



|                  |                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title            | A casual conversation system using modality and word associations retrieved from the web                                                                                                                                                                                                                                                                                                                        |
| Author(s)        | Higuchi, Shinsuke; Rzepka, Rafal; Araki, Kenji                                                                                                                                                                                                                                                                                                                                                                  |
| Citation         | EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing, 382-390<br><a href="https://doi.org/10.1145/1613715.1613765">https://doi.org/10.1145/1613715.1613765</a>                                                                                                                                                                                                           |
| Issue Date       | 2008-10-25                                                                                                                                                                                                                                                                                                                                                                                                      |
| Doc URL          | <a href="https://hdl.handle.net/2115/63992">https://hdl.handle.net/2115/63992</a>                                                                                                                                                                                                                                                                                                                               |
| Rights           | © 2008 ACM. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing, Pages 382-390, 2008-10-25. <a href="http://doi.acm.org/10.1145/1613715.1613765">http://doi.acm.org/10.1145/1613715.1613765</a> |
| Type             | conference paper                                                                                                                                                                                                                                                                                                                                                                                                |
| File Information | A Casual Conversation System Using Modality and Word Associations.pdf                                                                                                                                                                                                                                                                                                                                           |



# A Casual Conversation System Using Modality and Word Associations Retrieved from the Web

**Shinsuke Higuchi**

**Rafal Rzepka**

**Kenji Araki**

Graduate School of Information Science and Technology

Hokkaido University, Sapporo Japan 060-0814

{shin\_h,kabura,araki}@media.eng.hokudai.ac.jp

## Abstract

In this paper we present a textual dialogue system that uses word associations retrieved from the Web to create propositions. We also show experiment results for the role of modality generation. The proposed system automatically extracts sets of words related to a conversation topic set freely by a user. After the extraction process, it generates an utterance, adds a modality and verifies the semantic reliability of the proposed sentence. We evaluate word associations extracted from the Web, and the results of adding modality. Over 80% of the extracted word associations were evaluated as correct. Adding modality improved the system significantly for all evaluation criteria. We also show how our system can be used as a simple and expandable platform for almost any kind of experiment with human-computer textual conversation in Japanese. Two examples with affect analysis and humor generation are given.

## 1 Introduction

Many task-oriented dialogue systems (Liu et al., 2003; Reitter et al., 2006) have been developed. Research on non-task-oriented dialogue systems like casual conversation dialogue systems ("chatbots") is on the other hand not very common, perhaps due to the many amateurs who try to build naturally talking systems using sometimes very clever, but rather unscientific methods although there are systems with chatting abilities as (Bickmore and Cassell, 2001) but concentrate on applying strategies to casual conversation rather than their automatic generation of

those conversations. However, we believe that the main reason is that an unrestricted domain is disproportionately difficult compared to the possible use such a system could have. It is for example very hard to predict the contents and topics of user utterances, and therefore it is almost impossible to prepare conversational scenarios. Furthermore, scenarios need more or less specific goals to be useful. However in our opinion, sooner or later non-task-oriented dialogue systems will have to be combined with task oriented systems and used after recognizing that the user's utterance does not belong to a given task. This would lead to more natural interfaces for e.g. information kiosks or automatic guides placed in public places where anyone can talk to them about anything (Gustafson and Bell, 2000; Kopp et al., 2005) regardless of the role the developers intended. For this reason we have also started implementing emotiveness recognition and joke generation modules that are presented later in the paper.

Well-known examples of non-task-oriented dialogue systems are ELIZA (Weizenbaum, 1966) and A.L.I.C.E.<sup>1</sup>, though the former was built to parody a Rogerian therapist which can be regarded as a task. Both systems and their countless imitators<sup>2</sup> use a lot of rules coded by hand. ELIZA is able to make a response to any input, but these responses are only information requests without providing any new information to the user. In the case of A.L.I.C.E.,

<sup>1</sup>Wallace, R. The Anatomy of A.L.I.C.E.  
<http://www.alicebot.org/anatomy.html>.

