

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



## Automatic Generation of Appropriate Greeting Sentences using Association System

Eriko Yoshimura<sup>1</sup>, Seiji Tsuchiya<sup>2</sup>, Hirokazu Watabe<sup>1</sup>, Tsukasa Kawaoka<sup>1</sup>  
<sup>1</sup> Doshisha University, <sup>2</sup>The University of Tokushima  
<sup>1,2</sup>Japan

### 1. Introduction

When we humans start a conversation, we are greeting at first. If computer and robot are greeting like us, they can communicate smoothly with us because the next subject comes easily after greeting. That is to say, greeting conversation plays an important part to smooth communications in speaking. In this report, we describe a method of increase the number of appropriate greeting sentences for conversation and selecting sentences based on the situation by machine.

Many of conversation system tend to use templates. Lots of chatter bots (Eliza, A.L.I.C.E., Ractor, Verbot, Julia etc.) have been developed. For example, Eliza(Weizenbaum, J 1965) which is one of the well-known system acts for counselling by a personification therapist agent. Eliza does not evaluate an answer of a partner for the reply.

It memorizes only a part of the content that the partner spoke in the past and replies by using the word. It is prepared for several kinds patterns about the topic.

Like these, as for the natural language processing, task processing type conversation (e.g. automatic systems for tourist information and reservations) becomes the mainstream. However, even under the limited situation, it is known that it is difficult to make a knowledge base of all response case. Moreover, a method using only the prepared template makes monotonous reply and a reply except sentences made by a designer don't appear. So, to make various sentences automatically by machine is important, more than the method to select sentences designer prepared.

We herein propose an intelligent greeting processing by which a machine generates various reply sentences automatically by obtaining information about the surrounding state and then generating the best conversation response based on the situation.

All sentences are extended automatically from a small quantity of model sentences by using the concept base which is kind of natural-language ontology/concept networks. Simply mechanical extension of conversation sentences makes many improper sentences. So, the proposed method uses language statistics information to delete the improper sentences. In addition, for greetings conversation, we suggest "status space" expressing a certain situation. This is a model to give a weight to sentences automatically by taking the consequence at two

points of view that the appropriate selection that a human being performs unconsciously is classified in.

## 2. Requirements for Conversation Sentence

The greeting conversation is a “no insistence conversation” that does not cause argument or discussion. In the present paper, such greeting conversation sentences are synthesized automatically by machine.

The requirements for automatic conversation sentence synthesis by machine as follows:

- 1) Grammatical consistency
- 2) No contradiction in meaning
- 3) The use of usual words
- 4) Situation adaptability

“Grammatical consistency” refers to sentences in which no grammatical mistakes are found. “No contradiction in meaning” refers to sentences that have a reasonable meaning, e.g., the sentence “The sun is so bright tonight.” is not reasonable from the point of view of time. “The use of usual words” indicates words used in daily life, including colloquialisms. “Situation adaptability” refers to sentences that do not contradict reality. For example, “It’s a rainy today” contradicts the reality if the weather is fair.

Therefore, it is necessary to meet these requirements after a mechanical synthesis. The proposed system is constructed using the Japanese language and so is adapted to the characteristics of the Japanese language and Japanese culture.

## 3. Greeting Conversation System

Figure 1 shows the structure of the greeting conversation system proposed in the present paper. The greeting conversation system obtains inputs of the surrounding information and input sentence and then outputs greeting sentences. There are fixed pattern greetings, e.g., “Good morning” and “Hello” and greetings for starting a conversation, e.g., “It’s been raining all day.” and “It’s very hot, isn’t it?”:

Human: “Good morning.”

System: “Good morning.

It’s very hot, isn’t it?”

The former output is a fixed pattern greeting, the system can output sentence matching of the situation or input-sentences by a fixed pattern knowledge base. This problem can easily be solved. Thus, the latter greeting for starting a conversation is considered in the present paper. Greeting sentences indicate the latter.

