Emotional Information Retrieval for a Dialogue Agent

Rafal Rzepka and Kenji Araki
Graduate School of Engineering, Hokkaido University
Kita-ku, Kita 13-jo Nishi 8-chome, 060-8628 Sapporo, Japan
{kabura, araki}@media.eng.hokudai.ac.jp http://sig.media.eng.hokudai.ac.jp/index_e.html

Koji Tochinai
Graduate School of Business Administration, Hokkai-Gakuen University
Toyohira-ku, Asahi-machi 4-1-40, 062-8605 Sapporo, Japan
tochinai@econ.hokkai-s-u.ac.jp

Keywords: natural language processing, spoken dialog agents, affective computing

Received: May 12, 2001

In our project (GENTA - GENeral belief reTrieving Agent), we are trying to realize a conversational agent, which would be able to talk in any domain by using web-mining techniques to retrieve information that is impossible to obtain in usually used corpora. In our research we try to simulate reasoning processes based on Internet textual resources including chat logs. Our goal is a dialogue system which learns the linguistic behaviour of an interlocutor concentrating on the role of emotion during analysing discourse. The system is not using any databases of commonsensical word descriptions, they are being automatically retrieved from the WWW. We describe two values called Positiveness and Usualness and explain their role in the Inductive Learning that is used for achieving emotion-based reasoning skills. As this is a new approach to knowledge acquisition for dialogue agents we concentrate on the theoretical part of our project. Finally we introduce the results of the preliminary experiments.

1 Introduction

WWW is an enormous database, which is being used widely these days. Unfortunately it is said to be difficult to handle as it is full of informational garbage that makes web mining or knowledge-base creation a hard task. However when we started to rummage through those personal homepages with very similar contents that are seemingly useless for AI purposes, we imagined that human brain cells might look exactly the same. Not only are the stored pieces of semantic information important but also the number of how many times such similar data was stored. We assumed that the Internet is interesting material for retrieving the common sense, beliefs, opinions and emotional information for various types of agents. Without any sophisticated method our system is able to easily discover that in most cases "being cold" is not pleasant and cold beer almost always "sounds nice" or that one movie star is being loved and another hated. In this paper we introduce ideas for our project (GENTA - GENeral belief reTrieving Agent(1)) and the results of initial experiments with implementing a primitive method of retrieving basic feelings towards human user's utterances and applying this emotional information whilst inductively learning the speech acts. The earliest ideas for our project began whilst observing foreign students' linguistic behaviour while speaking the Japanese language, which was not their mother tongue. Although their language abilities and cultural background happened to be very different, their conversations always seemed interesting, which agrees with intercultural will of communication theories(2). We noticed several dependencies, for example that the keywords triggering the conversations were mostly of two types - topics that clearly have a positive or negative emotional load as "a present" or "the red tape" and that the conversational flow concentrates on what the interlocutors have in common or conversely - how they differ. Even if it is quite obvious to human beings, these conditions for what we call interesting conversation are very difficult to be met by machines. First of all, computers have problems recognizing what would be interesting and what could bore its conversational partner. Secondly, they have difficulties with spoken language, especially when there are plenty of grammatical mistakes not to mention interlocutors who use broken language. Therefore we looked for a conversational environment where the computer system could face-up to these challenges and chose IRC (Internet Relay Chat) because of its international character and "world simplicity"(3). Although it is a multi-user environment we concentrated on one-on-one conversations, as the program does not handle multi-thread yet. This time we chose the English language as the initial implementation was done in Japanese(1).