<sup>2</sup>Many of them have been quite successful in the Loebner Prize and the Chatterbox Challenge (competitions only for English-speaking bots) but explanations of their algorithms are not available.

the knowledge resource is limited to the existing database. Creating such databases is costly and a programmer must learn the AIML mark-up language to build it. Although there have been attempts at updating AIML databases automatically (Pietro et al., 2005), the scale was rather limited.

As mentioned above, these examples and many other "chatbots" need hand-crafted rules, and are thus often ignored by computer scientists and rarely become a research topic. However, they have proved to be useful for e-learning (Pietro et al., 2005) and machine learning (Araki and Kuroda, 2006) support.

Building a system using automatic methods, like we do, seems to be the most realistic way for unrestricted domains. Considering the large cost of developing a program that can talk about any topic, it is appealing to turn to the huge and cheap textual source that is the Internet.

In this very moment millions of people (Kumar et al, 2003) are updating their blogs and writing articles on every possible topic. These are available on the Web which we can access any time, and in a faster and faster manner, the search engines grow more and more efficient. Thus, the Web is well suited to extracting word associations triggered by words from user utterances made in a topic-free dialogue system.. We present a system making use of this type of information. It automatically extracts word association lists using all keywords in a given utterance without choosing a specific one (which most other systems that ignore the context do) then generates a reply using the only one strongest association from the nouns, verbs and adjectives association groups. Modality is then added to the reply, and then it is output.

Our system is built upon the idea that human utterances consist of a proposition and a modality (Nitta et al., 1989). In this paper we present an algorithm for extracting word associations from the Web and a method for adding modality to statements. We evaluate both the word associations and the use of modality. We also suggest some future possible extensions of the system and show a small experiment with adding humor to the system.

In this paper, the system described works for Japanese and uses text as input and output. Though the final goal of our research is to help developing freely talking car navigation systems that by their

chatting abilities can help to avoid drowsiness while driving and so on. in this part of the development we concentrate on proposition generation and modality processing. Therefore, we work only with text now. We plan to combine this project with research on in car voice recognition and generation.

## 2 Extracting Word Associations

In this chapter, we present a method for automatic extraction of word associations based on keywords from user utterances. We use the Google<sup>3</sup> search engine snippets to extract word associations in real time without using earlier prepared resources, such as off-line databases.

### 2.1 Extracting Word Associations from the Web

In the first step, the system analyzed user utterances using the morphological analyzer MeCab<sup>4</sup> in order to spot query keywords for extracting word associations lists. We define nouns, verbs, adjectives, and unknown words as query keywords. The reason we chose these word classes is that these word classes can be treated as important and, to some extent, describe the context. We define a noun as the longest set of nouns in a compound noun. For example, the compound noun *shizen gengo shori*<sup>5</sup> (natural language processing) is treated by MeCab as three words: (*shizen* - natural), (*gengo* - language) and (*shori* - processing). Our system, however, treats it as one noun.

In the next step, the system uses these keywords as query words for the Google search engine. The system extracts the nouns from the search results and sorts them in frequency order. This process is based on the idea that words which co-occur frequently with the input words are of high relevance to them. The number of extracted snippets is 500. This value was set experimentally, taking the processing time and output quality into account. The top ten words of a list are treated as word associations, see Table 1 for an example.

<sup>3</sup>Google, <http://www.google.co.jp/>

<sup>4</sup>MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <http://mecab.sourceforge.jp/>

<sup>5</sup>All Japanese transcriptions will be written in italics.