The proposed method extends the small-scale template database of greeting sentences. Suitable sentences are then selected automatically from the extension template considering the situation. This is achieved by using an association knowledge system that will be described later herein.

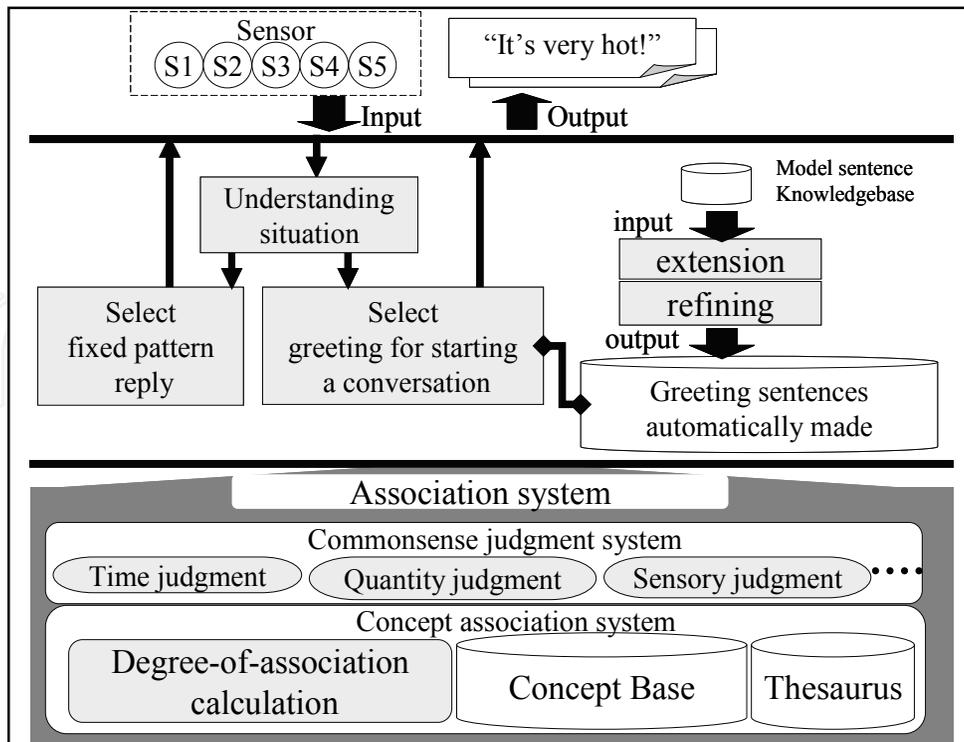


Fig. 1. Structure of greeting conversation system

#### 4. Association System

Humans can interpret received information appropriately because we accumulate basic knowledge about the language and we have empirical common sense knowledge related to words. In other words, the ability to relate images to words is thought to be important. In order to have such common sense judgment ability, a machine must understand the basic knowledge related to words.

Therefore, the machine must generate modelling knowledge about conversations and words. This model is useful for achieving a human-like conversation mechanism. The association system is composed of a concept association system and a common-sense judgment system. The concept association system defines the common meanings related to words, and the common-sense judgment system defines common sense related to words.

##### 4.1 Concept Association System

The Concept Association Mechanism incorporates word-to-word relationships as common knowledge. The Concept Association Mechanism is a structure that includes a mechanism for capturing various word relationships. In this section, we describe the concept base (Kojima et al., 2002) and a method of calculating the degree of association (Watabe & Kawaoka, 2001) using this base.

The concept base is a knowledge base consisting of words (concepts) and word clusters (attributes) that express the meaning of these words, which are mechanically and automatically constructed from multiple sources, such as Japanese dictionaries and newspaper texts. The concept base used in the present study contains approximately 90,000 registered words organized in sets of concepts and attributes. These concept and attribute sets are assigned weights to denote their degree of importance. For example, an arbitrary

concept,  $A$ , is defined as a cluster of paired values, consisting of attribute,  $a_i$ , which expresses the meaning and features of the concept, and weight,  $w_i$ , which expresses the importance of attribute  $a_i$ , in expressing concept  $A$ :

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_N, w_N)\} \quad (1)$$

Attribute  $a_i$  is called the first-order attribute of concept  $A$ . In turn, an attribute of  $a_i$  (taking  $a_i$  as a concept) is called a second-order attribute of concept  $A$ .

