9

Towards Socialized Machines: Emotions and Sense of Humour in Conversational Agents

Michal Ptaszynski, Pawel Dybala, Shinsuke Higuhi, Wenhan Shi,
Rafal Rzepka and Kenji Araki
Hokkaido University
Japan

1. Introduction

From the beginning of computer era over half a century ago, humanity was fascinated by the idea of creating a machine substituting their mental capabilities. This New Age version of Mary Shelley's *Frankenstein* gave birth to S-F literature and was one of the motors for development of our civilisation. The mental functions digitalized as the first ones were fast processing of large numbers or sophisticated formulas for specialized fields like mathematics or physics. These functions were the most troublesome for humans, but the easiest to process mechanically. Ironically, the human mental functions said to be the most human-like, and thought of as the ones which make up a grown well-socialized man, such as a sense of humour or understanding emotions of others, were neglected in Computer Science for a long time as too subjective and therefore unscientific. With the development of the Artificial Intelligence research and the related fields, like Human-Computer Interaction (HCI) or Human Factors Design, shortly before the new millennium the door opened to the fields of research of what had been unscientific till then – Affective Computing (Picard, 1997), and Humour Processing (Binsted, 1996). When Kerstin Dautenhahn and colleagues talked about the Socially Intelligent Agents (SIA) on the AAAI Fall Symposium in 2000 (Dautenhahn et al., 2002), they signalised the need for the attempts to incorporate multiple human factors into conversational agents. However, completing the task of creating a user-friendly and human-like machine was still far ahead.

In this chapter we present some of the first practical experiments on enhancing Japanese speaking conversational agent with human factors. In our research we focused on the two important features, said to make up an intelligent and socialized man: understanding emotions of others, and a sense of humour to evoke positive attitudes in other people for better socialization (Yip & Martin, 2006). These two features are also said to be the most creative and difficult to process by machines human factors (Boden, 1998). In our research we undertake the task to incorporate these two features in a conversational agent to make it more human like. A conversational agent is enhanced with a pun generator, and a system for affect analysis. The affect analysis system uses a novel method of estimating not only the valence and type of the conveyed emotions, but also, supported with a Web-mining procedure, verifies whether the emotion is appropriate for the present context of the

conversation. The pun generator is using Web contents to generate fresh jokes each time used. We perform a number of experiments concerning the incorporation of those two features. We verify the accuracy of affect analysis system in laboratory settings, as well as in the field, during a chat of users with two conversational agents - first one using modality to enhance utterance generation based on propositions gathered from the Web, and the second one - using also automatically generated puns to better socialize with the user. We check the influence of using puns on human-computer interaction.
The outline of this chapter is as follows. First, we present the conversational agent used as a base for further experiments. Next, we describe the pun-generator, and present the ideas for its combination with the conversational agent. One of the combination methods assumes using an affect analysis system to recognize user's emotions and on its basis decide whether to generate a joke or not. We present a system for affect analysis of textual input. Finally we describe experiments with implementing those two systems – pun generator and affect analysis system - into the baseline conversational agent. The implementation of those two systems is performed first separately, and then we present the first attempt to implement both of the systems. At the end concluding remarks are presented and perspectives for further research are discussed.

2. Modalin - conversational agent as a platform for experiments

Many task-oriented conversational agents (Liu et al., 2003; Reitter et al., 2006) have been developed. Research on non-task-oriented conversational agents like casual conversation dialogue systems ("chat-bots") is on the other hand not very common, perhaps due to many amateurs who try to build naturally talking systems using sometimes very clever, but rather unscientific methods. Although there are systems with chatting abilities (Bickmore & Cassell, 2001), they concentrate on applying strategies to casual conversations rather than on their automatic generation. We believe, that the main reason is that an unrestricted domain is disproportionately difficult compared to the possible uses 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 conversational agents 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, such as information kiosks or automatic guides placed in public places where anyone can talk to them about anything (Gustafson & Bell, 2000; Kopp et al., 2005) regardless of the role the developers intended. Well-known examples of non-task-oriented conversational agents are ELIZA (Weizenbaum, 1966) and A.L.I.C.E. Both systems and their countless imitators[1] 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., the knowledge resource is limited to the existing database. These examples and many other "chat-bots" need handcrafted 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

[1] 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.

learning (Araki & Kuroda, 2006) support. Therefore, building a system using automatic methods seems to be the most realistic way for unrestricted domains. Considering the large cost of developing a program capable to talk about any topic, it is reasonable to turn to the Internet - a huge and cheap source of text.
The baseline system described in this section is built upon the idea that human utterances consist of a proposition and modality (Nitta & Masuoka, 1989). The system uses an algorithm for extracting word associations from the Web and a method for adding modality to statements. The system described here works for Japanese and uses text as input and output. Though we plan to combine this project with research on voice recognition and generation, e.g., to help developing freely talking car navigation systems that by their chatting abilities could help avoiding drowsiness while driving. The general description of the system procedures in order is as follows: **1.** Extraction of keywords from user utterance; **2.** Extraction of word associations from the Web; **3.** Generation of sentence proposition using the extracted associations; **4.** Addition of modality to the sentence proposition

2.1 Extraction of keywords from user utterance

Every second millions of people update their blogs and write articles on every possible topic (Kumar et al., 2003). These are available on the Web, which can be accessed any time in a faster manner every day because of the growing efficiency of search engines. Thus, the Web is well suited to extracting word associations triggered by words from user utterance with a conversational agent. We use the Google [2] search engine snippets to extract word associations in real time without using pre-prepared resources, such as off-line databases. First, the system analyses user's utterances using the morphological analyser MeCab (Kudo, 2001) in order to spot query keywords for extracting word association lists. We define nouns, verbs, adjectives, and unknown words as query keywords. The reason we chose these word classes is that they, 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*[3] (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.

2.2 Extraction of word associations from the Web

The extracted keywords are used as query words in the Google search engine. The system extracts nouns from the search results and sorts them in frequency order. This process is based on the idea that words co-occuring frequently with the input words are of high relevance to them. The number of extracted snippets is 500 (value set experimentally, taking into account the processing time and output quality). The top five words of a list are treated as word associations (see Table 1). Approximately 81% of the word associations obtained using this method were judged as valid (Higuchi at al., 2008). 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.

[2] Google, http://www.google.co.jp/
[3] All Japanese transcriptions will be written in italics.

Sapporo wa samui. (Sapporo city is cold.)		
Association frequency ranking:		
1	*yuki* (snow)	52
2	*fuyu* (winter)	50
3	*kion* (temperature)	16
4	*jiki* (season)	12
5	*Tokyo* (Tokyo)	12

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

(noun) (*wa*) (adjective)
(noun) (*ga*) (adjective)
(noun) (*ga*) (verb)
(noun) (*wa*) (verb)
(*so-re*) (*wa*) (verb)
(noun)
(adjective)
(verb)

Table 2. Proposition templates.

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

Table 3. Examples of informative expression modality

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

Table 4. Examples of question modality sentence endings

2.3 Generation of proposition using word associations

Using the associations, the system generates the proposition of a sentence reply to the user input. A proposition is an expression representing an objective statement. It is generated by applying associations to a proposition template like [(noun) (particle *wa* indicating topic) (adjective)]. We prepared 8 proposition templates manually (see Table 2). The templates were chosen subjectively after examining statistics from IRC chat logs. Our criteria for choosing the templates was that they should belong to the 20 most frequent modality patterns and to be flexible enough to fit a range of grammatical constructions, e.g., 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: e.g., 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. In case of discarding an unnatural proposition, 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 in Google. 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. If a generated proposition is judged as valid (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 associations is prohibitively long processing time.

2.4 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 and a pair of sentence head and sentence ending auxiliary verb. This kind of modality is common in Japanese (Nitta & Masuoka, 1989).

2.4.1 Extracting Modality

There is no standard definition of what constitutes modality in Japanese. In this research we classify modality of casual conversation into questions and informative expressions. Questions are defiend as expressions that request information from the user. Informative expressions are transmitting 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 the ways as below:

- 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?* ″ (question marker) 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...). After obtaining the patterns this way, 668 for informative expressions and 396 for questions, they were filtered manually to discard the ones extracted incorrectly. The overall number of patterns obtained was 550 of the former (80%) and 292 of the latter (73%). The candidates were sorted in frequency order. The examples of modality patterns are presented in Table 3 for informative expressions and in Table 4 for questions.

2.4.2 Adding Modality

The system adds the modality from section 2.4.1 to the proposition from section 2.3 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 ″*iyaa ... desu-yo* (Ooh ... isn't it?)″ as modality pattern, the generated output will be ″*iyaa, fuyu-wa samui desu-yo* (Winter is cold, you know)″. However, there is a possibility that the output is unnatural, like ″*fuyu-wa samui dayo-ne* (Winter is cold, aren't it?)″, depending on the pair of proposition and modality. To solve 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 hits is lower than the threshold, the output is judged as incorrect and discarded, and a new reply is generated. We experimentally set the threshold to 100 hits.

2.5 Evaluation of Modalin

We used system α, generating only the proposition, and system β, generating both proposition and modality. 5 participants used each system for 10-turn conversations and evaluated the conversations on a 5-point scale. Evaluation criteria were ″will to continue the conversation″ (A), ″grammatical naturalness of dialogues″ (B), ″semantic naturalness of dialogues″ (C), ″vocabulary richness″ (D), ″knowledge richness″ (E), and ″human-likness of the system″ (F). Table 6 shows average scores for the evaluations of each system. System β that uses modality scored much higher than system α. In the evaluation, the participants expressed the opinion that an utterance like (xx *wa* 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. The results were considered to be very statistically significant with P value = .0032.

	System α (proposition)						System β (proposition + modality)					
Evaluation criteria	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	3	1.2	2	1.8	1.4	3.2	3	2.2	3.6	3.2	3.4

Table 6. Modalin evaluation results.

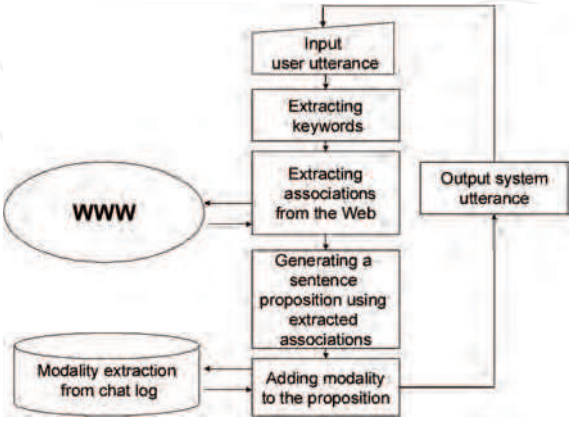

Fig. 1. Flow chart of the Modalin system.

user	*Nani-ka suki-na tabemono aru?*
	(What food do you like ?)
system	*Maa, tabemono-wa oishii-desu.*
	(Well, food tastes good.)
user	*Saikin-wa o-sake-mo suki-ni natte kita-n-da.*
	(Recently, I've become quite a alcohol lover)
system	*Demo, sake-wa yowai-no-yo-ne.*
	(But I can't drink much, oh no.)
user	*Yowai-n-da. Sugu yocchau-no?*
	(Oh, I see, you get loaded quickly?)
system	*Maa, sore-wa sugo-wa-ne.*
	(Well, it's amazing.)
user	*Boku-wa wari-to o-sake-wa tsuyoi hou-nan-da.*
	(In my case I guess I can drink quite a lot)
system	*Iya, sore-wa tsuyoi-n-de.*
	(Ooh, strong, one can say.)