Table 1: Examples of noun associations triggered by a user utterance

| <i>Sapporo wa samui.</i> (Sapporo city is cold.) |                           |    |
|--------------------------------------------------|---------------------------|----|
| Association frequency ranking:                   |                           |    |
| 1                                                | <i>yuki</i> (snow)        | 52 |
| 2                                                | <i>fuyu</i> (winter)      | 50 |
| 3                                                | <i>kion</i> (temperature) | 16 |
| 4                                                | <i>jiki</i> (season)      | 12 |
| 5                                                | <i>Tokyo</i> (Tokyo)      | 12 |
| 6                                                | <i>tenki</i> (weather)    | 11 |
| 7                                                | <i>chiiki</i> (area)      | 10 |
| 8                                                | <i>heya</i> (room)        | 10 |

## 2.2 Evaluation

We asked volunteers to use our system and to evaluate the correctness of word lists generated by the system. First, a participant freely inputs an utterance, for which the system retrieves ten association words. Next, a participant rated these words using a scale of one to three with 3 meaning "perfectly correct", 2 - "partially correct" and 1 - "incorrect". In this experiment we consider words that receive a 2 or 3 as usable. The reason associations rated 2 or 3 are considered as usable is that the definition of what makes a good word association here is difficult to specify. When it comes to topic-free conversations we have observed that associations have an effect on a certain context. Three volunteers repeated the experiment ten times, so the final amount of evaluated words was 300. Table 2 shows the results of the top 10 words, sorted by the frequency of appearance. Table 3 shows the results of the top 5 words.

What constitutes a correct word association was left to each volunteer to decide subjectively since in a casual conversation setting associations are hard to define strictly.

Table 2: Top 10 word associations

| score         | participant(A, B, C) | total |
|---------------|----------------------|-------|
| 3             | 40, 52, 57           | 149   |
| 2             | 37, 17, 27           | 81    |
| 1             | 23, 31, 16           | 70    |
| usability (%) | 77, 69, 84           | 77    |

As shown in Table 2 approximately 77% of the word associations were judged as usable but there

Table 3: Top 5 word associations

| score         | participant (A, B, C) | total |
|---------------|-----------------------|-------|
| 3             | 20, 29, 36            | 85    |
| 2             | 17, 9, 10             | 36    |
| 1             | 13, 12, 4             | 29    |
| usability (%) | 74, 76, 92            | 81    |

were individual differences between the evaluators. This shows that the definition of word associations is different for each participant. Table 3 shows that approximately 80% of the word associations were judged as usable. It is thus highly likely that the top words from the frequency lists are correct associations. The results show that automatic extracting of word associations using a Web search engine is feasible. The main reason for extracting word associations from the Web is that thanks to this method, the system can handle new information, proper names, technical terms and so on. by using only the snippets from the search engine. The word association extraction takes no more than few seconds. For the evaluation we used only nouns but we expect although verbs and adjectives are often more abstract than nouns, the word associations for them will improve the results.

## 3 General Description of the System

The system generates replies in the following way:

- extraction of keywords from user utterance
- extraction of word associations from the Web
- generation of sentence proposition using the extracted associations
- addition of modality to the sentence proposition

### 3.1 Extraction of Keywords from User Utterances

The system applies morphological analysis to the use utterances in the same way as described in section 2.1 and extracts keywords based on part of speech.

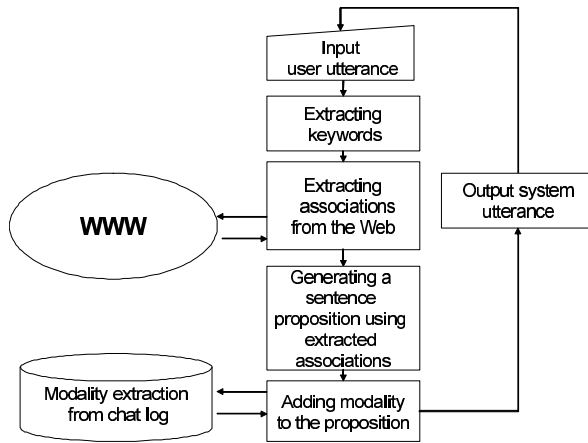


Figure 1: System flow

### 3.2 Extraction of Words Association from the Web

The system performs a Google search using the extracted keywords as a query. The system sorts the results obtained from the query by their frequency as in section 2.1. In section 2.1 only nouns were extracted but here we also extract verbs and adjectives. After sorting all words in adjective, verb and noun lists the system uses the ones with the highest frequency as word associations.