The degree of association is a parameter that quantitatively evaluates the strength of the association between one concept and another. The method for calculating the degree of association involves developing each concept up to the second-order attributes, determining the optimum combination of first-order attributes by a process of calculation using weights, and evaluating the number of these matching attributes. The value of the degree of association is a real number between 0 and 1. The higher the number, the greater the association of the word. Table 1 lists examples of the degree of association.

| Concept A | Concept B      | Degree of association between A and B |
|-----------|----------------|---------------------------------------|
| Flower    | Cherry blossom | 0.208                                 |
| Flower    | Plant          | 0.027                                 |
| Flower    | Car            | 0.0008                                |
| Car       | Bicycle        | 0.23                                  |

Table 1. Examples of the degree of association

#### 4.2 Common-Sense Judgement System

The common-sense judgment system derives common-sense associations from words in terms of various factors (e.g., quantity, time, and physical sense). These associations are constructed using the Concept Association Mechanism. In this section, we describe a time judgment system, a part of the common-sense judgment system.

The time judgment system (Tsuchiya et al., 2005) assesses elements of time, such as season and time of day, from nouns, using a knowledge base (Time Judgment KB) of words indicating the time (time word). The system sorts the relationships between a noun and time through the construction of the Time Judgment KB and extracts the necessary time words.

We identified a set of basic representative time words – “spring, summer, autumn, winter, rainy season,” and “morning, daytime, evening, night” – and applied these words to all of the time words registered in the system.

| Input       | Output    |            |           |
|-------------|-----------|------------|-----------|
|             | Time word | Start time | End time  |
| Sunset      | Evening   | 4 p.m.     | 5 p.m.    |
| Snow        | Winter    | December   | February  |
| Sea bathing | Summer    | June       | September |

Table 2. Examples of the time judgment system

The system can also handle time words not contained in the time judgment knowledge base (unknown words) through the use of the concept base of common knowledge. Table 2 lists examples of this system.

## 5. Extension of Conversation Sentences

Upon greeting to start a conversation, not all conversation sentences can be gathered in a database manually. Therefore, all conversation sentences are extended automatically from a small quantity of model sentences by using the association system. That is to say, if a machine obtains several model sentences, then the machine can produce several new sentences by association (Yoshimura et al., 2006).

First, parts (noun) that change are selected from the model sentences. These words are called element words. Second, all words that have the opposite meaning to these element words are extracted for getting words of the same affiliate as this element. Third, the number of these words is increased by synonyms and the attributes of the concept base.

Using this method, the machine associates several words that are related to the element word. This association word is returned to the beginning model sentence. Thus, several sentences can be produced.

## 6. Removal of Erroneous Sentences

Simple mechanical extension of conversation sentences produces several improper sentences. Therefore, the proposed method uses the association system to delete the improper sentences.

First, parts of speech are used to refine the sentence. For example, the original model sentence "It's a season of cherry-blossoms" produces the extension sentence "It's a season of open blossom." This seems strange. Therefore, association words other than the part of speech of the original element word.

Second, a thesaurus (NTT Communication Science Laboratory, 1997) is used to refine the sentence. For example, the original model sentence "It's a beautiful mountain" produces the extension sentence "It's a beautiful climax." In Japanese, the word *mountain* has many meanings, including mountain, climax, and important event. In this case, the original element word, *mountain* expresses the meaning 'mountain'. *Climax* is an extension of *mountain*, but the extension is to a meaning that differs from the original element word. The thesaurus is used to remove such ambiguity.

Next, the degree of association is used to refine the sentence. This is especially important for words strongly related to the original element word. Therefore, high degree words are extracted using the degree of association.