Table 5. Example of a dialogue with Modalin (language errors translated close to original).

3. Pundalin – implementing humour to conversational agent

3.1 Humour as a positive factor

There are, many proofs for beneficial influence of humour on our life. For example, humour is often used as a measure to cope with negative emotions and moods, such as stress (Cann

et al., 1999), anxiety, or depression (Moran, 1996). Vilaythong et al. (2003) showed that exposing people to humorous contents, like funny videos, increased their feeling of hope and made them feel better in general. Cook & Rice (2003) provided proofs for social benefits of humour, by showing that a sense of humour in another person increases the perceived benefits of a relationship. According to Sprecher & Regan (2002), humour is also one of the main characteristics people use when choosing a partner, which means we like to interact with people with a sense of humour. Finally, Mulkay (1988) proved that we tend to use jokes when discussing difficult matters, which leads to the conclusion that humour makes conversation easier in general.

3.2 Necessity of humour in talking agents

It has been demonstrated that humans treat computers as social actors. According to SRCT (Social Response to Communication Technologies) theory, people respond to computers using the same social attitudes and behaviours they apply to humans (Reeves and Nass, 1996). This also means we expect our interaction with them go smoothly and in a natural way. Therefore, if humour enhances the interaction between humans, a similar effect should be obtained in interaction with machines.

The necessity of creating a joking conversational agent was pointed out and motivated by Nijholt (2007). However, not much has been done to actually construct such an agent. The first known attempt of this kind was made by Loehr (1996), who combined Binsted's joking system JAPE (1996) and talking agent Elmo. The results of the evaluation experiment were relatively poor, for there was barely any relevance between the user's input and the agent's humorous output. Another attempt at creating a humour-equipped agent was made by Tinholt & Nijholt (2007), who implemented a cross-reference ambiguity-based joke generator into an AIML based chat-bot. However, the opportunities for generating cross-reference jokes in daily conversation turned out to be rather rare and the impact on human involvement in the conversation could not be evaluated properly. Also, Morkes et al. (1999), checked the impact of pre-programmed (not generated) humour on a task-oriented conversation. The results showed that a humour-equipped agent was evaluated as better and easier to socialize with by human participants.

3.3 Humoroids – new class of conversational agents

Although not completely untouched (see above), the research field on humour-equipped talking agents needed to be precisely defined. The first consistent definition of such agents was proposed by Dybala et al. (2009a). His definition of this new class of agents says that humour-equipped agents, or "humoroids", are agents that are able to use humour during a conversation. He also defined two major subclasses of humoroids: task-oriented (Loehr, 1996; Morkes et al., 1999) and non-task-oriented (Tinholt & Nijholt, 2007). The agent presented here belongs to the latter type. The presence of humour is of higher importance in non-task-oriented agents, for their main purpose is to entertain human interlocutors and socialize with them during the conversation.

3.4 Punda – a pun generator for Japanese

Considering the NLP methodology, the most "computable" genre of jokes is puns. They can be found in most of the existing languages. In some, however, puns are easier to create and

thus their amount is much bigger than in others. One of such languages is Japanese, in which puns (called *dajare*) are one of main humour genres. This makes Japanese a perfect environment for pun processing research. However, although some attempts of constructing pun generating engine have been made, also in Japanese, creating a funny joking conversational system have been an unfulfilled challenge in NLP field for a long time.
PUNDA research project (Dybala et al., 2008b) is a project aiming to create a Japanese joking conversational system. As a part of this project, we developed a simple pun generating system - PUNDA Simple. This system is a simplified version of the algorithm of the main PUNDA system, which, although still under development, at its current state can be used as a pun generating support tool. Although PUNDA Simple was created for the need of this research, the main part of the algorithm is similar to the one used in the main system.

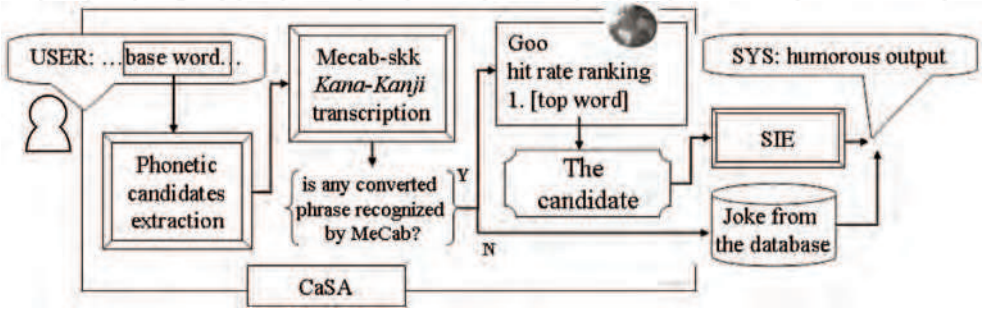

Fig. 2. Algorithm outline for PUNDA Simple joke generating engine.

3.4.1 Algorithm

The PUNDA Simple algorithm consists of two parts: Candidate Selection Algorithm (CaSA) and Sentence Integration Engine (SIE) – see Figure 2.

CaSA. In this step, the system generates a candidate for a pun. The input is a sentence, from which a base word for a pun (a word that will be transformed into a pun) is selected. The input is analysed by morphological analyser MeCab (Kudo, 2001), and if any element is recognized as an ordinary noun, it becomes the base word (a preliminary experiment proved that most of *dajare* base words are ordinary nouns). If no ordinary noun is found, one words with medium number of characters is selected randomly. Then, for the base word, pun candidates are generated using 4 generation patterns: homophony, initial mora addition, internal mora addition and final mora addition. For example, for the word *katana* (a Japanese sabre), the process goes as follows (* means one single mora):

base word: {katana}

candidates:

1. **homophony**: {katana}
2. **initial mora addition**: {*katana } (akatana, ikatana, ukatana...)
3. **final mora addition**: { katana*} (katanaa, katanai, katanau...)
4. **internal mora addition**: {ka*tana}, {kata*na} (kaatana, kaitana, kautana...)

The candidates are generated in Kana characters (one character = one mora). In the next step, for each candidate a list of possible Kanji (Japanese ideograms) transcriptions is

extracted using MeCab-skkserv Kana-Kanji Converter[4]. Then, if any of the converted character sequences of was recognized by the morphological analyser as an existing word, its hit rate was checked in the Internet. The candidate with highest Goo[5] hit rate was extracted as a pun candidate. For example, for the word katana the candidate would be *takatana* (a high shelf).

SIE. In this step, a sentence including the candidate extracted by CaSA is generated. To make system's response more related to the user's input, each sentence that included joke started with the pattern "[base word] *to ieba*" ("Speaking of [base word]"). The following part of the sentence is extracted from KWIC on WEB online Keyword-in-context sentences database (Yoshihira et al., 2004) - the candidate is used as a query word and the list of sentences including this word is extracted. All sentences that contain non-Japanese characters (such as Latin letters, brackets, numbers etc.) are deleted, as they may be difficult to understand without the context. Then, the shortest sentence with exclamation marks is selected (we assumed that most jokes convey some emotions, and exclamation marks are its visible sign). If no such sentence is found, the shortest sentence that ended with a period is selected. Finally, the part following the candidate is extracted and added as the second part to the system's response – see the example below:

User: - *Natsu wa atsui desu kedo, Toukyou hodo ja nai desu.* (The summer is hot, but not as hot as in Tokyo); [base word: *natsu* (summer), pun candidate: *natsukashii* (emotionally attached; loved)]

System: -*Natsu to ieba, natsukashii Nose de, kyuuyuu to waiwai!* (Speaking of summer, it will be fun to meet some old friends in my beloved Nose! <a town near Osaka>)

If no appropriate sentence is found on KWIC, the response ends with "[candidate] *da yo ne*" ("it's the [candidate], isn't it") – see the example below.

User: -*Ame da kara kaimono iku no wa mendokusai ne.* (It's raining so going shopping is a bad idea) [base word: *ame* (rain), pun candidate: *ame* (a candy)]

System: -*Ame to ieba ame da yo ne.* (Rain reminds me of candies.)

When the candidates list is empty, the system selects one random pun from Sjöbergh's pun database (Sjöbergh & Araki, 2008b) – see the example below.

User: - *Sou, atarashii baito ne.* (Yeah, it's my new part-time job) [base word: *baito* (a part-time job), pun candidate: *none*]

System's response (from the pun corpus): -*Kureyon wo katte kureyon!* (Buy the pen, man!).

3.5 Pundalin – joking conversational agent

The conversational agent Modalin and the pun generator were merged to create a joking conversational system Pundalin. The merging rule was – in every third dialogue turn Modalin's output was replaced by PUNDA Simple's joke. In other words, every third user's utterance became PUNDA's input and an appropriate pun for it was generated, using the algorithm described above. This method, albeit quite simple, allowed us to check if the usage of humour improved the reception of the conversational agent in the eyes of users, and therefore system's overall performance as a companion for conversation.

[4] MeCab-skkserv Kanji-Kana converter, http://chasen.org/~taku/software/mecab-skkserv/
[5] Goo search engine, http://www.goo.ne.jp/

3.6 Evaluation of Pundalin

To check if humour can enhance the non-task oriented dialogue, we conducted an evaluation experiment, using Modalin as the baseline system and Pundalin as the main, humour-equipped system. As is argued by Dybala et al., (2009c), since it is users who are the "clients" of our product, in the research on dialogue systems the first person oriented evaluation is of the highest importance. It allows checking the user's impressions of the interaction with the system in the most direct way. In the experiment, users were asked to perform a 10-turn dialogue with Modalin, and then with Pundalin. No topic restrictions were made. The utterance variety was big, however, the beginning of conversations by the users was usually very normal, like: "What did you do yesterday?", "May I ask you a question?" or "It's hot today, isn't it?" There were 13 participants, 11 male and 2 female; all of them were university undergraduate students. After talking with both systems, they were asked to fill out a questionnaire about each system's performance. The questions concerned both linguistic (B-D) and non-linguistic (A, E-H) areas of interaction: **A**) Do you want to continue the dialogue with the system?; **B**) Was the system's output grammatically natural?; **C**) Was the system's output 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 interesting?; **H**) Did you find the conversation with the system interesting?