### 3.3 Generation of Proposition Using Word Associations

Using the associations, the system generates the proposition of a sentence to be used as a reply to the user input. A proposition is an expression representing an objective statement. The proposition is generated by applying associations to a proposition template like [(noun) (topic indicating particle *wa*) (adjective)]. We prepared 8 proposition templates manually (see Table 4). The templates were chosen subjectively after examining statistics from IRC<sup>6</sup> chat logs. Our criteria for choosing templates from the chat logs was that they should belong to the 20 most frequent modality patterns and to be flexible enough to fit a range of grammatical constructions, for example in English, "isn't it" cannot follow verbs while "I guess" can follow nouns, adjectives, and verbs. The proposition templates are applied in a

predetermined order: for example, first a template "(noun) (*wa*) (adjective)" is used; next a template "(noun) (*ga*) (adjective)" is used. However, since the generated proposition is not always a natural statement, the system uses exact matching searches of the whole phrases in a search engine to check the naturalness of each proposition. If the frequency of occurrence of the proposition is low, it is defined as unnatural and deleted. This processing is based on the idea that the phrases existing on the Web in large numbers are most probably correct grammatically and semantically. If an unnatural proposition is generated, the system generates another proposition in the same way. In this experiment the system used propositions for which the hit number exceeded 1,000 hits using Google. Thus, the processing proceeds as follows. The system first selects the top noun, top verb, and top adjective word associations. These are applied to the templates in a predetermined order. If a generated proposition is judged as valid (using Google, occurrence on the web indicates validity), it is used. If not, another template is tried until a valid proposition is found. The reason for not trying every possible combination of associated words is prohibitively long processing time.

Table 4: Proposition templates

|                                       |
|---------------------------------------|
| (noun) ( <i>wa</i> ) (adjective)      |
| (noun) ( <i>ga</i> ) (adjective)      |
| (noun) ( <i>ga</i> ) (verb)           |
| (noun) ( <i>wa</i> ) (verb)           |
| ( <i>so-re</i> ) ( <i>wa</i> ) (verb) |
| (noun)                                |
| (adjective)                           |
| (verb)                                |

### 3.4 Adding Modality to the Propositions

Finally, the system adds modality to the generated proposition. By modality we mean a set of grammatical and pragmatic rules to express subjective judgments and attitudes. In our system, modality is realized through adverbs at the end of a sentence which is common in Japanese (Nitta et al., 1989). In our system, a pair of sentence head and sentence end auxiliary verb are defined as "modality".

<sup>6</sup>Internet Relay Chat Protocol,  
<http://www.irchelp.org/irchelp/rfc/rfc.html>

### 3.4.1 Extracting Modality

There is no standard definition of what constitutes modality in Japanese. In this paper modality of casual conversation is classified into questions and informative expressions. Questions are expressions that request information from the user. Informative expressions are expressions that transmit information to the user. Patterns for these modalities are extracted automatically from IRC chat logs (100,000 utterances) in advance. Modality patterns are extracted in these ways:

- pairs of grammatical particles and an auxiliary verbs placed at the end of sentences are defined as ending patterns
- sentences with question marks are defined as questions
- adverbs, emotive words, and connectives at the beginning of sentences are defined as informative expressions
- candidate patterns thus obtained are sorted by frequency

First the system extracts sentence ending patterns from IRC chat logs. If an expression contains question marks, it is classified as a question. Next, the system extracts adverbs, emotive words, and connectives from the beginning and end of sentences from the IRC logs. These pairs (beginning and end) of expressions are classified as "informative expressions". For example question expression "desu-ka?" is extracted from a human utterance like "Kyou-wa samui desu-ka?" (Is it cold today?). An informative expression "maa \*\*\* kedo" is extracted from a human utterance as "Maa sore-wa ureshii kedo" (Well, I'm glad, but you know...).