Finally, the common sense-judgement system is used to refine the sentence. In the present paper, strange sentences with respect to time are excluded. For example, simply mechanical extension of conversation sentences includes improper sentences such as "Tonight is cool daytime." This sentence is improper from the point of view of time. The time judgement system in the common sense judgment system judges the adaptability of *tonight* and *daytime* from the point of view of time.

This technique makes it possible to increase the appropriateness of sentences as greeting. These sentences include some sentences that can be used under in suitable situations. For

this reason, we must consider the situation related to greeting when we use these extension sentences.

## 7. A State about Greeting Conversation

Situation is important for some topics in greeting when the speaker recognizes some information. Information sources are divided into verbalized sources and non-verbalized sources. Verbalized sources are expressed using words such as news and weather. Non-verbalized sources are measurement sources that can measure: temperature, humidity, brightness, and volume of a sound, etc. When a computer uses a non-verbalized source, measured information obtained using the computer's sense organ (sensor) should change into some word. A situation related to greeting is expressed as mostly non-verbalized source. This paper pays attention to this non-verbalized source and verbalized source is used for a part.

Computers do not have sensory organs like humans. Computers recognize surroundings using mechanical sensors. In the present paper, the proposed method uses sensors such as thermometers, hygrometers, sound meters (microphones), luminance meters, and clocks with built-in computers. These are non-verbalized sources. After these non-verbalized sources are obtained, they are verbalized according to decided rules. Moreover, weather as a verbalized source is important information for greetings. Humans judge weather by looking at the sky. However, it is difficult for a machine to judge the weather based on photographs. Therefore, sources of weather obtain information by verbalized sources from the Web.

Using this situational information, appropriate sentences are selected from among the extension sentences. When discussing a topic related your situation, you unconsciously select the most suitable word. In the present paper, we separate this unconscious selection into two types, based on which the sentences obtain a weight indicating the importance of the greeting.

The first type is based on degree of peculiarity. When a certain state is considered to be special, the state is mentioned. For example, if you feel that "It is very hot.", you will mention the heat. This means that a signal that is different from daily life indicates a noteworthy topic. The weight given based on this idea is called the "degree of peculiarity".

Simple and comprehensible words are often used for greetings in spoken language. For example, "It is cloudy." is used more often than "It is cloudy weather." as a greeting. This means that words generally used in daily conversation, i.e., words of high utilization, are important. The weight given based on this conceit is called the "degree of importance" of a word.

Using the degree of peculiarity and the degree of importance of word, a certain situation is expressed. A space storing relation between a status word and its weight is referred to as a *status space*. Figure 2 shows a status space image, including the sensor, the status word, and its weight. The status space includes all words used in greeting conversation, and the weight of a word changes according to the state. All words are related to a status word. In a status space, all words are categorized based on words obtained from status words. A word group related to a certain status word is called a status word group of the word.

