Doctoral Dissertation

Affect Analysis of Textual Input Utterance in Japanese and its Application in Human-Computer Interaction

日本語のテキスト入力文の感情解析及び ヒューマンコンピュータインタラクションへの応用

by

Michal E. Ptaszynski

ミハウ・E・プタシンスキ

Research Group of Information Media Science and Technology

Division of Media and Network Technologies

Graduate School of Information Science and Technology

Hokkaido University

September 2010

To Hanna

motto:
"I understand a fury in your words
but not your words."
(- William Shakespeare, *Othello*, 4.2)

Table of Contents

Page

List of Tables

List of Figures

Abstract

This dissertation presents the development of my ideas on enhancing machines with Emotional Intelligence. I argue that equipping machines with computable means for processing user emotions is a practical need requiring implementation of a set of abilities included in an Emotional Intelligence framework. To achieve this I develop a set of affect analysis tools and propose methods for efficient utilization of the emotive information obtained by those tools.

Firstly, I develop a system for affect analysis of textual input utterance in Japanese, ML-Ask. The system is based on a linguistic assumption that emotional states of a speaker are conveyed by emotional expressions used in emotive utterances. ML-Ask firstly separates emotive utterances from non-emotive and in the emotive utterances seeks for expressions of specific emotion types. To verify the system performance I perform a series of experiments, based on a training set and several types of test sets: separate utterances, a whole conversation and a large corpus of online discussions.

The second system developed, CAO, is a system for analysis of emoticons in Japanese online communication. Emoticons are strings of symbols widely used in text-based online communication to convey user emotions. The presented system extracts emoticons from input and determines the specific emotion types they express. Firstly, it matches the extracted emoticons to a predetermined raw emoticon database containing over ten thousand emoticon samples extracted from the Web and annotated automatically. The emoticons, for which emotion types could not be determined using only this database, are automatically divided into semantic areas representing "mouths" or "eyes". These areas are automatically annotated according to their co-occurrence in the database. The evaluation, performed on both training and test sets, confirmed the system's capability to sufficiently detect, extract and analyze emoticons.

The above systems are then utilized in two methods for enhancing Human-Computer Interaction. The first is a method for automatic evaluation of conversational agents. The affect analysis systems are used to analyze users' emotional engagement during conversation. This data is reinterpreted to specify general attitudes to the conversational agent and its performance. In the evaluation, the method is used as a background procedure during conversations with two Japanese-speaking conversational agents. The users' attitudes to the agents are determined automatically during the conversations and compared to the results of a questionnaire taken after the conversations. The results provided by the method revealed similar tendencies to the questionnaire, proving the method as applicable in automatic evaluation of Japanese-speaking conversational agents.

Next, I present a method for determining whether emotions expressed by speaker are appropriate for the context of the conversation. In this method, affect analysis system estimates the speaker's affective states and a Web mining technique gathers from the Internet emotive associations consisting of a list of emotions that should be expressed at the moment. Implementing this method to a conversational agent could allow it choose appropriate conversational procedures, and therefore enhance human-computer interaction.

I conclude the dissertation with a discussion on possible further applications for the proposed systems and methods, and describe further work needed to implement the complete scope of Emotional Intelligence in machines.

Abstract (Japanese) / 博士論文概要

著者は自然言語処理分野における研究を行ってきた．特に注目してきたのは人間の会話に含まれる感情の言語表現をコンピュータに理解させ，それを基に話者（ユーザ）の感情状態を推定することである．人工知能における感情（喜怒哀楽など）の研究（Affective Computing，感情処理）は，15 年ほど前より行われている．その中には顔の表情や音声変動から感情認知を行う試みはあるが，言語における感情表現の研究はまだ初期段階である．

人間の感情のコミュニケーションの大部分は言語以外の媒体で伝達されるという考えが一般的である．しかし，言語の感情表現こそが，社会的関わりを表現していると考えられている．例えば，天気のいい日に友達と散歩に出かけた人は「今日はなんて気持ちいい日なんでしょう！」と感動を表すことで相手の注意を引き会話を始める．感情文では，話者の感情状態が表出されるのみならず，会話の流れも整理される．

著者は日本語における感情表現の研究を行って来た．第一段階として小規模なテキストデータの手作業での分析を行った．その結果，日本語における感情表現が人間同士のコミュニケーションを円滑に行うために非常に重要であるという結論が得られた．また，感情表現を２種類に分類することができた．一つは，感情が伝えられていることを聞き手に知らせ，発話の感情的コンテキストを設定する感情要素である．もう一つは，必ず感情的コンテキストで使われるわけではないが，感情的コンテキストで使われた場合，話者の感情状態を表す感情表現である．これらの発見を大規模データで確認する必要があった．そのため，上記の日本語の感情表現の働きを自然言語処理の分析方法を用いて確認するために必要なツールを開発し，実験システムを用いて評価実験を行った．

まずは文章内の感情認知・解析システム ML-Ask の開発を行った．ML-Ask では，ユーザの入力文を手作業で収集した感情要素・感情表現のデータベースに照らし，順番にマッチングを行う．感情要素がマッチングできた文では感情的コンテキストが決定される．さらに感情表現のデータベースを参照し，抽出された感情を話者の感情状態とする．

ML-Ask は大規模なテキストデータに感情タグ付けを自動的に付与ことができる．現在，日本語を豊富に含む大規模テキストデータとしてはインターネットが考えられる．しかし，インターネット上の言語リソース（ブログ，チャットルーム，掲示板など）には顔文字など，一般の辞書に存在しない表現が頻繁に使用されている．その処理を行うために顔文字解析システム CAO の構築を行った．

CAO システムは入力文から顔文字を抽出し，それらが表す感情を推定する．推定プロシージャではインターネットから１万以上の顔文字を抽出し，自動的に感情のグループ分けを行った．さらに，Kinesics（動作学）理論に基づき，顔文字を「口」や「目」などを表す部分に自動的に分け，システムのカバレッジ（顔文字の組み合わせ数）を約１万から３００万以上に拡大した．CAO システムの性能は９８％を超えた．

これらのツールを利用しインターネット上で感情表現や感情文の働きに関する研究をさらに進めた．これらの研究において，会話中の話者の感情状態を分析することで聞き手や会話対象に対する話者の態度を計算することができることが確認された．その成果を対話エージェントの自動評価手法として応用した．

本手法では感情情報論（Affect-as-Information）に基づき，エージェントとの会話中にユーザが表出した感情を基にして，ユーザが受けたエージェントの印象について推定する．評価実験では，２つの会話エージェントを利用し，ユーザはそれらと会話を行い，その後エージェントとの会話から得られた感情情報を印象評価実験のアンケート結果との比較を行った．本手法が，アンケート結果と類似した傾向を示し，本手法を自動評価手法として応用ができることが確認された．

また，WEB マイニング手法を用いて，認知した話者の感情状態が会話の場面に合っているかどうかを計算する手法を提案した．本手法では，以前に構築した感情解析システム（ML-Ask，CAO）が文中の感情の種類・感情極性を判断した後，その文に出現した感情の原因フレーズをインターネットで検索し，それと頻繁に出現する感情表現を ML-Ask の結果と照合し，一致した場合に文中の感情が文脈に適していると判断する．これらのシステム及び手法を対話エージェントに応用することで，ユーザの感情が理解でき適宜反応ができるロボットの開発に貢献ができると考えられる．また，本研究は現在日本語を用いて行われているが，開発してきた手法やシステムには統計的計算方法を用いているので研究成果は他の言語にも応用できると考えられる．

Chapter 1

Introduction

Scientists have been fascinated by emotions for centuries. There are remarkable works trying to describe the phenomenon of emotions, such as the ones by Darwin [22], or many others [51, 37, 92, 131, 132]. However, for a long time emotions were treated rather as an idol to worship, worthy of the attention of philosophers but not material or tangible enough to be accurately described in detail or processed by machines. Recent years have brought research on emotions into focus in Computer Science and Artificial Intelligence [96], and its sub-fields like Natural Language Processing [87]. The subjectivity of emotions however, drives researchers into a corner of ambiguity and often becomes an impediment to research in this field. However, I assumed automatic analysis of emotions, narrowed down to specified bounds, should give satisfying results.

Technological development has led to the creation of a new dimension of communication, where a machine is one part, referred to as Man-Machine Communication (MMC) [39], or Human-Computer Interaction (HCI) [27]. A rush in development of talking robots and conversational systems [45], indicates that the functional implementation of agents like intelligent car navigation systems [138] or talking furniture [43] in our everyday lives has already become a current process, and a need for humanized interfaces in MMC grows rapidly. In fact, recent years have showed a rising tendency

in Computer Science and Artificial Intelligence research to enhance Human-Computer Interaction by humanizing machines to make them more user-friendly [146]. The humanizing process can consist of making robots look like humans [73], but if a robot capable to act and talk with a user on the human level is to be created, it also needs to be equipped with procedures allowing it to understand human cognitive behaviors. Such robots could be very useful, playing roles of intelligent companions for humans, for example helping children in their development [119] or helping drivers not to fall asleep during a long ride home [138].

One of the most important cognitive human behaviors present in everyday communication is expressing and understanding emotions. Emotional states influence the decision making process in humans [124, 166] and are a vital part of human intelligence [121]. Therefore one of the current issues in Artificial Intelligence is to produce methods for efficient processing of emotional states of users. This concerns not only recognition and analysis of emotive information obtained from the users, but also efficient utilization of this information in the process of enhancement of peoples lives.

The field embracing this subject, called Affective Computing, although being a rather young discipline of study has been gathering popularity of researchers since being initiated only a little over ten years ago [96]. The interest in such research is usually focused on recognizing the emotions of users in human-computer interaction. In the most popular methods the emotions are recognized from: facial expressions [42], voice [144] or biometric data [141]. However, these methods, usually based on behavioral approaches, ignore the semantic context of emotions. Therefore, although achieving good results in laboratory conditions, such methods are often inapplicable in real

life tasks. For example, 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 recent discoveries prove that affective states should be analyzed as emotion-specific rather than divided simply into positive or negative valence [69, 132], most of the behavioral approach methods are incapable to distinguish emotions in more subtle manner than the two-part classification to pairs like joy-anger, or happiness-sadness (see for example [141]). Furthermore, biometric methods, like functional Magnetic Resonance Imaging (fMRI) or Electroencephalography (EEG), are usually laborious, time consuming and expensive. A problem is also how much the performing of an fMRI experiment in itself would influence the participants' emotional states. For example, a subject might feel nervous because of being plugged into the fMRI apparatus, which would obviously influence the results of an emotion detection experiment.

This led to the formation of Affect Analysis - a field focused on developing natural language processing techniques for estimating the emotive aspect of text. There have been several attempts to achieve this goal for the Japanese language. For example, Tsuchiya et al. [147] tried to estimate emotive aspect of utterances with a use of an association mechanism. On the other hand, Tokuhisa et al. [145] used a large number of examples gathered from the Web. Although the number of previous methods for text-based affect analysis is small, it indicates a positive tendency in natural language processing and text mining approaches. However, as the development of the field is still in progress, a large set of problems still needs to be recognized and solved. One

of such problems is the lack of standardization of emotion types, which often causes inconsistencies in emotion classification (compare for example [147] and [145]). Secondly, although there have been some methods for dealing with simple sentences, not much attention has been paid to the processing of elements of natural language appearing usually in text-based communication (e-mails, text messengers, Internet forums), such as informal language, jargon or emoticons, popular in online communication. Another problem still not addressed appropriately was, in general, the practical utilization of the emotive information recognized in the user.

In this dissertation I tried to address the above problems to develop tools and methods capable of specifying user emotional states in a more sophisticated way. Firstly, I restrain from using only a simple two-side classification of emotions into positive and negative. To contribute to the standardization of classification of emotions in affect analysis research I apply the most reliable classification available today for the Japanese language. Secondly, I focus on the development of tools and systems for affect analysis able to deal with the natural language used widely today in online communication, including informal language and emoticons.

Finally, I utilized the emotive information obtained through affect analysis in two ways. Firstly, to perform an automatic evaluation of conversational agents. Using this method in development of agents interacting with humans saves time and labor consumed on usability questionnaires. Secondly, I developed a deeper affect analysis method that not only specifies what type of emotion was expressed, but, applying contextual information processing and Web-mining techniques, determines whether the expressed emotion is appropriate for the context it appears in.

1.1 Note on the Language of Focus in this Dissertation

This dissertation describes research on natural language processing methods. The main language of focus in this dissertation is Japanese. All of the performed processing and all tools and methods are developed for this language. The topic of this dissertation, "emotions in language", is culture and language dependent, and therefore author did not focus on developing language-independent methods. However, all of the methods presented here are theoretically applicable within other languages, although their performance may vary from the one presented here.

All examples showing the performance of the systems and methods are given in Japanese, however the author always provides as close English translation of the examples as possible.

1.2 Transcription of Sentence Examples

To transcribe the examples in the Japanese language into the Latin alphabet, in this dissertation I used the Hepburn romanization system (Hebon-shiki Rōmaji). The system was first used by James Curtis Hepburn in the third edition of Japanese–English dictionary, published in 1887 [44]. The system was officially approved by the Society for the Romanization of the Japanese Alphabet [1] in 1885. In particular, I applied the revised version of the system called Shūsei Hebon-shiki Rōmaji.

[1] http://www.roomazi.org/

1.3 Annotation of Grammatical Information in Examples

In this dissertation, for annotation of grammatical information in sentence examples, I use the Leipzig Glossing Rules standard developed by the Max Planck Institute for Evolutionary Anthropology and the University of Leipzig [78].

Chapter 2

Background

2.1 Definitions

In this section I present the definitions of terms frequently used in this dissertation or closely related to the present research. I also briefly introduce the general ideas borrowed from psychology and explain how I applied them to the present research in Natural Language Processing. Further detailed explanations will appear along with the description of a system or a method in further chapters.

2.1.1 Definition and Classification of Emotions

As mentioned in section 1.1, the main language of focus in this dissertation is Japanese. Therefore I needed to apply the definition and classification of emotions proved to be the most appropriate for this language.

The simplified general definition of emotions says that emotions are every temporary state of mind, feeling or affective state evoked by experiencing different sensations [70]. This definition is also applied in the Dictionary of Emotive Expressions [85], a repository of words describing states of emotions developed by Akira Nakamura. This repository is also utilized in this research.

I complemented the above definition with the claims of Robert C. Solomon, who argues that people are not passive participants in their emotions, but rather emotions are strategies by which people engage with the world [131]. The assumption that emotions are strategies indicates there are also such strategies within the language. The strategies are evoked with the use of specific language patterns, such as sentences (utterances) or expressions. Therefore there is also a need to complement the above definition of emotions with a definition of emotive utterances. I applied Fabian Beijer's definition of emotive utterances, which says the emotive utterance is every utterance in which the speaker is emotionally involved, and this involvement, expressed linguistically, is informative for the listener [9].

As for the classification of emotions, I applied the one proposed by Nakamura [85], who after over 30 years of thorough study in lexicography of the Japanese language and emotive expressions, distinguishes 10 emotion types as the most appropriate for the Japanese language and culture. These types include: *ki / yorokobi*[1] (joy, delight), *do / ikari* (anger), *ai / aware* (sorrow, sadness, gloom), *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).

I used this classification instead of proposing an original one, which has been a common practice in affect analysis research, as Nakamura's several decades-long research on emotive expressions makes his classification the most reliable for the Japanese language.

[1]In this dissertation italics are used mostly to indicate expressions in Japanese.

2.1.2 Clarifying the Related Nomenclature

In this subsection I briefly clarify the differences between some of the emotion-related terms used in this dissertation.

Emotion. The classic definition of **emotion** says that it is a mental and physiological state caused by subjective experiences. However, in modern psychology and cognitive science [131] it is perceived more as a process in time including various specifically defined phenomena, such as affective states, sentiments, moods, or changes in attitudes (see also above for my working definition of emotion).

Feeling is defined in psychology as a conscious subjective experience of any physical sensation [151]. In common sense understanding it is used not only in terms of emotions, but includes also other sensations, such as "warm", "cold" or "soft" - also subjective and evaluative, but not directly emotional. In non-scientific, everyday language use the word "feeling" is also used in the meaning of "intuition".

Affect is often referred to as the experience of feeling [46] and represents an organism's reaction to stimuli. **Affective state** is the state caused by the experience of feeling (affect) and includes a process during which the organism interacts with and responds to the stimuli. The linguistic part of this phenomenon, on which I focus in particular, includes expressing one's emotions in a way informative to the environment (other interlocutors, such as people or an agent).

Mood is usually distinguished from affect on the basis of time and intentionality. It is said to be a relatively long-lasting emotional state not caused by any easily determinable stimuli. It is known that moods can be caused by changes of weather or diet. It was also discovered that moods influence people's tendencies in decision-making [124, 166].

Sentiment is defined as a person's conscious opinion, or attitudinal tendency towards an object. In the context of Sentiment Analysis it refers to attitudes (positive or negative sentiments) or opinions (specific).

Attitude in psychology refers to a person's perspective toward a specified object, in particular one's degree of liking or disliking of the object [13].

2.1.3 Two-dimensional Model of Affect

According to Solomon [131], people sometimes misinterpret specific emotion types, but rarely their valence. One might, for example, confuse such emotions as anger and irritation, but it is unlikely they would confuse admiration with detestation. Therefore, in my research I checked whether the general features of the emotion types, extracted by affect analysis systems (described in chapter 3), were in agreement. By "general features" I mean those proposed in the theory of a two-dimensional model of affect.

The idea of a two-dimensional model of affect was first proposed by Schlosberg [123] and developed further by Russell [114]. Its main assumption is that all emotions can be described in a space of two dimensions: the emotion's valence (positive vs. negative) and activation (active or activated vs. passive or deactivated). An example of positive-activated emotion would be "ex-

citement"; a positive-deactivated emotion is, for example, "relief"; negative-activated and negative-deactivated emotions would be "anger" and "gloom" respectively. In this way, four areas of emotions are distinguished: activated-positive, activated-negative, deactivated-positive and deactivated-negative.

Nakamura's emotion types (for reference see section 2.1.1) were mapped on this two-dimensional model of affect, and their affiliation to one of the spaces was determined. For some emotion types the affiliation is straightforward, e.g. gloom is never positive or activated. However, for other emotion types the emotion affiliation is not that obvious, e.g., surprise can be both positive as well as negative; dislike can be either activated or deactivated, etc. The emotion types with uncertain affiliation were mapped on all groups they could belong to. However, no emotion type was mapped on more than two adjacent fields. For the details of the mapping of the emotion types, see Figure 2.1.