685 patterns were obtained for informative expressions. 550 of these informative expression patterns were considered by authors as correct (80%). For questions 396 patterns were obtained, and 292 patterns (73%) were evaluated as correct. We sorted these candidates in frequency order. The words appearing at the top of the list were correct, but even the ones appearing only once were still deemed as usable. For example, the question expression "janakatta deshita-kke?" is a correct expression,

but appeared only once in the 100,000 utterances. Hence, we confirmed that chat logs include various modality expressions, and only a few of them are incorrect. Tables 5 and 6 show some examples of modality patterns.

Table 5: Examples of informative expression modality

| informative expression                                 | frequency |
|--------------------------------------------------------|-----------|
| <i>maa - kedo</i><br>(Well , it can be said - but -)   | 21        |
| <i>maa - dana</i><br>(Well , it can be said -)         | 16        |
| <i>maa - desu-ga</i><br>(Well , it appears that -)     | 16        |
| <i>soko-de - desu-yo</i><br>(Here , it is said that -) | 15        |
| <i>maa - da-ga</i><br>(Well , it can be said - but -)  | 14        |
| <i>maa - desu-yo</i><br>(Well , it is that -)          | 12        |

Table 6: Examples of question modality sentence endings

| question                                      | frequency |
|-----------------------------------------------|-----------|
| <i>...desuka?</i><br>(Is it that ... ?)       | 232       |
| <i>...kana?</i><br>(Maybe ... ?)              | 90        |
| <i>...da-kke?</i><br>(Is it right that ... ?) | 87        |
| <i>...masu-ka?</i><br>(Is it that ... ?)      | 69        |
| <i>...nano?</i><br>(Is it that ... ?)         | 68        |
| <i>...toka?</i><br>( ... , isn't it ?)        | 55        |

### 3.4.2 Adding Modality

The system adds the modality from section 3.4.1 to the proposition from section 3.3 to generate the system output. This process is based on the idea that human utterance consists of proposition and modality. A modality pattern is selected randomly. For example, if the system generates the proposition "fuyu wa samui (Winter is cold.)" and selects the modality "iyaa ... desu-yo (Ooh ... isn't it?)", the gen-

erated output will be "iyaa, fuyu-wa samui desu-yo (Winter is cold, you know)". However, there is a possibility that the system generates unnatural output like "fuyu-wa samui dayo-ne (Winter is cold, aren't it?)", depending on the pair of proposition and modality. To this problem, the system uses the Google search engine to filter out unnatural output. The system performs a phrase search on the end of the sentence. If the number of search hits is higher than threshold, the output is judged as correct. If the number of a search hits is lower than the threshold, the output is judged as incorrect and discarded, and a new reply is generated. Here, we experimentally set the threshold to 100 hits.

#### 4 Experimental Results

We used system  $\alpha$ , generating only the proposition, and system  $\beta$ , generating both proposition and modality. 5 participants used each systems for conversations of 10 turns and evaluated the conversations on a 5-point scale. Evaluation criteria were "will to continue the conversation" (A), "grammatical naturalness of dialogues" (B), "semantical naturalness of dialogues" (C), "vocabulary richness" (D), "knowledge richness" (E), and "humanity of the system" (F). Table 7 shows average scores for the evaluations of each system. System  $\beta$  that uses modality scored much higher than system  $\alpha$ . Table 8 shows examples of actual dialogue. In the evaluation, the participants expressed the opinion that an utterance like (*xx ha yy*) is unnatural and using a modality like (*maa*)("well"), (*moo*)("anyway") is very natural. Thus we can say that the modality expressions make the utterances of the system seem more natural.

#### 5 The System as a Platform

The simplicity, real-time processing capabilities and promising results showing that users do not get bored so quickly encouraged us to perform trials with other ongoing projects and experiment with the system working as a platform for adding various modules and algorithms. By using our system it is possible to perform tests to see if a new idea will support or improve human-computer interaction or not. Here we will briefly describe two such trials - one on guessing emotive values of utterances and