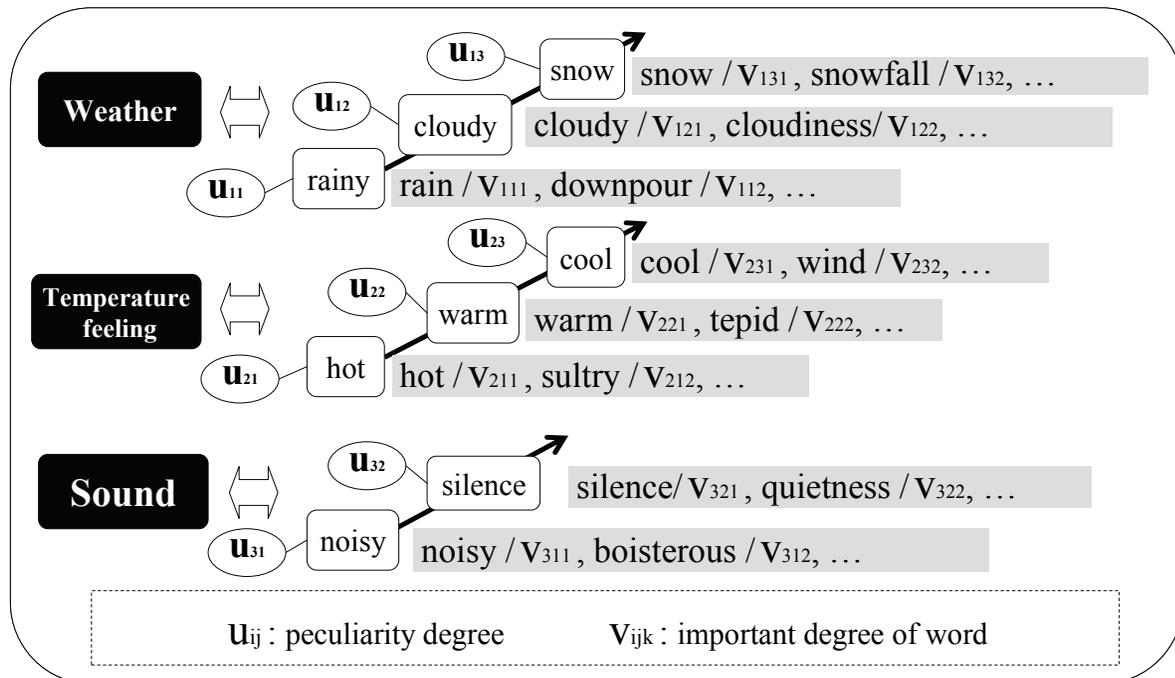


Fig. 2. Status space image

## 8. Weight for situation in Greeting

Simple and comprehensible words are often used for greetings in spoken language. The weight based on this idea is referred to as the *degree of importance* of the word. Each word has a weight in status space. To show the degree of usefulness of a word, we use the concept base Inverted Document Frequency (IDF). The concept base IDF is the weight related to the frequency of use of a word in the concept base. This technique means that low-frequency words in the concept base are not used frequently in daily conversation. Such words are excluded by using the concept base IDF. The concept IDF can be expressed as follows:

$$idf(t) = \log N_{ALL} / df(t)$$

$t$  : object word  
 $N_{ALL}$  : number of all concept in the concept-base  
 $df(t)$  : number of  $t$  in an attribute in the concept-base

(2)

Another weight for the situation in greeting is expressed as the *degree of peculiarity*. When we feel that a certain state is special, the matter is mentioned. The degree of peculiarity expresses a noteworthy level as topic, and each status word is assigned a degree of peculiarity, which takes a discontinuous value of  $\{-1, 0, 1, 2\}$ . A minus value is assigned to a status word that should not be mentioned as topic. If the state is "hot", the status word "hot" is assigned the value of 1, but the status words "cold", "cool", and "warm" are assigned the value of -1. If the state is "very hot", then status word "hot" is assigned the value of 2 in order to increase the noteworthy level as the topic. Moreover, if a state is not hot, cool, warm, or cold, these status words are assigned the value of 0, which expresses a

state of no feeling and is also used when the machine has no related sensor. All words of a status word group have the same value as that status word.

These weights depend on the state and are expressed as status-space. The weight of a word is decided depending on the multiplication of these weights. Extension sentences can have such weights. If there is some word having a negative value in a sentence, the sentence is given a weight of negative value. However, when a sentence is given a mean value of all word weights in the sentence, the sentences corresponding to the corresponding situation can be selected.

## 9. Evaluation

This section mentions evaluation of proposal system in this present paper. First, we evaluate the accuracy and number of extended and refined sentences, and then we evaluate the appropriateness of sentences based on the situation.

### 9.1 Increasing the Number of Greeting Sentences for Conversation

This section evaluates the accuracy and the number of sentences that are extended and refined.

A total of 300 evaluation sentences are chosen at random from among the extension sentences using the technique described herein. Three individuals evaluated these sentences based on common sense correctness of the greeting.

Accuracy is the mean value of common sense correctness of the word as judged by these three individuals. Using the method described herein, we showed that the proposed method was able to increase the accuracy and number of the greeting sentences, as shown in Fig. 3.