This grouping is then used in my research for several purposes. Firstly, in a system for affect analysis of textual input utterances, ML-Ask, it is used in a procedure for determination of emotion types after a valence shifting procedure. Here, the grouping is used to specify which emotions correspond to the one negated by a phrase causing the shifting of a valence (for further explanations see sections 2.2.3 and 3.1.4). Secondly, the grouping of emotions on the two-dimensional affect space is used in an emoticon analysis system, CAO. In the evaluation of this system I used the grouping to verify whether the emotion types extracted by CAO belong to the same quarter, even if they do not match perfectly the gold standard of emotion types (for further explanations see section 3.2.5). Thirdly, the emotion grouping is used in a method for automatic evaluation of conversational agents, to estimate the

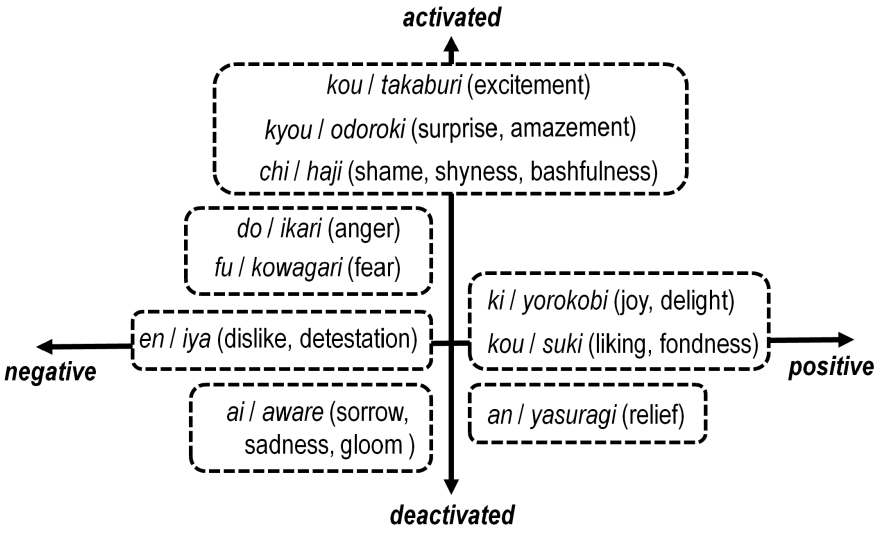


Figure 2.1:
Grouping Nakamura's classification of emotions on Russell's space.

attitude towards an agent by determining the valence polarity of emotions conveyed during a conversation (for further explanations see section 4.1). Finally, in emotion appropriateness verification procedure, the grouping helps estimating whether the emotion types tagged by affect analysis systems belong to the same Russell's space, even if they do not perfectly match the emotive associations gathered from the Web (for further explanations see section 4.2.3).

2.2 Linguistic Approach to Emotions, or How is it Possible to Recognize Emotions from Text?

It has been argued that the semantic and pragmatic diversity of emotions is best conveyed in language [131]. There are different linguistic and paralinguistic means used to inform interlocutors of emotional states in an everyday conversation. The emotive meaning is conveyed verbally and lexically through exclamations [9, 90], hypocoristics (endearments) [56], vulgar language [19] or, especially in Japanese, through mimetic expressions (*gitaigo*) [6]. A key role in expressing emotions is also played by the lexicon of words describing the states of emotions [85]. The analysis of elements of language such as intonation or tone of voice as well as nonverbal elements, like gestures or facial expressions, is also important in the task of recognizing emotions. However, in research such as mine, where the realization of language and therefore communication channel is limited to the transmission of lexical symbols, all nonverbal information is represented by its textual manifestations, like exclamation marks or ellipsis.

The function of language realized by the elements of language used to convey emotive meaning is called the **emotive function of language**. It was first distinguished by Bühler [14] in his theory of language as one of the three basic functions realised by language[2]. Bühler's theory was developed further by Jakobson [52], who distinguished three other functions providing the basis to structural linguistics and communication studies. The realisa-

[2]Other two functions were *descriptive* and *impressive*.

tion of the emotive function in language enriches the uttered language with a feature called **emotiveness**. This feature was widely discussed by Stevenson [135], who defined it as a strong and persistent linguistic tendency used to inform interlocutors about the speaker's emotions and evoke corresponding emotions in those to whom the speaker's remarks are addressed [135]. Bahameed [7] argues after Shunnaq [126], that emotiveness is the speaker's emotive intention embedded in the text through specific language procedures or strategies, some of which convey neutral/objective meaning, whereas others convey emotive/subjective meaning.

To grasp the view on how emotiveness is realized within language, I performed a literature review on the general subject of studying emotions within a language. The summary of this literature review is presented in the section 2.2.1. This study provided me with a view on how emotiveness is realized in languages in general. Further, I have performed a study on the procedures of how emotions appear in the Japanese language, especially wihtin the new medium, the Internet, a rich source of a natural up-to-date language [99]. Section 2.2.2 presents a short summary of the discoveries of this study.

2.2.1 Study of Emotions in Language: Literature Review

Research on emotions from a linguistic point of view, although still a young discipline, has already been widely done. For example, Wierzbicka's [156] works mark out a fresh track in research on cognitive linguistics of emotions among different cultures. Fussell and colleagues [38] approached the emotions from a wide cross-disciplinary perspective, trying to investigate the

emotion phenomena form three broad areas: background theory of emotions, figurative language use, and social/cultural aspects of emotional communication. Weigand and colleagues [155] tried to formulate a model of emotions in dialogic interactions proposing a rare attempt to explain emotions from a communicativist point of view. As for the Japanese language, which this research focuses on, Ptaszynski [99], made an attempt to explain communicative, as well as semiotic functions of emotive expressions in Japanese.

Apart from research generalizing about emotions, such as the above, there is also a wide range of study in the expressions of particular emotion types, or specific expressions of emotions. As for the former, Nakamura's [85] lifetime research in lexicography and rhetology of the Japanese language resulted in the creation of the first dictionary of expressions describing states of emotions in Japanese. As for the latter, for example, Baba [6] studied Japanese mimetics in spoken discourse, Ono [90] studied emphatic functions of Japanese particle *-da*, and Sasai [122] examined *nanto*-type exclamatory sentences.

However, there have been little cross-sectional research gathering the knowledge regarding the structures and functions of emotions and their expressions in language from points of view of semiotics, communication studies or pragmatics. The rare examples that exist, such as Ptaszynski's [99], still only emphasize the need for further research in these fields.

The lack of such research is most likely caused by the limitation that linguists usually perform the analysis manually. A great help here could be offered by computer supported corpora analysis. Some of the first attempts in this direction were made by Bednarek [8], who performed corpus studies of emotion terms and their patterns in English. Wilson and Wiebe [157]

started a long term project of manual annotation of English corpora for subjectivity analysis. By the subjectivity they mean not only emotions, but also sentiments, speculations, evaluations and other private states. Such a diversity of features to process makes their effort praiseworthy. As for Japanese, there is a wide range of research on sentiment analysis and opinion mining [48, 61]. However, there have not been much done in the topic of another feature of subjective language, namely expressing emotions. As I found out, this feature is realized in Japanese very systematically. As for corpora analysis for emotion research in Japanese, there are still no large corpora annotated with emotive information, nor is there a reliable system for automatic annotation of available corpora. The need of such is strongly highlighted in the little research done in this field, which still depend on small manually annotated corpora [80].

2.2.2 Emotiveness in the Japanese Language

The interdisciplinary research on the emotive function of language in Japanese [99], performed with regards the linguistic approach to emotions described on the beginning of this section, allowed me to distinguish two general types of realizations of this function in the Japanese language. The first realization is the use of elements indicating that emotions have been conveyed, but not detailing what specific emotions there are, or expressing different emotion type according to different context. I named this group emotive elements, or, shortly, **emotemes**. Although the same emotive element can express different emotion types depending on context, their use indicates undoubtedly that the speaker performed an emotively emphasized utterance. This group is linguistically realized by interjections, exclamations, mimetic expressions,

or vulgar language. Examples are: *sugee* (great!), *wakuwaku* (heart pounding), -*yagaru* (a vulgarization of a verb), respectively. As for the second type of realization of emotive function, I distinguished parts of speech, or phrases, that in emotive sentences describe specific emotions. This type, generally referred to as **emotive expressions**, includes parts of speech like nouns, verbs, adjectives, etc. Examples are: *aijou* (love), *kanashimu* (to feel sad), *ureshii* (happy).

Some emotemes are used to express only a limited number of emotion types. For example, an exclamation mark "!" is used to express excitement, or anger, rather than relief, or gloom. There are also emotemes that are used to express only one emotion type, like mimetic expressions in Japanese (*gitaigo*). *Gitaigo*, when used in an utterance introduce informal speech making the utterance emotive, like in:

Hoomu ni tsukiotosarenaika hiya-hiya shiteta.
A platform DAT push-PASSIVE-NEG-QUO be afraid-CONT-PAST.
I was afraid [he/she/someone] would push me under the train.

Therefore it can be said that some emotemes fulfill both functions - of emotemes and emotive expressions. This way the working definition of emotemes has to be supplemented with the statement saying that: emotemes either do not detail what specific emotions are conveyed, or express different emotion type according to different contexts, while the number of contexts is larger than zero and at least one.

Examples of sentences containing emotemes and/or emotive expressions are shown in Table 2.1. Examples (1) and (2) show emotive sentences. (1)

Table 2.1:
Examples of sentences containing emotemes (underlined) and/or emotive expressions (bold type font).

Example of a sentence Grammatical information English translation	emotemes	**emotive expressions**
(1) *Kyo wa nante* ***kimochi ii*** *hi nanda !* Today TOP [*emoteme*] [**pleasant**] day [*emoteme* x2] Today is such a nice day!	yes	yes
(2) *Iyaa, sore wa sugoi desu ne !* [*emoteme*] that TOP [*emoteme*] is [*emoteme* x2] Whoa, that's great!	yes	no
(3) *Ryoushin wa minna jibun no kodomo wo* ***aishiteiru****.* Parents TOP all one's own child(/ren) ACC [**love**] GER are. All parents love their children.	no	yes
(4) *Kore wa hon desu.* This TOP book COP. This is a book.	no	no

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.

2.2.3 Contextual Valence Shifters

In the task of analyzing emotions in language, an idea that becomes helpful is the idea of Contextual Valence Shifters.

The idea of Contextual Valence Shifters (CVS) as an application in Natural Language Processing was first proposed by Polanyi and Zaenen [97] for the task of Sentiment Analysis[3]. They distinguished two kinds of CVS: negations and intensifiers. The group of negations contains words and phrases like "not", "never", and "not quite", which change the polarity of valence, or the semantic orientation, of an evaluative word they refer 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 the field of Sentiment Analysis of texts in English [59]. A few attempts for the Japanese language [81] indicate the idea is applicable for this language as well.

Examples of CVS negations in the Japanese language are grammatical structures such as *amari -nai* (not very-), *-to wa ienai* (cannot say it is-), *mattaku -nai* (not at all-), or *sukoshi mo -nai* (not even a bit-). Intensifiers are represented by such grammatical structures as *totemo-* (very much-), *sugoku-* (-a lot), or *kiwamete-* (extremely).

In this research I applied CVS to Affect Analysis. I focused mostly on negations, since they have an immediate and significant influence on the meaning of emotive expressions. I manually collected a database of CVS containing 71 negation structures. These structures are used to shift the valence of the recognized emotive expressions. Valence shifting is necessary to avoid confusion in determination of emotion types. A detailed explanation

[3]For the definition of Sentiment Analysis see sections 2.1.2 and 4.1.1

of the CVS procedure in affect analysis system is presented in section 3.1.4.

2.3 On the Need for Context Processing in Affective Computing

Research on Emotions within Artificial Intelligence and related fields has flourished rapidly through several years. Unfortunately, in much research the contextuality of emotions is disregarded. Based only on behavioral approaches, methods for emotion recognition ignore the context of emotional expression. Therefore, although achieving good results in laboratory conditions, such methods are often inapplicable in real world tasks. For example, a system for recognition of emotions from facial expressions, assigning "sadness" when user is crying would be critically mistaken if the user was, e.g., cutting an onion in the kitchen.

In this section I argue, that recognizing emotions without recognizing their context is incomplete and cannot be sufficient for real-world applications. I present logical underpinnings of this claim and describe some consequences of how disregarding the context of emotion could cause a fallacy in system performance.

In this dissertation I will present my approach, in which I focused both on the expression of emotion and the context it appears in. In particular, I apply context processing to affect analysis in two ways.

Firstly, one of the common problems in the keyword-based systems for affect analysis is confusing the valence of emotion types, since the emotive expression keywords are extracted without their grammatical context. An idea aiming to solve this problem is the idea of Contextual Valence Shifters

(CVS). As the first step towards contextual processing of emotions I applied CVS as a supporting procedure for affect analysis system for Japanese (for details se sections 2.2.3 and 3.1.4).

As another realization of contextuality in affect analysis I have developed a method making use of the wider context an emotion is expressed in. The method, using a Web mining techinque, determines, whether the expressed emotion is appropriate for its context. It introduces an idea of Contextual Appropriateness of Emotions to Affective Computing research. This idea adds a new dimension in emotion recognition, since it assumes that both positive and negative emotions can be appropriate, or inappropriate, depending on their contexts (for details se section 4.2). This method is based on the assumption that the Internet can be considered as a database of experiences people describe on their homepages or weblogs. Since context of emotions is formulated through collecting experiences (see section 2.3.2), these experiences could be as well "borrowed" from the Internet [118].

In cocnlusions of this dissertation I present a discussion on future directions and applications of context processing within Affective Computing.

2.3.1 Emotions and Intentionality

View on emotion phenomena has evolved in time. In Middle Ages, emotions were considered as biological disturbances, passive states with no relation to rational thinking or cognition [26, 53]. This approach has been proved wrong in neurobiology where it was showed that emotions and rationality are not separable entities but stem from each other as equally important processes in decision-making [20, 21, 88], and cognitive processes [66, 37, 92, 68, 67]. It was thus reassured that emotions are conscious and intentional mental

phenomena [47, 24]. Oxford English Dictionary defines intentionality as "the distinguishing property of mental phenomena of being necessarily directed upon an object, whether real or imaginary" [127]. In other words, intentional phenomena are always "about something". As a property of emotional processes, the idea of intentionality implies that emotions necessarily need to be assigned a [formal/intentional] object [67, 131, 60]. The linguistic-pragmatic reality proves this. When people express emotions they often express them in terms of specifying their objects. For example, people are *afraid/proud* **of** something, or *happy* **about** something[4], etc. A function of all specific emotion objects forms a formal object of emotion. The formal objects of emotions have been defined as "axiological properties[5] which individuate emotions, make them intelligible and give them correctness conditions" [158, 83, 142]. Moreover, Solomon, in his theory of emotions as "engagements with the world" argues that emotions are not only intentional, but they are conscious choices and strategies by which people manage the world. The targets of those strategies are formal objects. Moreover, emotions and their formal objects are necessarily in a causal relation [5]. Formal objects, as sets of axioms defining the emotions, can be further reformulated as properties determining the context of emotions.

[4]It has been argued that "moods" are not about anything specific and therefore are not intentional, which would introduce an inconsistency in the definition of emotion and emotion-related processes. Solomon solves this problem by noticing that the objects of moods are not unspecified, but rather moods take as their objects the whole world [131, 133].

[5]Or "properties derived from axioms [= here, specific objects of emotions]".

2.3.2 Emotions and Contextuality

The idea of contextuality with an application in logics, as proposed by Gershenson, assumes that "concepts are determined [...] by the context they are used in. Gershenson gives a relative notion of a context as follows.

> *"A **context** consists of the set of circumstances and conditions which surround and determine an idea, theory, proposition, or concept. These circumstances and conditions can be spatial, temporal, situational, personal, social, cultural, ecological, etc."*
>
> - Gershenson, 2002 [40]

Gershenson gives the following general example: "the concept 'cat' will be determined by the context in which it is used. It can be a context of veterinary medicine, naughty pets, violent cartoons, cute animals, Broadway musicals, etc. The way we refer to 'cat' will change considerably depending on the specific context that we are using." [40]. He formalizes this idea as follows: "Every proposition P can only have a truth value (or vector) **in dependence** of a context C. This truth value is **relative to** the context C." Gershenson argues further that, since people learn concepts socially, incongruencies within context of any concept are verified by experience.

All the above, when expanded to the emotion phenomena provides a set of conditions for an emotion to take place.

- Emotion E, takes form of an expression e and has an object O_E;
- E is in causal relation with O_E ($O_E \rightarrow E$);
- Emotion object O_E defines (partially) E, and

- Gives the [correctness/truth] condition to E;
- O_E (or set of O_E's) makes up a context C_E for E, and
- $O_E \in C_E$;
- General C_E is formulated through collecting experiences X, and
- Value of E changes with the change of C_E;

With this set of conditions I propose a simplified statement that *emotion* is a function of *expression* appearing within *context*, where *context*, learned by *experience*, is constituted by *object(s)*. As one can see, the function is solvable only when a certain expression appears within a matching context. The function is not solvable when the context either does not match the expression or is not given, or computed. Below I present a set of examples of situations, where [not providing/providing false] context for an expression ends in error in computation.

Consequences of Ignoring Emotion Context

Generally perceived emotion recognition can be defined as "using some (behavioral[6]) assumptions to determine emotional state (of a human)". The assumption that emotions can be sufficiently analyzed looking only at the behavior comes from William James [53, 54]. James gives an example of what happens when people see a bear. When one sees a bear, hair on one's head stand up, he feels shivers, opens his eyes wide and runs. In James's interpretation the mind perceives the behavior (adrenaline, fast heartbeat,

[6]We use "behavioral" in a wide meaning, including body or face movements (bodily behavior), physiological processes (inner behavior of human body as a system), speech signals and language (language behavior).

eyes open wide) as the emotion (fear). Although this theory has been proved wrong (see for example Ellsworth in [34]), in emotion recognition it is still a usual approach. In examples below, we show how looking only at the behavior/expression and not taking into consideration the context/object of emotion might cause critical errors in emotion recognition.

Example 1: Recognition of emotion from facial expressions

- **Expression**: User is crying (presence of tears and facial expression);
- **Assumption**: User is sad (?);
- **Context**: The user is cutting an onion in the kitchen;

It can be easily noticed that a system based on the assumption that, when the user is crying, he must be sad, will be critically wrong when not processing the context of this behavior.

Example 2: Recognition of emotion from speech signals

- **Expression**: User speaks loudly;
- **Assumption**: User is angry (?);
- **Context**: The user is listening to the music with his headphones on and cannot hear well;

Example 3: Recognition of emotion from gestures

- **Expression**: User waving hands above his head;
- **Assumption**: User is angry (?);

- **Context**: The user has won a lottery or is in trouble and is waving for help;

In the worst scenario, a robot designed to draw back when the user is angry will not help the user eventually causing the user's death.

Example 4: Recognition of emotion from physiological signals

- **Expression**: User has a high blood pressure;
- **Assumption**: User is excited (?);
- **Context**: The user has a hypertension or arrhythmia;

Not knowing the context that the user has a hypertension might bring serious consequences. In the worst scenario one can imagine a grotesque situation when a robot, designed to familiarize with a user, starts showing an expression of excitement, while the actual need is to give the user a medicine.

