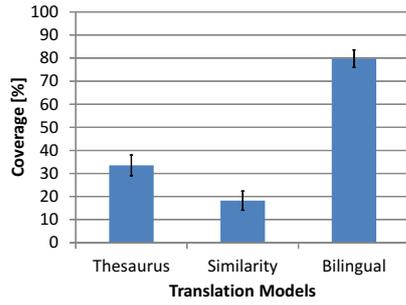
In this evaluation, three subjects evaluate result for 3 speakers, each with 9 utterances, 27 conversion results in total. We find the confidence interval of each evaluation measure using bootstrap resampling [27] with significance level  $p < 0.05$ .

## 6.2 Experiment Result

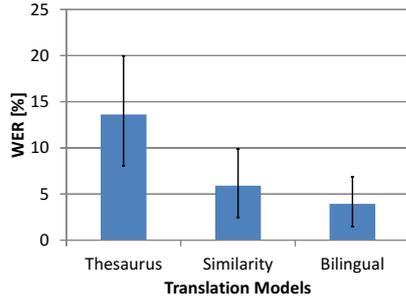
In this section, we describe the results of our evaluation of the proposed method for transforming individuality of text. We first discuss the results for the automatic evaluation measures. In Figure 1 we show the coverage, in Figure 2 we show the LM ratio.

In this evaluation, when we used BILINGUAL, coverage improved most, with a total of 80% of words being possible candidates for replacement. However, when

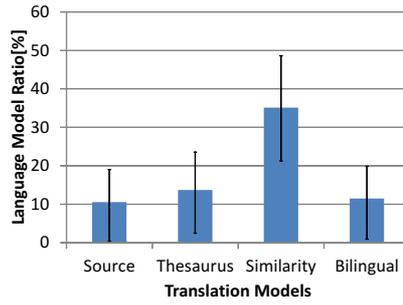
<sup>3</sup> Verbosity is one component of individuality, so setting  $\lambda_{wp}$  to a different value for each source/target speaker pair is more appropriate, but we leave this to future work.



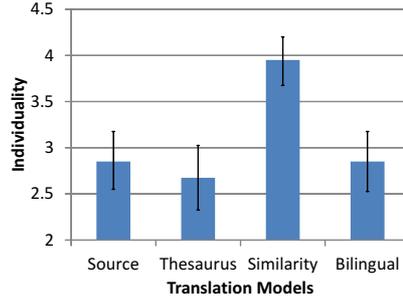
**Fig. 1** Coverage for each model.



**Fig. 3** WER for each model.



**Fig. 2** Language Model Ratio for each model.



**Fig. 4** Individuality for each model.

we used BILINGUAL, the percentage of words changed was 7.6%, lower than that of SIMILARITY, with a total of 13.0%. This is because function words (acquired by SIMILARITY) are more easily replaceable than content words or mixed phrases containing both function and content words. In addition, when we used the SIMILARITY TM, LM ratio improved most, with a total of 35% from 10% in SOURCE.

We show the results of manual evaluation of WER in Figure 3, and individuality in Figure 4. The first result to be noted is that transformation using SIMILARITY is able to raise the individuality to 3.9 from the SOURCE of 2.8, a significant difference. If we compare it with Figure 4, LM ratio and individuality understand a similar thing evaluating. This demonstrates that our proposed method of transforming the individuality of speakers is able to successfully do so to a noticeable degree.

However, the results given the other two methods were mixed. For THESAURUS, we can see the individuality actually unchanged. This is due to the fact that the WER of this method was high, and often the meaning of the sentence was lost due to mistaken conversions of content words. This unnaturalness resulted in very low evaluations of individuality.

When we used BILINGUAL, coverage and WER generally improved, and change rate improved over THESAURUS (which was 4.3%). However, LM ratio and individuality didn't improve over the SOURCE. The reason for this is that function words are very important for expressing individuality [17, 18], but were not well enough represented in the paraphrases acquired from bilingual data. The reason for this is twofold. First, data that is translated between languages usually contains few fillers

(as they are deleted before translation), and other common spoken expressions. Second, in our case we used English and Japanese for the pivot languages, but these languages diverge in their use of function words (for example, English does not use explicit case markers, and Japanese does not use articles), making acquiring good transformations for these words difficult. This is illustrated by the fact that BILINGUAL only covers a total of 11% of the transformations covered by SIMILARITY.

## 7 Conclusion

In this paper, we proposed a method for transforming individuality. We performed an evaluation of the effectiveness of TMs acquired using  $n$ -gram similarity, thesauri, and bilingual text in this context. We found that function word transformations based on  $n$ -gram similarity were the most effective in improving the individuality of text.

While the experimental results showed that the proposed technique is able to successfully convert speaker individuality to some extent, there are still a number of future challenges related to refining the language and TMs to convert speaker individuality more precisely. The main area for improvement lies in improvements of the TM, particularly the handling of function words in paraphrasing models acquired from bilingual text. We also plan on construct LMs that can evaluate speaker individuality in consideration of conversation context, and experimenting on larger data from the web.

## References

1. Florian Metze, Roman Englert, Udo Bub, Felix Burkhardt, and Joachim Stegmann. Getting closer: tailored human-computer speech dialog. *Universal Access in the Information Society*, Vol. 8, No. 2, pp. 97–108, 2009.
2. William Yang Wang, Samantha Finkelstein, Amy Ogan, Alan W Black, and Justine Cassell. Love ya, jerkface: using sparse log-linear models to build positive (and impolite) relationships with teens. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 20–29, 2012.
3. Amy Isard, Carsten Brockmann, and Jon Oberlander. Individuality and alignment in generated dialogues. In *Proceedings of the Fourth International Natural Language Generation Conference*, pp. 25–32, 2006.
4. François Mairesse and Marilyn A Walker. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, Vol. 37, No. 3, pp. 455–488, 2011.
5. Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in personality*, Vol. 37, No. 6, pp. 504–528, 2003.
6. Katherine Isbister and Clifford Nass. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *Int. J. Hum.-Comput. Stud.*, Vol. 53, No. 2, pp. 251–267, August 2000.
7. Masanobu Abe, Satoshi Nakamura, Kiyohiro Shikano, and Hisao Kuwabara. Voice conversion through vector quantization. In *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, pp. 655–658, 1988.
8. Junichi Yamagishi, Bela Usabaev, Simon King, Oliver Watts, John Dines, Jilei Tian, Yong Guan, Rile Hu, Keiichiro Ooura, Yi-Jian Wu, et al. Thousands of voices for HMM-based speech synthesis—analysis and application of TTS systems built on various ASR corpora. *IEEE Transactions on, Audio, Speech, and Language Processing.*, Vol. 18, No. 5, pp. 984–1004, 2010.

9. Yao Qian, Frank K. Soong, and Zhi-Jie Yan. A unified trajectory tiling approach to high quality speech rendering. *IEEE Transactions on Audio, Speech & Language Processing*, Vol. 21, No. 2, pp. 280–290, 2013.
10. Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, and Colin Cherry. Paraphrasing for style. In *Proceedings of Computational Linguistics 2012*, pp. 2899–2914, December 2012.
11. Eric Brill and Robert C Moore. An improved error model for noisy channel spelling correction. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pp. 286–293, 2000.
12. Graham Neubig, Yuya Akita, Shinsuke Mori, and Tatsuya Kawahara. A monotonic statistical machine translation approach to speaking style transformation. *Computer Speech & Language*, Vol. 26, No. 5, pp. 349–370, 2012.
13. Franz Josef Och and Hermann Ney. Discriminative training and maximum entropy models for statistical machine translation. In *Proc. Association for Computational Linguistics*, 2002.
14. George A. Miller. Wordnet: A lexical database for English. *Communications of the ACM*, Vol. 38, pp. 39–41, 1995.
15. Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. Enhancing the Japanese wordnet. In *Proceedings of the 7th Workshop on Asian Language Resources*, pp. 1–8, 2009.
16. Kentaro Inui and Atsushi Fujita. A survey on paraphrase generation and recognition. *Journal of Natural Language Processing*, Vol. 11, No. 5, pp. 151–198, 2004.
17. Cindy Chung and James W Pennebaker. The psychological functions of function words. *Social communication*, pp. 343–359, 2007.
18. Mihoko Teshigawara and Satoshi Kinsui. Modern Japanese “Role Language”(Yakuwarigo): fictionalised orality in Japanese literature and popular culture. In *Sociolinguistic Studies Vol 5-1*, 2012.
19. Ido Dagan, Lillian Lee, and Fernando CN Pereira. Similarity-based models of word cooccurrence probabilities. *Machine Learning*, Vol. 34, No. 1-3, pp. 43–69, 1999.
20. Regina Barzilay and Lillian Lee. Learning to paraphrase: an unsupervised approach using multiple-sequence alignment. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, pp. 16–23, 2003.
21. Colin Bannard and Chris Callison-Burch. Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 597–604, 2005.
22. Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 48–54, 2003.
23. Jason Riesa, Ann Irvine, and Daniel Marcu. Feature-rich language-independent syntax-based alignment for statistical machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 497–507, 2011.
24. Graham Neubig, Yosuke Nakata, and Shinsuke Mori. Pointwise prediction for robust, adaptable Japanese morphological analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, pp. 529–533, 2011.
25. Takuya Hiraoka, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. Construction and analysis of a persuasive dialogue corpus. In *5th International Workshop on Spoken Dialog Systems (IWSDS)*, 2014.
26. Toshiyuki Takezawa, Eiichiro Sumita, Fumiaki Sugaya, Hirofumi Yamamoto, and Seiichi Yamamoto. Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world. In *Proceedings of Language Resources and Evaluation Conference*, pp. 147–152, 2002.
27. Philipp Koehn. Statistical significance tests for machine translation evaluation. In *Conference on Empirical Methods on Natural Language Processing*, pp. 388–395, 2004.