Table 8: Examples of dialogues with system  $\beta$

|         |                                                                                                                              |
|---------|------------------------------------------------------------------------------------------------------------------------------|
| user:   | <i>Nanika suki-na tabemono aru?</i><br>(What food do you like ?)                                                             |
| system: | <i>Maa, tabemono-wa oishii desu.</i><br>(Well, food tastes good.)                                                            |
| user:   | <i>Saikin-wa osake-mo</i><br><i>suki-ni natte-kita-n-da.</i><br>(Recently, I begin to like alcohol, too.)                    |
| system: | <i>Demo, sake-wa yowai-no-yo-ne.</i><br>(But, I can't drink much.)                                                           |
| user:   | <i>Yowai-n-da... Sugu yochau-no?</i><br>(Really...?<br>Do you get drunk<br>quickly?)                                         |
| system: | <i>Maa, sore-wa sugoi-wa-ne.</i><br>(Whow, that's amazing.)                                                                  |
| user:   | <i>Boku-wa wari-to osake-wa</i><br><i>tsuyoi hou-nan-da.</i><br>(I have a rather high<br>tolerance for alcohol.)             |
| system: | <i>Iya, sore-wa tsuyoi-n-de.</i><br>(Well, it is strong, I can say.)                                                         |
| user:   | <i>Kyou-wa sukoshi hadazamui-yo-ne.</i><br><i>Natsu-nanoni.</i><br>(It is a little cold today.<br>Even though it is summer.) |
| system: | <i>Asa-wa hadazamui-toka ?</i><br>(Isn't it chilly in morning?)                                                              |
| user:   | <i>Asa-kara hadazamukatta-ne.</i><br>(It was chilly since morning.)                                                          |
| system: | <i>Aa kyou-wa hayai-na.</i><br>(Ah, it's early. )                                                                            |

one on improving the system's overall evaluation by adding a pun generator.

##### 5.1 Testing Affect Analysis

Ptaszynski et al.(Ptaszynski et al., 2008) have developed a method for affect analysis of Japanese text. Their method is based on cross-referencing lexical emotive elements with emotive expressions appearing in text. In the process of analysis first a general emotive context is determined and then the specific types of emotional states conveyed in an utterance are extracted. They support this method with a

Table 7: Evaluation Results

| Evaluation criteria | System $\alpha$ (proposition) |     |     |     |     |     | system $\beta$ (proposition + modality) |     |     |     |     |     |
|---------------------|-------------------------------|-----|-----|-----|-----|-----|-----------------------------------------|-----|-----|-----|-----|-----|
|                     | A                             | B   | C   | D   | E   | F   | A                                       | B   | C   | D   | E   | F   |
| Participant a       | 1                             | 3   | 2   | 2   | 4   | 2   | 4                                       | 4   | 3   | 4   | 3   | 5   |
| Participant b       | 1                             | 3   | 1   | 2   | 1   | 1   | 4                                       | 4   | 4   | 5   | 4   | 3   |
| Participant c       | 1                             | 2   | 1   | 2   | 1   | 1   | 1                                       | 2   | 1   | 2   | 1   | 1   |
| Participant d       | 1                             | 3   | 1   | 3   | 1   | 2   | 4                                       | 3   | 1   | 3   | 3   | 4   |
| Participant e       | 1                             | 4   | 1   | 1   | 2   | 1   | 3                                       | 2   | 2   | 4   | 5   | 4   |
| Average             | 1.0                           | 3.0 | 1.2 | 2.0 | 1.8 | 1.4 | 3.2                                     | 3.0 | 2.2 | 3.6 | 3.2 | 3.4 |

Web-mining technique to improve the performance of the emotional state type extraction. A system constructed on the basis of their method achieved human level performance in determining the emotiveness of utterances, and 65% of human level performance in extracting the specific types of emotions. Also, the supporting Web mining technique improved the performance of the emotional state type extraction to 85% of the human level (Shi et al, 2008). As these are very promising figures we are currently in the phase of implementing their ideas in our system and testing how emotion recognition can influence speech act analysis and the automatic choice of proper modality.

## 5.2 Improving the System Using Humor

In this trial, an experiment showing that humor can improve a non-task oriented conversational system's overall performance was conducted.