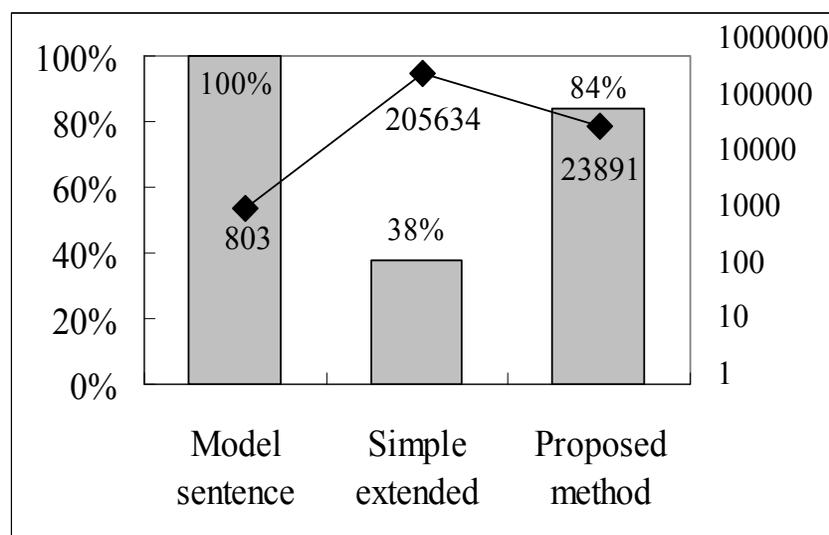


Fig. 3. Evaluation of extended greeting sentences

### 9.2 Greeting Sentence Selection Based on the Situation

This section evaluates the accuracy of the appropriateness of a sentence based on the situation. An evaluation set is prepared for 20 situations (weather, temperature and humidity, brightness, time, and volume of sound). In each situation, the status space is

produced and then extension sentences are assigned a value using the proposed technique. Six sentences having higher values are used as evaluation sentences. The three individuals evaluated the sentences and the situation patterns. They evaluated whether the sentences were appropriate to the situation. In addition, the accuracy is compared with the accuracy of randomly outputting sentences in order to verify the effectiveness of the proposed method. This result is shown in Fig. 4.

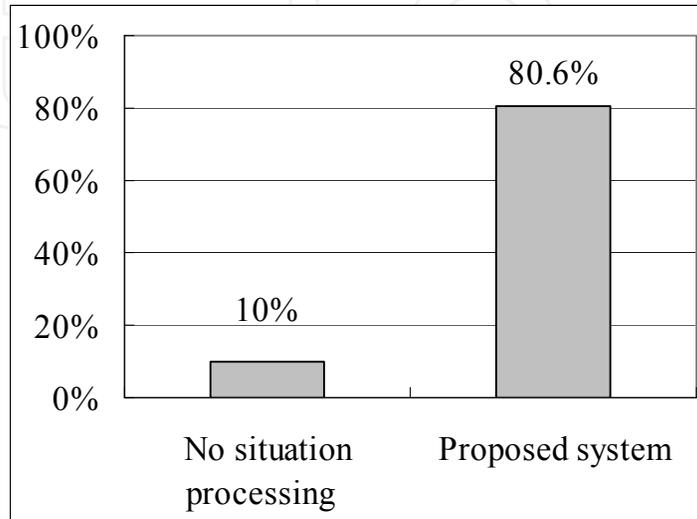


Fig. 4. Evaluation of Greeting Sentence Selection Based on Situation

## 10. Discussion

With respect to the technique of increasing the accuracy of greeting sentences, the proposed technique could improve the accuracy to 84%, compared to a simple extension technique, for which the accuracy is 38%. Moreover, the number of sentences increased from 803 sentences to 23,891 sentences. Table 3 shows examples of extension sentences.

| Template            |                  | It's a beautiful river.        |
|---------------------|------------------|--------------------------------|
| Extension Sentences | Refined sentence | It's a beautiful flow.         |
|                     |                  | It's a beautiful spectacle.    |
|                     | Deleted sentence | It's beautiful scenery.        |
|                     |                  | It's a beautiful water imp     |
|                     |                  | It's a beautiful muddy stream. |
|                     |                  | It's a beautiful crime case.   |

Fig. 5. Example of Extension Sentences

"It's a beautiful flow" and "It's a beautiful spectacle" are used as greeting sentences, but "It's a beautiful water imp" and "It's a beautiful muddy stream" are not used. These sentences were deleted automatically using the proposed method. Therefore, the proposed method of generating extended greeting sentence is shown to be effective.