2 Basic assumptions

From the very beginning of human-computer interaction, the purpose of communication was almost always clear - the machine had to understand an order from the human user and help him in some way. That approach is obviously being forced by the pressure that industry puts on the world of science. Even if the talk itself was a purpose of a program(6), it was always supposed to help the user somehow and to be socially useful. Pure "chat for chat's sake" agents are not widely developed and scientifically neglected because of their low usability and problems with evaluating such systems. In our opinion, concentrating on problem-solving or helping agents makes HMC (Human-Machine Conversations) research unbalanced and we argue against the importance of usability while developing programs that can communicate with human beings. We have noticed that nowadays approaches become more and more sophisticated, machines learn how to analyse metaphors and to stochastically use very large corpora but there are still too few agents based on affective computing(7)(8)(9), that can react not only in a "store-clerk-human-way" but also a "human-way" if a user says "I'm sad" or just "I want to buy a dog". Since Damasio(10) has underlined the meaning of emotions in reasoning, the number of AI or cognitive science researchers who search for the methods of implementing emotions into machines grows rapidly. Virtual and physical architectures are built and agents or robots are taught to analyse emotional state as the additional information for perceiving. In our approach we have assumed that addition is not enough and that this is crucial, since we understand human conversation as a cycle of exchanging emotion-based information. By that we do not mean that pure information exchange does not exist, we want to make a machine remember that its reaction depends also on the emotional load of an input. Hence our plan is to implement Pavlovian-like reactions into the computer's conversation abilities. For the simplicity of our system we also simplified an idea on the utterance structure. We imagined a human language interaction as an electron jumping through the layers shown in Fig. 1. It goes up when making an utterance and down when receiving it. We assume that the emotional layer includes intentions that motivate to start, to continue, to change or to stop a conversation flow. We imagine that if anyone puts an electron to any area of semantic layer, for example by saying "a kingdom for a soda!" and an interlocutor who does not know what this cliché exactly means hears it, the electron jumps down to the emotional layer instead of wandering along the complicated semantic network. The interlocutor *feels* that "kingdom" and "soda" are not abusive, "soda" even sounds nice. The nicer it sounds to him, the greater the probability of choosing it for a conversation topic, or as we call it triggering keyword. An aggressive reaction becomes less possible in such case. Guided by the aforementioned assumption, we decided to realize a method that would allow us to simplify semantic layer processing by limiting choices in lexical and emotional layers. We also made an assumption based on Fletcher division of blind people's mental imagery(11) and we believe that machines could gain not deficient or inefficient but different imagination. Our idea is that the electron having fewer choices in the emotional and lexical layers may naturally decrease the possibilities of exploring the semantic layer as we illustrate in Fig. 2. Since we are also interested in the aspects of modelling an artificial imagination, we decided to base our system on pure written word-level conversation without audio-visual stimuli, although we used their simplified substitute, as they are necessary for obtaining the basic emotional information. This substitute is the usage of emoticons (facemarks) widely employed in keyboard-based chat, which suggest that a given utterance was, for instance, ironic or supposed to be a joke. For example, if the utterance from Fig. 2. is done with a smiley ":)" the electron should not hit the emotional layer at the fear spot because this would not be natural behaviour.

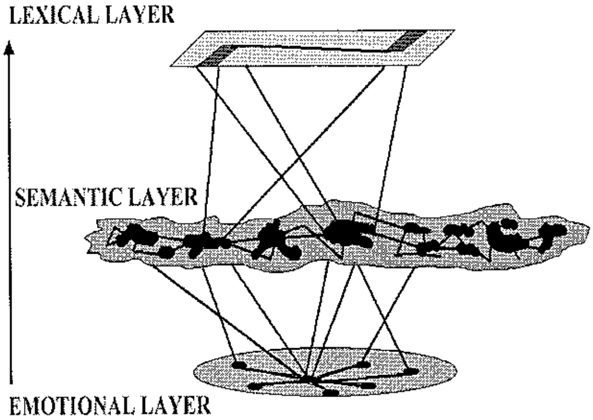

Figure 1: The idea of the utterance structure before simplifying.

3 Other kinds of retrieval

In order to see what kind of information could be retrieved from Internet resources and used for general belief and emotion processing, we made several tests with different search engines. For retrieving common emotions we were observing the hit numbers of the results of searched frames as: "I am afraid of" [N] (for nouns usually causing fear) "I always wanted a" [N] and "I always wanted to" [V] (for what people may dream about) or "Usually people [V] at [N]" (for retrieving verbs for building a script of given noun). We discovered that it is possible to retrieve thousands of sentences that could be processed for what we call emotional common-sense statistics. The more specific given frame was, for instance "I always wanted to be" instead of "I always wanted to", the less ambiguous sentences were retrieved and results as "I always wanted

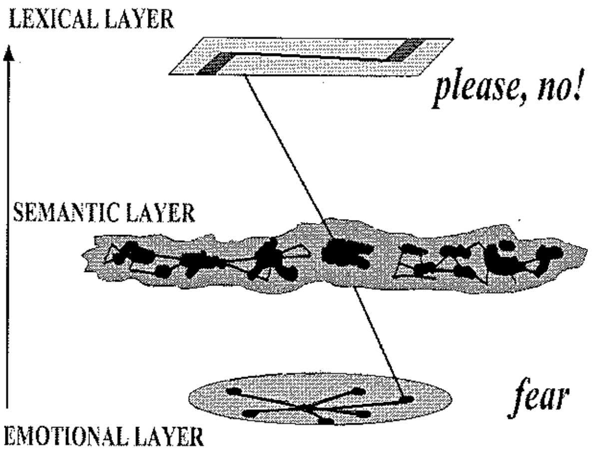

Figure 2: The idea of the utterance structure after simplifying.

to do it" were apparently decreasing. As we noticed in previous works(1)(4)(5), the Internet can also provide basic semantic tagging by counting frames including prefixes (or particles as in Japanese). For example if a parser receives an unknown word, let's say "Sapporo", it can easily declare that it is a place by comparing result numbers of given searching frames: "I talked to [Sapporo] and", "I felt [Sapporo] then", "I ate [Sapporo] with", "I went to [Sapporo] by", etc. Such frames must be chosen very carefully and one should use stop words as "and", "then", "with" or "by" to avoid counting expressions as "I talked to Sapporo Council" but treating such expressions as "animates" also produced some interesting effects. Search engines also provide misspell detection what can be useful in fast typing chat environments with non-native speakers. Unfortunately automation of such searching requires the usage of crawlers (web robots) and complicated filtering which still takes too much time to keep up with interactive conversation processing. Therefore we decided to limit the retrieved information only to the simple test of whether something is liked or disliked and how common it is.

4 Description of GENTA system

4.1 General Belief

By this expression we mean a mixture of common sense based on retrieved opinions and simplified environmental knowledge. The system's knowledge of a user is assumed as almost none - GENTA does not know the nationality, age or sex of its conversation partner; he or she is not necessarily a native speaker of English. After exchanging greetings, the system waits for a user's initial utterance and if there is none it starts a conversation using the learning data of "Conversation Keeper", which will be described later. While detecting the speech act, which is also explained below, GENTA tries to guess the leading keyword(s) from the first user's utterance since the domain of conversation is still unknown. Next, GENTA searches the Internet for the whole utterance and its grammatically connected parts previously parsed by a parser(13) trying to establish what can be associated with given verbs, nouns, noun phrases, adjectives or conditional expressions concentrating on feelings-based opinions. For example when the input is, "Do you like playing soccer when it rains?" GENTA counts how many sentences "I like playing soccer when it rains", "I love playing soccer", "I hate playing soccer", and "I love it when it rains" and "I hate it when it rains", etc. appear on the Web. This lets our system achieve "own" opinion about playing soccer when it rains because this knowledge is assumed as "general" or "common". Then, paraphrasing Shannon's information theory(14), we assume that the keyword with less frequency is more interesting for interlocutors and GENTA chooses playing soccer as a leading topic (27268 vs 33985 hits). The system believes that the discourse should be continued in this "semantic direction". But before that, "Conversation Keeper" must establish which linguistic behaviour (called "a dialogue act" here) will be proper for a reply, which is our current task.

4.2 Conversation Keeper

Taking last decade research results into consideration (15)(16)(17)(18)(19)(20)(21)(22) we decided to find precisely the combination of web-mining and learning methods that would help us to create a dialogue system that does not require a large amount of initial data prepared by hand and does not need sophisticated modules. As our first step, before creating the real dialogue manager, NLG module, etc., we decided to confirm that our system is able to learn from above-word level information. As already mentioned, we assumed that peculiarity and emotional load of given expressions could support intention recognition, which is one of the most important tasks of human discourse management. Therefore here we divided General Beliefs into two above word-level values that we call Positiveness and Usualness, which are also measured by counting the aforementioned string frequencies upon the WWW. Positiveness value is calculated with following formula:

$$Positiveness = \frac{C_{\alpha_1} + C_{\alpha_2} * \gamma}{C_{\beta_1} + C_{\beta_2} * \gamma}$$

$$\alpha_1 = disliked, \alpha_2 = hated$$

$$\beta_1 = liked, \beta_2 = loved, \gamma = 1.3$$

Where γ is to strengthen the "love" and "hate" opinions. We prepared dialogue act tags, as handing or demanding of information, opinion and reason; advising, warning, greeting and nodding. GENTA can automatically declare Usualness and Positiveness for utterances, as in the following example:
Do you like playing soccer when it rains?
becomes a *DAPU string* (Dialogue + Act + Positiveness + Usualness):

OD P5U4 cond P3U5
which means that it was a Demand of Opinion consisting of two positive and usual expressions connected by subordinate clause conjunctor (SCC) "cond" (conditional clause). What is characteristic for our method, even if the Positiveness of expression seems to be doubtful (that most Web page creators like it when it rains that does not necessarily mean that most human beings also do) it does not disturb the process since the opinion remains logical. Values of Usualness and Positiveness are calculated by comparing frequencies of ("I don't like ..." / "I hate ...") and ("I like..." / "I love ...") searching frames. The frequency thresholds are different depending on the length of the searched string.

4.3 Inductive Learning

GENTA system has the ability to learn from spontaneous human conversations. We use the Inductive Learning method(16) to predict which utterance should be used and to make new rules while talking. The system represents dialogue discourse as connected *DAPU strings*

OH P5U2 : OD P5 U3 : ND : ...

(A1 B1 A2 B2 A3 B3 ...)

which are divided into double rules

(A1 B1) (B1 A2) (A2 B2) (B2 A3) (A3 B3) ...

stored in a Dictionary (Fig. 3,4). When new input is done, GENTA parses the utterance to *DAPU string* by recognizing a dialogue act determinator, which are words attributed to every dialogue act tag. For example, should determines advising tag. If subordinate clause conjunctor (SCC) is detected, both clauses are parsed into *DAPU strings* and they become an individual element for learning. When there is more than one sentence during one turn, GENTA confirms if they are of the same dialogue act. If not, the input is divided - the last rule is changed and a new one is created, for example:

A1: Do you care?

B1: Well, I don't care. What about ya?

A2: Me neither, man!

creates (A1:B1a) (B1b: A2) instead of (A1:B1) (B1: A2).

Learning concentrates on dialogue acts tagging and conjunctors, and their coexistence with Positiveness and Usualness. For example, if an unknown dialogue act determinator appears, our system decides the most probable tag and unless a user cancels a computer's output by using one of the cancelling expressions, such as "???" or "What are you talking about?" and so on, a new rule is created in the rule dictionary (Fig. 4.). If an emoticon is detected, the Positiveness value is decreased or increased depending on the kind of a facemark.

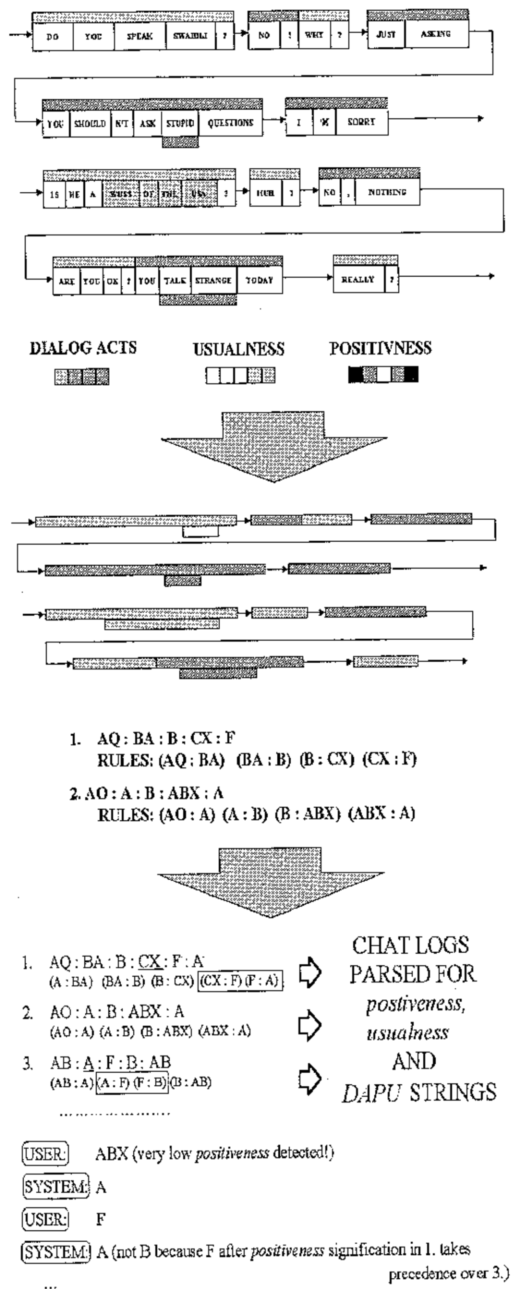

Figure 3: A simple example of the learning process from the utterances already parsed for Positiveness, Usualness and *DAPU strings*.

5 Experiment and evaluation

5.1 Method

Existing standards (23) in HMC evaluation, which concentrate on semantic quality of output, were not useful for evaluating spontaneous chat only with above-word information supported dialogue acts. Hence we had to prepare our own evaluation method:

- There are two human interlocutors A and B.

- They converse through IRC channel, which is monitored by our system (G).

- G listens to A's utterances and proposes its own answers (as *DAPU strings*).

- *DAPU strings* of B's utterances are compared with G's ones.

- Afterward the third person evaluates the naturalness of the strings when a system chose a different dialogue act, as there is more than one possibility.

5.2 Results

As we are not particular about perfect language, two non-native English speakers took part in our experiment. There was no particular topic of conversation. Subjects made 128 turns and they mostly talked about sports. GENTA's dictionaries were empty in the initial phase and we taught the system only one determinator for every dialogue act and only two to three basic conjunctors for every kind of subordinate clauses. We decided on empty dictionaries to see when a system starts to learn and because we want GENTA to retrieve what is needed from the Internet. The Web as a corpus that is constantly changing and creating dictionaries for Positiveness or Usualness would not reflect those changes. By comparing user B and GENTA's *DAPU strings* we understood that:

- The systems already started to use learned rules by the eighth turn, as the chat was mostly question-answer style but finally less than half (37.5%) of dialogue acts were chosen the same way by a user. Although 81.25% of those different ones were evaluated as natural by human being.

- Positiveness (On a five point scale: 1-negative, 2-slightly negative, 3-neutral, 4-slightly positive, 5-positive) of systems output that had the same dialogue act tag as a human was in 59.1% of cases the same as the user's. By this we mean there are three levels: positive, indifferent, negative.

- Usualness (On a five point scale: 5-very usual, 4-usual, 3-slightly peculiar, 2-peculiar, 1-very peculiar) of a system's output was only in 20.8% of cases the same as a human user's, because all parser errors due to misspells were detected as the most peculiar expressions.

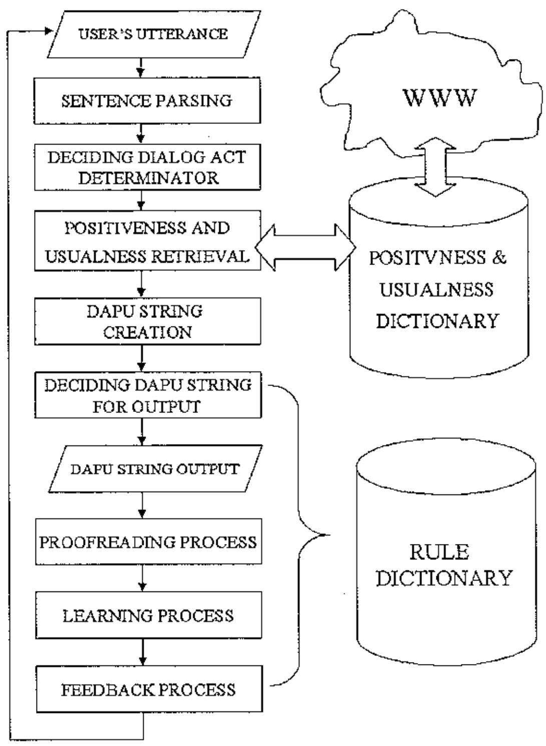

Figure 4: GENTA's Inductive Learning process.

Positiveness and Usualness were compared only in cases where dialogue acts were the same in human and machine outputs, as the dialogue acts choice significantly influences those two values. Because these two values were depending on the Internet connection speed (about 15-20 seconds for 1 calculation), the computer's propositions were given with growing time difference but it had no influence on our experiment's results.

6 Conclusion and discussion

We have described a new approach to WWW statistical information usage in a dialogue system, which is able to achieve information that is not obvious to the machine without using logic programming and other sophisticated methods. As we are in the preliminary stages of our project, we could only indirectly evaluate our ideas as grammatically built sentences were not output by the system. However, the results are convincing enough to continue walking on our chosen path - even if the system was not guessing an interlocutor's intentions properly, it proposed its own dialogue acts, which were not against the logical flow of conversation. In most cases too few turns made learning material inefficient but it is quite difficult to evaluate GENTA system before implementing further modules that will lead to generating more understandable output. The following is an example of this difficulty:

A: They prefer drinking beer. (IH P5U4)

G: P5U4 ex (that could be: "a beer!!!" or "I love beer!")

B: C'mon, they're watching games too. (IH P4U3)

It seems obvious that at this stage it is too early to evaluate our system as a talking system and it is rather impossible to see its abilities in context management. Although when we add the knowledge retrieval and representation modules we foresee that such outputs will be useful. We believe that tuning up the parsing methods and increasing data for learning will help to achieve better results in the future. We made a first step in the creation of an agent that should be able to chat about any topic with proper human-like reactions. A promising factor for future tasks is that our program was only based on automatically retrieved knowledge of common opinions and the peculiarity of user's utterance, which could be used in many interesting ways, as manipulating GENTA's "personality" for example by decreasing its Positiveness when, for instance, the weather is bad. Another idea is that one could use retrieved information as a model imagination of an interlocutor. We want GENTA to know what his partner probably thinks while for instance saying "I need a girlfriend". Thus there is a need to experiment with different parsers and to create mechanisms which allow GENTA to learn other things from the Internet - the largest and most rapidly growing database in the world, and try to apply those methods to commonly explored areas as for example in qualitative spatial reasoning. It must also be able to answer "wh-" questions, so we plan to concentrate on implementing a substitution of imagination which should be an elastic plan retrieval mechanism supported by commonsensical libraries created through search frames as "I always (verb) when it rains" or "usually people buy (noun) when they want to (verb + noun)"and also on the automatic creation of such frames. Our method could also be interesting from a sociological point of view, since GENTA can become a "mirror personality" of an average wired English or Japanese speaker in his original version(1), which could make it a much more interesting conversation partner than its predecessors. By using searching frames as "computers never", "computers can" we can also model a basic knowledge base for a machine that could also be conscious of its possibilities. We also understood that adopting GENTA to other languages would only be limited to translating the search frames. We strongly believe that with the constantly improving computer and network abilities, the Internet will become the main source of any kind of knowledge for future AI systems. We also predict that millions of private homepages sharing users' feelings and opinions will be crucial information to helping machines to "understand" what we typically think and what average humans know.

Acknowledgement

This work is partially supported by the Grants from the Government subsidy for aiding scientific researches (No.14658097) of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] R. Rzepka, K. Araki, K. Tochinai (2001) Basic Idea of General Belief Retrieving Agent "GENTA" *2001 IEEE Hokkaido Chapters Joint Convention Record*, IEEE Sapporo Section, pp. 414-415.

[2] B. Malinowski (1922) *Argonauts of the Western Pacific*, Routledge & Kegan Paul.

[3] R. Rzepka (1999) Communication methods in Japanese Internet chat - new semiotic changes of the natural language structure, *Master's Thesis, Department of Linguistics*, Poznan University.

[4] R. Rzepka, K. Araki, K. Tochinai (2002) Prediction of the User's Reply Using Emotional Information Retrieved from Internet Resources *2002 IEEE Hokkaido Chapters Joint Convention Record*, IEEE Sapporo Section, pp. 229-230.

[5] R. Rzepka, K. Araki, K. Tochinai (2002) Is It Out There? The Perspectives of Emotional Information Retrieval from the Internet Resources, *Proceedings of the IASTED Artificial Intelligence and Applications Conference*, ACTA Press, Malaga, pp. 22-27.

[6] J. Weizenbaum (1966) ELIZA - a computer program for the study of natural language communication between man and machine, *Communications of the Association for Computing Machinery 9*, ACM, pp. 36-44.

[7] J. Bates (1994) The role of emotion in believable agents, *Communications of the Association for Computing Machinery 37*, ACM, pp. 122-125.

[8] R.W. Picard, J. Klein (2002) Computers that Recognise and Respond to User Emotion: Theoretical and Practical Implications, *MIT Media Lab Tech Report 538*, to appear in Interacting with Computers.

[9] M. Seif El-Nasr, T.R. Ioerger, J. Yen (1999) A Web of Emotions, *Proceedings of Workshop on Emotion-Based Agent Architectures part of Autonomous Agents '99*.

[10] A.R. Damasio, (1994) *Descartes' error – emotion, reason, and the human brain*, Avon, New York.

[11] J.F. Fletcher (1980) Spatial representation in blind children, Development compared to sighted children, *Journal of Visual Impairment and Blindness 74*, pp. 381-385.

[12] J. Groenendijk, M. Stokhof, F. Veltman (1996) Coreference and modality, *In S. Lappin, editor, Handbook of Contemporary Semantic Theory*, Blackwell, Oxford, pp. 179-213.

[13] D.D. Sleator, D. Temperley (1994) Parsing English with a link grammar, *Third International Workshop on Parsing Technologies.*

[14] C.E. Shannon (1948) A mathematical theory of communication, *Bell System Technical Journal, vol. 27* pp. 379-423 and 623-656.

[15] J. Alexanderson (1996) Some ideas for the automatic acquisition of dialogue structure, *Vermobil Verbundvorhaben, Report 157.*

[16] K. Araki, K. Tochinai (2001) Effectiveness of Natural Language Processing Method Using Inductive Learning, *Proceedings of the IASTED International Conference Artificial Intelligence and Soft Computing*, ACTA Press, Cancun.

[17] D. Jurafsky, R. Bates, N. Coccaro, R. Martin, M. Meteer, K. Ries, E. Shriberg, A. Stolcke, P. Taylor, C. Van Ess-Dykema (1997) Automatic Detection of Discourse Structure for Speech Recognition and Understanding, *Proceedings of 1997 IEEE Workshop on Speech Recognition and Understanding*, IEEE.

[18] J.R. Glass (1999) Challenges for spoken dialogue systems, *Proceedings of the 1999 IEEE ASRU Workshop*, IEEE.

[19] J. Kreutel, C. Matheson (1999) Modelling questions and assertions in dialogue using obligations. *Proceedings of Amselogue 99, the 3rd Workshop on the Semantics and Pragmatics of Dialogue*, University of Amsterdam.

[20] A. Stockle (1998) Dialog act modeling for conversational speech, *In Papers from the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing.*

[21] L. Bell, J. Gustafson (1999) Utterance types in the August dialogues, *Proceedings of IDS'99, ESCA Workshop on Interactive Dialogue in Multi-Modal Systems.*

[22] J. Gustafson, M. Lundeberg, J. Liljencrants (1999) Experiences from the development of August - a multi modal spoken dialogue system, *Proceedings of IDS'99, ESCA workshop on Interactive Dialogue in Multi-Modal Systems.*

[23] W. Minker (1998) Evaluation Methodologies for Interactive Speech Systems, *Proceedings of LREC'98*, Granada.