### 5.2.1 Implementing PUNDA system

By using a simplified version of Dybala's PUNDA system (Dybala et al., 2008), a pun-generation was added to our baseline system. The PUNDA algorithm consists of two parts: A Candidate Selection Algorithm and a Sentence Integration Engine. The former generates a candidate for a pun analyzing an input utterance and selects words or phrases that could be transformed into a pun by one of four generation patterns: homophony, initial mora addition, internal mora addition or final mora addition. The latter part generates a sentence including the candidate extracted in the previous step. To make the system's response more related to the user's input, each sentence that included a joke started with the pattern "[base phrase] to iebe"

("Speaking of [base phrase]"). The remaining part of the sentence was extracted from the Web and the candidate was used as a query word and the list of sentences including this word was retrieved. Then the shortest sentence with an exclamation mark is selected as most jokes convey some emotions. When the candidate list was empty, the system selected one random pun from a pun database.

### 5.2.2 Experiment results

In the first experiment, 5 participants were asked to perform a 10-turn dialogue with two systems. After using both systems (baseline and humor-equipped), users were asked to evaluate both systems's performances by answering the following questions: A) Do you want to continue the dialogue?; B) Was the system's utterances grammatically natural?; C) Was the system's utterances semantically natural?; D) Was the system's vocabulary rich?; E) Did you get an impression that the system possesses any knowledge?; F) Did you get an impression that the system was human-like?; G) Do you think the system tried to make the dialogue more funny and interesting? and H) Did you find the system's utterances interesting and funny? Answers were given on a 5-point scale and the results are shown in Table 9.

A third-person evaluation experiment was also performed and again the humor-equipped system scored higher than the non-humor one. The question asked in this evaluation was: "Which dialogue do you find most interesting and funny?". Evaluators could choose between 3 options: Dialogue 1 (Baseline system first 3 turns), Dialogue 2 (Humor-equipped system, first 3 turns with system's third response replaced by pun generator's output) and Dia-

Table 9: Results of humor experiments

| Evaluation Criteria | A   | B   | C   | D   | E   | F   | G   | H   |
|---------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Baseline System     | 3.0 | 2.2 | 2.4 | 2.4 | 2.0 | 2.8 | 2.2 | 2.8 |
| With pun generator  | 3.2 | 3.0 | 2.8 | 2.8 | 2.2 | 3.0 | 3.4 | 3.6 |

logue 3 (the first 3 turns of the baseline system with joking ability). Dialogue 1 and Dialogue 2 have the same input. Among 25 evaluators, only 5 (20%) responded that Dialogue 1 was most interesting and funny. 10 chose Dialogue 2 and the other 10 chose Dialogue 3 (40% respectively). This means that each of humor equipped dialogues received evaluations two times higher than non-humor dialogue.

### 5.3 A Toolkit for Conversation-Related Experiments

Our system can be also disassembled into a set of flexible tools which help students to experiment with dialogue processing. By using simple web-mining techniques we described, this dialogue engine is capable of automatic retrieval of associations which can be used to produce a whole range of utterances - for example by using the bottom, not the top of the associations list, one can examine how interesting or provocative the dialogue becomes. As the system has a cgi interface, the experiments are easy and any new feature (for instance a speech act choice menu) can be easily added. Such toolkit gives students an opportunity to experiment on a given aspect of dialogue processing without the need of building a conversation system from the scratch. There is also no need of laborious knowledge input and, as such open-domain oriented system generates new "on topic" utterances, experiment subjects do not get bored quickly, which is always a problem while collecting conversation logs of human-machine interaction. A programmer also can freely choose between thousands of IRC logs utterances and Internet resources for the statistical trials, grammar patterns retrieval, speech acts analysis.

## 6 Conclusion and Future Work

In this research we investigated if word associations extracted automatically from the Web are reasonable (semantically on topic) and if they can be successfully used in non-task-oriented dialogue systems.

We also implemented such a system extraction module. It is able to automatically generate in real-time responses to user utterances by generating a proposition and adding modality retrieved from IRC chat logs. We conducted evaluation experiments on the overall influence of the modality usage and it improved the system. Therefore we showed that it is possible to construct a dialogue system that automatically generates understandable on-topic utterances without the need of creating vast amounts of rules and data beforehand. We also confirmed that our system can be used as a experimental platform which can be easily used by other researchers to test their algorithms with a more unpredictable (and less boring) "chatbot", an important factor for long tiring sessions of human-computer conversation. Currently there are several projects which use the system described here as a platform for experiments and we introduced two of them - on joke generation and affect analysis.

There is still a lot of work left to be done. It is necessary for a non-task-oriented dialogue system to obtain not only word associations, but also different kinds of knowledge - of user's preferences or of dialogue itself - for example conversational strategies. At this moment the system generates utterances by applying word associations to the proposition templates and adding modality. We also need to more deeply consider semantics, speech acts and context to create a more advanced system. Finally, the system needs to recognize not only keywords, but also user's modality. We assume that the affect recognition mentioned above will help us to achieve this goal in near future and this is our next step. By opening the system's code and giving others the opportunity of adding their own modules and changes we hope to solve remaining problems. In this paper we focus on the impact of adding modality to a system. Comparing the system to Japanese versions of ELIZA (already available) and ALICE (not available in Japanese yet) is also one of our next steps.

## Acknowledgments

This work was partially supported by the Research Grant from the Nissan Science Foundation.

Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki. 2008. Extracting Dajare Candidates from the Web - Japanese Puns Generating System as a Part of Humor Processing Research. Proceedings of LIBM'08 First International Workshop on Laughter in Interaction and Body Movement:46-51.

## References

- Bei Liu, Limin Du, Shuiyuan Yu. 2003 The method of building expectation model in task-oriented dialogue systems and its realization algorithms. Proceedings of Natural Language Processing and Knowledge Engineering:174-179
- David Reitter, Johanna D. Moore, and Frank Keller. 2006. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In Proc. 28th Annual Conference of the Cognitive Science Society (CogSci), Vancouver, Canada.
- Timothy Bickmore and Justine Cassell. 2001 Relational Agents: A Model and Implementation of Building User Trust. Proceedings of Human Factors Computing Systems (SIGCHI'01): 396-403.
- Joakim Gustafson and Linda Bell. 2000. Speech technology on trial: Experiences from the August system. In *Natural Language Engineering*, 1(1):1-15.
- Stefan Kopp, Lars Gesellensetter, Nicole C. Kramer, and Ipke Wachsmuth. 2005. *A Conversational Agent as Museum Guide- Design and Evaluation of a Real-World Application*. Intelligent Virtual Agents, LNAI 3661:329-343.
- Joseph Weizenbaum. 1966. ELIZA - computer program for the study of natural language communication between man and machine. *Commun. ACM*, vol.9, no.1:36-45.
- Orlando De Pietro, Maurizio De Rose, and Giovanni Frontera. 2005. Automatic Update of AIML Knowledge Base in E-Learning Environment. In Proceedings of Computers and Advanced Technology in Education. , Oranjestad, Aruba, August:29-31.
- Kenji Araki and Michitomo Kuroda. 2006. Generality of a Spoken Dialogue System Using SeGA-IL for Different Languages, Proceedings of the IASTED International Conference COMPUTER INTELLIGENCE:70-75.
- Ravi Kumar, Jasmine Novak, Prabhakar Raghavan, and Andrew Tomkins. 2003. On the Bursty Evolution of Blogspace. Proceedings of The Twelfth International World Wide Web Conference:568-257
- Yoshio Nitta and Takashi Masuoka, *Japanese modality(Nihongo no modality)* Kuroshio.
- Michal Ptaszynski, Pawel Dybala, Rafal Rzepka, and Kenji Araki. 2008. Double Standpoint Evaluation Method for Affect Analysis System. The 22nd Annual Conference of Japanese Society for Artificial Intelligence (JSAI 2008).
- Wenhan Shi, Rafal Rzepka and Kenji Araki. 2008. Emotive Information Discovery from User Textual Input Using Causal Associations from the Internet (in Japanese)", Proceedings of the 7th Forum of Information Technology(Vol2):267-268