The replies to the questions were given on 5-point scales with explanations added. Each evaluator filled out two questionnaires, one for each system. The final, summarizing question was "Which system do you think was better?" Statistical significance of the results was calculated using the student's t-test. The results are summarized in Table 7. The results show that the system with humour received higher scores in both linguistic and non-linguistic areas. As for the former, it may seem unusual that the presence of humour improved the system's linguistic skills – this fact, however, could have been caused by the fact that Pundalin uses fragments of human created sentences and jokes from a data base, which naturally are more correct than those generated automatically. Also in the non-linguistic area all results point at the humour-equipped system. Users wanted to continue the conversation with Pundalin more than with Modalin; Pundalin was perceived as more human-like, knowledgeable, funny and generally better than Modalin (Dybala et al., 2008a). Results for questions A and B were found to be significant on 6% level, and for remaining questions – on 5% level. The overall compared results of Modalin and Pundalin were extremely statistically significant, with P value = .0002.

Questions	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**Which is better?**
Modalin	2.62	2.15	1.85	2.08	2.15	2.38	1.92	2.46	15%
Pundalin	3.38	2.92	2.69	3.00	2.85	3.31	4.15	4.08	**85%**
Difference	**0.76**	**0.77**	**0.84**	**0.92**	**0.70**	**0.93**	**2.23**	**1.62**	
P value	> .05	> .05	**< .05**	**< .05**	**< .05**	**< .05**	**< .05**	**< .05**	

Table 7. User evaluation results for Modalin and Pundalin for detailed questions. Answers were given on a 5-point scale.

4. Implementing Emotional Intelligence in conversational agents

Developing methods for processing human emotions is one of the current issues in Artificial Intelligence. The field embracing this subject, Affective Computing, has been gathering popularity of researchers since being initiated only a little over ten years ago (Picard, 1997). The interest in this field is usually focused on recognizing the human emotions in human-computer interaction. In the popular methods, emotion recognition is focused on: facial expressions (Hager et al., 2002), voice (Kang et al., 2000) or biometric data (Teixeira et al., 2008). However, these methods, based on a behavioural approach, ignore the semantic and pragmatic context of emotions. Therefore, although they achieve good results in laboratory settings, such methods lack usability in real life. A system for recognition of emotions from facial expressions, assigning "sadness" when a user is crying would be critically mistaken, if the user was, e.g., cutting an onion in the kitchen. This leads to the need of applying contextual analysis to emotion processing. Furthermore, although it was proved that affective states should be analysed as emotion specific (Lerner & Kelter, 2000), most of the behavioural approach methods simply classify them to opposing pairs such as joy-anger, or happiness-sadness (Teixeira et al., 2008). A positive change in this tendency can be seen in text mining and information extraction approaches to emotion estimation (Tokuhisa et al., 2008; Ptaszynski et al., 2009b). However, the lack of standardization often causes inconsistencies in emotion classification. As one of the recent advances in affect analysis, it was shown that Web mining methods can improve the performance of language-based affect analysis systems (Tokuhisa et al., 2008; Ptaszynski et al., 2009b). However, in such methods, although the results of experiments appear to be positive, the two different approaches, the language-syntax based and Web mining based, are mixed. The former, comparing the information provided by the user to the existing lexicons and sets of rules, is responsible for recognizing the particular emotion expression conveyed by the user. The latter is based on gathering from the Internet large numbers of examples and deriving from them an approximated reasoning about what emotions usually associate with a certain contents. Using the Web simply as complementary mean for the language based approach, although achieving reasonable results, means not fully exploiting the great potential lying in the Web (Rzepka and Araki, 2007).

In this research we present a method capable of specifying users' emotional states in a more sophisticated way than simple valence classification. The method also contributes to standardization of the emotion classification for the Japanese language since instead of creating a new classification we apply the most reliable and coherent one available today, mentioned firstly by Ptaszynski et al. (2008) and developed further by Ptaszynski et al. (2009b), who base their classification on Nakamura's (1993) research in lexicology of emotive expressions in the Japanese language. Finally, our method does not only specify what type of emotion was expressed, but also determines whether the expressed emotion is appropriate for the context it appears in. In the method we use Ptaszynski's et al., (2009f) system for affect analysis and annotation of utterances and Shi's et al. (2008) method for gathering emotive associations from the Web. The baseline of the system presented here was first proposed by Ptaszynski et al. (2009c) and evaluated at 45% of accuracy. We improved the system in two ways. Firstly, Ptaszynski's system for affect analysis was improved with Contextual Valence Shifters to avoid confusing the valence polarity of emotive expressions. Secondly, we improved Shi's Web mining technique. The problem was it was gathering too much noise from the Internet. To solve this problem we referred to the proof provided by

Abbasi and Chen (2007), who showed that public Web services, such as forums or blogs, are rich in emotive information and thus ideal for affect analysis. Therefore we restricted the mining scope of Shi's technique from the whole Web to the contents of *Yahoo!Japan–Blogs* (blogs.yahoo.co.jp) a robust weblog service.

4.1 Definitions

Emotional Intelligence

The idea of *Emotional Intelligence* (EI) was first officially proposed by Salovey & Mayer (1990), who defined it as a part of human intelligence consisting of the ability to: I) perceive emotions; II) integrate emotions to facilitate thoughts; III) understand emotions; IV) regulate emotions. In the EI Framework (Mayer & Salovey, 1997) the first step consists of the abilities to a) identify emotions and b) discriminate between appropriate and inappropriate expressions of emotion. Salovey and Mayer (1990) argue that recognizing emotions is only the first step to acquire full scope of Emotional Intelligence and does not tell us anything about whether it is appropriate for a given situation or what reactions should be undertaken. According to Solomon (1993), the valence of emotions is determined by the context they are expressed in. For example, anger can be positive, when warranted (e.g. a reaction to a direct and deliberate offence) and negative, when unwarranted (scolding one's own children unjustly) and the reactions should be different for the two different contexts of anger. The attempts to implement the EI Framework usually do not go beyond theory (Andre et al., 2004), and the few practical attempts eventually still do not surmount the first step of recognition (Picard et al., 2001). The research presented here is an attempt to go beyond this simple approach. Following emotion recognition, their appropriateness is verified against their contexts. By providing an agent means to determine the appropriateness of emotions, we make a step towards the full implementation of EI framework in machines.

Definition and classification of emotions

Our working definition of emotions is based on Nakamura's (1993), who defines them as every temporary state of mind, feeling, or affective state evoked by experiencing different sensations. This definition is complemented by Solomon's (1993), who argues that people are not passive participants in their emotions, but rather the emotions are strategies by which people engage with the world. Since we operate on language, the above is further complemented by Beijer's (2002) definition of emotive utterances, which he describes as every utterance in which the speaker is emotionally involved, and this involvement, expressed linguistically, is informative for the listener. Nakamura (1993), proposed also a 10 type emotion classification, the most appropriate for the Japanese language: *ki / yorokobi* (joy, delight), *do / ikari* (anger), *ai / aware* (gloom, sorrow, sadness), *fu / kowagari* (fear), *chi / haji* (shame, shyness, bashfulness), *kou/ suki* (liking, fondness), *en / iya* (dislike, detestation), *kou / takaburi* (excitement), *an / yasuragi* (relief) and *kyou / odoroki* (surprise, amazement).

Contextual Valence Shifters

The idea of Contextual Valence Shifters (CVS) as an application in Sentiment Analysis was first proposed by Polanyi & Zaenen (2004). They distinguish two kinds of CVS: negations and intensifiers. The group of negations contains words like "not", "never", and "not quite", which change the valence polarity of semantic orientation of an evaluative word they are attached to. The group of intensifiers contains words like "very", "very much", and "deeply", which intensify the semantic orientation of an evaluative word. So far the idea of CVS

analysis was successfully applied to Sentiment Analysis of English texts (Kennedy & Inkpen, 2005). Successful attempts on Japanese ground (Miyoshi & Nakagami, 2007) show that it is also applicable for the Japanese language. Examples of CVS negations in Japanese are grammatical structures like: *amari -nai* (not quite-), *-to wa ienai* (cannot say it is-), or *mattaku - nai* (not at all-). Intensifiers are represented by: *totemo-* (very much-), *sugoku-* (-a lot), or *kiwamete-* (extremely). The idea of CVS is applied in line with Ptaszynski's et al. (2009a) research on improving affect analysis by valence shifting. The Contextual Valence Shifting Procedure (details below) is supported further with Russell's (1980) 2-dimmensional model of affect.

Two-dimensional model of affect

The idea of a two-dimensional model of affect was first proposed by Schlosberg (1952) and developed further by Russell (1980). Its main assumption is that all emotions can be described in a space of two-dimensions: valence polarity (positive/negative) and activation (activated/deactivated). An example of positive-activated emotion is excitement; positive-deactivated emotion is, e.g., relief; negative-activated and deactivated emotions are anger and gloom respectively. This way 4 emotion areas are distinguished: activated-positive, activated-negative, deactivated-positive and deactivated-negative. Nakamura's emotion types were mapped on this model and their affiliation to one of the spaces was determined. Those emotions for which the affiliation was not obvious (e.g. surprise can be both positive as well as negative, etc.) were mapped on all of the areas they could belong to. However, no emotion type was mapped on more than two adjacent fields. This grouping is then used in our system for two reasons. Firstly, in the CVS analysis procedure to specify which emotion corresponds to the one negated by a CVS phrase. Secondly, in emotion appropriateness verification procedure, for estimating whether the emotion types belong to the same area, even if not perfectly matching with the emotive associations gathered from the Web.

Example of a sentence (English translation)	Emotemes	**Emotive expressions**
(1) *Kyo wa nante* ***kimochi ii*** *hi nanda!* (Today is such a nice day!)	yes	yes
(2) *Iyaa, sore wa sugoi desu ne!* (Woa, that's great!)	yes	no
(3) *Ryoushin wa minna jibun no kodomo wo* ***aishiteiru.*** (All parents love their children.)	no	yes
(4) *Kore wa hon desu.* (This is a book.)	no	no

Table 8. Examples of sentences containing emotemes (underlined) and/or emotive expressions (bold type font).

4.2 Linguistic approach to emotions – the emotive function of language

The semantic and pragmatic diversity of emotions is best conveyed in language (Solomon, 1993). Therefore we designed our method to be language-based. There are different linguistic means used to inform other interlocutors of emotional states. The elements of speech used to convey emotive meaning are described by the emotive function of language (Jakobson, 1960). In Japanese it is realized lexically through such parts of speech as exclamations (Beijer, 2002), hypocoristics (endearments), vulgar language (Crystal, 1989;

Potts & Kawahara, 2004) and mimetic expressions (in Japanese: *gitaigo*) (Baba, 2003). A key role in expressing emotions is also played by the lexicon of words describing emotional states (Nakamura, 1993). The para-linguistic elements, like intonation, are represented lexically by exclamation marks or ellipsis. Ptaszynski (2006) classified the realizations of emotive function in Japanese in two general types. The first one, emotive elements (or **emotemes**), indicate that emotions have been conveyed, but not detailing their specificity. This group is linguistically realized by interjections, exclamations, mimetic expressions, or vulgarities. The second type, **emotive expressions**, are parts of speech like nouns, verbs, adjectives or metaphors describing affective states. Examples of sentences containing emotemes and/or emotive expressions are shown in Table 8. Examples (1) and (2) are emotive sentences. (1) is an exclamative sentence, which is determined by the use of exclamative constructions *nante* (how/such a) and *nanda!* (exclamative sentence ending), and contains an emotive expression *kimochi ii* (to feel good). (2) is also an exclamative. It is easily recognizable by the use of an interjection *iyaa*, an adjective in the function of interjection *sugoi* (great), and by the emphatic particle *-ne*. However, it does not contain any emotive expressions and therefore it is ambiguous whether the emotions conveyed by the speaker are positive or negative. The examples (3) and (4) show non-emotive sentences. (3), although containing an emotive verb *aishiteiru* (to love), is a generic statement and, if not put in a specific context, does not convey any emotions. Finally, (4) is a simple declarative sentence without any emotive value.