The proposed technique of weighting corresponding to the situation was shown to have an accuracy of 80.6%, whereas randomly output sentences have an accuracy of 10%. The proposed technique of weight selection corresponding to the situation was shown to be effective.

## 11. Conclusion

Using the method described herein, we show that the proposed method increased the accuracy of greeting sentences to 84% and increased the number of greeting sentences to 23,891 sentences from 803 model sentences. Moreover, the technique used to select an appropriate sentence based on the situation was able to achieve an accuracy of 80.6%. When the situation was not considered, the accuracy was 10%. Thus, the proposed method is shown to be effective.

When the proposed technique is applied, even for the same situation, the machine does not reply using the same sentences in different conversations because a different reply can be chosen for high-weight sentences. As a result, the machine will be able to have an intelligent conversation with a high level of satisfaction.

## References

- Kojima, K.; Watabe, H. & Kawaoka, T.(2002). A Method of a Concept-base Construction for an Association System: Deciding Attribute Weights Based on the Degree of Attribute Reliability. *Journal of Natural Language Processing*, Vol.9, No.5, 2002.10, 93-110, ISSN1340-7619.
- NTT Communication Science Laboratory.(1997). *NIHONGOGOITAIKEI*, Iwanami Shoten, ISBN4-00-009884-5 C3581.
- Tsuchiya, S.; Watabe, H. & Kawaoka, T.(2005). A Time Judgement System Based on an Association Mechanism. *Proc. of KES2005, Part3*, 2005.9, 742-748
- Watabe, H. & Kawaoka, T. (2001). Measuring Degree of Association between Concepts for Commonsense Judgements. *Journal of Natural Language Processing*, Vol.8, No.2, 2001.4, 39-54, ISSN1340-7619.
- Weizenbaum, J. (1965). ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the Association For Computing Machinery*, Vol.9, No.1, 36-45
- Yoshimura, E.; Watabe, H. & Kawaoka, T. (2006). An Automatic Enhancing Method of Greeting Sentences Using Association Knowledge Mechanism. *Journal of Natural Language Processing*, Vol.13, No.1, 2006.1, 117-141, ISSN1340-7619.



## **Frontiers in Robotics, Automation and Control**

Edited by Alexander Zemliak

ISBN 978-953-7619-17-6

Hard cover, 450 pages

**Publisher** InTech

**Published online** 01, October, 2008

**Published in print edition** October, 2008

This book includes 23 chapters introducing basic research, advanced developments and applications. The book covers topics such as modeling and practical realization of robotic control for different applications, researching of the problems of stability and robustness, automation in algorithm and program developments with application in speech signal processing and linguistic research, system's applied control, computations, and control theory application in mechanics and electronics.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Eriko Yoshimura, Seiji Tsuchiya, Hirokazu Watabe and Tsukasa Kawaoka (2008). Automatic Generation of Appropriate Greeting Sentences Using Association System, *Frontiers in Robotics, Automation and Control*, Alexander Zemliak (Ed.), ISBN: 978-953-7619-17-6, InTech, Available from:  
[http://www.intechopen.com/books/frontiers\\_in\\_robotics\\_automation\\_and\\_control/automatic\\_generation\\_of\\_appropriate\\_greeting\\_sentences\\_using\\_association\\_system](http://www.intechopen.com/books/frontiers_in_robotics_automation_and_control/automatic_generation_of_appropriate_greeting_sentences_using_association_system)

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2008 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen