Automatic conversion of sentence-end expressions for utterance characterization of dialogue systems

Chiaki Miyazaki Toru Hirano Ryuichiro Higashinaka
Toshiro Makino Yoshihiro Matsuo
NTT Media Intelligence Laboratories
1-1 Hikarinooka, Yokosuka, Kanagawa, Japan
{miyazaki.chiaki, hirano.tohru, higashinaka.ryuichiro,
makino.toshiro, matsuo.yoshihiro}@lab.ntt.co.jp

Abstract

Building characters for dialogue agents is important in making the agents more friendly and human-like. To build such characters, utterances suitable for the designated characters are usually manually prepared. However, it is expensive to do this for a large number of utterances for various types of characters. We propose a method for automatically converting system utterances into those that are characteristic of designated personal attributes, such as gender, age and area of residence, to characterize agents. In particular, we focus on converting sentence-end expressions, which are considered to greatly affect personal attributes in Japanese. Conversion is done by (i) automatically collecting conversion candidates from various utterances on the Web (e.g., Twitter postings), and (ii) using syntactic and semantic filters to suppress the generation of ill-formed utterances. Experimental results show that our method can convert approximately 95% of utterances into those that are grammatically and semantically acceptable and approximately 90% of utterances into those that are perceived to be acceptable for designated personal attributes.

1 Introduction

Dialogue agents, which can carry out various tasks according to user demand, have been gaining in popularity due to their convenience and potential in casual conversations with humans. To make the agents more attractive as conversation partners, characterization is important since it makes the agents more friendly and human-like. *Characterization* here means adding particular personal characteristics to agent utterances; for example, adding the characteristics of a particular person (Mizukami et al., 2015), Big Five personalities (Mairesse and Walker, 2007), or personal attributes such as gender, age and area of residence (which is closely related to dialects). To characterize agents, utterances suitable for the designated characteristics are usually manually prepared. However, it is expensive to do this for a large number of utterances.

To reduce this cost, we propose a method for automatically converting utterances into those that are suitable for various characters. In particular, the method automatically modifies 'how to say it' (i.e., linguistic expressions) without changing 'what to say' (i.e., contents of the utterances). Conversion is done by (i) collecting conversion candidates from various utterances on the Web (e.g., Twitter postings), which are annotated with their authors' personal attributes (this paper deals especially with gender, age, and area of residence), and (ii) using syntactic and semantic filters to suppress the generation of ill-formed utterances.

The rest of the paper is organized as follows. Section 2 introduces studies related to characterization, Section 3 discusses the features of Japanese sentence-end expressions, Section 4 presents our method for converting sentence-end expressions, Section 5 shows our experimental results, and Section 6 concludes the paper and refers to future work.

2 Related work

Studies related to characterization of dialogue agent utterances have been conducted. For example, a

29th Pacific Asia Conference on Language, Information and Computation pages 307 - 314
Shanghai, China, October 30 - November 1, 2015

method for transforming individual characteristics in dialogue agent utterances (Mizukami et al., 2015) and a language generator that can control parameters related to speakers' Big Five personalities (PERSONAGE) (Mairesse and Walker, 2007) have been proposed. There is also a method for automatically adjusting the language generation parameters of PERSONAGE by using movie scripts (Walker et al., 2011) and a method for automatically adjusting the parameters so that they suit characters or stories of role playing games (Reed et al., 2011).

These studies, including ours, share the same motivation to control personal characteristics of utterances. However, there have not been any studies on converting utterances from the viewpoint of personal attributes. This is mainly because there has been few resources containing utterances together with the personal attributes of interlocutors. The novelty of our work lies in using Twitter as such a resource to mine sentence-end expressions anchored to particular personal attributes.

3 Sentence-end expressions in Japanese

We focus on sentence-end expression since, in Japanese, it is a core element of *role language* (Kinsui, 2003), which relates to stereotypical or characteristic wordings of particular personal attributes such as *feminine language* and *youth language*. We assume that converting sentence-end expressions can be effective in modifying the characteristics of agent utterances. For example, though the utterances shown below differ only in sentence-end expressions, Japanese native speakers can detect the differences in assumed writer/speaker personal attributes.

- gakkoo -ni iki **-tai -no -kayo** [masculine]
- gakkoo -ni iki **-tai -no -kashira** [feminine]
- gakkoo -ni iki **-tai -n -kaina** [western dialect-like]

In these utterances, function words are marked with '-' and those that correspond to sentence-end expressions are in bold. These utterances all convey the meaning that corresponds to 'Do you want to go to school?' in English.

We define a sentence-end expression as a sequence of function words that occurs at the end of a sentence. Function words are those except for content words, such as nouns, verbs, adjectives, and adverbs. The basic role of function words is to denote relations between words, phrases, and clauses, such as case markers (e.g., subject markers and object markers) and connectives (i.e., conjunctions and conjunctive particles).

Japanese sentence-end expressions also play an important role in interaction. Japanese sentence-end expressions contain *interactional particles* (Maynard, 1997), which express speaker judgment and attitude toward the message and the hearer. For instance, 'ne' (a marker of the speaker's assumption that he/she has less information than the hearer; an English counterpart would be "isn't it?") occurs at the end of utterances. In addition, Japanese sentence-end expressions contain auxiliary verbs (e.g., 'mitai' (like) and 'souda' (it seems)), which express speaker attitudes.

4 Method for converting sentence-end expressions

We propose a method for converting sentence-end expressions to characterize dialogue agent utterances. Figure 1 shows the process of the sentence-end expression conversion. First, as preparation, sentence-end expressions, which are characteristic of target characters, are collected through processes (1) and (2) shown in Figure 1 (details are given in Section 4.1). Second, each input utterance is processed in process (1) to find a sentence-end expression to be converted. Here, sequences of function words at the end of sentences are detected as sentence-end expressions according to the part-of-speech (POS) tags obtained using a morphological analyzer (Fuchi and Takagi, 1998). Third, through process (3), appropriate candidates to be substituted for the original sentence-end expression are selected using two filters: POS adjacency and semantic label. Finally, a converted utterance whose sentence-end expression is substituted with one of the candidates is generated as an output.

4.1 Extracting characteristic sentence-end expressions

This section explains a corpus from which the characteristic sentence-end expressions are extracted and the method for extracting the expressions.

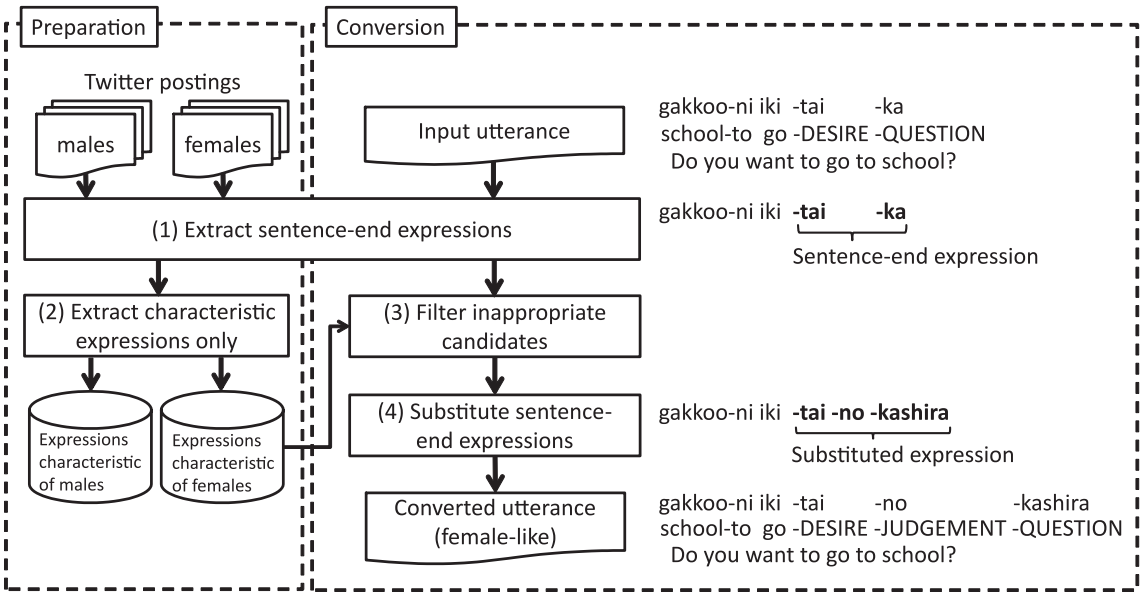

Figure 1: Flow of sentence-end conversion

Attribute	Value	# of authors
gender	female	810
	male	870
age	under 40	1070
	40 and over	610
area of residence	eastern Japan	979
	western Japan	701

Table 1: Author attributes and number

4.1.1 Twitter corpus

For collecting sentence-end expressions, which are characteristic of targeted characters, we use a corpus consisting of Twitter postings that are annotated with their authors' personal attributes (Hirano et al., 2013). The corpus includes two million postings written by 1680 authors. The annotation of the authors' personal attributes to the postings was done based on the self-declarations by the authors. The number of authors for each personal attribute-value is shown in Table 1.

4.1.2 Method for extracting characteristic sentence-end expressions

From each posting of the Twitter corpus, we extract the sequences of function words at the end of the sentences as sentence-end expressions (sentences are period-delimited sequences of words). Then, for each expression, we count the numbers of authors who used them. The numbers of authors are counted separately according to their gender, age, and area of residence. Table 2 lists examples of sentence-end expressions and number of authors who used the corresponding expressions. Then, to extract characteristic sentence-end expressions, the numbers of authors who used each expression are compared. For example, when extracting expressions that are characteristic of female authors, the number of female and male authors who used the expression are compared. With our method, this comparison is done using the G-test. We regard a sentence-end expression as being characteristic of a specific attribute-value if (i) the p-value for the expression is less than a significance level of 5%, which means the number of authors who use the expression is not independent of their attribute, and (ii) if the proportion of authors who used the expression for the specific attribute-value is larger than that for the other value. For example, the expression 'いのだー (i-no-da)' in Table 2 is listed in Table 3 as a characteristic expression of females because its p-value is less than a significant level of 5% and the proportion of female authors who used the expression (14/810) is larger than that for male authors (1/870). Table 3 lists the examples of characteristic sentence-end expressions of females, western Japan, and under 40. In Table 3, some of the characteristic sentence-end expressions of females include the

Expressions	# of authors			
	female		male	
	used	not used	used	not used
いのだー (i-no-da)	14	796	1	869
いのだが (i-no-da-ga)	64	746	132	738
いのだけれど (i-no-da-keredo)	14	796	38	832
いのだし (i-no-da-shi)	0	810	4	866
いのだなぁ (i-no-da-naa)	0	810	4	866

Table 2: Examples of sentence-end expressions and number of authors who used corresponding expressions

	Expressions	G
females	いのよー (i-no-yo)	26.10
	いのよ (i-no-yo)	22.88
	いのー (i-no)	18.37
	いのよね (i-no-yo-ne)	16.20
	いのだー (i-no-da)	14.50
western Japan	いんやけど (i-n-ya-kedo)	19.24
	いんや (i-n-ya)	15.49
	いんやね (i-n-ya-ne)	9.93
	いんやけどね (i-n-ya-kedo-ne)	8.89
	いんやけどな (i-n-ya-kedo-na)	8.83
under 40	いんじゃね (i-n-ja-ne)	23.15
	いよなー (i-yo-na)	16.11
	いよぉー (i-yoo)	7.24
	いょ (i-yo)	6.59
	いよぅ (i-you)	6.53

Table 3: Examples of sentence-end expressions characteristic of females, western Japan, and under 40

expression 'のよ (no-yo)', which is a stereotypical feminine conversational wording in Japanese. In addition, all of the characteristic sentence-end expressions of western japan include the expression 'や (ya)', which is used as a copula in western dialect.

4.2 Part-of-speech adjacency filter

The POS adjacency filter is one of the filters that are used in process (3) in Figure 1. This filter works as a constraint for suppressing the conversion into ungrammatical utterances. This filter removes candidates that are not allowed to be adjacent to a content word on the left of the original sentence-end expression. In particular, the filter removes the candidates whose left adjacent POS is different from that of the original sentence-end expression. The left adjacent POSs of the candidate sentence-end expressions are also extracted and stored together with the candidates, as shown in Table 4.

Adjacent POS on left	Expressions
noun	だからな (da-kara-na)
noun	だからなぁ (da-kara-naa)
noun	だが (da-ga)
verb	ないが (nai-ga)
verb	ないし (nai-shi)
verb	ないじゃないか (nai-ja-nai-ka)
adjective	いです (i-desu)
adjective	いでした (i-deshi-ta)
adjective	いですか (i-desu-ka)

Table 4: Examples of adjacent content word's part-of-speech (POS)

Category	Sub-category	Semantic labels	Examples
factuality	polarity	negation	ない (nai)
	degree of certainty	question	か (ka)
		guess	だろう (darou)
	tense (aspect)	completion	た (ta)
		continuation	ている (te-iru)
intention		desire	たい (tai)
		volition	う (u)
		invitation	うか (u-ka)
		request	てください (te-kudasai)

Table 5: Semantic labels that should be consistent before and after conversion

4.3 Semantic label filter

The semantic label filter is another type of filter that is used in process (3) in Figure 1. We define a set of semantic elements that must be included in both sentence-end expressions before and after conversion. To this end, we use the nine semantic labels listed in Table 5, which were selected from 435 labels corresponding to the meaning categories for functional expressions (Matsuyoshi et al., 2006). From these, we select nine labels regarding the following two aspects: (i) factuality and (ii) intention, since we regard them as the key components of dialogue content.

(i) Semantic labels related to factuality Event factuality refers to the distinction whether

event-denoting expressions are presented as corresponding to real-world situations, situations that have not occurred, or situations of uncertain status (Saurí and Pustejovsky, 2007). According to them, event factuality is impacted by polarity (positive vs. negative) and degree of certainty of what is asserted (e.g., possible vs. certain). Tense (aspect) is also often discussed in relation to the meaning of an event (Izumi et al., 2010). Taking these into account, we select five semantic labels, *negation* for polarity, *question* and *guess* for degree of certainty, and *completion* and *continuation* for tense (aspect) to keep the factuality consistent before and after conversion.

(ii) Semantic labels related to intention

Intentions are defined here as what a speaker wants (Sidner and Israel, 1981) or as a discourse purpose (Grosz and Sidner, 1986). To keep the intention consistent before and after conversion, we select four labels, namely, *desire*, *volition*, *invitation*, and *request*. These labels are important for expressing what a speaker wants (to do) or wants his/her hearer to do.

The input utterances and postings in the Twitter corpus, from which the candidates are extracted, are automatically tagged with the semantic labels by using a method that selects the best sequence of semantic labels by a discriminative model (Imamura et al., 2011).

4.4 Conversion of sentence-end expressions

A sentence-end expression of the input utterance is converted through the steps shown in Table 1. First, the input utterance is processed to find a sentence-end expression along with the POS of its adjacent content word and the semantic labels included in it. Second, the pool of sentence-end expressions that are characteristic of a designated personal attribute is filtered with the syntactic and semantic filters (See Sections 4.2 and 4.3). Finally, the sentence-end expression of the input utterance is substituted with the conversion candidates that passed the filters.

When filtering the candidates, the POS of the last content word (the adjacent content word of the sentence-end expression) in the input utterance is used for removing the candidates whose left adjacent POS is different from the last content word of the input utterance. In addition, the semantic labels, which are included in the sentence-end expression of the input utterance, and those of the candidates are compared. If a candidate contains exactly the same set of labels, it remains a candidate; otherwise, the candidate is abandoned.

Adjacent POS	Semantic labels	Expressions	G
verb	DESIRE, JUDGMENT, QUESTION	たいのかしら (tai-no-kashira)	33.00
verb	DESIRE, QUESTION	たいかしら (tai-kashira)	31.19
verb	DESIRE, JUDGMENT, QUESTION, EXCLAMATION	たいのかなあー (tai-no-kanaa)	13.15
verb	DESIRE, QUESTION	たいですかっ (tai-desu-ka)	6.45
verb	DESIRE, QUESTION, EXCLAMATION	たいかなあー (tai-kanaa)	5.90

Table 6: Examples of candidates that passed filters

Consider the following utterance as an example of an input.

gakkoo -ni iki -tai -ka
school -GOAL go -DESIRE -QUESTION
N Particle V Aux Particle
'Do you want to go to school?'

In this utterance, the first line is the alphabetical transcription of the input utterance, and the second line is the semantic denotation that corresponds to the first line. In the semantic denotation, the meanings of content words are denoted in English counterparts and those of function words are denoted with semantic labels written in uppercase. The third line shows the POS of each word, and the fourth line shows the English translation of the input utterance.

In this utterance, a sequence of function words at the end of the utterance 'tai ka' is the sentence-end expression, which is to be converted. Since the sentence-end expression is adjacent to a verb, only

the candidates that are also capable of being adjacent to verbs can pass the POS adjacency filter. In addition, the input sentence-end expression includes two kinds of semantic labels, DESIRE and QUESTION. Therefore, only the candidates that also include both labels can pass the semantic label filter. Table 6 lists the examples of the surviving candidates that are characteristic of the female attribute.

In the example in Table 6, there are some semantic labels that are not included in the original sentence-end expression, such as JUDGMENT and EXCLAMATION. Since these labels are not considered with the semantic label filter, it does not matter if they are included in the candidates.

5 Experiments

We conducted two experiments to investigate the performance of our proposed method of converting sentence-end expressions. In particular, we asked a subject to score the converted utterances from the two perspectives of (i) grammatical and semantic acceptability, and (ii) appropriateness for desired personal attributes. The subject was a person who had been working as a linguistic annotator for more than three years. To evaluate inter-rater agreement, we also asked another subject to rate half the utterances.

5.1 Data for collecting conversion candidates and testing

For collecting candidates to be used for conversion, we used the Twitter corpus introduced in Section 4.1.1. The target personal attributes (and values) were gender (male/female), age (under 40/40 and over), and area of residence (eastern/western Japan), and the number of authors for each attribute-value is shown in Table 1.

As input utterances, we used 100 Japanese utterances, which were randomly extracted from a database consisting of manually created utterances (in the form of text) for a dialogue system, which we created. Examples of input utterances are shown below.

水族館が大好きです
suizokukan-ga daisuki-desu
'I like aquariums very much.'

占いって信じますか？
uranai-tte shinji-masu-ka?
'Do you believe in astrology?'

あなたの部屋から星が見えますか？
anata-no heya-kara hoshi-ga mie-masu-ka?
'Can you see stars from your window?'

These utterances were converted so that they would be characterized with six different personal attributes, i.e., male, female, under 40, 40 and over, living in eastern Japan, and living in western Japan. Though various sentence-end expressions were collected as the conversion candidates, we used only one expression whose G-value was the highest among the candidates.

5.2 Procedure and evaluation indices

We randomly presented the converted utterances and the original input utterances to the subjects and asked them to score the utterances regarding the following two aspects.

Grammatical and semantic acceptability Whether an utterance is acceptable in Japanese with respect to grammar and meaning (1: very unacceptable - 5: very acceptable).

Character acceptability Whether an utterance is acceptable regarding a desired characteristic (1: very unacceptable - 5: very acceptable).

Since it is difficult to clearly separate acceptability of grammar from semantics, we evaluated them together. We calculated the inter-rater agreement rate as the percentage of utterances for which the two subjects gave identical judgments.

5.3 Results

Figures 2 and 3 show the average scores of 100 utterances for each personal attribute. In the figures, ***, ** and * indicate statistical significance at the 0.001, 0.01 and 0.05 levels, respectively, and n.s. indicates "not significant". The average scores of characteristic acceptability of the converted utterances were

significantly higher than those of the original utterances (paired samples t-test; $p<0.05$) except for the case of 40 and over. In particular, the scores for the cases of under 40, male, female, and western Japan drastically improved (by approximately 0.8-2.0 points) due to the conversion.

Moreover, for the cases of female and western Japan, there were no significant differences in the average scores of grammatical and semantic acceptability between before and after conversion according to paired samples t-test. For the cases of the other attributes, the average scores of grammatical and semantic acceptability of the converted utterances were significantly lower than those of the original utterances (paired samples t-test; $p<0.05$). However, the average scores exceeded 4 (acceptable) except for the case of male. Therefore, we argue that our proposed method can convert utterances without severe malformation of grammar and semantics.

Tables 7 and 8 show the breakdown of scores of the two evaluation indices. For the evaluation of grammatical and semantic acceptability, unacceptable utterances scored 1 (very unacceptable) or 2 (unacceptable) were only 5% or less (The inter-rater agreement rate was 95% when distinguishing unacceptable utterances (2 or below) from acceptable ones) except for the case of male. For the evaluation of characteristic acceptability, the average percentage of unacceptable utterances scored 1 or 2 was 10%, which we believe is good (the inter-rater agreement rate was 85% when distinguishing unacceptable utterances (2 or below) from acceptable ones). However, unacceptable utterances of 40 and over and western Japan scored 1 or 2 exceeded 20%. Considering the practical use in dialogue systems, the results suggest that utterances that are not appropriate for a designated attribute are generated in approximately one in five utterances. Thus, we believe that the characterization is still not sufficient for certain personal attributes, and further investigation and improvement are needed.

6 Conclusion and future work

To build characters for dialogue agents, we proposed a method for automatically converting sentence-end expressions. Our contributions are as follows:

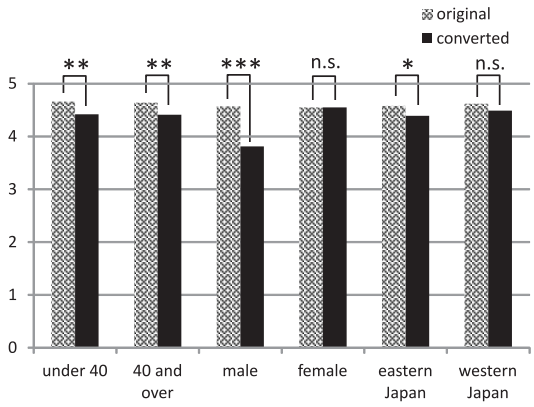

Figure 2: Average scores of grammatical and semantic acceptability

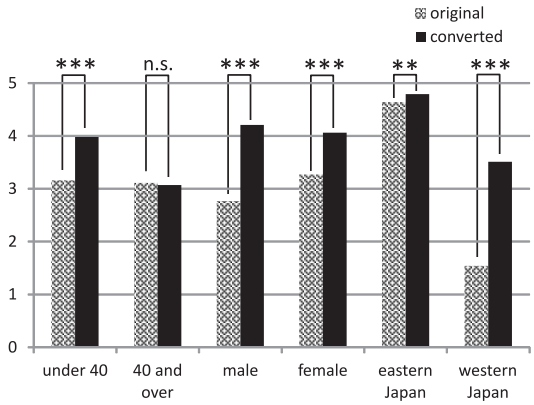

Figure 3: Average scores of character acceptability

	% of utts. for each score				
	1	2	3	4	5
under 40	**3%**	**2%**	1%	38%	56%
40 and over	**2%**	**1%**	8%	32%	57%
male	10%	13%	7%	26%	44%
female	**2%**	**0%**	4%	29%	65%
eastern Japan	**3%**	**1%**	7%	32%	57%
western Japan	**2%**	**3%**	4%	26%	65%

Table 7: Breakdown of scores of grammatical and semantic acceptability

	% of utts. for each score				
	1	2	3	4	5
under 40	**1%**	**3%**	9%	71%	16%
40 and over	2%	28%	31%	39%	0%
male	**0%**	**2%**	16%	41%	41%
female	**0%**	**0%**	21%	52%	27%
eastern Japan	**0%**	**0%**	3%	15%	82%
western Japan	12%	11%	34%	0%	43%

Table 8: Breakdown of scores of character acceptability

- We introduced an effective way of characterization for dialogue agent utterances in Japanese, i.e., conversion of sentence-end expressions.
- We presented a method for converting the sentence-end expressions with limited risk of being syntactically or semantically ill-formed.

These contributions are supported by the experimental results, which show that our method can, except for the case of male, convert approximately 95% of utterances into those that are grammatically and semantically acceptable, and approximately 90% of the converted utterances are perceived to be acceptable for designated personal attributes.

There are still limitations to our proposed method. For instance, conversion of content words is not possible. Since we assume that lexical choice of content words would also be an important component of characterization, we would like to investigate this as future work. In addition, the attributes dealt with in this study were limited to gender, age, and area of residence. The values for each of the attributes were also limited to binary distinctions, such as male/female, under 40/40 and over, and eastern/western Japan. As far as these attributes are concerned, characteristic expressions were successfully extracted from Twitter postings, but this might not be in the case of other attributes and values. Investigating how our proposed method works on different types of attributes and different sizes of data will also be necessary.

References

Takeshi Fuchi and Shinichiro Takagi. 1998. Japanese morphological analyzer using word co-occurrence - JTAG. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, volume 1, pages 409–413.

Barbara J Grosz and Candace L Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3):175–204.

Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. 2013. Inferring user profile from text using markov logic. In *Proceedings of the 27th Annual Conference of the Japanese Society for Artificial Intelligence (in Japanese)*.

Kenji Imamura, Tomoko Izumi, Genichiro Kikui, and Satoshi Sato. 2011. Semantic label tagging to functional expressions in predicate phrases. In *Proceedings of the 17th Annual Meeting of Association for Natural Language Processing (in Japanese)*, pages 308–311.

Tomoko Izumi, Kenji Imamura, Genichiro Kikui, and Satoshi Sato. 2010. Standardizing complex functional expressions in Japanese predicates: Applying theoretically-based paraphrasing rules. In *Proceedings of the 2010 Workshop on Multiword Expressions: from Theory to Applications*, pages 64–72.

Satoshi Kinsui. 2003. Vaacharu nihongo: Yakuwarigo no nazo (in japanese) [Virtual Japanese: The mystery of role language]. *Iwanami Shoten*.

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.

Suguru Matsuyoshi, Satoshi Sato, and Takehito Utsuro. 2006. Compilation of a dictionary of Japanese functional expressions with hierarchical organization. In *Proceedings of the 21st International Conference on Computer Processing of Oriental Languages*, pages 395–402.

S. K. Maynard. 1997. *Japanese Communication: language and thought in context*. University of Hawai'i Press.

Masahiro Mizukami, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. 2015. Linguistic individuality transformation for spoken language. In *Proceedings of the 6th International Workshop On Spoken Dialogue Systems*.

Aaron A Reed, Ben Samuel, Anne Sullivan, Ricky Grant, April Grow, Justin Lazaro, Jennifer Mahal, Sri Kurniawan, Marilyn A Walker, and Noah Wardrip-Fruin. 2011. A step towards the future of role-playing games: The SpyFeet Mobile RPG Project. In *Proceedings of the 7th Conference on Artificial Intelligence and Interactive Digital Entertainment*, pages 182–188.

Roser Saurí and James Pustejovsky. 2007. Determining modality and factuality for text entailment. In *Proceedings of the 1st IEEE International Conference on Semantic Computing*, pages 509–516.

Candace L Sidner and David J Israel. 1981. Recognizing intended meaning and speakers' plans. In *Proceedings of 7th International Joint Conference on Artificial Intelligence*, pages 203–208.

Marilyn A Walker, Ricky Grant, Jennifer Sawyer, Grace I Lin, Noah Wardrip-Fruin, and Michael Buell. 2011. Perceived or not perceived: Film character models for expressive NLG. In *Proceedings of the 4th International Conference on Interactive Digital Storytelling*, pages 109–121.