4.2.1 Defining emotive linguistic features

We defined emotemes and emotive expressions according to Ptaszynski's two-part classification. The feature set was defined in a way similar to the one proposed by Alm et al. (2005), by using multiple features to handle emotive sentences. Alm however, designed their research for English children's stories, whereas we focus on utterances in Japanese, and therefore used Ptaszynski's classification as more appropriate for our research.

Emotemes

Into the group of emotive elements, formally visualisable as textual representations of speech, Ptaszynski (2006) includes the following lexical and syntactical structures.
Exclamative utterance. The research on exclamatives in Japanese (Ono, 2002; Sasai, 2006) provides a wide scope of topics useful as features in our system. Some of the exclamative structures are: *nan(te/to/ka)-*, *-darou*, or *-da(yo/ne)*, partially corresponding to *wh-* exclamatives in English (see the first sentence in Table 8).
Interjections are typical emotems. Some of the most representative Japanese interjections are *waa*, *yare-yare* or *iyaa* (see the second sentence in Table 8).
Casual Speech. Casual speech is not an emotem per se, however, many structures of casual speech are used when expressing emotions. Examples of casual language use are modifications of adjective and verb endings *-ai* to *-ee*, like in the example: *Ha ga itee!* (My tooth hurts!), or abbreviations of forms *-noda* into *-nda*, like in the example: *Nani yattenda yo!?* (What the hell are you doing!?).
Gitaigo. Baba (2003) distinguishes *gitaigo* (mimetic expressions) as emotems specific for the Japanese language. Not all *gitaigo* are emotive, but rather they can be classified into emotive mimetics (describing one's emotions), and sensation/state mimetics (describing manner and appearance). Examples of emotive *gitaigo* are: *iraira* (be irritated), like in the sentence:

Omoidasenkute iraira shita yo. (I was so irritated, 'cause I couldn't remember.), or *hiyahiya* (be in fear, nervous), like in the sentence: *Juugeki demo sareru n janai ka to omotte, hiyahiya shita ze*. (I thought he was gonna shoot me - I was petrified.)

Emotive marks. This group contains punctuation marks used as textual representations of emotive intonation features. The most obvious example is exclamation mark „!" (see Table 8). In Japanese, marks like "..." (ellipsis), or prolongation marks, like "—" or "~" are also used to inform interlocutors that emotions have been conveyed

Hypocoristics (endearments) in Japanese express emotions and attitudes towards an object by the use of diminutive forms of a name or status of the object (*Hanako* [girl's name] vs *Hanako-chan* [/endearment/]; *o-nee-san* [older sister] vs *o-nee-chan* [sis /endearment/], *inu* [a dog] vs *wanko* [doggy /endearment/]). Sentence example: *Saikin Oo-chan to Mit-chan ga boku-ra to karamu youni nattekita!!* (Oo-chan and Mit-chan has been palling around with us lately!!)

Vulgarisms. The use of vulgarisms usually accompanies expressing emotions. However, despite a general belief that vulgarisms express only negative meaning, Ptaszynski (2006) notices that they can be also used as expressions of strong positive feelings, and Sjöbergh (2006) showed, that they can also be funny, when used in jokes, like in the example: *Mono wa mono dakedo, fuete komarimasu mono wa nanda-? Bakamono.* (A thing (*mono*) is a thing, but what kind of thing is bothersome if they increase? Idiots (*bakamono*).)

Emotive expressions

A lexicon of expressions describing emotional states contains words, phrases or idioms. Such a lexicon can be used to express emotions, like in the first example in Table 8, however, it can also be used to formulate, not emphasized emotively, generic or declarative statements (third example in Table 8). Some examples are:

adjectives: *sabishii* (sad), *ureshii* (happy);

nouns: *aijou* (love), *kyofu* (fear);

verbs: *yorokobu* (to feel happy), *aisuru* (to love);

fixed phrases/idioms: *mushizu ga hashiru* (give one the creeps [of hate]), *kokoro ga odoru* (one's heart is dancing [of joy]);

proverbs: *dohatsuten wo tsuku* (be in a towering rage), *ashi wo fumu tokoro wo shirazu* (be with one's heart up the sky [of happiness]);

metaphors/similes: *itai hodo kanashii* (pain of sadness), *aijou wa eien no honoo da* (love is an eternal flame);

4.3 ML-Ask

Based on the linguistic approach towards emotions as well as the classification of emotions, Ptaszynski et al. (2009f) constructed ML-Ask (eMotive eLements–SeeK & Analyse) system for automatic annotation of utterances with emotive information. The emotem database was gathered manually from other research and grouped into five types (code, reference research and number of gathered items in square, round and curly brackets, respectively):

1. **[EX]** Interjections and structures of exclamative and emotive-casual utterances (Nakamura, 1993; Oshima-Takane et al., 1995-1998; Tsuchiya, 1999; Ono, 2002). {477}
2. **[GI]** *Gitaigo* (Nakamura, 1993; Oshima-Takane et al., 1995-1998; Baba, 2003). {213}
3. **[HY]** Hypocorystrics (Kamei et al., 1996). {8}
4. **[VU]** Vulgarisms (Sjöbergh, 2008a). {200}

5. **[EM]** Emotive marks (Kamei et al., 1996). {9}

These databases were used as a core for ML-Ask. We also added Nakamura's (1993) dictionary as a database of emotive expressions (code: **[EMO-X]**, 2100 items in total). The breakdown with number of items per emotion type was as follows: *yorokobi* {224}, *ikari* {199}, *aware* {232}, *kowagari* {147}, *haji* {65}, *suki* {197}, *iya* {532}, *takaburi* {269}, *yasuragi* {106}, *odoroki* {129}.

4.3.1 Emotems analysis procedure

Based on the databases described above, a textual input utterance is analysed and emotive information is annotated. The system first determines whether an utterance is emotive (appearance of at least one emotive feature), extracts all features from the sentence, and analyses the structure of the emotive utterance. This is the system's main procedure. Examples of analysis are shown below (from top line: example in Japanese, emotive information annotation, English translation; emotems-underlined, emotive expressions bold type font, n-noun, ptl-particle, AUX-auxiliary verb, the system flow is shown on Figure 3).

(1)	*Kyo*	*wa*	*nante*	***kimochi ii***	*hi*		*nanda*	*!*
	Today	ptl:THEM	EX:nante	**EMO-X:joy**	day:SUBJ		EX:nanda	EM:!
(2)	*Iyaa,*	*sore*	*wa*	*sugoi*	*desu*	*ne*	*!*	
	EX:iyaa	that	ptl:THEM	EX:sugoi	AUX	EX:ne	EM:!	
(5)	***Akirame***		*cha*	*ikenai*		*yo*	*!*	
	EMO-X:dislike		EX:cha	|CVS:cha-ikenai{→joy}		EX:yo	EM:!	

Translation: Don't cha give up!

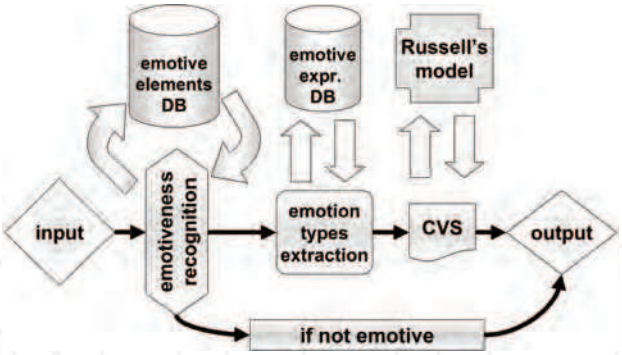

Fig. 3. ML-Ask system flow chart

4.3.2 Emotive expressions analysis procedure

In all utterances determined as emotive, the system searches for emotive expressions from the databases (Nakamura's dictionary). This procedure is used to verify:

1) How many of all determined emotive utterances contain emotive expressions;

2) If the system is capable of determining specific types of emotions in human-computer interaction. However, keyword-based extraction allowed mismatching the specific emotion types. To avoid this we applied Contextual Valence Shifters.

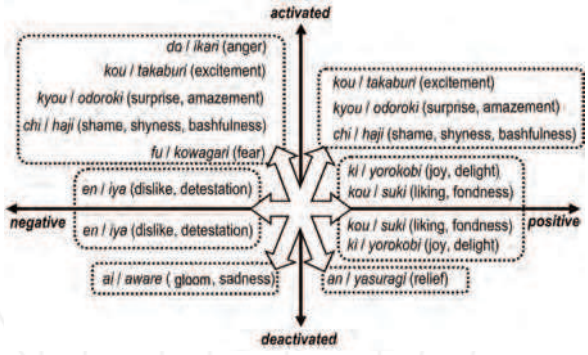

Fig. 4. Nakamura's emotion types mapped on Russell's space and gathered in 6 contrasting groups: positive-activated↔negative-deactivated; negative-activated↔positive- deactivated; positive↔negative (for emotion types with two possible activation parameters).

4.3.3 Contextual Valence Shifters in ML-Ask

When a CVS structure is discovered, ML-Ask changes the valence polarity of the detected emotion. To specify the emotion types afterwards, we applied the 2-dimensional model of affect (Russell, 1980). After valence shifting the emotion type is determined as the one with valence polarity and activation parameters opposite to the contrasted emotion (see Figure 4 and example (5) above). Nakamura's emotion types mapped on Russell's model specify a restricted group of potential emotions.

4.3.4 Evaluation of ML-Ask in laboratory settings

In the evaluation we contrasted the emotive information annotation by ML-Ask and laypeople. We gathered a corpus of natural utterances through an anonymous survey in which we asked people of different ages and social groups to remember a conversation with a friend and write three sentences from that conversation: 1) free, 2) emotive, and 3) non-emotive. From this collection we used only the utterances meant to be either emotive or non-emotive. The participants also annotated the specific emotion types on emotive utterances. Since laypeople, not possessing any linguistic knowledge, are unable to describe the structure of emotive utterances, we checked whether ML-Ask could distinguish between emotive and non-emotive utterances. As a result, ML-Ask annotated correctly 72 from 80 utterances (90%). In 2 cases the system wrongly annotated utterances as "emotive", in 6 cases it was the opposite. The system's agreement with annotators was indicated as very high (kappa=.8). Therefore, ML-Ask proved its reliability in identifying emotive utterances.

We also checked whether the system could determine about the specified emotion types. ML-Ask can potentially extract up to ten emotion types for one utterance. However, some of them could be extracted wrongly or unextracted at all. Therefore we calculated the accuracy as balanced F-score with emotive tags added by authors of the utterances as gold standard. The system's accuracy in estimating the specific types of emotions including "non-emotive" reached F=.45 (Precision=.62, Recall=.35) of balanced F-score. Then we asked a group of third party human annotators (but different to the first group) to annotate the emotion types to the utterances (12 people for one utterance) to check the human level in this task. For the third party human annotators the average F-score was .72 (Precision=.84, Recall=.64),

therefore system's accuracy was 62.5% (.45/.72) of the human level. This experiment verified the reliability of Nakamura's dictionary as emotive expressions database. The coverage was not high (Recall=.35), which has two reasons: 1) Nakamura stopped updating his dictionary in 1993 and the lexicon is out-of-date (only 2100 expressions); 2) instead of using straight forward emotive expressions, people often use ambiguous emotive utterances in which valence of emphasis is based on the situation the sentence is used in (see example (2) in Table 8).

4.4 Web mining technique

To verify the appropriateness of the speaker's affective states we applied Shi's et al. (2008) Web mining technique for extracting emotive associations from the Web. Ptaszynski et al. (2009b) already showed that ML-Ask and Shi's technique are compatible and can be used as complementary means to improve the emotion recognition task. However, these two methods are based on different assumptions. ML-Ask is a language based affect analysis system and can recognize the particular emotion expression conveyed by a user. On the other hand, Shi's technique is gathering from the Internet large number of examples and deriving from such data an approximated reasoning about what emotion types usually associate with an input contents. Therefore it is more reasonable to use the former system as emotion detector, and the latter one as verifier of naturalness, or appropriateness of user emotions.

Shi's technique performs common-sense reasoning about what emotions are the most natural to appear in a context of an utterance, or, which emotions should be associated with it. Emotions expressed, which are unnatural for the context (low or not on the list) are perceived as inappropriate. The technique is composed of three steps: 1) phrase extraction from an utterance; 2) morpheme modification; 3) extraction of emotion associations.

4.4.1 Phrase extraction procedure

An utterance is first processed by MeCab (Kudo, 2001). Every element separated by MeCab is treated as a unigram. All unigrams are grouped into larger n-gram groups preserving their word order in the utterance. The groups are arranged from the longest n-gram (the whole sentence) down to all groups of trigrams.

4.4.2 Morpheme modification procedure

On the list of n-gram phrases the ones ending with a verb or an adjective are then modified grammatically in line with Yamashita's argument (Yamahsita, 1999) that Japanese people tend to convey emotive meaning after causality morphemes. This was independently confirmed experimentally by Shi et al., (2008). They distinguished eleven emotively stigmatised morphemes for the Japanese language using statistical analysis of Web contents and performed a cross reference of appearance of the eleven morphemes with the emotive expression database using the Google search engine. This provided the results (hit-rate) showing which of the eleven causality morphemes were the most frequently used to express emotions. For the five most frequent morphemes, the coverage of Web mining procedure still exceeded 90%. Therefore for Web mining they decided to use only those five ones, namely: *-te*, *-to*, *-node*, *-kara* and *-tara* (see Table 9).

morpheme	*-te*	*-to*	*-node*	*-kara*	*-tara*	*-ba*	*-nowa*	*-noga*	*-nara*	*-kotoga*	*-kotowa*
result	**41.97%**	**31.97%**	**7.20%**	**6.32%**	**5.94%**	3.19%	2.30%	2.15%	1.17%	0.35%	0.30%

Table 9. Hit-rate results for each of the eleven morphemes

4.4.3 Emotion type extraction procedure

In this step the modified phrases are used as a query in Google search engine with 100 snippets for one morpheme modification per query phrase. This way a maximum of 500 snippets for each queried phrase is extracted and cross-referenced with emotive expression database (see Figure 5). The emotive expressions extracted from the snippets are collected, and the results for every emotion type are sorted in descending order. This way we obtain a list of emotions associated with the queried sentence - the approximated emotive common sense used further as an appropriateness indicator (an example is shown in Table 10).

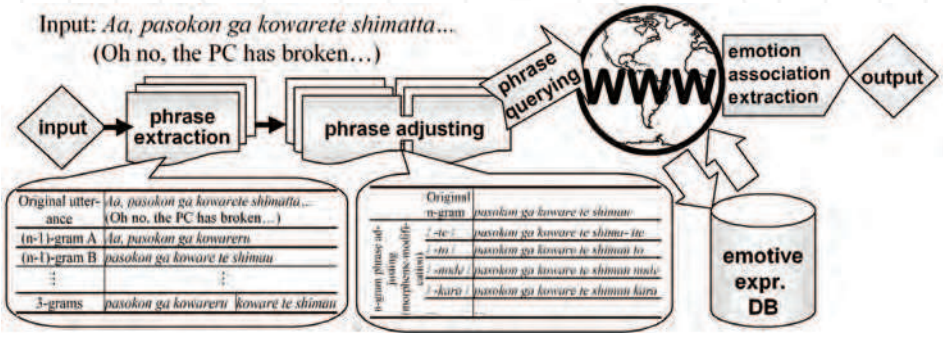

Fig. 5. Flow chart of the Web mining technique

4.4.4 Blog mining

The baseline Web mining method, using Google to search through the whole Web, was gathering a large amount of noise. To solve this problem we made two modifications. Firstly, we added a command stopping the search if any emotions were found using the longer n-grams. This assures the extraction of only the closest emotive associations and speeds up the extraction process. Secondly, since, as mentioned before, people convey on blogs their opinions and emotions, we restricted the mining to blog contents to assure extraction of more accurate emotive associations. The blog mining procedure performs the query first on the public blogs from *Yahoo!Japan-Blogs*. The paragraphs of each blog containing query phrases are co-referenced with emotive expression database to gather the emotive associations. If no information was gathered from the blog contents, the same search is performed with the baseline conditions - on the whole Web. An example of improvement is presented in Table 10.

4.4.5 Evaluation of Web mining technique

To evaluate the Web mining technique we used the same collection of utterances as in evaluation of ML-Ask. However, as mentioned above, the Web mining technique is meant not to recognize the emotions of a particular user, but rather to find a general common sense

about what emotion should be expressed in a particular utterance. Therefore, here, we used the emotions tagged by the third party evaluators as the gold standard. The correct result had to fulfil at least one of the two conditions: **1)** one or more of the extracted emotive associations belonged to the group of emotion types tagged by the third party annotators; **2)** the extracted emotive associations agreed with the majority of the human annotations. Under these conditions, Shi's Web mining technique obtained an accuracy rate of 72%.

Sentence: *Konpyuuta wa omoshiroi desu ne.* (Computers are so interesting.)			
Extracted emotion type	**Baseline: Type extracted / all extracted types(Ratio)**	**Extracted emotion type**	**Blogs: Type extracted / all extracted types(Ratio)**
liking	79/284(0.287)	liking	601/610(0.985)
surprise	30/284(0.105)	excitement	1/610 (0.001) [rejected as noise]
excitement	30/284(0.105)	fear	1/610 (0.001) [rejected as noise]
fear	29/284(0.102)	relief	1/610 (0.001) [rejected as noise]

Table 10. Example of emotion association extraction from the Web and its improvement by blog mining procedure.

4.5 Contextual appropriateness of emotion verification (CAEV) procedure

The two above described systems are combined as follows: when ML-Ask discovers an emotive utterance and successfully specifies emotion types, the Web mining technique begins the process of verification of whether the expressed emotions are appropriate to be used in such context. The emotions commonly associated with the context of the utterance are extracted from the Web and the emotions reported by ML-Ask from the user utterances are compared to the Web mining results. If the two match, the emotions expressed by the user are determined to be appropriate for the context. In such situations, a conversational agent equipped with this system could choose a dialogue strategy that sympathizes with the user (two examples are presented below).

Positive-appropriate emotion:
User: I'm so happy I passed the exam!
`ML-Ask: joy;  Web mining: joy, excitement`
Agent: Yeah! That's great! `[sympathy]`

Negative-appropriate emotion:
User: I hate him for making a fool of me in front of everyone.
`ML-Ask: dislike;  Web mining: anger, dislike`
Agent: You have a reason to be angry. `[empathy]`
However if the verification procedure indicates that the expressed emotion is inappropriate for the context, the agent could undertake different measures, e.g. helping the user to manage their emotions. Two examples are shown below.

Positive-inappropriate emotion:
User: I'd be happy if he was hit by a car!
`ML-Ask: joy;  Web mining: fear, sadness`
Agent: Are you sure that is what you are really feeling? `[counselling]`

Negative-inappropriate emotion:
User: I won the prize, but I feel so bored.
`ML-Ask: dislike, depression; Web mining: excitement, joy`
Agent: You should be happy! `[consolation]`

4.9.1 Two-dimensional Model of Affect in CAEV procedure

According to Solomon (1993), people sometimes misattribute the specific emotion types, but they rarely misattribute their valence. One could, e.g., confuse anger with irritation, but not admiration with detestation. Therefore, we checked if at least the general features matched even when specific emotion types did not match perfectly with the emotive associations. By general features we refer to those proposed by Russell (1980) in the theory of the two-dimensional model of affect. Using the mapping of Nakamura's emotion types on Russell's model we checked whether the emotion types tagged by ML-Ask belonged to the same space, even if they did not perfectly match the emotive associations gathered from the Web.

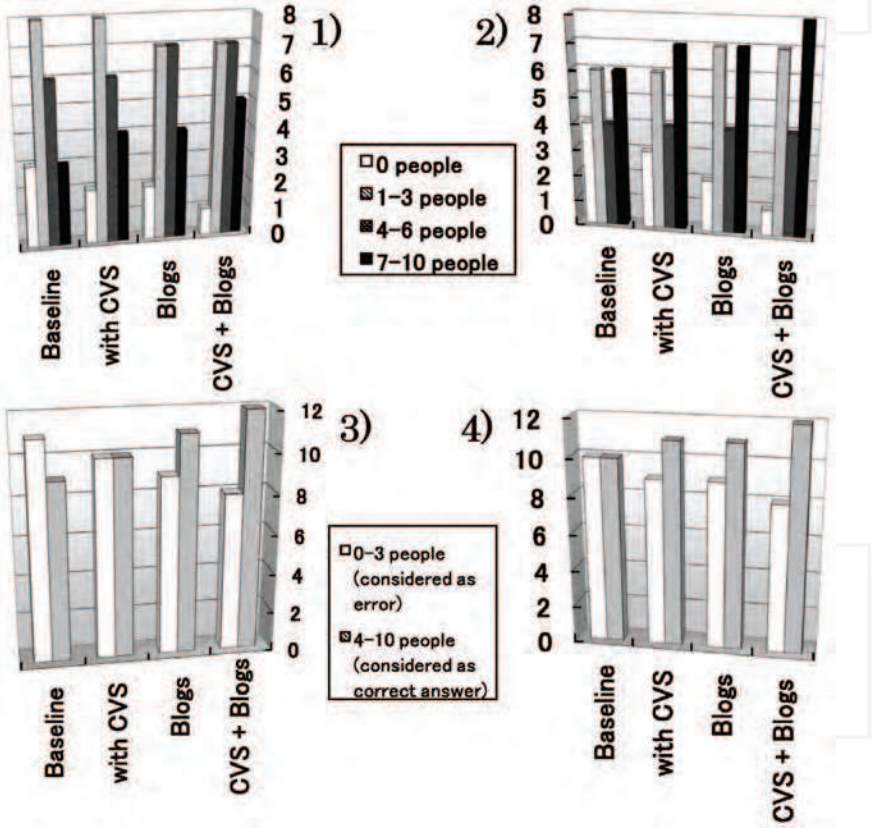

Fig. 6. Results of CAEV Procedure evaluation in estimating appropriateness of: specific emotion types - 1), 3) and valence - 2), 4). Summarization in four - 1), 2) and two - 3), 4) groups of results.

4.10 Evaluation experiment

To evaluate the method we performed an experiment. In the experiment we used the chat logs from the evaluation experiment of Modalin and Pundalin (see section **3**). All 26 conversations were analysed by ML-Ask. 6 out of all 26 conversations contained no specified emotional states and were excluded from the further evaluation process. For the rest the Web mining procedure was carried out to determine whether the emotions expressed by the user were contextually appropriate. We compared four versions of the method: 1) ML-Ask and Web mining baseline; 2) ML-Ask supported only with CVS, Web mining baseline; 3) ML-Ask baseline and Blog mining; 4) supported with both - CVS and blog mining. The difference in results appeared in 5 conversation sets. Then a questionnaire was designed to evaluate how close the results were to human thinking. One questionnaire set consisted of one conversation record and questions inquiring what were: 1) the valence and 2) the specific type of emotions conveyed in the conversation, and 3) whether they were contextually appropriate. Every questionnaire set was filled out by 10 people (undergraduate students, but different from the users who performed the conversations with the agents). The conversations where differences in results appeared for the two compared procedures, were evaluated separately for each version of the method. Therefore there were 20 questionnaire sets for the baseline method and additional 5 for the conversations which results changed after improvements. With every questionnaire set filled by 10 human evaluators we obtained a total number of 250 different evaluations performed by different people. For every conversation set we calculated how many of the human evaluators confirmed the system's results. The evaluated items were: A) general valence determination and B) specific emotion types determination accuracies of ML-Ask; the accuracy of the system as a whole to determine the contextual appropriateness of C) specific emotion types and D) valence. The results for A) and B) are provided in Table 12. The results for C) and D) are given in Table 11, 12 and in Figure 6: 1)-4).

4.10.1 Results

In the majority of research on emotion processing in Japanese, the process of evaluation is usually performed by a small number of evaluators, sometimes only one or two people (Tokuhisa et al., 2008). This might cause the problem of small statistical significance of the results, or, simply of subjectivisation of the results and its small reliability. To overcome this tendency, in our research we used 10 independent evaluators for each agent conversation and assumed that at least four people out of ten have to confirm the system's results. This is a fair ratio when we consider that it means that at least four people of ten had to provide exactly the same results as the system. The survey provided many positive results. Improving the method with both CVS procedure and restricting the query scope in the Web mining procedure to blog contents improved the performance of the appropriateness verification procedure both on the level of valence and specific emotion types. The highest accuracy and therefore the most efficient one was the version of the system with both improvements applied, by which the system's performance was improved from 45.0% to 60.0% for the specific emotion types and from 50.0% to 60.0% for the valence. Moreover, for the conversations without humour (with Modalin) the system’s performance reached satisfying 70%. The contextual appropriateness of emotions was more difficult to determine in the conversations containing puns, which is reasonable, since humour is said to be one of the most creative and therefore difficult tasks in Artificial Intelligence (Boden, 1998). In most

cases the results changed after the improvement were statistically significant (see Table 13) on a 5% level. The only version in which the change of the results was not significant was the baseline method with only CVS improvement (P value= 0.1599). Improving the system with blog mining, when compared to both – baseline version of the system and with CVS, were statistically significant (0.0274) and, what is the most important, the results of the version fully improved were the most significant from all (P=0.0119). Some of the successful examples of the CAEV Procedure are shown in Table 14. Aside from this, in most cases the results of affect analysis performed by ML-Ask were confirmed by humans. ML-Ask supported with the CVS procedure acquired 90.0% of accuracy in emotion valence recognition and 85% in specific emotion types recognition. This confirmed the system's high performance "in the field" although not ideal accuracy was achieved in laboratory settings (see section 4.3.4). This shows that language behaviour causing the system errors occurs rarely in the real conversation. As an interesting remark we should add that the survey participants sometimes determined the valence and the specific types of emotions in a non-standard way, e.g. for some contexts "fear" was determined as positive, or "joy" as negative, etc. As we assume, it is another proof that emotions are not only constituted of valence, but rather the valence of an emotion is made up by a certain context it is expressed in.

C) Emotion types	**Modalin**				**Pundalin**				**Modalin + Pundalin**			
Version of the system	Baseline system	With CVS	Blog mining	CVS + Blogs	Baseline system	With CVS	Blog mining	CVS + Blogs	Baseline system	With CVS	Blog mining	CVS + Blogs
0 people	1	1	1	1	2	1	1	0	3	2	2	1
1-3 people	3	3	2	2	5	5	5	5	8	8	7	7
4-6 people	**4**	**4**	**5**	**5**	**2**	**2**	**2**	**2**	**6**	**6**	**7**	**7**
7-10 people	**2**	**2**	**2**	**2**	**1**	**2**	**2**	**3**	**3**	**4**	**4**	**5**
Summary (10-4 people)	60%	60%	70%	70%	30%	40%	40%	50%	45%	50%	55%	60%
D) Valence	**Modalin**				**Pundalin**				**Modalin + Pundalin**			
Version of the system	Baseline system	With CVS	Blog mining	CVS + Blogs	Baseline system	With CVS	Blog mining	CVS + Blogs	Baseline system	With CVS	Blog mining	CVS + Blogs
0 people	2	2	1	1	2	1	1	0	4	3	2	1
1-3 people	2	2	3	3	4	4	4	4	6	6	7	7
4-6 people	**1**	**1**	**1**	**1**	**3**	**3**	**3**	**3**	**4**	**4**	**4**	**4**
7-10 people	**5**	**5**	**5**	**5**	**1**	**2**	**2**	**3**	**6**	**7**	**7**	**8**
Summary (10-4 people)	60%	60%	60%	60%	40%	50%	50%	60%	50%	55%	55%	60%

Table 11. The number of people that agreed with all four system versions when analysing one agent at a time for evaluated items C) specified emotion types and D) valence; summarized results for Modalin (upper), Pundalin (middle) and for both systems (lower).

Since one of the agents was using humorous responses we also checked whether the jokes influenced the human-computer interaction. Most of the emotions expressed in the conversations with Pundalin were positive (67%) whereas for Modalin most of the emotions were negative (75%), which confirms that users tend to be positively influenced by the use of jokes in conversational agents (Dybala et al., 2009a). With all improvements applied there were as much as 8 cases of a perfect agreement with all 10 human evaluators for one questionnaire set. In conversations with Modalin ML-Ask reached perfect agreement two times for both - valence and emotion-specific determination. As for Pundalin, the perfect

agreement cases appeared twice in valence determination and, in both kinds of determining about appropriateness of emotions. Although the method presented here is still not ideal, the numerous improvements show, that it is easily improvable (Ptaszynski et al., 2009c, 2009d, 2009e). Considering the further enhancements that are already in plans, we are expecting a very high improvement in the near future.

	Modalin				**Pundalin**				**Overall results**				**Summary (10-4ppl.)**	
No. of people	10-7 ppl.	6-4 ppl.	3-1 ppl.	0 ppl.	10-7 ppl.	6-4 ppl.	3-1 ppl.	0 ppl.	10-7 ppl.	6-4 ppl.	3-1 ppl.	0 ppl.	Baseline system	Blog mining + CVS
A	**7**	2	0	1	**7**	2	1	0	**14**	4	1	1	75%	90%
B	4	**5**	0	1	**6**	2	2	0	**10**	7	2	1	75%	85%
C	2	**5**	2	1	3	2	**5**	0	5	**7**	**7**	1	45%	60%
D	**5**	1	3	1	3	3	**4**	0	**8**	4	7	1	50%	60%

Table 12. The number of people that agreed with the fully improved system when analysing one agent at a time for evaluated items A)-D); summarized results for both agents separately (left; centre-left) and together (centre-right); the results compared for the baseline and blog-mining with CVS versions of the system (right).

Versions of the system	**Baseline vs CVS**	**Baseline vs Blogs**	**CVS vs CVS+Blogs**	**Baseline vs CVS+Blogs**
Statistical significance (p value)	0.1599	0.0274	0.0274	**0.0119**

Table 13. Statistical significance of the results for different versions of the system

Part of conversation in Japanese (English translation)	**ML-Ask output**	**Web mining**	**CAEV**
USER: *Konpyuuta wa omoshiroi desu ne.* (Computers are so interesting!)	**positive** [joy]	**positive** [joy]	**appro-priate**
SYSTEM: *Sore wa oishii desu ka.* (Is it tasty?) [about instant noodles]	x	x	x
USER: *Oishii kedo, ore wa akita kana.* (Its tasty, but I've grown tired of it.)	**negative** [dislike]	**negative** [dislike]	**appro-priate**

Part of conversation in Japanese (English translation)	**ML-Ask baseline**	**with CVS**	**Web Mining**	**CAEV**
SYSTEM: *Sore wa omoshiroi tte* (Its so interesting!) [about conversation]	x	x	x	X
USER: *Sore hodo omoshiroku mo nakatta yo.* (It wasn't that interesting.)	positive [joy]	**negative** [dislike]	**negative** [fear], [sad]	**appro-priate**

Table 14. 3 examples of the results provided by the emotion appropriateness verification procedure (CAEV) with a separate display of CVS improvement of the procedure.

5. Discussions and work in progress

5.1 The socialized conversational agent algorithm

With already started attempts of combining the systems described above, we decided to combine them all and make the next step on the path to a naturally chatting agent. As the first step we decided to implement the following algorithm (see Figure 7): first, the emotive recogniser tries to categorize the emotion included inside the user's utterance. Here,

experimentally we checked, after which emotions it was appropriate to say a joke. We tried different possibilities and the decision for this step is still unsettled. However, for now the tendency is that the use of jokes is more effective when the user's attitude is negative or neutral. When the system decides that it is possible to say a joke, the Punda Generator tries to find a pun fitting the keywords. If it is not possible, the Modalin creates a preposition and statistically chooses a modality using the Internet Relay Chat logs while the original program adds modality almost randomly. The chat-logs are automatically tagged by affect analyser for emotion types and for sentence endings usually describing grammatical functions and dialog acts[6]. This version of the algorithm, although still not evaluated, was presented by Rzepka et al. (2009a). The system is still in its test phase, during which different versions of the algorithm are proposed and evaluated. However, the first preliminary experiments, which results will be published in the near future, made us rethink the procedures deciding about the joke generation according to users' emotional states and its appropriateness to the situation.

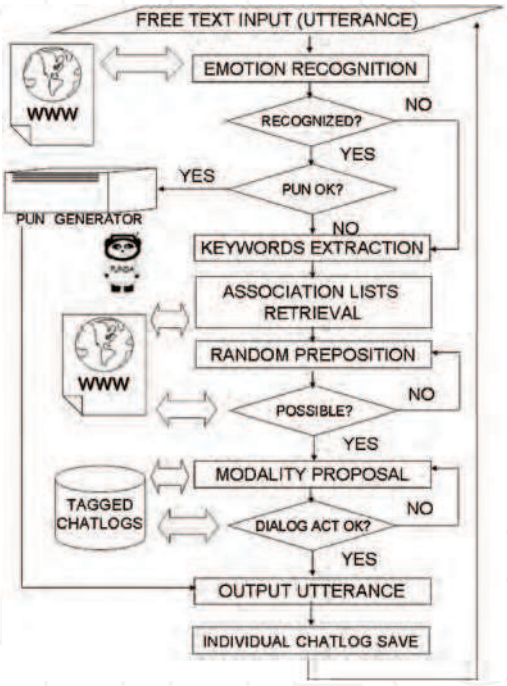

Fig. 7. Three systems combined into one multi-agent system.

5.2 Evolutionary algorithm for modelling individualized sense of humour

ML-Ask, the affect analysis system described in section 4, determines the user's emotional states. However, the applicability of this system exceeds the emotion recognition task. As it was proposed by Dybala et al. (2009b), the results of analysis can also be used in a sense of

[6] For determining the ending types we used a grammar book for Japanese language learners.

humour evolution algorithm. During the dialogue, the system can analyse each user's utterance, and check how he or she reacts to particular types of jokes. On this basis, the system can build a model of the user's sense of humour, determining the types of jokes the user likes or dislikes. For example, if the user reacts with positive emotions to jokes concerning politics, the system can assume that this type of joke matches his/her sense of humour. In this manner, the longer the system talks to the user, the more accurate "tags" of humour sense it can attach – and this, in effect, shall lead to more personalized, more individualized jokes that with a high probability would be appreciated by the user.

5.3 Emotional appropriateness as "conscience calculus" - implications towards computational conscience.

"*Public opinion is a second conscience.*" (William R. Alger)

As mentioned above, expressing and understanding emotions is one of the most important cognitive human behaviours present in everyday communication. In particular, Salovey & Mayer (1990) showed that emotions are a vital part of human intelligence, and Schwarz (2000) showed, that emotional states influence the decision making process in humans. When we define the process of decision making as distinguishing between good and bad, or appropriate and inappropriate, the emotions appear as an influential part of human conscience. The thesis that emotions strongly influence the development of human conscience was proved by Thompson and colleagues (2003) who showed, that children acquire the conscience by learning the emotional patterns from other people. The significance of the society was pointed out also by Rzepka et al. (2006), who define the Internet, being a collection of other people's ideas and experiences, as an approximation of general human common sense. Since conscience can be also defined as a part of common sense, this statement can be expanded further to that the Web can also be used to determine human conscience. The need for research in this matter, was pointed out inter alia by Rzepka et al. (2008), who raised the matter not of creating an artificial human being, as it is popularly ventured in Artificial Intelligence research, but rather an intelligent agent in the form of a toy or a companion, designed to support humans in everyday life. To perform that, the agent needs to be equipped, not only in procedures for recognizing phenomena concerning the user, in which emotions play a great role, but it also needs to be equipped with evaluative procedures distinguishing about whether the phenomena are appropriate or not for a situation the user is in. This is an up to date matter in fields such as Roboethics (Veruggio & Operto, 2006), Human Aspects in Ambient Intelligence (Treur, 2007), and in Artificial Intelligence in general. In our research we perform that by verifying emotions expressed by the user with a Web mining technique for gathering an emotional common sense, which could be also defined as an approximated vector of conscience. We understand, that the idea of conscience is far more sophisticated, but when defined narrowly as the ability to distinguish between what is appropriate and what is not, our method for verifying contextual appropriateness of emotions could be applied to obtain simplified conscience calculus for machines. We plan to develop further this idea and introduce it as a complementary algorithm for the novel research on discovering morality level in text utterances presented by Rzepka et al., (2009b).

6. Conclusions

In this chapter we presented a series of experiments aiming to implement into machines human factors and therefore making them more human-like and therefore user-friendly. In the first experiment on a conversational agent, Modalin, we showed, that adding modality to the propositions gathered from the Web is a good mean of improving the overall impression of the machine interlocutor. In the second experiment, we compared Modalin to its modified version Pundalin, which uses humorous output. In this experiment we showed that especially for non-task oriented conversational agents, which goal is mainly to entertain the user, the use of humour in the conversation greatly improves not only the impression about the agent's general performance, but also the user's attitudes towards the agent, making it easier to familiarize. In the last experiment we presented a novel method for estimating the contextual appropriateness of emotions. Two systems are used in this method, ML-Ask, a language based affect analysis system, and Web mining technique for extracting from the Internet a generalized emotive common sense. The first one used as the emotion detector and the second one as the emotion verificator provide a conversational agent computable means to determine whether the emotions expressed by the user are appropriate for the context they are expressed in. Enhancing a conversational agent with this method is the next step in implementing the full scope of Emotional Intelligence Framework in machines. We also presented some future implications of using this method, concluding that it could also be used to estimate human conscience. As the results of all experiments were satisfying enough, we developed a design of the full implementation of the two human factors, emotion estimation and sense of humour, into a conversational agent to be evaluated in the near future.

7. Acknowledgements

This research is partially supported by a Research Grant from the Nissan Science Foundation and The Global COE Program founded by Japan's Ministry of Education, Culture, Sports, Science and Technology.

8. References

Abbasi, A. & Chen, H. (2007). Affect Intensity Analysis of Dark Web Forums, *Intelligence and Security Informatics 2007*, pp. 282-288

Alm, C. O.; Roth, D. & Sproat, R. (2005). Emotions from text: machine learning for text based emotion prediction, In *Proceedings of HLT/EMNLP*, pp. 579-586

Andre, E.; Rehm M.; Minker W. & Buhler D. (2004). Endowing Spoken Language Dialogue Systems with Emotional Intelligence, *LNCS*, Vol. 3068, pp. 178-187

Araki K. & Kuroda, M. (2006). Generality of a Spoken Dialogue System Using SeGAIL for Different Languages, In *Proceedings of the IASTED International Conference Computer Intelligence*, pp. 70-75

Baba, J. (2003). Pragmatic function of Japanese mimetics in the spoken discourse of varying emotive intensity levels. *Journal of Pragmatics*, Vol.35, No. 12, pp. 1861-1889, Elsevier

Beijer, F. (2002). The syntax and pragmatics of exclamations and other expressive/emotional utterances, *Working Papers in Linguistics* 2, The Dept. of English in Lund

Bickmore, T. & Cassell, J. (2001). Relational Agents: A Model and Implementation of Building User Trust, In *Proceedings of Human Factors in Computing Systems (SIGCHI'01)*, pp. 396–403

Binsted, K. (1996). *Machine humour: An implemented model of puns*, University of Edinburgh

Boden, M. A. (1998). Creativity and artificial intelligence, *Artificial Intelligence*, Vol. 103, No. 1-2, pp. 347-356

Cann, A.; Holt, K. & Calhoun, L. G. (1999). The roles of humor and sense of humor in responses to stressors. *Humor*, Vol. 12, No. 2, pp. 177–193

Cook, K.S. & Rice, E. (2003). Social exchange theory, In Delamater, J. (ed.) *Handbook of social psychology*, Plenum, New York, pp. 53–76

Crystal, D. (1989). *The Cambridge Encyclopedia of Language*, Cambridge University Press

Dautenhahn, K.; Bond, A. H.; Cañamero, L. & Edmonds, B. (2002). *Socially intelligent agents: creating relationships with computers and robots*, Springer

Dybala, P.; Ptaszynski, M.; Higuchi, S.; Rzepka, R. & Araki, K. (2008a). Humor Prevails! - Implementing a Joke Generator into a Conversational System, *LNAI*, Vol. 5360, pp. 214-225, Berlin-Heidelberg, 2008.

Dybala, P.; Ptaszynski, M.; Rzepka, R. & Araki, K. (2008b). Extracting Dajare Candidates from the Web - Japanese Puns Generating System as a Part of Humor Processing Research, In *Proceedings of LIBM'08*, pp. 46-51

Dybala, P.; Ptaszynski, M.; Rzepka, R. & Araki, K. (2009a). Humoroids - Conversational Agents That Induce Positive Emotions with Humor, In *Proceedings of 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, May, 2009, Budapest, Hungary, pp. 1171-1172

Dybala, P.; Ptaszynski, M.; Rzepka, R. & Araki, K. (2009b). Humorized Computational Intelligence: Towards User-Adapted Systems with a Sense of Humor, *LNCS*, Vol. 5484, pp. 452–461, Springer-Verlag Berlin Heidelberg, 2009.

Dybala, P.; Ptaszynski, M.; Rzepka, R. & Araki, K. (2009c). Subjective, But Not Worthless - Non-linguistic Features of Chatterbot Evaluations, The IJCAI-09 Workshop on Knowledge and Reasoning in Practical Dialogue Systems, In *Working Notes of Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*, California, Pasadena

Gustafson, J. & Bell, L. (2000). Speech technology on trial: Experiences from the August system, In *Natural Language Engineering*, Vol. 1. No. 1, pp. 1-15

Hager, J. C.; Ekman, P.; Friesen, W. V. (2002). *Facial action coding system*, Salt Lake City, UT: A Human Face

Higuchi, S.; Rzepka, R. & Araki, K. (2008). A Casual Conversation System Using Modality and Word Associations Retrieved from the Web, *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP'08)*, Honolulu, USA, October 2008, pp. 382-390

Jakobson, R. (1960). Closing Statement: Linguistics and Poetics, *Style in Language*, pp. 350-377, The MIT Press

Kamei, T.; Kouno, R. & Chino, E. [Eds.] (1996). *The Sanseido Encyclopedia of Linguistics*, VI, Sanseido

Kang, B.S.; Han, C.H.; Lee, S.T.; Youn, D.H. & Lee, C. (2000). Speaker dependent emotion recognition using speech signals. In *Proceedings of ICSLP*, pp. 383-386

Kennedy, A. & Inkpen, D. (2005). Sentiment classification of movie and product reviews using contextual valence shifters, *Workshop on the Analysis of Informal and Formal Information Exchange during Negotiations (FINEXIN-2005)*

Kopp, S.; Gesellensetter, L.; Kramer, N. C. & Wachsmuth, I. (2005). A Conversational Agent as Museum Guide–Design and Evaluation of a Real-World Application, *Intelligent Virtual Agents, LNAI*, 3661, pp. 329-343

Kudo, T. (2001). MeCab: Yet another part-of-speech and morphological analyzer, http://mecab.sourceforge.ne

Kumar, R.; Novak,J.; Raghavan, P.& Tomkins,A.(2003).On the Bursty Evolution of Blogspace, In *Proceedings of The Twelfth International World Wide Web Conference*, pp. 568-257

Lerner, J.S. & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgment and choice, *Cognition & Emotion*

Liu, B; Du, L. & Yu, S. (2003). The method of building expectation model in task-oriented dialogue systems and its realization algorithms, In *Proceedings of Natural Language Processing and Knowledge Engineering*, pp. 174-179

Loehr, D. (1996). An integration of a pun generator with a natural language robot. In *Proceedings of International Workshop on Computational Humor*, pp. 161-172

Mayer, J. D. & Salovey, P. (1997). What is emotional intelligence?, In Salovey, P. & Sluyter, D. [Eds.] (1997). *Emotional Development and Emotional Intelligence*, Basic Books, pp. 3-31

Miyoshi, T. & Nakagami, Y. (2007). Sentiment classification of customer reviews on electric products, *IEEE International Conference on Systems, Man and Cybernetics*

Moran, C.C. (1996). Short-term mood change, perceived funniness, and the effect of humor stimuli, *Behavioural Medicine*, Vol. 22, No. 1, pp. 32–38

Morkes, J.; Kernal, H. K. & Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction*, Vol. 14, No. 4, pp. 395-435

Mulkay, M. (1988). *On humor: Its nature and its place in modern society*, Basil Blackwell, NY

Nakamura, A. (1993). *Kanjo hyogen jiten* (Dictionary of Emotive Expressions) (in Japanese), Tokyodo Publishing, Tokyo

Nijholt, A. (2007). Conversational Agents and the Construction of Humorous Acts, In *Conversational Informatics: An Engineering Approach*, John Wiley & Sons, Chicester, England, pp. 21-47

Nitta, Y. & Masuoka, T. [Eds.] (1989). *Nihongo no modarithi* (Japanese Modality), Kuroshio, Tokyo, pp. 131–157

Ono, H. (2002). *An emphatic particle DA and exclamatory sentences in Japanese*, University of California, Irvine

Oshima-Takane, Y. & MacWhinney, B. [Eds.], Shirai, H.; Miyata, S. & Naka, N. [Rev.] (1995-1998). *CHILDES Manual for Japanese*, McGill University, The JCHAT Project

Picard, R. W. (1997). *Affective Computing*, The MIT Press, Cambridge

Picard, R. W.; Vyzas, E.; Healey, J. (2001). Toward Machine Emotional Intelligence: Analysis of Affective Physiological State, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 10, pp. 1175-1191

Pietro, O.; Rose, M. & Frontera, G. (2005). Automatic Update of AIML Knowledge Base in E-Learning Environment, In *Proceedings of Computers and Advanced Technology in Education*, Oranjestad, Aruba, August, pp. 29-31

Polanyi, L. & Zaenen, A. (2004). Contextual Valence Shifters, *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*

Potts, C. & Kawahara, S. (2004). Japanese honorifics as emotive definite descriptions, In *Proceedings of Semantics and Linguistic Theory 14*, pp. 235-254

Ptaszynski, M. (2006). *Boisterous language. Analysis of structures and semiotic functions of emotive expressions in conversation on Japanese Internet bulletin board forum '2channel'* (in Japanese), M.A. Dissertation, UAM, Poznan, Poland

Ptaszynski, M.; Dybala, P.; Rzepka, R. & Araki, K. (2008). Effective Analysis of Emotive-ness in Utterances Based on Features of Lexical and Non-Lexical Layers of Speech, In *Proceedings of NLP-2008*, pp. 171-174

Ptaszynski, M.; Dybala, P.; Rzepka, R. & Araki, K. (2009a). Contextual Valence Shifters Supporting Affect Analysis of Utterances in Japanese, In *Proceedings of NLP-2009*, pp. 825-828

Ptaszynski, M.; Dybala, P.; Shi, W.; Rzepka, R. & Araki, K. (2009b). A System for Affect Analysis of Utterances in Japanese Supported with Web Mining, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Special Issue on Kansei Retrieval*, Vol. 21, No. 2 (April), pp. 30-49 (194-213)

Ptaszynski, M.; Dybala, P.; Shi, W.; Rzepka, R. & Araki, K. (2009c). Towards Context Aware Emotional Intelligence in Machines: Computing Contextual Appropriateness of Affective States, In *Proceedings of IJCAI 2009*, California, Pasadena

Ptaszynski, M.; Dybala, P.; Shi, W.; Rzepka, R. & Araki, K. (2009d). Shifting Valence Helps Verify Contextual Appropriateness of Emotions, The IJCAI-09 Workshop on Automated Reasoning about Context and Ontology Evolution (ARCOE-09), In *Working Notes of Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*, California, Pasadena

Ptaszynski, M.; Dybala, P.; Shi, W.; Rzepka, R. & Araki, K. (2009e). Conscience of Blogs: Verifying Contextual Appropriateness of Emotions Basing on Blog Contents, In *Proceedings of The Fourth IASTED International Conference on Computational Intelligence (CI-2009)*, August, 2009, Honolulu, Hawaii, USA

Ptaszynski, M.; Dybala, P.; Rzepka, R. & Araki, K. (2009f). Affecting Corpora: Experiments with Automatic Affect Annotation System - A Case Study of the 2channel Forum -, In *Proceedings of PACLING 2009*, Sapporo, Japan

Reeves, B. & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*, Cambridge University Press

Reitter, D.; Moore, J. D. & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation, In *Proceedings of 28th Annual Conference of the Cognitive Science Society (CogSci)*, Vancouver, Canada

Russell, J. A. (1980). A circumplex model of affect, *Journal of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161-1178

Rzepka, R.; Ge, Y. and Araki, K. (2006). Common Sense from the Web? Naturalness of Everyday Knowledge Retrieved from WWW. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 10 No. 6, pp. 868-875

Rzepka, R. & Araki, K. (2007). Consciousness of Crowds - The Internet As a Knowledge Source of Human's Conscious Behavior and Machine Self-Understanding, *AI and Consciousness: Theoretical Foundations and Current Approaches*, Papers from AAAI Fall Symposium, Technical Report, November, 2007, Arlington, USA, pp. 127-128

Rzepka, R.; Higuchi, S.; Ptaszynski, M. & Araki, K. (2008). Straight Thinking Straight from the Net - On the Web-based Intelligent Talking Toy Development, In *The Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC'08)*, Singapore, October 2008, pp. 2172-2176

Rzepka, R.; Dybala, P.; Shi, W.; Higuchi, S.; Ptaszynski, M. & Araki, K. (2009a). Serious Processing for Frivolous Purpose - A Chatbot Using Web-mining Supported Affect Analysis and Pun Generation, In *Proceedings of IUI'09 - International Conference on Intelligent User Interfaces*, Sanibel Island, FL, USA. (ACM Press), pp 487-488

Rzepka, R.; Masui, F. and Araki, R. (2009b). The First Challenge to Discover Morality Level In Text Utterances by Using Web Resources, In *Proceedings of The 23rd Annual Conference of the Japanese Society for Artificial Intelligence*, Takamatsu, June 2009

Salovey, P. & Mayer, J. D. (1990). Emotional intelligence, *Imagination, Cognition, and Personality*, Vol. 9, pp. 185-211

Sasai, K. (2006). The Structure of Modern Japanese Exclamatory Sentences: On the Structure of the Nanto-Type Sentence, *Studies in the Japanese language*, Vol. 2, No. 1, pp. 16-31

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions, *Journal of Experimental Psychology*, Vol. 44, pp. 229-237

Schwarz, N. (2000). Emotion, cognition, and decision making, *Cognition & Emotion*, Vol. 14, No. 4, pp. 433-440

Shelley, M. W. (1818). *Frankenstein, or, The Modern Prometheus*, Oxford University Press, republished in 1998

Shi, W.; Rzepka, R. & Araki, K. (2008). Emotive Information Discovery from User Textual Input Using Causal Associations from the Internet, *FIT2008*, pp. 267-268

Sjöbergh, J. (2006). Vulgarities are fucking funny, or at least make things a little bit funnier, *Technical Report of KTH*, Stockholm

Sjöbergh, J. & Araki, K. (2008a). A Multi-Lingual Dictionary of Dirty Words, *LREC*

Sjöbergh, J. & Araki, K. (2008b). Robots Make Things Funnier, In *Proceedings of LIBM'08*, pp. 46-51

Solomon, R. C. (1993). *The Passions: Emotions and the Meaning of Life*, Hackett Publishing

Sprecher, S. & Regan, P.C. (2002). Liking some things (in some people) more than others: Partner preferences in romantic relationships and friendships. *Journal of Social and Personal Relationships*, Vol. 19, No. 4, pp. 463-481

Tinholt, H. W. & Nijholt, A. (2007). Computational Humour: Utilizing Cross-Reference Ambiguity for Conversational Jokes, *LNAI*, Springer-Verlag, Berlin, pp. 477-483

Teixeira, J.; Vinhas, V.; Oliveira, E. & Reis, L. (2008). A New Approach to Emotion Assessment Based on Biometric Data, HAI'08 Workshop, In *Proceedings of the WI-IAT'08*, pp. 459-500

Thompson, R. A.; Laible, D. J; Ontai, L. L. (2003). Early understandings of emotion, morality, and self: Developing a working model, *Advances in child development and behavior*, Vol. 31, pp. 137-171

Tokuhisa, R.; Inui, K. & Matsumoto, Y. (2008). Emotion Classification Using Massive Examples Extracted from the Web, In *Proceedings of Coling 2008*, pp. 881-888

Treur, J. (2007). On Human Aspects in Ambient Intelligence, In *Proceedings of The First International Workshop on Human Aspects in Ambient Intelligence*, pp. 5-10

Tsuchiya, N. (1999). Taiwa ni okeru kandoshi, iiyodomi no togoteki seishitsu ni tsuite no kosatsu (Statistical observations of interjections and faltering in discourse) (in Japanese), *SIG-SLUD-9903-11*

Veruggio, G.; Operto, F. (2006). Roboethics: a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics, *International Review of Information Ethics, Ethics in Robotics*, Vol. 6 (12)

Vilathong, A. P.; Arnau, R. C.; Rosen, D. H. & Mascaro, N. (2003). Humor and hope: Can humor increase hope?, *Humor*, Vol. 16, No. 1, pp. 79–89

Weizenbaum, J. (1966). ELIZA-computer program for the study of natural language communication between man and machine, *Communications of the ACM*, Vol. 9, No. 1, pp. 36-45

Yamashita, Y. (1999). Kara, node, te-conjunctions which express cause or reason in Japanese (in Japanese). *Journal of the International Student Center, Hokkaido University*, Vol. 3

Yip, J. A.; Martin, R.A. (2006). Sense of humor, emotional intelligence, and social competence. *Journal of Research in Personality*, Vol. 40, Issue 6, December 2006, pp. 1202-1208

Yoshihira, K.; Takeda, T. & Sekine, S. (2004). KWIC system for Web Documents (in Japanese) In *Proceedings of NLP-2004*, pp. 137-139