Example 5a: Recognition of emotion from language

- **Expression**: User has used vulgar language, such as "f*ck";
- **Assumption**: User is irritated (?);
- **Context**: The user is actually saying it like "Oh, f*ck, yeah!" (positively excited);

Example 5b: Recognition of emotion from language

- **Expression**: User has used the word "happy";
- **Assumption**: User is happy (?);

- **Context**: The user might be actually saying: "I'm not happy", or "I'm so happy that bastard was hit by a car!";

The examples above show that determining only the expression for an emotion does not yet provide a sufficient conditions for the computation to take place and the need for considering the context is clearly visible.

Processing the context of emotions, or Contextual Affect Analysis [108], is a newly recognized field. During its fifteen years of history, Affective Computing was in great part focused on recognizing user emotions. However, little research addressed the need for computing the context of the expressed emotions. In the age of information explosion, with an easy access to very large sources of data (such as Internet), the time has come to finally address this burning need. My research is focused on only one type of emotion processing, affect analysis of text. The future challenge will be to develop methods for processing the context in more general meaning, making the machines aware of the sophisticated environment humans live in. Contextual affect analysis is a feasible task and I believe much research will be done in this matter in the near future.

2.4 Pragmatic Approach to Implementation of Emotions in Machines

2.4.1 Affective Computing or Defective Computing?

The man-made rapid evolution of machines made computers equal or even surpass humans in features, like processing speed[7] or memory capacity[8]. Still, whether machines could obtain more human-like features, like emotions, remains a riddle. A motivation for Affective Computing [96] is to create a machine able to understand the emotions of users and adapt its behavior according to these emotions. Two approaches to fulfill this goal has emerged: recognizing user emotions; and implementing the actual emotion procedures in machines. Emotion recognition, the main stream in the field, has a fairly long history of attempts to recognize emotions from facial expressions, voice and language. However, in research focused on recognizing the emotions, questions like "How to use the recognized information?" or "Is it enough to recognize the emotions, or is there something more?" are often neglected. Although the machine is meant to respond appropriately for user emotions, the actual research on how would these responses look like is rare or simplified [145]. The new stream, focused on implementation of the emotions as agent-focused procedures, assumes that by providing the agent architectures for simulating emotional experience its reactions will be more human. This however leads to a more profound problem. The scientific description of

[7]Neuron switching speed is known to be on the order of 10^{-3} seconds, whereas computer switching speed is of 10^{-10} seconds.

[8]Current estimates of brain capacity range from 1 to 1000 terabytes, whereas 64-bit computer architecture is estimated to be capable to process effectively 16.8 million terabytes.

human emotions is still incomplete and their implementation might lead to undesirable effects. An agent with implemented the procedure of "fear" activated in the case of failing to solve a task [134], after approaching too many unmanageable tasks could become paranoid or depressive. Since paranoia is a typically human illness, such an agent would surely be human-like, but this kind of human-likeness should be considered rather as a defect, since there will be no practical application for such an agent. Although architectures simulating emotional experience used as models for simulating human behavior, might help understand emotional processes and contribute to curing psychosomatic diseases, such research should be performed with caution and attention of psychologists as much as computer scientists.

2.4.2 Agent-companion for Emotion Management

In my research I focused on exploiting emotional information in user-agent interaction to enhance human lives. As the semantic and pragmatic diversity of emotions is said to be best conveyed in language [131], I decided to focus on natural language processing methods for emotion recognition and their use in conversational agents, in particular the ones which perform a non-goal-oriented free conversation, or small talk. As is argued in the literature, small talk has important social functions [18] and, e.g., in the form of humorous conversations, is a necessary mean of emotion management in counseling [36]. It also has a great influence on children's acquisition of moral rules [89]. Therefore, in my assumption, in the development of an agent-companion/counselor, emotional information conveyed during the small talk with the user could be of good use. The human-likeness of such an agent is thus an important issue. It was already proved that conversational agent's re-

sponses are perceived as more natural when modality is added to the agent's response [45]. However, the naturalness is not the only issue. The agent-counselor[9] should be able to recognize user's negative emotions and induce positive moods, e.g., by a humorous response. It was showed by Dybala et al., that implementation of a joke generator in a conversational agent greatly improves its impression [31]. However, a humorous response is not always the desirable one, and, although meant to cheer-up the user with a joke, might cause the opposite effect, especially when he or she requires other responses, like a counsel or a consolation. Therefore the agent should be able to evaluate the user's emotions towards the context of the conversation to choose the appropriate response. Also, supported with morality rules [115], it should be able to detect the potentially inappropriate user utterances and react by pointing out the potentially undesirable consequences. To achieve this the agent must obtain a certain level of Emotional Intelligence.

2.4.3 Computing Emotional Intelligence

The idea of *Emotional Intelligence* (EI) was first officially proposed by Salovey and Mayer [121] who defined it as a part of human intelligence consisting of a set of 16-20 abilities grouped in four general groups labelled as: **I)** perceiving emotions, **II)** integrating emotions in facilitation of thoughts, **III)** understanding emotions and **IV)** regulating emotions to promote personal growth. This set of abilities is assembled in an EI framework [79].

After close investigation of the framework, I found out that managing emotions was set as the final ability requiring the presence of all others.

[9] I apply the general definition of the term counselor as 'a conversational partner able to help people understand and manage their emotions.'

A surprising discovery was that recognizing emotions, on which Affective Computing has been focused for over fifteen years was only the first, basic ability from a dozen or so. The attempts to implement the EI Framework often do not go beyond theory [4], and the few practical attempts eventually still do not go beyond the first basic step, namely recognition of emotions [95].

Mayer and Salovey generally divide the first step in the EI Framework, perception of emotions, into **a)** the ability to identify or recognize emotions and **b)** discriminate between accurate (or appropriate) and inaccurate (or inappropriate) expression of emotions.

According to Salovey and Mayer, recognizing emotions is only the basic step to acquire full scope of Emotional Intelligence and does not yet reveal anything about whether it is appropriate to express such an emotion in a given situation, and what actions should be undertaken as a reaction [121]. Solomon [131] argues further, that the valence of emotions is determined by the context they are expressed in. For example, anger can be warranted (a reaction to a direct offense) or unwarranted (scolding one's children for one's own mistakes), and the reactions should be different for the two different contexts of anger. Unfortunately, in the grand majority of the emotion processing research, the above fact is not taken into consideration. It is often assumed that positive emotion is always desirable and therefore appropriate, and negative emotion is always undesirable and therefore inappropriate.

A remedy for this misunderstanding could be the second ability in the EI framework, namely 'discriminating whether the expression of emotion is accurate for the situation it is expressed in'. In other words, whether it is appropriate or for its context. This introduces a new dimension in emotion

recognition, since it assumes that both positive and negative emotions can be appropriate, or inappropriate, depending on their contexts. See, e.g., the examples below.

1. I'm so happy I passed the exam! `[happiness/positive: appropriate]`
2. I'm so happy that bastard was hit by a car! `[happiness/ positive:inappropriate]`
3. I'm so depressed since my girlfriend left me... `[depression/negative:appropriate]`
4. I'm so depressed for the Easter is coming... `[depression/negative:inappropriate]`

The structure of these particular examples consists of: expression of emotion (here: the beggining of the sentence), and its context (the latter part of the sentence).

With the research presented in this dissertation I made an attempt to go one step further on the long way of implementation of Emotional Intelligence into machines and developed a prototype method for verification of appropriateness of emotions to their contexts, which takes advantage of the type of sentences as the above [105]. Following recognition of emotions expressed by a speaker, the appropriateness of those states is verified against their contexts. See the details of the method in section 4.2.

The method paves the way to the implementation of other EI abilities, such as understanding emotions, and regulating emotions. I believe the implementation of the whole scope of EI framework into machines is possible and will greatly contribute to the research on human-computer interaction. The research presented in this dissertation is the first step to achieve this goal.

Chapter 3

Tools for Affect Analysis of Textual Input

In this chapter I introduce two systems for affect analysis I developed. The first system, ML-Ask, performs affect analysis of textual input utterances in Japanese and provides information on emotive structure of utterances in Japanese. The second system, CAO, performs analysis of emoticons, representations of body language in online communication and provides the classification of potential emotion types represented by the analyzed emoticons.

3.1 ML-Ask: A System for Affect Analysis of Utterances in Japanese

3.1.1 Related Work

Recent years have brought research on emotions into focus in Computer Science, Artificial Intelligence, and Natural Language Processing [96]. One of the tasks recognized within the scope of interest in this field is called Affect Analysis. Affect Analysis, as defined by Grefenstette [41] is a natural language processing technique for recognizing the emotive aspect of text.

There have been many attempts to analyze the emotive aspect of text, with much divesrity in the results. For axample, Read [112] was not able to gather an appropriate emotive lexicon and used an unsystemized classification of emotion types, which caused a rather low final result of 33% of accuracy in recognition. Alm et al. [3] achieved a better accuracy score, 69%, however they too wrestled with the lack of emotive databases and with an inappropriate evaluation corpus consisting of children's stories full of ambiguities arising from mixing styles and means of expression (e.g. dialogues mixed with narrative style and descriptions). Wu et al. [160] achieved 72% of average accuracy, but ran into a problem of ambiguity of emotional rules.

There have been also some attempts to affect analysis for the Japanese language. For example, Tsuchiya et al. [147] tried to estimate emotive aspect of utterances with a use of an association mechanism. On the other hand, Tokuhisa et al. [145] used a large number of examples gathred from the Web. This positive tendency seen in natural language processing and text mining approaches lead however to a new set of problems in emotion estimation.

The lack of standardization often causes inconsistencies in emotion classification (compare [147] and [145]). Moreover, there is still a lack of computer supported linguistic studies on emotions. The systems developed using NLP methods, usually focused on detecting emotional states in Human-Computer Interaction (HCI), reraly perform any linguistic analysis of the content of utterances (for details see [147, 145]). This makes such methods unable to provide answers for questions like: "What makes an utterance emotive?", "What is the structure of an emotive expression used in an utterance", or "How does the emotive utterance function in context?".

Another problem of such systems is that they often distinguish only between positive and negative valence of the text [149], although recent discoveries show that affective states should be analyzed as emotion-specific [69, 132]. Moreover, the ones capable of more fine-grained emotion type classification [147, 145] do not base the classification of emotion types on any standards, but propose a non-standard classification tailored to their own needs. Furthermore, present NLP methods for Affect Analysis are incapable of distinguishing between emotionally emphasized and neutral contents [145]. Finally, the methods using Web mining, like the ones mentioned above, are grappling with the problem of noise contained in the contents gathered from the Web. All of the above makes the usual NLP methods less than ideal in tasks where affect analysis would be of great use, namely, automatic corpus annotation or linguistic studies on emotions.

The problems and needs described above encouraged me to undertake the research described in this chapter. I aimed to create a system capable to help in linguistic research on expressing emotions in the Japanese language. As mentioned above, one of the problems in this kind of research, is that re-

searchers need to annotate manually the corpora they wish to analyse. This is also the reason for the lack of large corpora annotated with emotive information. Therefore the creation of a system for automatic affect annotation of corpora was an urgent need. Allowing for quick annotation of large corpora, such a system would significantly speed up the research on emotions in language.

To develop the system I first performed a study on how the expressions of emotions are represented in linguistics. The details of this study are described in section 2.2. A large number of linguistic research provided me a list of related features which should be regarded in the creation of the system. These features are described below in section 3.1.2.

3.1.2 Defining Emotive Linguistic Features

I gathered databases of emotemes and emotive expressions according to the previous two-part classification into emotemes and emotive expressions. The feature set was collected in a way similar to the one proposed by Alm et al. [3], by using multiple features to handle emotive sentences. Alm et al. however, designed their research for English children's stories, whereas I focus on utterances in Japanese, and therefore used Ptaszynski's classification as more appropriate for the task.

Emotemes

Into the group of emotive elements, formally visualisable as textual representations of speech, Ptaszynski [99] includes the following lexical and syntactical structures.

Exclamative utterance. I agree with Beijer's [9] definition of exclamative/emotive utterances, as every utterance in which the speaker is emotionally involved, and this involvement is linguistically expressed. The research on exclamatives in Japanese [90, 122] provides a wide scope of topics useful as features in my system. Some of the exclamative structures are: *nan(te/to/ka)∼darō*, or *–da(yo/ne)*, partially corresponding to wh-exclamatives in English (see the first sentence in Table 2.1).

Interjections are typical emotemes. Some of the most representative Japanese interjections are *waa*, *yare-yare* or *iyaa* (see the second sentence in Table 2.1).

Casual Speech. Casual, or colloquial speech (COL) is not an emoteme 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 *-ē*, like in the example:

Ha ga itē!
Tooth NOMINATIVE hurts[COL] !
My tooth hurts!

or abbreviations of forms *-noda* into *-nda*, like in the example:

Nani yattenda yo!?
What do[COL] COP SFP!?
What the hell are you doing!?.

Gitaigo. Baba [6] distinguishes *gitaigo* (mimetic expressions) as emotemes specific for the Japanese language. Not all mimetics 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), or *hiyahiya* (be in fear, nervous), like in the sentences:

Omoidasenkute iraira shita yo.
Recall NEG GER be irritated-PAST-SFP.
I was so irritated, 'cause I couldn't remember [what I wanted].

Jūgeki demo sareru n janai ka to omotte, hiyahiya shita ze.
Shoot even do-PSV PARTICLES QUOT think GER, be afraid-PAST SFP.
I thought he was gonna shoot me - I was petrified.

Emotive marks. This group contains punctuation marks used as a textual representations of emotive intonation features. The most obvious example is exclamation mark "!". In Japanese, marks like ellipsis "...", prolongation marks, like "–", or "∼", are also used to inform interlocutors that emotions have been conveyed (see examples (2) and (6) Table 3.1).

Hypocoristics (HY, 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-nē-san* [older sister] vs *o-nē-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!!
Lately Oo[HY] and Mit[HY] NOM me-PL become involved with PAST !!
Oo-chan and Mit-chan has been palling around with us lately!!

Vulgarisms. The use of vulgarisms usually accompanies expressing emotions. However, despite the general belief that vulgarisms express only negative meaning, Ptaszynski [99] notices that they can be used also as expressions of strong positive feelings, and Sjöbergh [129] 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 contain words, phrases or idioms. Such a lexicon can be used to express emotions, like in the first example in the Table 1, however, it can also be used to formulate, not emphasized emotively, generic or declarative statements (third example in the same table). Some examples are:

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

nouns: *aijō* (love), *kyōfu* (fear);

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

fixed phrases/idioms: *mushizu ga hashiru* (to 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* (saddness like a physical pain), *aijō wa eien no honoo da* (love is an eternal flame);

3.1.3 Database Collection for Affect Analysis System

Based on the linguistic approach to emotions, as well as the definitions of emotive linguistic features, I constructed ML-Ask (eMotive eLements / Emotive Expressions Analysis System). ML-Ask was developed for analyzing the emotive contents of utterances and automatic annotation of corpora with the emotive information. The system uses a two-step procedure: 1) Analyzing the general emotiveness of an utterance by detecting emotive elements, or emotemes, expressed by the speaker and classifying the utterance as emotive or non-emotive; 2) Recognizing the particular emotion types by extracting expressions of particular emotions from the utterance.

The databases for each step of the procedure were gathered manually. The emoteme databases for the system were gathered from other research and grouped into five types. Code, reference research and number of gathered items are presented below in square, round and curly brackets, respectively:

1. **[EX]** Interjections and structures of exclamative and emotive-casual utterances ([85, 93, 148, 90]). {477}
2. **[GI]** *Gitaigo* ([85, 93, 6]). {213}
3. **[HY]** Hypocorystrics ([56]). {8}
4. **[VU]** Vulgarisms ([130]). {200}
5. **[EM]** Emotive marks ([56]). {9}

These databases were used as a core for ML-Ask. The databases of emotive expressions contain Nakamura's dictionary [85] (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}.

3.1.4 Affect Analysis Procedure

On textual input provided by the user, two features are computed in order: the emotiveness of an utterance and the specific type of emotion.

To determine the first feature, the system searches for emotive elements in the utterance to determine whether it is emotive or non-emotive. In order to do this, the system uses MeCab [63] for morphological analysis and separates every part of speech. MeCab recognizes some parts of speech I define as emotemes, namely, interjections, exclamations or sentence-final particles, like *-zo*, *-yo*, or *-ne*. If these appear, they are extracted from the utterance as emotemes. Next, the system searches and extracts every emoteme based on the system's emoteme databases (907 items in total). Finally, a simple emoticon detector informs about the presence of emoticons in the utterance. This is performed by detecting the appearance of at least three symbols in a row, used usually in emoticons. A set of 455 of those symbols was selected as being the most frequent symbols appearing in emoticons analyzed by Ptaszynski [99]. This simple emoticon detector activates further emoticon analysis system CAO described later in section 3.2.

All of the extracted elements mentioned above (exclamations from MeCab, emotemes and emoticons) indicate the emotional level of utterance.

Secondly, in utterances classified as emotive, the system uses a database of emotive expressions to search for expressions describing emotional states. This determines the specific emotion type conveyed in the utterance. Some examples of output, analysis performed by ML-Ask, are shown in Table 3.1.

Table 3.1:
Examples of the ML-Ask system analysis. From the top line: example in Japanese, emotive information annotation, English translation. Emotemes-underlined, emotive expressions-bold type font.

(1)	*Kyo*	*wa*	*nante*	***kimochi ii***	*hi*	*nanda*	*!*
	Today	TOP	EX:nante	**EMO-X:joy**	day:SUB	EX:nanda	EM:!
	'Today is such a nice day!'						
(2)	*Iya∼,*	*sore*	*wa*	*sugoi*	*desu*	*ne–*	*!*
	EX:iya∼	this	TOP	EX:sugoi	COP	EX:ne–	EM:!
	'Whoa, that's great!'						
(3)	***Hitoribocchi***	*nante*		***iya***		*da*	*∼∼*
	EMO-X:sadness	EX:nante–da		**EMO-X:dislike**		COP	EM:∼∼
	'Being alone sucks...'						
(4)	*Kanojo-wa*	*nante*		*kirei-na*	*josei*	*nanoda.*	
	she-TOP	EX:nante-nanoda		beautiful	lady	COP	
	'What a beautiful lady she is!' [Ono, 2002, [90]]						

CVS Procedure in ML-Ask

One problem in the procedure described above was confusing the valence polarity of emotive expressions in some sentences. The cause of this problem was extracting from the utterance only the emotive expression keywords without their grammatical context. Two utterance showing such cases are presented in Table 3.2 in examples (1) and (2). For example, in (1) the emotive expression is the verb *akirameru* (to give up [verb]), however, a CVS phrase, *-cha ikenai* (Don't- [particle+verb]) suggests that the speaker is in fact negating and forbidding the emotion expressed literally. Analysis of phrases like that allows automatic shifting in the valence polarity of emotive expressions in utterances containing Contextual Valence Shifter structures and solves the problem of confusing the value of an emotive expression. In this research I focused mostly on negation-type of CVS, since they have an

Table 3.2:
Two examples of change in emotion determination in ML-Ask by CVS procedure. Emotemes - underlined; emotive expressions - bold type font.

(1)	***Akirame***	*cha* *ikenai*	*yo*	*!*
	EMO-X:dislike	EX:cha|CVS:*cha-ikenai*{→**joy**}	EX:yo	EM:!
	'Don't ya give up!'			
(2)	*Sonnani*	***omoshiroku*** *mo nakatta*	*yo*	*...*
	So much	**EMO-X:joy**|CVS:*mo nakatta*{→**dislike**}	EX:yo	EM:...
	Oh, it wasn't that interesting...			

immediate and significant influence on the meaning of emotive expressions. My hand-crafted database of CVS contains 71 negation structures.

However, using only the CVS analysis, although it would be possible to find out about the appropriate valence of emotions conveyed in the utterance, an exact emotion type would be still unknown. Therefore, to specify the emotion types in such utterances I applied the idea of the two-dimensional model of affect to the CVS procedure.

Applying Two-dimensional Model of Affect to CVS Procedure The need to change the valences in emotion estimation research is a common problem. However, it is not uncommon for researchers to use valence changing patterns constructed by themselves without any scientific grounds. For example Tsuchiya and colleagues [147] used their own list of contrasting emotions. However, they do not consider that, as is argued by Solomon [131], the fact that two emotions are in contrast is not a matter of clear division, but is more complex and context dependent. I assumed this complexity could be specified with the help of the two-dimensional model of affect.

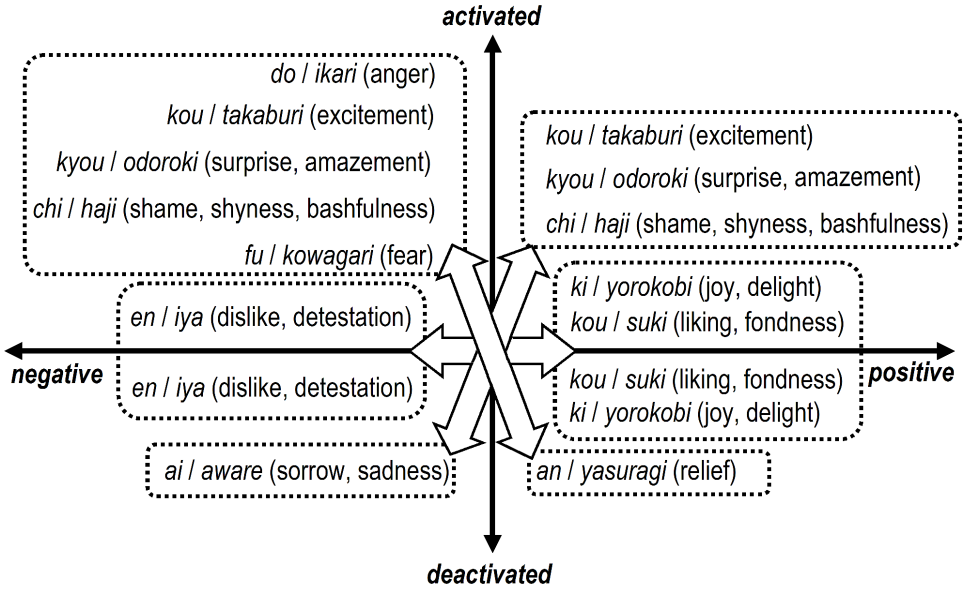


Figure 3.1: Grouping Nakamura's classification of emotions on Russell's space.

Description of CVS Procedure Analysis of Contextual Valence Shifters is a supplementary procedure for ML-Ask and works as follows. When a CVS structure is discovered, ML-Ask changes the valence polarity of the emotion conveyed in the sentence. Every emotion is placed in a suitable space according to Russell's model. The appropriate emotion is determined as belonging to the emotion space with both valence polarity and activation parameters opposite to those of the primary emotion (note arrows in Figure 3.1). If an emotion was located in only one quarter, e.g. positive-activated, the contrasting emotions would be determined as negative-deactivated. A difference in output is shown in Table 3.2 in examples (1) and (2). In the first example, originally ML-Ask selected [dislike]. This emotion is located in both quarters of the negative valence space. Therefore, after valence shifting, ML-Ask determines the new emotion types as positive and belonging to both of

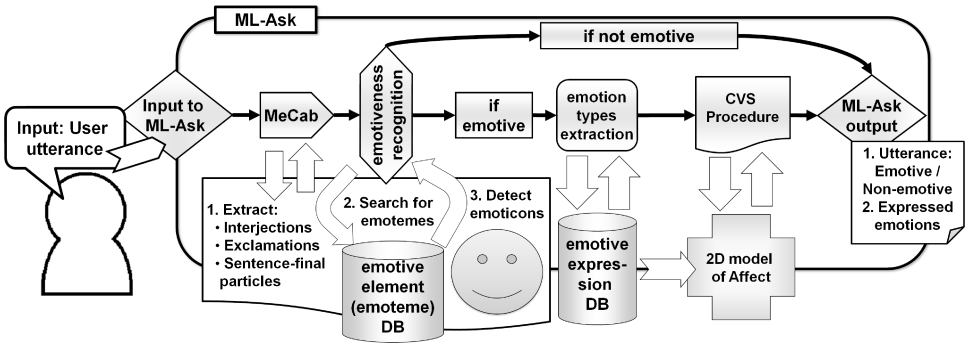


Figure 3.2: Flow chart of the ML-Ask system.

the positive quarters. The new proposed emotion types are: [joy] and [liking] belonging to both positive-activated and positive-deactivated quarters. The second example presents the opposite situation. The procedure, as described above, was shown to improve affect analysis in Japanese by Ptaszynski and colleagues [106]. The system flow chart including CVS procedure is shown in the Figure 3.2.

3.1.5 Information Provided in ML-Ask System Output

Analysis of utterances with ML-Ask provides several kinds of information, each useful in different tasks.

Firstly, ML-Ask determines whether an input utterance is emotive or not. This information is useful in analysis of emotional level of a speaker in either one utterance, or a set of utterances in a conversation.

Secondly, ML-Ask calculates an emotive value of an utterance, which is a sum of all emotemes appearing in the utterance. The emotive value can be further applied as another factor in determining emotional level of a speaker during a conversation.

Third information provided by the system is the emotive structure of an

utterance. This consists of a mapping of emotive layer on syntactic structure of the utterance, which is provided by MeCab [63]. The emotive layer consists of allocation of emotemes, emotive expressions and CVS structures on the syntactic structure of utterance. Examples are presented in Tables 3.1 and 3.2.

Another information is the emotion expressed in the utterance. Labelling of emotions is based on Nakamura's emotion classification (for details see section 2.1.1). This information is useful in general in traditional affect analysis to recognize the speaker's affective state.

Finally, mapping of emotion types on Russell's two-dimensional model of affect provides a classification of general features of the emotion types expressed in the utterance, namely valence (positive or negative) and activation (active or passive).

3.1.6 Evaluation Experiments

This section presents experiments performed to evaluate the system usability and verify whether the assumptions on which I built the system were correct. Firstly, since ML-Ask was built on linguistic approach to emotions I needed to confirm whether it provides enough linguistic analysis of utterances for the need of emotion research. This is verified in Experiment 1 performed on a training set. Secondly, since the system is meant to annotate natural language corpora, which usually consist of contents generated by laypeople, I needed to verify, whether the system's annotations are in agreement with laypeople. The experiments verifying this are presented in Experiments 2, 3 and 4, on three different test sets. Experiment 2 contains a detailed evaluation of the system on a collection of separate utterances. Experiment 3

verifies the system performance within a scope of one conversation taken from a natural conversation corpus. Experiment 4 presents evaluation of the system as an affect annotation tool for large-scale corpora with a set of online conversations used as the test set.

Experiment 1: Training Set Evaluation

Purpose The system had to be verified whether it provides enough linguistic analysis of emotive utterances for the need of research on emotions in language. If the verification was positive, the system would represents linguistic approach to affect annotation.

Design I manually gathered 214 emotive utterances to perform this experiment. The data was extracted from the research cited in sections 2.2.1 and 3.1.2 from which I gathered the emotive elements. ML-Ask was to determine the emotive structure of those utterances with a special focus on features described in the particular research (e.g. in a sentence borrowed from a research on exclamations, ML-Ask had to recognize at least the exclamations).

Results As a result, all of the sentences were recognized correctly, and in most of them ML-Ask could determine the emotive structure in a more detailed way, than it was described in the original research (see, e.g., example (4) in table 3.1). This ability of the system could help examine correlations between different emotive features. This could help finding emotive sentence patterns and therefore contribute to the research on pragmatics of emotive utterances, such as the one by performed by Beijer [9] of Potts et al. [98].

Experiment 2: Test Set (Separate Utterances)

After confirming the performance of the system on a training set I performed a set of experiments on several types of test sets. The first of the test set experiments was based on a collection textual utterances and was divided into two parts. In the first part (2a) I evaluated classification of sentences into emotive and non-emotive. In the second part (2b) I evaluated classification of particular emotion types on these sentences.

Experiment 2a: Emotiveness Classification

Purpose One of the planned applications for the ML-Ask system is annotation of natural language corpora. Such material usually consists of contents generated by laypeople (e.g., users participating in conversation). This experiment was conducted to verify, whether ML-Ask's annotations are in agreement with the laypeople perspective.

Design The experiment was based on a corpus of natural utterances gathered through an anonymous survey in which I asked 30 Japanese native speakers of different ages (19-35 years old) and social groups (students, businessmen, housewives) to imagine or remember a conversation with any person they know and write up to three sentences from that conversation: one free (optional), one emotive, and one non-emotive. The participants were also allowed to write more then one set of such sentences. This way I gathered 90 utterances: 10 written as free, 40 as emotive and 40 as non-emotive. From this collection I took only the utterances which were meant to be written as either emotive or non-emotive. Since laypeople are not capable to describe emotive structure of utterances, I checked whether ML-Ask could distinguish

between emotive and non-emotive utterances to verify how close to human thinking is the linguistically based discrimination between emotionally emphasized and neutral utterances. The participants also annotated specific emotion types conveyed in the emotive utterances written by themselves. This information was to be used in the experiment 2b.

Results As a result, ML-Ask annotated correctly 72 from the 80 utterances (90%). In 6 cases the system annotated the utterance wrongly as "non-emotive", in 2 cases it was the opposite. The kappa value was 0.8, which means that the system was in a very high agreement with human annotators. The results from the experiments 1 and 2 described above suggest that the system is capable of annotating corpora with detailed information on emotive structure of utterances and this annotation is reliable.

As mentioned on the begging of the section, I checked also the statistical significance of the difference between the linguistic and layperson approach with an assumption that if the difference is not significant, the linguistic material can be used in the creation of automatic affect annotation systems for other languages. Here, the layperson approach is represented by the set of 80 utterances and their classification into emotive or non-emotive. The linguistic approach is represented by the system's results in annotation of the 80 utterances, since the system was build on the base of this approach. The statistical significance of the difference between the two approaches was P value = .1586, which, by conventional criteria, means that the difference is considered to be not statistically significant. This means the differences that appeared might have been a matter of chance. Therefore it can be said that the examples from linguistic research can be efficiently used in the

creation of systems like the one presented here. This information is valuable for researchers performing similar research in other languages.

Experiment 2b: Emotion Type Determination

Purpose In this experiment I checked whether the system in its present form can be used in the task of affect recognition in textual input utterance.

Design The evaluation of affect recognition systems is usually performed by asking a third party if the system's results were correct [147, 145]. However, to perform the evaluation more objectively I assumed, agreeing with Solomon [131], that neither do people themselves understand their emotional states with perfect reliability, nor do other people perfectly perceive the affective states of their interlocutors (e.g., that is why people experience misunderstandings). Neither viewpoint can be thus used as a gold standard in affect recognition. An attention should be rather paid on a balance between the two viewpoints: speaker-specific and observer-specific. The collection of utterances used in experiment 2a is used here as a base for the experiment. The people asked to generate the utterances also annotated the specific emotions they expressed through the emotive utterances. The number emotion annotations per one sentence was not limited. One person could annotate more than one emotion type using his or her own vocabulary (this represents the speaker-specific viewpoint). These annotations were cross-referenced with Nakamura's dictionary to specify which of the 10 standard emotion types correspond to the expressions used by the annotators. In this evaluation I used all 90 utterances gathered in a way described in experiment 2a.

As mentioned above, in the evaluation process of affect recognition a view-

point of third-party observers is also important. Therefore I asked another 10 people (undergraduate students) to perform annotations of emotion types on the same collection (this represents the observer-specific viewpoint).

In the evaluation process I verified the system's performance for all emotion types with the type "non-emotive" included in the evaluation. The system is first evaluated on whether it can appropriately recognize the emotional states of authors of utterances (speaker-specific viewpoint). The result is calculated as an approximated F_1 score calculated separately for all emotion types, and "non-emotive".

This is compared to the system's result in evaluation based on the observer-specific viewpoint (the third-party annotations). A final score, balance of the recognition level between the speaker-specific and the observer-specific evaluation is calculated as a ratio of those two scores. The method is the more balanced the closer to 1 the ratio is.

As an additional result, I verified the general human level of recognition for the system. This is calculated as a ratio of scores in speaker-specific viewpoint evaluation reached by 1) the system and 2) the third-party (approximated for all annotators).

Results

Speaker-Specific Evaluation The conditions for the annotation to be perceived correct was correctly recognizing any and at least one emotion assigned by the author for the utterance, including non-emotive. However, people often misinterpret the specific types of emotions they experience, but sufficiently guess whether the emotion type they experience is positive or neg-

ative, and whether it is activated (aroused), or deactivated (passive). Therefore additionally, the result was considered as positive also when the extracted emotions were belonging to the same quarter of Russell's 2-dimensional space (see section 2.1.3), even if the extracted emotions were not exactly the same as the ones annotated by the authors of the utterance.

The system's result in estimating the specific types of emotions was a balanced F-score of 0.47 (P=.72, R=.35). As for human evaluators, the average result was 0.72 (P=.84, R=.64). Therefore the system's accuracy was approximately 65.3% (0.47/0.72) of the human level. In future research it is desirable to improve the algorithm for determining the specific emotion types and to perform the evaluation on a larger number of tagged utterances. I also calculated processing times to check whether the system is capable to operate in real time. The approximate time of processing one utterance in the basic procedure was 0.143 s, which is sufficient enough to annotate large corpora (e.g., a corpus of two thousand sentences is processed in less than five minutes).

Observer-Specific Evaluation As for the observer-specific evaluation, in many cases the results differed significantly between human annotators and there were sentences in which the annotators were not able to identify any emotions. With this in mind the following allowances were made. If ML-Ask extracted from a sentence at least one of the emotion types classified by annotators (see examples in Table 3.3) or the system's classification coincided with the majority (the condition applicable mostly when the majority of the annotators judged no emotion types), the result was positive.

In the observer-specific evaluation the basic method for emotion extrac-

tion reached 0.45 of the observer-specific human level. Three examples of successful outputs are shown in Table 3.3. The result is not ideal, and in the future I plan to improve the accuracy in determining the particular emotion types. The not ideal score of the system has two predictable reasons. Firstly, as the emotive expression database, ML-Ask uses Nakamura's dictionary [85]. Unfortunately, Nakamura stopped updating his dictionary in 1993 and the lexicon is out-of-date (only 2100 expressions). Secondly, instead of using straight forward emotive expressions, people would rather use ambiguous emotive utterances in which emphasis is based on the context. As implications for future work, the first problem could be solved by updating the lexicon, and the second one, by assigning potential emotional affiliation to emotemes (E.g. exclamation mark "!" is used to express anger or excitement rather than gloom or relief).

However, the system performance proved to be very balanced. The accuracy in estimating the specific types of emotions in speaker-specific evaluation reached a balanced F-score of 0.47. For the observer-specific evaluation the system's score was 0.45. The balance was 0.957 (0.45/0.47), which is close to 1. This confirms that the method, although still not perfect is well balanced.

Experiment 3: Test Set (Conversation Annotation)

Purpose ML-Ask is meant to perform annotations of conversations for the studies of emotions in language (see for example [98]). In such studies often two sets of utterances are compared to show which should be accounted as more emotive. I performed an evaluation experiment to verify whether the system is capable to perform annotation with a similar tendency to human evaluators.

Table 3.3:
Three examples of successful recognition of emotion types in the observer-specific evaluation of ML-Ask.

Utterance in Japanese / ***romanized reading*** **/ meaning in English**	**ML-Ask (basic procedure)**	**Human evaluators**
あなたの事が好きなんです。/ *Anata no koto ga suki nan desu.* / I love you.	好 [liking, fondness]	好 [liking, fondness] (7), 怖 [fear] (3), 喜 [joy] (2), 恥 [shame, shyness, bashfulness] (1), 昂 [excitement] (1)
この本さー、すげーやばかったよ。まじ怖すぎ。 / *Kono hon saa, sugee yabakatta yo. Maji kowa sugi.* / That book, ya know, 't was a killer. It was just too scary.	怖 [fear]	怖 [fear] (6), 驚 [surprise] (2), 昂 [excitement] (2), 喜 [joy] (1), 無 [nothing] (1)
うん、うまい、感激だ。/ *Un, umai, kangeki da.*/ Yeah, it's great, I'm impressed.	昂 [excitement]	喜 [joy] (10), 昂 [excitement] (5), 驚 [surprise] (1), 好 [liking, fondness] (1)

Design From a corpus of Japanese human-human dialogs [150], I chose two conversations, the first one being a record of a first meeting of two company workers and the second one being a small talk between two female high school students. The criterion for choosing these two conversations was that I wanted them to differ in terms of emotiveness. Thus, the dialog between company workers was assumed to be much less emotive than the one between schoolgirls. To verify this assumption, I performed the analysis of emotiveness of first twenty turns of each conversation.

Next, I prepared a questionnaire containing the two dialogs and a question: "Which dialog was generally more emotive?" The questionnaire was filled by 30 evaluators (university students). The results are summarized in Table 3.4.

Results As shown in Table 3.4, the results of this experiment showed that the conversation assessed as more emotive by 90.0% of evaluators was also containing more emotive utterances (85%) and its emotive value was higher

Table 3.4:
Results of the conversation annotation experiment compared to ML-Ask annotations of emotive utterances.

	Which more emotive?	Number of emotive utterances by ML-Ask	Emotive values (Ratio per sentence)
Dialogue 1	10.0%	6 (30%)	6 (0.3)
Dialogue 2	90.0%	17 (85%)	21 (1.05)

(21). As both dialogs were of the same length (20 turns each) and there were no other visible differences between them, it can be stated that the system properly annotated the emotiveness of conversations.

Experiment 4: Test Set (Corpus Annotation)

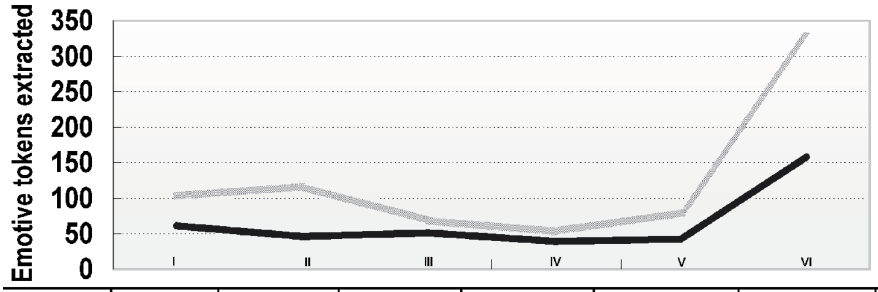
Purpose After confirming the system's usability I performed an annotation of a larger corpus of Internet forum discussions. By this experiment I aimed to verify whether the system is applicable in corpus linguistics and corpus statistics tasks within the research on emotions in language. The experiment was designed to verify several things. Firstly, to provide a quantitative proof for the thesis saying that frequent use of emotive utterances is an important part of conversation. Secondly, to verify whether the system can provide reliable statistics of emotive expressions and emotion types appearing in the corpus. Thirdly, to verify whether the system can provide a reliable frequency ranking of the expressed emotion types and show general emotive tendencies of a corpus. Except from the applicability in linguistic research, especially the two latter tasks are important in Security Informatics, in tasks like monitoring Internet forums for undesirable tendencies (see for example Abbasi, 2007 [1]).

Design: *2channel* Forum as a Corpus As the corpus I used a collection of discussions from a popular Japanese forum *2channel* [1], where every day millions of people discuss about current topics. The popularity of this forum was one of the reasons I decided to use it as a corpus base for my research. Although lately Blog contents gain on popularity as corpora for Computational Linguistic research, the contents of Blogs is mainly written by one person whereas ML-Ask was designed to perform better in conversation-like environment. Therefore an Internet forum was a reasonable choice. Especially on *2channel*, where the on-line communication often turns into a live and expressive chat-like conversations about present hot topics. Another reason to chose this corpus source was that *2channel*, although being a rich source of natural language, is usually disregarded in NLP research as difficult to process. This difficulty comes from the expressivity of its users, which in the case of ML-Ask case was in fact an advantage. Moreover, being able to process *2channel* contents efficiently one would obtain a useful tool for monitoring changes in Japanese society.

The special feature of this forum is the anonymity of speech. Utterances without signature appear on the board as uttered by *Nanashi-san* (Mr. Nameless). As Ptaszynski [99] argues, such restrictions forced users to create communication strategies compensating for the limitations of the medium. One kind of such strategy is a frequent usage of emotive expressions. Matsumura et al. [77] confirm this and argue that expressing even negative emotions helps keeping the discussion up. Providing a quantitative proof for the above statements was another goal of this experiment.

For the corpus base in the annotation experiment I chose a collection

[1] http://www.2ch.net/



Parts of corpus	I: 1-299 (299)	II: 300-535 (236)	III: 536-770 (235)	IV: 771-924 (154)	V: 925-1131 (207)	VI: 1132-1840 (709)	Summary
ML-Ask	61(20%)	46 (19%)	51 (22%)	39 (25%)	42 (20%)	158 (22%)	397 (22%)
Human annotator	103 (34%)	116 (49%)	68 (29%)	53 (34%)	79 (38%)	328 (46%)	747 (41%)
Accuracy ratio	59%	40%	75%	74%	53%	48%	58% P value=.0052

Figure 3.3:
ML-Ask results in calculating the number of emotive utterances containing emotive expression tokens. The table below contains detailed results.

of discussions that took place on the forum. The collection was officially published in Japan as an experimental novel under the title *Densha otoko* (Train man) [86], and contains 179,435 characters in 1,840 utterances divided into 6 parts. On this corpus Ptaszynski [99], with a help of a Japanese native speaker, annotated manually the emotive utterances. However, they focused only on the ones containing emotive expressions, since the ambiguously emotive utterances appeared too frequently. I annotated this corpus using ML-Ask and compared the results.

Annotation Results and Discussion

1. **Number of Emotive Utterances.** First, I used the system to calculate the number of emotive utterances in the corpus. Since the system proved its high reliability in identifying emotive utterances, I used its results as the correct ones. The result was 1506 emotive utterances

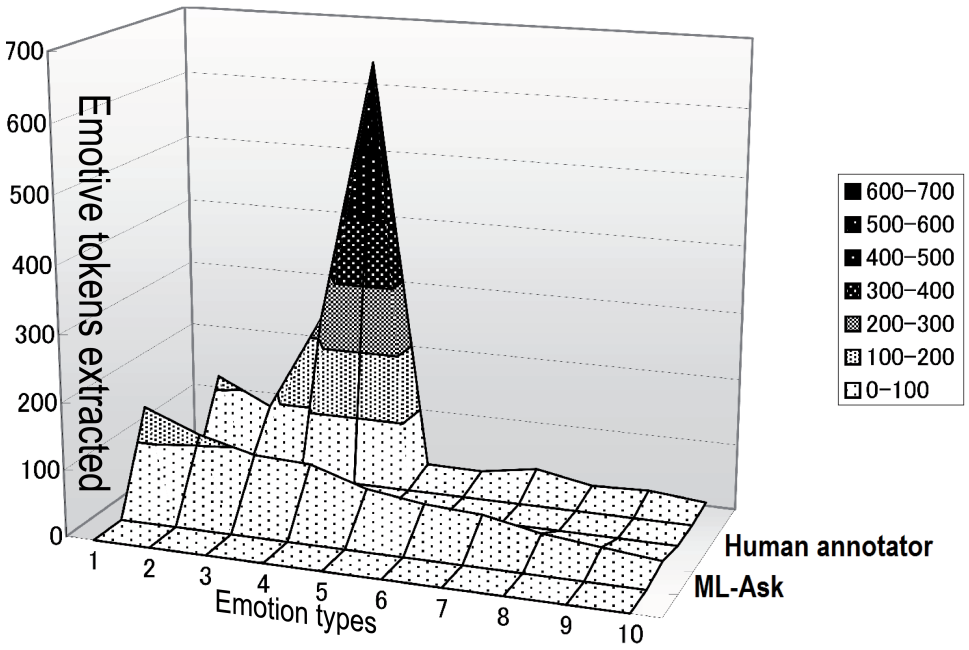


Figure 3.4:
ML-Ask results in extracting emotive expression tokens in comparison with human annotators. The graph is a visualization of the number of tokens per every emotion type (x-axis). The emotion types correspond to the order from the table 3.5

(81% of all corpus). It is reasonable, knowing the forum's reputation. This large number also confirms Ptaszynski's thesis that expressing emotions frequently on *2channel* has an important function of sustaining and supporting the discussion.

2. **Number of Emotive Utterances Containing Emotive Expressions.** Next, I calculated in how many of the emotive utterances the system found emotive expressions and specified emotion types, and compared the results to manual annotations. For the six different parts of the corpora the system annotated from 19% to 25% of the whole con-

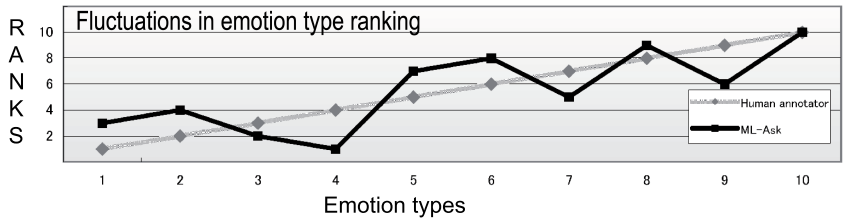


Figure 3.5:

ML-Ask results in calculating the rank setting fluctuations for emotion types when compared to the human annotators. The order of emotion types corresponds to the order from the table under this figure. The table below contains detailed results of fluctuations in emotion type ranking (right part) and the results of emotive token extraction (left part).

tents of corpora parts as emotive utterances. This corresponds to the general scope of 40% to 75% (Average of 58%) comparing to the human annotations, which was considered as accuracy ratio). The results were considered to be very statistically significant with P value = .0052 and, when contrasted, showed similar tendency - emotive expressions appeared in the largest number in the last part of the corpus (for details see Figure 3.3 and Table 3.5).

3. **Emotion Type Annotations.** The similarities appeared also in the emotion type annotations. Although the overall number of the emotive tokens extracted by the system and annotated manually differed, the detailed analysis revealed that the system had some problems only with determining about 'dislike' and 'excitement' (see Figure 3.4). The analysis of errors revealed that the expressions ML-Ask was not able to process, or in other words, that were not included in Nakamura's lexicon, were usually specific jargon used only on 2channel, consisting

Table 3.5:
Table represents the results of comparison of emotive expression token extraction and emotion type ranking assignment.

Comparison of emotive expression tokens extraction					Comparison of emotion types ranking		
ML-Ask		Human annotator		Absolute value of differences	Human annotator	ML-Ask	Fluctuation in ranking
1 joy	146	1 joy	118	28	1	3	3
2 fondness	113	2 fondness	78	35	2	4	1
3 dislike	92	3 dislike	229	137	3	2	2
4 excitement	88	4 excitement	633	545	4	1	2
5 relief	61	5 relief	13	48	5	7	2
6 gloom	49	6 gloom	12	37	6	8	3
7 fear	44	7 fear	26	18	7	5	2
8 surprize	27	8 surprise	10	17	8	9	2
9 anger	18	9 anger	13	5	9	6	1
10 shame	9	10 shame	5	4	10	10	0

of sophisticated ASCII-art-like pictures and emoticons made only with the use of punctuation marks. Since ML-Ask's database of emotemes contains also emotive marks, which include some punctuation marks, the system was able to determine that an utterance containing such an expression is emotive, but it could not specify the emotion type. This kind of jargon is difficult to process and there have not been constructed yet any system to deal completely with all 2channel-like emoticons and ASCII art. However, I have developed a system capable of processing one-line emoticons, incluging *2channel*-like emoticons. This system, CAO, is described in details in section 3.2.

Except of the two troublesome emotion types the system could extract other emotive tokens similarly to human annotator. There were 90% of agreements observed with the strength of agreement coefficient considered to be good (Kappa = .681) and very statistically significant (P value = .0035).

4. **Rank Setting Tendency.** Finally, I checked the tendency in providing the ranking of all emotion types appearing in the corpus. The ranking was based on the number of tokens extracted. The system's ranking was compared to the human annotations considered as the gold standard. The assigning of ranking places was scored from 10 points (a perfect hit, ranking place fluctuation = 0) to 0 (ranking place for an emotion type not assigned at all, fluctuation = 10). A maximum number of points to be gathered this way was 100 (all ranking places perfectly assigned), minimum was 0 points (none of the emotion type specified/assigned). ML-Ask's fluctuations were from 0 (for 'shame') to 3 (for 'joy' and 'gloom') (see Figure 3.5). With such results ML-Ask acquired a high score of 82 points. The difference in ranking places fluctuations and therefore the score in emotion type ranking place assessment was considered extremely statistically significant by conventional criteria (P value = .0002). To confirm the rank fluctuation test I calculated correlation between the rank setting by humans and ML-Ask using Spearman's rank correlation coefficient (Spearman's ρ [rho]). There was a middle correlation for all emotion types (ρ=0.4145), but after excluding the two troublesome emotion types (dislike and excitement), the system reached almost ideal correlation with human annotators (ρ=0.9487), which confirmed the rank fluctuation test. The statistical significance test calculated for the correlation test results was extremely high (P value = 0.00032). The general view on the ranking results revealed that they are divided into two groups: ranked more similarly (emotion types in bold type font in the Table 3.5 and the majority of tokens) and less.

3.2 CAO: A System for Analysis of Emoticons

One of the primary functions of the Internet is to connect people online. The first developed online communication media, such as e-mail or BBS forums, were based on text messages. Although later improvement and popularization of Internet connection allowed for phone calls or video conferences, the text-based message did not lose its popularity. However, its sensory limitations in communication channels (no view or sound of the interlocutors) prompted users to develop communication strategies compensating for these limitations. One such strategy is the use of emoticons, strings of symbols imitating body language (faces or gestures). Today, the use of emoticons in online conversation contributes to the facilitation of the online communication process in e-mails, BBS, instant messaging applications, or blogs [137, 25, 15]. Obtaining a sufficient level of computation for this kind of communication would improve machine understanding of language used online, and contribute to the creation of more natural human-machine interfaces. Therefore, analysis of emoticons is of great importance in such fields as Human-Computer Interaction (HCI), Computational Linguistics (CL) or Artificial Intelligence (AI).

Emoticons are virtual representations of body language and their main function is similar, namely to convey information about the speaker's emotional state. Therefore the analysis of emoticons appearing in online communication can be considered as a task for affect analysis, a sub-field of AI, focusing on classifying users' emotional expressions (e.g. anger, excitement, joy, etc.). There have been several attempts to analyze emotive information

conveyed by emoticons. For example, Tanaka et al. [140] used kernel methods for extraction and classification of emoticons, Yamada et al. [161] used statistics of n-grams, and Kawakami [58] gathered and thoroughly analyzed a database of 31 emoticons. However, all of these methods struggle with numerous problems, such as the lack of ability to precisely extract an emoticon from a sentence, incoherent emotion classification, manual and inconsistent emoticon sample annotation, inability to divide emoticons into semantic areas, or small sample base resulting in high vulnerability to user creativity in generating new emoticons.

This section presents a system dealing with all of those problems. The system extracts emoticons from input and classifies them automatically, taking into consideration semantic areas (representations of mouth, eyes, etc.). It is based on a large database collected from the Internet and improved automatically to a coverage exceeding 3 million possibilities. The performance of the system is thoroughly verified with a training set and a test set based on a corpus of 350 million sentences in Japanese.

3.2.1 Previous Research on Emoticons

Research on emoticons has developed in three general directions. Firstly, research in the fields of social sciences and communication studies have investigated the effects of emoticons on social interaction. There are several examples worth mentioning. The research of Ip [49] investigates the impact of emoticons on affect interpretation in Instant Messaging. She concludes that the use of emoticons helps the interlocutors in conveying their emotions during the online conversation. Wolf [159] showed further, in her study on newsgroups, that there are significant differences in the use of emoticons by

men and women. Derks et al. [25] investigated the influence of social context on the use of emoticons in Internet communication. Finally, Maness [75] performed linguistic analysis of chat conversations between college students, showing that the use of emoticons is an important means of communication in everyday online conversations. The above research is important in its investigation of the pragmatics of emoticons concerned as expressions of the language used online. However, most of such research focuses on Western-type emoticons.

Two practical applications of emoticon research in the field of Artificial Intelligence are to generate and analyze emoticons in online conversations in order to improve computer-related text-based communication, in fields such as Computer-Mediated Communication (CMC) or Human-Computer Interaction (HCI).

One of the first significant attempts to the first problem, emoticon generation, was done by Nakamura and colleagues [84]. They used a Neural Networks-based algorithm to learn a set of emoticon areas (mouths, faces, etc.) and used them later in a dialog agent. Unfortunately, the lack of a firm formalization of the semantic areas made the choice of emoticons eventually random, and the final performance far from ideal. This was one of the reasons for abandoning the idea of exploiting parts of emoticons as base elements for emoticon-related systems. From that time most of the research on emoticon generation focused mostly on preprogrammed emoticons [137, 136, 139]. In my research I revived the idea of exploiting the emoticon areas, although not in the research on emoticon generation, but in emoticon extraction and analysis.

There have been several attempts to analyze emoticons or use them in

affect analysis of sentences. For example, Reed [113] showed that the use of preprogrammed emoticons can be useful in sentiment classification. Yang et al. [163] made an attempt to automatically build a lexicon of emotional expressions using preprogrammed emoticons as seeds. However, both of the above research focus only on preprogrammed Western-type emoticons, which are simple in structure. In my research I focused on more challenging Eastern-type emoticons (for the description of types of emoticons, see the definition of emoticon below in section 3.2.2).

There have been three significant attempts to analyze Eastern emoticons. Tanaka et al. [140] used kernel methods for extraction and classification of emoticons. However, their extraction was incomplete and the classification of emotions incoherent and eventually set manually. Yamada et al. [161] used statistics of n-grams. Unfortunately, their method was unable to extract emoticons from sentences. Moreover, as they based their method on simple occurrence statistics of all characters in emoticons, they struggled with errors, as some characters were calculated as "eyes", although they represented "mouths", etc. Finally, Kawakami [58] gathered and thoroughly analyzed a database of 31 emoticons. Unfortunately, his analysis was done manually. Moreover, the small number of samples made his research inapplicable in affect analysis of the large numbers of original emoticons appearing on the Internet. All of the previous systems strictly depend on their primary emoticon databases and therefore are highly vulnerable to user creativity in generating new emoticons.

In my research I dealt with all of the above problems. The system I developed is capable of extraction of emoticons from input and fully automatic affect analysis based on a coherent emotion classification. It also takes into

Table 3.6:
Previous research on emoticon analysis with comparison to CAO.

Research → **(approach)** **Capability ↓**	**Tanaka et al. (2005) (kernel methods)**	**Yamada et al. (2007) (n-grams)**	**Kawakami (2008) (database)**	**CAO (theory of kinesics)**
1. Detection whether input equals emoticon	X	X	X	O
2. Detection of emoticon in sentence input	O (included in 3.)	X	X	O
3. Extraction of emoticon from any string of characters	O	X	X	O
4. Division into semantic areas	X	X	X	O
5. Database coverage	1,075	693	31	10,137 (expanded automatically to over 3 million)
6. Classification of emotion types	6 types (Amateur emoticon dictionary-based)	7 types (Subjective)	6 types (Subjective)	10 types (Language/ Culture Based)
7. Emotion estimation of separate emoticons	O	O	O	O
8. Affect Analysis of sentences with emoticons	△ (Possible, but not evaluated)	X	X	O

consideration semantic areas (representations of mouth, eyes, etc.). The system is based on a large emoticon database collected from the Internet and enlarged automatically, providing coverage of over 3 million possibilities. The system is thoroughly evaluated with a training set (the database) and a test set (a corpus of over 350 million sentences in Japanese). I summarized all of the previous research with comparison to my system in Table 3.6.

3.2.2 Definition of Emoticon

Emoticons have been used in online communication for many years and their numbers have developed depending on the language of use, letter input sys-

tem, the kind of community they are used in, etc. However, they can be roughly divided into three types: a) Western one-line type; b) Eastern one-line type; and c) Multi-line ASCII art type.

Western emoticons are characteristic as being rotated by 90 degrees, such as "`:-)`" (smiling face), or "`:-D`" (laughing face). They are the simplest of the three as they are usually made of two to four characters and are of a relatively small number. Therefore, I excluded them from my research as not being challenging enough to be a part of language processing. Moreover, my research focuses on the use of emoticons by Japanese users, and this type of emoticons is rarely used in Japanese online communities. However, as the Western-type emoticons can be gathered in a list of about fifty, such a list could be simply added to the system at the end in a sub-procedure.

Multi-line ASCII art type emoticons, on the other hand, consist of a number of characters written in several, or even up to several dozens of lines, which, when looked at from a distance, make up a picture, often representing a face or several faces. Their multi-line structure leads their analysis to be considered more as a task for image processing than language processing, as this would be the only way for the computer to obtain an impression of the emoticon from a point of view similar to a user looking at the computer screen. Because of the above, I do not include multi-line ASCII art emoticons in my research.

Finally, Eastern emoticons, in contrast to the Western ones are usually unrotated and present faces, gestures or postures from a point of view easily comprehensible to the reader. Some examples are: "`(^o^)`" (laughing face), "`(^_^)`" (smiling face), "`(ToT)`" (crying face). They arose in Japan, where they are called *kaomoji*, in the 1980s and since then have been developed

Table 3.7:
Examples of emoticon division into sets of semantic areas: [M] - mouth, [E_L], [E_R] - eyes, [B_1], [B_2] - emoticon borders, [S_1] - [S_4] - additional areas.

No. of sets	Emoticon	S_1	B_1	S_2	E_L	M	E_R	S_3	B_2	S_4	...
1	ヾ(。·ω·)ﾉ	ヾ	(	。	·	ω	·	N/A	)	ﾉ	
1	(——;)	N/A	(	N/A	—	N/A	—	;	)	N/A	
				SET 01							SET 02
2	(^^)人(^^)	N/A	(	N/A	^	N/A	^	N/A	)	人	(^^)
2	☆- (●≧ ▿)人(▿ ≦ ●)- ☆	☆-	(	●	≧	▿	N/A	N/A	)	人	(▿ ≦ ●)- ☆
				SET 01		SET 02		SET 03		SET 04	
4	(∇ ´○)皿´★)ω´☆)∀ ´●)			(∇ ´○)		皿´★)		ω´☆)		∀ ´●)	

in a number of online communities. They are made up of three to over twenty characters written in one line and consist of a representation of at least one face or posture, up to a number of different face-marks. In the research described in this chapter I focused mainly on this type of emoticon, as they have a large variation of appearance and are sophisticated enough to express different meanings. See Table 3.7 for some examples of this type of emoticons.

Emoticons defined as above can be considered as representations of body language in text-based conversation, where the communication channel is limited to the transmission of letters and punctuation marks. Therefore I based my approach to analysis of emoticons on assumptions similar to those from research on body language. In particular I applies the theory of kinesics to define semantic areas as separate kinemes, and then automatically assign to them emotional affiliations.

3.2.3 Theory of Kinesics

The word *kinesics*, as defined by Vargas [152], refers to all non-verbal behavior related to movement, such as postures, gestures and facial expressions and functions as a term for body language in current anthropology. It is studied as an important component of nonverbal communication together with *paralanguage* (e.g. voice modulation) and *proxemics* (e.g. social distance). The term was first used by Birdwhistell [10, 11], who founded the theory of kinesics. The theory assumes that non-verbal behavior is used in everyday communication systematically and can be studied in a similar way to language. A minimal part distinguished in kinesics is a *kineme* - the smallest meaningful set of body movements, e.g., raising eyebrows, or moving the eyes upward. Birdwhistell developed a complex system of *kinegraphs* to annotate kinemes for the research on body language. Some examples of kinemes are given in Figure 3.6.

Emoticons from the Viewpoint of Kinesics

One of the current applications of kinesics is in annotation of affect display in psychology to determine which emotion is represented by which body movement or facial expression. Emoticons are representations of body language in online text-based communication. This suggests that the reasoning applied in kinesics is applicable to emoticons as well.

Therefore, for the purposes of this research I specified the definition of "emoticon" as a one-line string of symbols containing at least one set of semantic areas, which I classify as: "mouth" [M], "eyes" [E_L], [E_R], "emoticon borders" [B_1], [B_2], and "additional areas" [S_1] - [S_4] placed between

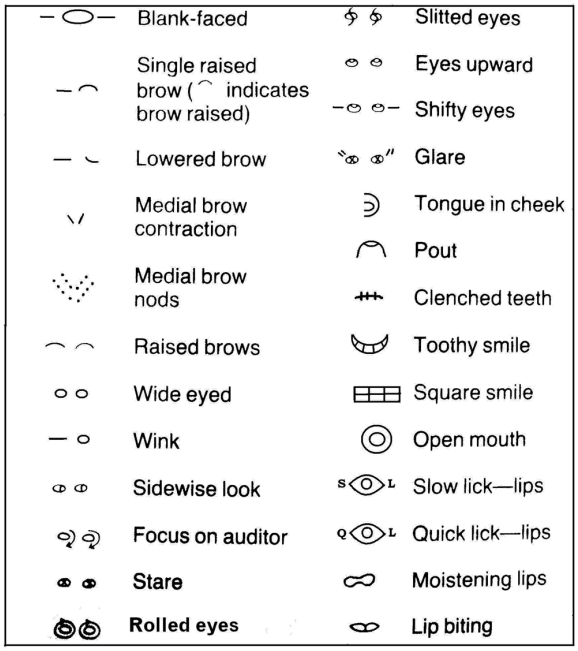


Figure 3.6:
Some examples of kinegraphs used by Birdwhistell to annotate body language.

the above. Each area can include any number of characters. We also allowed part of the set to be of empty value, which means that the system can analyze an emoticon precisely even if some of the areas are absent. The minimal emoticon set considered in this research contains of two eyes (a set represented as "E_L,E_R", e.g. "(^^)" (a happy face)), mouth and an eye ("E_L,M" or "M,E_R", e.g. "(^o)" (a laughing face) and "(_^)" (a smiling face) respectively), or mouth/eye with one element of the additional areas ("M/E_R,S_3/S_4" or "S_1/S_2,E_L/M", e.g. "(^)/∼" (a happy face) and "\(`)" (a sad face) respectively). However, many emoticons contain all or most of the areas, as in the following example showing a crying face

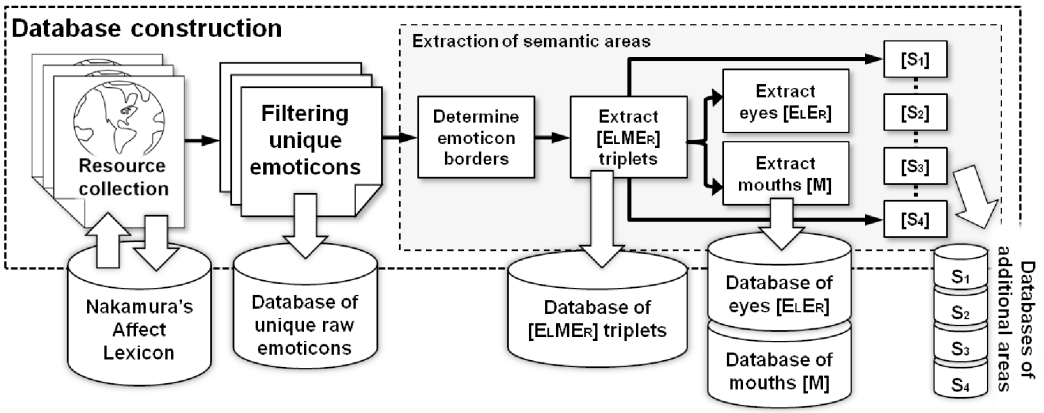


Figure 3.7: Flow chart of the database construction.

”·°·(/Д`;)·°·”. See Table 3.7 for some examples of emoticons and their semantic areas. The analysis of emotive information conveyed in emoticons can therefore be based on annotations of the particular semantic areas grouped in an automatically constructed emoticon database.

3.2.4 Database of Emoticons

To create a system for emoticon analysis I first needed a coherent database of emoticons classified according to the emotions they represent. The database development was performed in several steps. Firstly, raw emoticon samples were collected from the Internet. Then, the naming of emotion classes expressed by the emoticons was unified according to Nakamura's [85] classification of emotions. Next, the idea of kinemes was applied in order to divide the extracted emoticons into semantic areas. Finally, the emotive affiliations of the semantic areas were determined by calculating their occurrences in the database. The flow of the procedure of database generation is represented on Figure 3.7.

Resource Collection

The raw emoticons were extracted from seven online emoticon dictionaries available on seven popular Web pages dedicated to emoticons: Facemark Party, Kaomojiya, Kaomoji-toshokan, Kaomoji-cafe, Kaomoji Paradise, Kaomojisyo and Kaomoji Station[2]. The dictionaries are easily accessible from the Internet.

Database Naming Unification

The data in each dictionary is divided into numerous categories, such as "greetings", "affirmations", "actions", "hobbies", "expressing emotions", etc. However, the number of categories and their nomenclature is not unified. To unify them I used ML-Ask system for affect analysis, described above 3.1. One of the procedures in this system is to classify words according to the emotion type they express, based on Nakamura's emotion classification. Categories with names suggesting emotional content were selected and emoticons from those categories were extracted, giving a total of 11,416 emoticons. However, as some of them could appear in more than one collection, I performed filtering to extract only the unique ones. The number of unique emoticons after the filtering was 10,137 (89%). Most of the emoticons appearing in all seven collections were unique. Only for the emoticons annotated as expressions of "joy", a large amount, over one third, was repeated. This means that all of the dictionaries from which the emoticons were extracted provided emoticons that did not appear in other collections. On the

[2]Respectively: http://www.facemark.jp/facemark.htm, http://kaomojiya.com/, http://www.kaomoji.com/kao/text/, http://kaomoji-cafe.jp/, http://rsmz.net/kaopara/, http://matsucon.net/material/dic/, http://kaosute.net/jisyo/kanjou.shtml

Table 3.8:
Ratio of unique emoticons to all extracted emoticons and their distribution in the database according to emotion types.

Emotion types	All extracted emoticons	Unique emoticons	Ratio (Unique/All)
joy, delight	3,128	1,972	63%
liking, fondness	1,988	1,972	99%
anger	1,238	1,221	99%
surprise, amazement	1,227	1,196	97%
sadness, gloom	1,203	1,169	97%
excitement	1,124	1,120	99%
dislike	704	698	99%
shame, shyness	526	511	97%
fear	179	179	100%
relief	99	99	100%
Overall	**11,416**	**10,137**	**89%**

other hand, high repeating frequency of emoticons annotated as expressions of "joy" suggest that this emotion type is expressed by Internet users with a certain number of popular emoticons. The emotion types for which the number of extracted emoticons was the highest were in order: joy, fondness, anger, surprise, gloom, and excitement. This suggests that Internet users express these emotion types more often than the rest, which were in order: dislike, shame/bashfulness, fear and relief. The ratio of unique emoticons to all extracted ones and their distribution across the emotion types is shown in Table 3.8.

Extraction of Semantic Areas

After gathering the database of raw emoticons and classifying them according to emotion types, I performed an extraction of all semantic areas appearing

in unique emoticons. The extraction was done in agreement with the definition of emoticons and according to the following procedure. Firstly, possible emoticon borders are defined and all unique eye-mouth-eye triplets are extracted together ($E_L ME_R$). From those triplets I extracted mouths (M) and pairs of eyes (E_L,E_R). The rule for extracting eye-patterns from triplets goes as follows. If the eyes consist of multiple characters, each eye has the same pattern. If the eyes consist only of one character, they can be the same or different (this was always true among the 10,137 emoticons in the database). Finally, having extracted the $E_L ME_R$ triplets and defined the emoticon borders I extracted all existing additional areas (S_1,...,S_4). See Figure 3.8 for the details of this process.

Emotion Annotation of Semantic Areas

Having divided the emoticons into semantic areas, occurrence frequency of the areas in the emotion type database was calculated for every triplet, eyes, mouth and the additional areas. All unique areas were summarized in order of occurrence within the database for each emotion type. Each area's occurrence rate is considered as the probability of which emotion it tends to express.

Database Statistics

The number of unique combined areas of $E_L ME_R$ triplets was 6,185. The number of unique eyes (E_L,E_R) was 1,920. The number of unique mouth areas (M) was 1,654. The number of unique additional areas was respectively S_1= 5,169, S_2=2,986, S_3=3,192, S_4=8,837 (Overall 20,184). The distribution of all types of area elements across the whole database is shown in Table 3.9.

```
------------------------------------------------------------
1.Input:  (.*) - STRING OF CHARACTERS;
2.Determine emoticon borders:B1{NULL,|,(,<,[,...},B2{NULL,],>,),
|,...}:B1(.*)B2;
3.Localize ELMER triplet in the potential emoticon:
B1(.*)ELMER(.*)B2;
4.Separate eyes EL,ER and mouth M areas{
5.  from ELMER take n characters from the left nL and right
nR;
6.  if nL=nR, nL is EL and nR=ER; if no match take n-1
characters;
7.  if the above fails, take one character from the left as EL
and from the right as ER;
8.  mouth area M is what is left between EL and ER;}
9.Determine additional areas S1,...,S4 according to the regular
expression:  m/[S1?][B1?][S2?][ELMER][S3?][B2?][S4?]/;
10.  if no match, replace [ELMER] gradually with:[ELER],[MER],
[ELM];
11.Calculate the number of occurrences separately for triplet
ELMER, eye pair EL,ER, mouth M, and additional areas S1,...,S4,
for all emotion types;
------------------------------------------------------------
```

Figure 3.8: The flow of the procedure for semantic area extraction.

Database Coverage

In previous research on emoticon classification one of the most popular approaches was the assumption that every emoticon is a separate entity, and therefore is not divisible into separate areas or characters [58]. However, this approach is strongly dependent on the number of emoticons in the database and is heavily vulnerable to user creativity in generating new emoticons. I aimed in developing an approach as much immune to user creativity as possible. To verify that, I estimated the coverage of the raw emoticon database in

Table 3.9:
Distribution of all types of unique areas for which occurrence statistics was calculated across all emotion types in the database.

Area type	$E_L ME_R$	S_1	B_1	S_2	E_L,E_R	M	S_3	B_2	S_4
joy, delight	1298	1469	–	653	349	336	671	–	2449
anger	741	525	–	321	188	239	330	–	1014
sadness, gloom	702	350	–	303	291	170	358	–	730
fear	124	72	–	67	52	62	74	–	133
shame, shyness	315	169	–	121	110	85	123	–	343
liking, fondness	1079	1092	–	802	305	239	805	–	1633
dislike	527	337	–	209	161	179	201	–	562
excitement	670	700	–	268	243	164	324	–	1049
relief	81	50	–	11	38	26	27	–	64
surprise	648	405	–	231	183	154	279	–	860
Overall	6185	5169	–	2986	1920	1654	3192	–	8837

comparison to the database of all semantic areas separately. The number of all possible combinations of triplets calculated as E_L,E_R×M, even excluding the additional areas, is equal to 3,175,680 (over three million combinations[3]). Therefore the basic coverage of the raw emoticon database, which contains a somewhat large number of 10,137 unique samples, does not exceed 0.32% of the whole coverage of this method. This means that a method based only on a raw emoticon database would lose 99.68% of possible coverage, which in my approach is retained.

3.2.5 CAO - Emoticon Analaysis System

The databases of emoticons and their semantic areas described above were applied in CAO - *a system for emotiCon Analysis and decOding of affective information*. The system performs three main procedures. Firstly, it detects

[3]However, including the additional areas in the calculation gives an overall number of possibilities equal to at least $1.382613544823877 \times 10^{21}$

whether input contains any emoticons. Secondly, if emoticons were detected, the system extracts all emoticons from the input. Thirdly, the system estimates the expressed emotions by matching the extracted emoticon in stages until it finds a match in the databases of:

1. Raw emoticons;
2. $E_L ME_R$ triplets and additional areas S_1,...,S_4;
3. Separately for the eyes E_L,E_R, mouth M and the additional areas.

Emoticon Detection in Input

The first procedure after obtaining input is responsible for detecting the presence of emoticons. It is determined when at least three symbols usually used in emoticons appear in a row. A set of 455 symbols was statistically selected as symbols appearing most frequently in emoticons.

Emoticon Extraction from Input

In the emoticon extraction procedure the system extracts all emoticons from input. This is done in stages, looking for a match with: 1) the raw emoticon database; in case of no match, 2) any $E_L ME_R$ triplet from the triplet database. If a triplet is found the system matches the rest of the elements of the regular expression: `m/[`S_1`?][`B_1`?][`S_2`?][`$E_L ME_R$`][`S_3`?][`B_2`?][`S_4`?]/`, with the use of all databases of additional areas and emoticon borders; 3) in case the triplet match was not found, the system searches for: 3a) any triplet match from all 3 million $E_L ME_R$ combinations with one of the four possible mouth patterns matched gradually (`[`$E_L ME_R$`], [`$E_L E_R$`], [`ME_R`], [`$E_L M$`]`); or as a last resort 3b) a match for any of all the areas separately. The flow of

```
------------------------------------------------------------
1.Input:  "SOME CHARACTERS ·°·(/Д`;)·°·SOME CHARACTERS"
2.Find match in raw emoticon database:  ·°·(/Д`;)·°·
3.  If no match, localize E_LME_R triplet in the E_LME_R triplet
database:  /Д`
4.  If no triplet found, look for any E_LME_R combination;
5.  If no combination matched, find any E_LE_R or M from separate
semantic area database:  /`, Д
6.Localize emoticon borders B_1,B_2 :  (,)
7.Localize additional areas S_1,S_2,S_3,S_4:  ·°·,;,·°·
8.Determine the emoticon structure:  S_1:  ·°·, B_1:(, S_2:N/A,
E_LE_R:  /`, M: Д, S_3:;, B_2:), S_4:  ·°·
9.Look for next emoticon;
------------------------------------------------------------
```

Figure 3.9: The flow of the procedure for emoticon extraction.

this procedure is represented in Figure 3.9. Although the extraction procedure could function also as a detection procedure, it is more time consuming. The differences in processing time are not noticeable when the number of consecutive inputs is small. However, I plan to use CAO to annotate large corpora including over several million entries. With this code improvement the system skips sentences with no potential emoticons, which shortens the processing time.

Affect Analysis Procedure

In the affect analysis procedure, the system estimates which emotion types are the most probable for an emoticon to express. This is done by matching the recognized emoticon to the emotions annotated on the database elements and checking their occurrence statistics. This procedure is performed as an extension to the extraction procedure. The system first checks which

```
------------------------------------------------------------
1.Input; (e.g.:  ･ﾟ･(/Д`;)･ﾟ･)
2.Determine emotion types according to raw emoticon database;
(･ﾟ･(/Д`;)･ﾟ･ :  sorrow/sadness(3), excitement(2))
3.  If no match, determine emotion types for E_LME_R triplet;
(/Д`:excitement(14),anger(2),sorrow(1),fear(1),joy(1),fondness(1))
4.  If no emotion types for triplet found, find emotion types
for separate semantic areas E_LE_R and M;
(/`:sorrow(3),shame(3),joy(2),fondness(2),fear(1),excitement(1),...)
(Д:sorrow(53),excitement(52),anger(42),surprise(37),joy(28),...)
5.  Determine emotion types for additional areas;
(･ﾟ･:..., ;:..., ･ﾟ･:...)
6.  Proceed to next emoticon in the character string;
7.If no more emoticons, summarize scores;
------------------------------------------------------------
```

Figure 3.10: The flow of the procedure for affect analysis of emoticon.

emotion types were annotated on raw emoticons. If no emotion was found, it looks for a match with emotion annotations for E_LME_R triplet. If no match was found, the semantic area databases for eyes E_LE_R and mouth M are considered separately and the matching emotion types are extracted. Finally, emotion type annotations for additional areas are determined. The flow of this procedure is shown with an example in Figure 3.10. The flow chart of the whole system is presented in Figure 3.11.

Output Calculation

After extracting the emotion annotations of emoticons and/or semantic areas, the final emotion ranking output is calculated. In the process of evaluation I calculated the score in five different ways to specify the most effective method of result calculation.

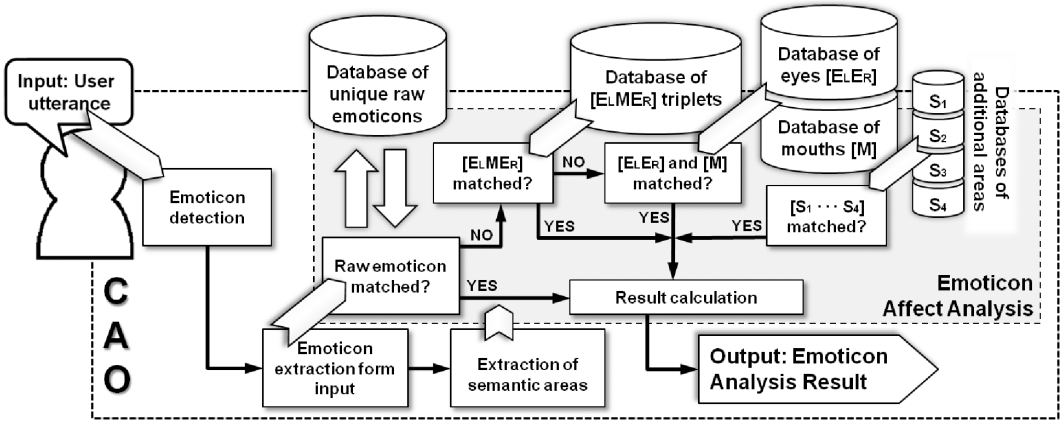


Figure 3.11: Flow chart of the CAO system.

Occurrence The processing of one emoticon provides a set of lists - one for each emoticon part (mouth, eyes, additional areas, etc.). Any part of emoticon may appear in databases belonging to different emotion types (e.g., in the crying emoticon ·°·(/Д`;)·°·, element representing "mouth" - Д appears 53 times in sorrow database, 52 times in excitement, 28 times in joy, etc., (see Figure 3.10 for details). Each of those lists contain emotion types with assigned numbers of occurrences for the element in the database of each emotion type. Having these lists it is possible to perform different calculations to summarize/generalize the scores. Firstly, all results can be added and then the emotion type appearing most often will be the most probable for the emoticon to express. In other words, occurrence is the straightforward number of occurrences of an element (emoticon/triplet/semantic area). The higher occurrence of an element in the emotion type database, the higher it scored. For more elements, the final score for an emotion type was calculated as the sum of all occurrence scores for all emotion types. The final emotion scores were placed in descending order of the final sums of their occurrences.

Frequency However, it might be said that to simply add the numbers is not a fair way of score summarization, since there is a different number of elements in each database (see Table 3.9), and a database with a small number of elements will have a tendency to lose. To avoid this bias I divided the emotion score by the number of all elements in the database. Therefore frequency is calculated as the occurrence number of a matched element (emoticon or semantic area) divided by the number of all elements in the particular emotion type database. The higher the frequency rate for a matched element in the emotion type database, the higher it scored. For more elements, the final score for an emotion type was calculated as the sum of all frequency scores of the matched elements for an emotion type. The final scores for each emotion type were placed in descending order of the final sums of their frequencies.

Unique Frequency It could be further said, that a simple division by the number of all elements is also not ideally fair, since there are elements appearing more often and therefore are stronger, which will also cause a bias in the results. To avoid this I also divided the occurrences by the number of all unique elements. Unique frequency is thus calculated similarly to the usual frequency. The difference is that the denominator (division basis) is not the number of all elements in the particular emotion type database, but the number of all unique ones.

Position Position is calculated in the following way. The strings of characters in all databases (raw emoticons, triplets, semantic areas) are sorted by their occurrence in descending order. By position, I mean the place of the matched string in the database. Position is determined by a number of

strings, occurrence of which was greater than the occurrence of a given string. For example, in a set of strings with the following occurrences: n_1=5, n_2=5, n_3=5, n_4=3, n_5=3, n_6=2, n_7=2, n_8=1, the strings n_6 and n_7 will be in sixth position. If the string was not matched in a given database, it is assigned a position of the last plus one element from this database.

Unique Position Unique Position is calculated in a similar way to the normal Position, with one difference. Since some strings in the databases have the same number of occurrences, they could be considered as appearing in the same position. Therefore, here I considered the strings with the same occurrences as the ones with the same position. For example, in a set of strings with the following occurrences: n_1=5, n_2=5, n_3=5, n_4=3, n_5=3, n_6=2, n_7=2, n_8=1, the strings n6 and n7 will be in third position. If the string was not matched in a given database it is assigned a position of the last plus one element from this database.

Two-dimensional Model of Affect Applied in CAO

I also checked whether the general features of the extracted emotion types were in agreement. By "general features", I mean those proposed by Russell in his theory of a two-dimensional model of affect [114]. For some emotion types the affiliation to a general feature-group is somewhat obvious, e.g. gloom is never positive or activated. However, for other emotion types the emotion affiliation is not that obvious, e.g., surprise can be both positive as well as negative; dislike can be either activated or deactivated, etc. The emotion types with uncertain affiliation were mapped on all groups they could belong to, according to the explanation in section 2.1.3. These groups

are then used for estimating whether the emotion types extracted by CAO belong to the same quarter. For the details of the mapping of the emotion types, see section 2.1.3 and Figure 2.1.

3.2.6 Evaluation of CAO

To fully verify the system's performance I carried out an exhaustive evaluation. The system was evaluated using a training set and a test set. The evaluated areas were: emoticon detection in a sentence, emoticon extraction from input, division of emoticons into semantic areas, and emotion classification of emoticons.

Training Set Evaluation

The training set for the evaluation included all 10,137 unique emoticons from the raw emoticon database. However, to avoid perfect matching with the database (and therefore scoring 100% accuracy) I made the system skip the first step, matching the raw emoticon database - and continue with further procedures (matching triplets and separate semantic areas).

The system's score was calculated as follows. If the system annotated an emoticon taken from a specific emotion type database with the name of the database as the highest one on the list of all annotated emotions, it counted as 1 point. Therefore, if the system annotated 5 emotion types on an emoticon taken from the "joy" database and the "joy" annotation appeared as the first one on the list of 5, the system's score was 5/5 (1 point). If the name of the emotion database from which the emoticon was taken did not appear in the first place, the score was calculated as the rank number the emotion achieved

divided by the number of all emotions annotated. Therefore, if the system annotated 5 emotion types on an emoticon taken from the "joy" database and the "joy" annotation appeared as the second one on the list of 5, the system's score was 4/5 (0.8 point), and so on. These calculations were further performed for all five ways of score calculation.

Test Set Evaluation

In the test set evaluation I used Yacis Blog Corpus.

Yacis Blog Corpus Yacis Blog Corpus is an unannotated corpus consisting of 354,288,529 Japanese sentences. Average sentence length is 28.17 Japanese characters, which fits in the definition of a short sentence in the Japanese language [62]. Yacis Corpus was assembled using data obtained automatically from the pages of Ameba Blog (ameblo.jp), one of the largest Japanese blogging services. The corpus consists of 12,938,606 downloaded and parsed web pages written by 60,658 unique bloggers. There were 6,421,577 pages containing 50,560,024 comments (7.873 comments per page that contains at least one comment). All pages were obtained between 3rd and 24th of December 2009. I used this corpus as it has been shown before that communication on blogs is rich in emoticons.

Experiment Settings From Yacis Blog Corpus I randomly extracted 1000 middle-sized sentences as the test set. 418 of those sentences included emoticons. Using Cohen's kappa agreement coefficient and balanced F-score I calculated CAO's performance in detecting emoticons in sentences (with Coehn's agreement coefficient, kappa), and emoticon extraction (including di-

vision of emoticons into semantic areas). In the evaluation of the emotion estimation procedure, I asked 42 people to annotate emotions on separate emoticons appearing in the sentences to verify the performance of CAO in specifying emotion types conveyed by particular emoticons (each person annotated 10 sentences/emoticons, except one person, who annotated 8 emoticons). Additionally, I asked the annotators to annotate emotions on the whole sentences with emoticons (however, the emoticon samples appearing in the sentences were different to the ones assigned in only emoticon annotation). This was used in an additional experiment not performed before in other research on emoticons. The usual evaluation considers only recognizing emotions of separate emoticons. I wanted to check how much of the emotive information encapsulated in a sentence could be conveyed with the addition of emoticons and whether it is possible to recognize the emotion expressed by the whole sentence looking only at the emoticons used in the sentence. Emoticons are something like an addition to this meaning. The question was how much does the emoticon match the meaning expressed by the sentence? I checked this looking on the emotion types and the general emotive features (valence and activation). However, since meaning of written/typed sentences is mostly understood on the basis of lexical information, I expected these results to be lower than those from only emoticon evaluation.

The system's results were calculated in a similar way to the training set, considering human annotations as a gold standard. Moreover, I checked the results of annotations for specific emotion types and groups of emotions belonging to the same quarters from Russell's two-dimensional affect space. The calculations were performed for the best three of the five ways of score calculation selected in training set evaluation.

Comparing CAO with Other Systems

I also compared CAO to other emoticon analysis systems where it was possible. The emoticon extraction was compared to the system developed by Tanaka et al. [140]. Emotion estimation of emoticons was compared to the system developed by Yamada et al. [161], as their approach is similar to mine in the way of exploiting the statistical occurrence of parts of emoticons. The two methods are described in detail below.

Kernel Method for Emoticon Extraction The system for extraction and analysis of emoticons with kernel methods was proposed by Tanaka and colleagues [140]. In their method they used popular tools for processing sentences in Japanese, a POS tagger ChaSen [76] and a Support Vector Machine-based chunker, *yamcha* [64] to chunk sentences and separate parts of speech from "other areas in the sentence", which they defined as potential emoticons. However, their method was significant as it was the first evaluated attempt to extract emoticons from input. Unfortunately, the method was unable to perform many important tasks. Firstly, as the method is based on a POS tagger, it could not extract emoticons from input other than a chunkable sentence. Therefore, if their system got a non-chunkable input (e.g. a sentence written in a hurry, with spelling mistakes, etc.), the method would not be able to proceed, or would give an erroneous output. Moreover, if a spelling mistake appeared inside a parenthesis, a non-emoticon contents could be recognized as a potential emoticon. All this made their method highly vulnerable to user creativity, although in a closed test on a set of prepared sentences their best result was somewhat high with 85.5% of Precision and 86.7% of Recall (balanced F-score = 86%).

Their classification of emoticons into emotion types however, was not ideal. The set of six emotion types was determined manually and the classification process was based on a small sample set. Therefore as the system for comparison of emotion type classification I used a later one developed by Yamada et al. [161].

N-gram Method for Emoticon Affect Estimation Yamada et al. [161] used statistics of n-grams to determine emotion types conveyed by emoticons. Although their method was not able to detect or extract emoticons from input, their set of emotion types was not set by the researchers, but borrowed from a classification appearing on BBS Web sites with emoticon dictionaries. Although not ideal, such classification was less subjective than their predecessors. To classify emoticons they used simple statistics of all characters occurring in emoticons without differentiating them into semantic areas. Eventually this caused errors, as some characters were calculated as "eyes", although they represented "mouths", etc. However, the accuracy of their method still achieved somewhat high scores of about 76% to 83%. For comparison with CAO I built a second system similar to theirs, but improved it with my emotion type classification (without this improvement, in my evaluation, their system would always score 0% for the lacking emotion types) and emoticon extraction from input, which capability the system of Yamada et al. did not posses. Moreover, I also used my database of raw emoticon samples, which improved the coverage of their system's database to 10,137 from 693 (6.8% of the improved database). Improved this way, I used this system in evaluation of CAO to verify the performance of my system in comparison with other methods in the fairest way possible. I also used three versions of

Yamada's system, based on unigrams, bigrams and trigrams.

3.2.7 Results and Discussion

Training Set Evaluation

Emoticon Extraction from Input The system extracted and divided into semantic areas a total number of 14,570 emoticons from the database of the original 10,137. The larger number of extracted emoticons on the output was caused by the fact that many emoticons contain more than one emoticon set (see example in Table 3.7). In primary evaluation of the system [109] approximately 82% of all extracted emoticons were extracted correctly. The problem appeared in erroneously extracting additional areas as separate emoticons. I solved this problem by detecting the erroneously extracted additional areas in a post-procedure, using the additional area database and reattaching the erroneously extracted areas with the actual emoticons they belonged to. This optimized the extraction procedure. There were still 73 cases (from 14,570) of erroneously extracting additional areas as emoticons. The analysis of errors showed that these erroneously extracted additional areas contained elements appearing in databases of semantic areas of eyes or mouths and emoticon borders. To solve this problem the error cases would have to be added as exceptions, however, this would prevent the extraction of such emoticons in the future if they actually appeared as emoticons. Therefore I agreed to this minimal error rate (0.5%), with which the extraction accuracy of CAO is still near ideal (99.5%). Finally, the results for the emoticon extraction and division into semantic areas, when represented by the notions of Precision and Recall, were as follows. CAO was able to ex-

tract and divide all of the emoticons, therefore the Recall rate was 100%. As for the Precision, 14,497 out of 14,570 were extracted and divided correctly, which gives the rate of 99.5%. The balanced F-score for these results equals 99.75%, which clearly outperforms the system of Tanaka et al. [140].

Affect Analysis of Emoticons Firstly, I calculated for how many of the extracted emoticons the system was able to annotate any emotions. This was done with a near ideal accuracy of 99.5%. The only emoticons for which the system could not find any emotions were the 73 errors appeared in the extraction evaluation. This means that the emotion annotation procedure was activated for all of the correctly extracted emoticons (100%).

Secondly, I calculated the accuracy in annotation of the particular emotion types on the extracted emoticons. From the five ways of result calculation two (Position and Unique Position) achieved much lower results than the other three, about 50%, and were discarded from further evaluation. All of the other three (Occurrence, Frequency and Unique Frequency) scored high, from over 80% to over 85%. The highest overall score in the training set evaluation was achieved in order by: Occurrence (85.2%), Unique Frequency (81.8%) and Frequency (80.4%). Comparison with the other emoticon analysis system showed, that even after the improvements I made, the best score it achieved (80.2%) still did not exceed my worst score (80.4%). For details see Table 3.10.

Test Set Evaluation

Emoticon Detection in Input The system correctly detected the presence or absence of emoticons in 976 out of 1000 sentences (97.6%). In 24

Table 3.10:
Training srt evaluation results for emotion estimation of emoticons for each emotion type with all five score calculations in comparison to another system.

Emotion type	Yamada et al [161] improved 1-gram	2-gram	3-gram	CAO: Occurrence	Freq-uency	Unique Frequency	Position	Unique Position
anger	0.702	0.815	0.877	0.811	0.771	0.767	0.476	0.476
dislike	0.661	0.809	0.919	0.631	0.800	0.719	0.556	0.591
excitement	0.700	0.789	0.846	0.786	0.769	0.797	0.560	0.516
fear	0.564	0.409	0.397	0.451	0.936	0.858	0.652	0.671
fondness	0.452	0.436	0.448	0.915	0.778	0.783	0.460	0.389
joy	0.623	0.792	0.873	0.944	0.802	0.860	0.522	0.421
relief	1.000	0.999	1.000	0.600	0.990	0.985	0.599	0.621
shame	0.921	0.949	0.976	0.706	0.922	0.910	0.538	0.566
sorrow	0.720	0.861	0.920	0.814	0.809	0.791	0.553	0.520
surprise	0.805	0.904	0.940	0.862	0.866	0.874	0.520	0.523
All approx.	0.675	0.751	0.802	0.852	0.804	0.818	0.517	0.469

cases (2.4% of all sentences) the system failed to detect that an emoticon appeared in the sentence. However, the system achieved an ideal score in detecting the absence of emoticons. This means that there are no errors in the detecting procedure itself, but that the database does not cover all possibilities of human creativity. However, it can be reasonably assumed that if CAO, with the database coverage of over 3 million possibilities still has 2.4% of error in emoticon detection, the methods based on smaller databases would fail even more often in similar tasks. The strength of the Coehn's coefficient of agreement with human annotators was considered to be very good (kappa=0.95). The results are summarized in Table 3.11.

Emoticon Extraction from Input From 418 sentences containing emoticons CAO extracted 394 (Recall=94.3%). All of them were correctly extracted and divided into semantic areas (Precision=100%), which gave an

Table 3.11:
Results of the CAO system in emoticon detection and extraction from input.

Detection				Extraction		
		System		R	P	F-score
		Emoticon	No emoticon	94.3%	100%	97.1%
Users	Emoticon	394	24			
	No emoticon	0	582	$(\frac{394}{418})$	$(\frac{394}{394})$	$2\frac{P*R}{P+R}$
No. of agreements=976 (97.6%), Kappa=0.95						

overall extraction score of over 97.1% of balanced F-score. With such results the system clearly outperformed Tanaka et al.'s [140] system in emoticon extraction and presented ideal performance in emoticon division into semantic areas, a capability not present in the compared system.

As an interesting remark, it should be noticed that in the evaluation on the training set, the Recall scored perfectly, but the Precision did not, and in the evaluation on the test set it was the opposite. This suggests that sophisticated emoticons, which CAO had problems detecting, do not appear very often in the corpora of natural language such as blog contents, and the database applied in CAO is sufficient for the tasks of emoticon extraction from input and emoticon division into semantic areas. However, as human creativity is never perfectly predictable, sporadically (at least in 2.4% of cases), there still appear new emoticons which the system is not able to extract correctly. This problem could be solved by frequent updates of the database. The race against human creativity is always an uphill task, although with close to ideal extraction (over 97%), CAO is already a large step forward. The results are summarized in Table 3.11.

Affect Analysis of Separate Emoticons The highest score was achieved in order by: Unique Frequency (93.5% for specific emotion types and 97.4% for estimating groups of emotions mapped on Russell's affect space model), Frequency (93.4% and 97.1%) and Occurrence (89.1% and 96.7%). The compared system by Yamada et al. [161], despite the numerous improvements I made to this system, did not score well, achieving its best score (for trigrams) far below the worst score obtained by CAO (Occurrence/Types). The scores are shown in the top part of Table 3.12. The best score was achieved by Unique Frequency, which in training set evaluation achieved the second highest score. This method of score calculation will be therefore used as default score calculation in the system. However, to confirm this, I also checked the results of evaluation of affect analysis of sentences with CAO.

Affect Analysis of Emoticons in Sentences The highest score was achieved in order by: Unique Frequency (80.2% for specific emotion types and 94.6% for estimating groups of emotions mapped on Russell's affect space model), Frequency (80% and 94%) and Occurrence (75.5% and 90.8%). It is the same score order, although the evaluation was not of estimating emotions of separate emoticons, but of the whole sentences. This proves that Unique Frequency is the most efficient method of output calculation for the CAO system. The compared system scored poorly here as well, achieving only one score (for bigrams) higher than CAO's worst score (Occurrence/Types). The scores are shown in the bottom part of Table 3.12.

The score for specific emotion type determination was, as I expected, not ideal (from 75.5% to 80.2%). This confirms that, using only emoticons, affect analysis of sentences can be performed at a reasonable level (80.2%).

Table 3.12:
Results of the CAO system in Affect Analysis of emoticons. The results summarize three ways of score calculation, specific emotion types and two-dimensional affect space. The CAO system showed in comparison to another system.

Emotion Estimation on Separate Emoticons								
Yamada et al. (2007)			CAO					
1-gram	2-gram	3-gram	Occurrence		Frequency		Unique Frequency	
			Types	2D space	Types	2D space	Types	2D space
.721	.865	**.877**	.891	.967	.934	.971	**.935**	**.974**
Emotion Estimation on Sentences								
Yamada et al. (2007)			CAO					
1-gram	2-gram	3-gram	Occurrence		Frequency		Unique Frequency	
			Types	2D space	Types	2D space	Types	2D space
.686	**.798**	.715	.755	.909	.801	.941	**.802**	**.946**

However, as the emotive information conveyed in sentences consists also of other lexical and contextual information, it is difficult to achieve a result close to ideal. Although, the results for 2-dimensional affect space were close to ideal (up to nearly 95%), which means that the emotion types for which human annotators and the system did not agree still had the same general features (valence polarity and activation). This also confirms the statement that people sometimes misinterpret (or use interchangeably) the specific emotion types of which general features remain the same (in the test data people annotated, e.g., "fondness" on sentences with emoticons expressing "joy"; or "surprise" on "excitement", etc., but never, e.g., "joy" on "fear"). The above can be also interpreted as further proof for the statement from section 3.2.2, where emoticons are defined as expressions used in online communication as representations of body language. In direct communication, body language is also often used to convey a supportive meaning for the contents conveyed through language. Moreover, some sets of behavior (or kinemes) can be used

to express different specific meanings for which the general emotive feature remains the same. For example, wide opened eyes and mouth might suggest emotions like fear, surprise or excitement; although the specificity of the emotion is determined by the context of a situation, the main feature (activation) remains the same. In the evaluation, the differences in the results for specific emotions types and two-dimensional affect model prove this phenomenon. Some examples illustrating this have been presented in Table 3.13.

3.2.8 Conclusions

In this section I presented CAO, a prototype system for automatic affect analysis of Eastern style emoticons. The system was created using a database of emoticons containing over ten thousand of unique emoticons collected from the Internet. These emoticons were automatically distributed into emotion type databases with the use of an affect analysis system developed previously (see section 3.1). Finally, the emoticons were automatically divided into semantic areas, such as mouths or eyes and their emotion affiliations were calculated based on occurrence statistics. The division of emoticons into semantic areas was based on Birdwhistell's [10, 11] idea of kinemes as minimal meaningful elements in body language. The database applied in CAO contains over ten thousand raw emoticons and several thousands of elements for each unique semantic area (mouths, eyes, etc.). This gave the system coverage of over three million combinations. With such coverage the system is capable of automatically annotating potential emotion types of any emoticon. There is a finite number of semantic areas used by users in emoticons generated during online communication. The number CAO can match,

over three million emoticon face (eye-mouth-eye) triplets, is sufficient enough to cover most possibilities.

The evaluation on both the training set and the test set showed that the system outperforms previous methods, achieving results close to ideal, and has other capabilities not present in its predecessors: detecting emoticons in input with a very strong agreement coefficient (kappa = 0.95); and extracting emoticons from input and dividing them into semantic areas, which, calculated using balanced F-score, reached over 97%. Among the five methods of calculating emotion rank score I compared in evaluation of emotion estimation of emoticons, the highest and the most balanced score was based on Unique Frequency and this method of score calculation will be used as a default setting in CAO. Using Unique Frequency, the system estimated emotions of separate emoticons with an accuracy of 93.5% for the specific emotion types and 97.3% for groups of emotions belonging to the same two dimensional affect space [114]. There were some minor errors, however not exceeding the standard error level, which can be solved by optimization of CAO's procedures during future usage. Also, in affect analysis of whole sentences CAO annotated the expressed emotions with a high accuracy of over 80% for specific emotion types and nearly 95% for two dimensional affect space.

Table 3.13:
Examples of analysis performed by CAO. Presented abilities include: emoticon extraction, division into semantic areas, and emotion estimation in comparison with human annotations of separate emoticons and whole sentences. Emotion estimation (only the highest scores) given for Unique Frequency.

Example 1: *Chakku-shime wasure-san ga ooi desu ne, watashi mo tama ni yarakashite hitori sekimen* (;^_^A
Translation: Many people forget to close their fly. I sometimes do that too and when I notice, I get all red (;^_^A

S_1	B_1	S_2	$E_L ME_R$	S_3	B_2	S_4
N/A	(	;	^_^	A	N/A	N/A

CAO	Human Annotation	
fear / anxiety (0.06450746)	emoticon	sentence
...	fear / anxiety	fear / anxiety, shame

Example 2: *Itsumo, "Mac, ne——" tte shibui kao sareru n desu. Windows to kurabete meccha katami ga semai desu* (/ Д`):･ﾟ･+:･
Translation: People would pull a wry face on me saying "Oh, you're using a Mac...?" . It makes me feel so down when compared to Windows (/ Д`):･ﾟ･+:･

S_1	B_1	S_2	$E_L ME_R$	S_3	B_2	S_4
N/A	(	N/A	/ Д`	N/A	)	:･ﾟ･+:･

CAO	Human Annotation	
sadness / sorrow (0.00698324)	emoticon	sentence
excitement (0.004484305)	sadness / sorrow	sadness / sorrow, dislike
dislike (0.001897533)		
...		

Example 3: *>Aki-san, eee,* (/°o°)*/ipod wa nai to iya dakara sugu ni juden da yo!!*
Translation: >>Aki-san, What!? (/°o°)/I couldn't imagine a day without my ipod! Recharge your battery at once!

S_1	B_1	S_2	$E_L ME_R$	S_3	B_2	S_4
N/A	(	/	°o°	N/A	)	/

CAO	Human Annotation	
surprise (0.02686763)	emoticon	sentence
joy (0.02679939)	surprise	surprise
excitement (0.02238806)		
...		

Example 4: *2000 bon anda wo tassei shita ato ni iroiro to sainan tsuzuita node nandaka o-ki no doku...* (°. °)
Translation: All these sudden troubles, after scoring 2000 safe hits. Unbelievable pity ... (°. °)

S_1	B_1	S_2	E_L	M	E_R	S_3	B_2	S_4
...	(	N/A	°	.	°	N/A	)	N/A

CAO	Human Annotation	
surprise (0.4215457)	emoticon	sentence
...	surprise	surprise , dislike

Chapter 4

Application of Emotive Information in Human-Computer Interaction

In this chapter I present two methods developed for enhancing Human-Computer Interaction. The first is a method for automatic evaluation of conversational agents. The affect analysis systems described in precious chapter are used to analyze users' emotional engagement during conversation. This data is reinterpreted to specify general attitudes toward the conversational agent and its performance. The second method is determining whether emotions expressed by speaker are appropriate for the context of the conversation. In this method, affect analysis system estimates the speaker's affective states and a Web mining technique gathers from the Internet emotive associations consisting of a list of emotions that should be expressed at the moment.

4.1 Method of Automatic Evaluation of Conversational Agents

Technological development focused on enhancing and facilitating human lives has led to a need for intelligent environments meeting all human needs. Some examples are long-term projects, such as MIT's *House_n*[1], *MavHome*[2] or *Living Tomorrow*[3] Smart Home Projects. Explorations in the field of Ambient Intelligence [28] brought to light a new dimension of communication, where humans and machines become interlocutors, Human-Computer Interaction (HCI) [27]. With this came a rush in development of intelligent conversational agents, beginning with freely talking chat-bots [45], through car navigation systems [138] to talking furniture [43]. Their functional implementation into our lives has already become a current process. A need for an environment not only intelligent, but also humanized, is growing rapidly [146].

Along with this, researchers focused on agent development have found themselves with an urgent need to develop fast automatic evaluation methods for such agents. The usual methods used to evaluate conversational agents are based on subjective questionnaires in which user-testers express their opinions about the agent, their satisfaction during interaction with it, their will to continue the conversation, the naturalness of the agent's utterance generation, etc. There have been some attempts to automatically evaluate spoken task-oriented dialog systems, such as those by Litman, Walker and colleagues [154, 71]. However, these apply only to task-oriented spo-

[1] http://architecture.mit.edu/house_n
[2] http://ailab.wsu.edu/mavhome/index.html
[3] http://www.livtom.com/

ken dialog agents, and therefore are based on simple detection of keywords appropriate to the task performed by the English-speaking agent. A different approach was presented by Isomura and colleagues [50], who made an attempt to evaluate a non-task-oriented Japanese-speaking dialog agent using the Hidden Markov Model. However, their results were rather low (54%). Moreover, their method was able to evaluate only the naturalness of the agent's utterance, whereas in a usual subjective questionnaire there are many other dimensions than naturalness in which the agent is evaluated. One could, for example, imagine a conversational agent that has very natural utterance generation, but through a lack of, e.g., rules of politeness, inappropriate proposition generation, or not keeping up the topic, would make the user irritated or even angry with the agent. Assuming that Isomura's method worked (54% of accuracy), such an agent would be evaluated in their method as being very good (very natural utterance generation); however, as the utterances eventually made the user dissatisfied, the overall evaluation would be rather negative.

To obtain a satisfying automatic evaluation method for conversational agents, there is a need for something to act as a substitute for the subjective questionnaire. In questionnaire-type evaluation the users make decisions about how highly to mark the agent, and as such the process of questionnaire evaluation could be perceived from a typical decision-making perspective. The acts of decision-making and expressing opinions in humans strongly depend on features like emotional states or experience [72, 117]. Therefore I assumed that it should be useful to analyze the attitudes of user-testers towards agents. Another problem with the questionnaire is that, as it is carried out after the conversations (sometimes an hour or more later, if the testing

is time-consuming), the users' attitude may change from the time of the conversation. This change may be caused by the passing of time gradually obscuring the impression of the agent; or, in the evaluation of two or more agents, the impression of the former may be altered by the performance (better or worse) of the latter; also, as is argued by Clore and colleagues [16, 17], changes in attitudes may be influenced by mood fluctuations caused by different factors, such as weather or, for example, news seen on television in the time between the actual experiment and filling in the questionnaire. All the above makes it most desirable to gather the attitudinal information from the users during the time of the conversation with the evaluated agent.

In this section I propose such a method. In this method, during user conversations with two non-task-oriented Japanese-speaking conversational agents, users' current attitudes and sentiments towards the agents are estimated automatically. I based this idea on "Affect-as-Information" [124] reasoning about the emotions expressed in the users' utterances.

4.1.1 General Approach: Attitude From Affect

Sentiment Analysis for Agent Evaluation

As mentioned above, in order to evaluate a conversational agent there is a need to obtain information about the user's attitudes toward the agent. The field focused on gathering such information is called Sentiment Analysis. It is a sub-field of Information Extraction that has only recently captured the interest of scientists [149]. The general idea of sentiment analysis is to gather and classify (into positive and negative) sentiments and attitudes about particular topics or entities. Sentiment Analysis is important for marketing

research [94], monitoring of chat-room content for security reasons [1], and customer feedback on particular products [149]. Since conversational agents can be considered as products as well, it would be desirable to acquire objective information about the agents' performance before putting them on the market, as failure may cause a substantial loss of funds and human effort. Tests, where people are hired to verify the performance of market-destined agents, are burdened with heavy use of effort and funds. Moreover, paying user-testers high sums of money for the evaluation undermines the objectivity of such a test. Although there is no other way of performing the test than making a human talk to the agent, in my assumption there is a better way to gather more objective information for the evaluation than a typical questionnaire performed after the test phase. Namely, information about the tester's sentiment towards the product (agent) could be gathered during the test phase (conversation with the agent). This should provide the objective information. However, in the usual sentiment analysis methods, the attitudinal information is extracted from the text with regards to a particular object (product). This means that such methods are applicable only if the user explicitly expresses his/her attitude towards the product. Unfortunately, in a free, non-task-oriented conversation, users usually do not express their attitudes directly towards their machine interlocutors. However, it can be assumed that the users' attitude should be revealed in how they respond to the agents' utterances. Therefore, analyzing the emotional level of users' utterances and applying some kind of function transforming this data into attitudinal information should provide information about what users think about the agents. This, if mapped efficiently on a set of questions from a usual subjective questionnaire, would in effect provide a substitute for the

questionnaire.

I decided to gather information about users' emotional states during conversations using the techniques for Affect Analysis developed previously and described in chapter 3, and transform the data obtained this way using reasoning based on the "Affect-as-Information" Theory to determine the user's attitude toward the agent interlocutor.

Affect Analysis for Attitude Estimation

Affect Analysis, as defined in section 3.1, is also a relatively new sub-field of Information Extraction, and focuses on classifying users' expressions of emotions. However, in contrast to Sentiment Analysis, where the goal is to determine the user's general attitude (positive or negative) to a specific object (movie review, or a product), this field takes as an object the user himself, and its goal is to estimate human emotional states in a more detailed manner. While attitude could be either positive or negative, the expression of emotion could represent a wide scope of emotional states, from fear, anger, or excitement to joy, pleasure, or relief. There is some research on affect analysis, also for the Japanese language [147, 145]. However, there have been only a few approaches to apply affect analysis to gather information about sentiments and attitudes [41], and no significant work has been done on applying such an approach to the evaluation of conversational agents speaking Japanese. This section presents the first research attempt of this kind.

Affect-as-Information Theory

The theory of Affect-as-Information was introduced in 1983 by Schwarz and Clore [124] and is widely studied in the field of psychology and social psychology. Schwarz and Clore claimed that people use affect in the same way as any other criterion, by applying the informational value of their affective reactions to form their judgments, attitudes and opinions.

Schwarz, Clore and colleagues studied this phenomenon thoroughly in numerous experiments [124, 16, 17]. They reached the conclusion that people's choices and evaluations, and therefore attitudes, change according to the changes in their current moods. This change could be caused naturally (e.g. weather), or induced by various factors. For example, watching a sad movie induces sad moods in a person, which could be further used as a factor to cancel a party with friends. As another example, talking to someone one hates may spoil the whole day. Similarly, talking to someone interesting and friendly could induce positive mood, and the overall estimation of one's relationship with this person could be even better.

Using the same reasoning, I assumed Schwarz and Clore's findings to be useful in transforming the results of affect analysis of user utterances carried out during conversation with an agent into information about the users' attitudes towards the evaluated agents. The subsequent filling in of a subjective questionnaire about the interlocutor after the conversation can be perceived as a typical decision-making process (people make decisions about how to evaluate the agent). Therefore, if the approach is correct, the automatic estimation of users' attitudes through the conversation should indicate similar tendencies to the results acquired through the questionnaire.

Proving this to be true would be a step towards the practical realization

of the idea of affective human factors design [55], where the information about a product (agent) is derived from information about dynamic changes of the user's affective states during usage. If proved, this would provide strong evidence that in the process of product design, affective factors are not only as important as usability [55], but that affect itself provides valuable information about usability, and can thus be a source of information for continuous improvement of the product.

4.1.2 Information Derived from Affect Analysis

The affect analysis system employed in the automatic evaluation method described in this chapter is ML-Ask developed previously by me in [102, 107]. To realize the method, one can use any reliable affect analysis system available in the field. However, as mentioned in section 4.1.1, the information about users' affective states needs to be extracted and analyzed in real time. Therefore, I used ML-Ask system as it is fast (analysis of one utterance takes less than 0.15 seconds) and reliable (different evaluations confirmed the system's reliability in laboratory conditions as well as in the field; for details see: [102, 107, 105]).

ML-Ask is used to analyze utterances of a user talking to a conversational agent, during the conversation. The results of analysis of each utterance provide information on how many user utterances were emotive. Furthermore, the emotions extracted from the user's emotive utterances form a vector on which the emotional states of the user changed during the conversation. This is then processed as follows.

Firstly, if many[4] of the user's utterances were determined as emotive,

[4]I do not specify here the value of "many". I compare the results for two different

I assume the user was emotionally involved in the conversation. Emotional involvement in a conversation suggests a tendency towards easier familiarization with the interlocutor [167]. Therefore I can further assume that during a user's conversation with an agent, the machine interlocutor is considered to be more human-like the more emotionally emphasized the user's utterances are. However, this does not yet mean a positive familiarization. The conversation could become emotional also when the interlocutors quarrel. This could happen in a case where the agent makes the user angry. However, if the user agrees to participate in a quarrel with an agent, this could also mean that the user finds the agent's linguistic capabilities to be comparable to himself. Therefore, the information obtained about the general emotiveness of the conversation could be interpreted as signifying how much the user finds the agent worth talking to, including familiarity and the user's opinion about the agent's linguistic skills.

Secondly, analysis of specified emotion types conveyed by the user in the whole conversation provides information on the user's particular emotions during the conversation. If the emotions, according to the Russell's model (see [114] and section 2.1.3), were positive or changing from negative to positive while talking, the general attitude towards the agent is considered to be positive. If the emotions were negative or changing from positive to negative, the attitude is classified as negative. The general attitude towards an agent is calculated as the ratio of conversations with positive tendency to the conversations with negative tendency. The flow of the procedure is

conversational agents to verify for which the tendency was higher. However, I assume it is possible to set a threshold in the results for evaluation of only one agent. Such a threshold could be obtained statistically, after performing several evaluations of different agents. It could also be set arbitrarily, as, for example, in [65].

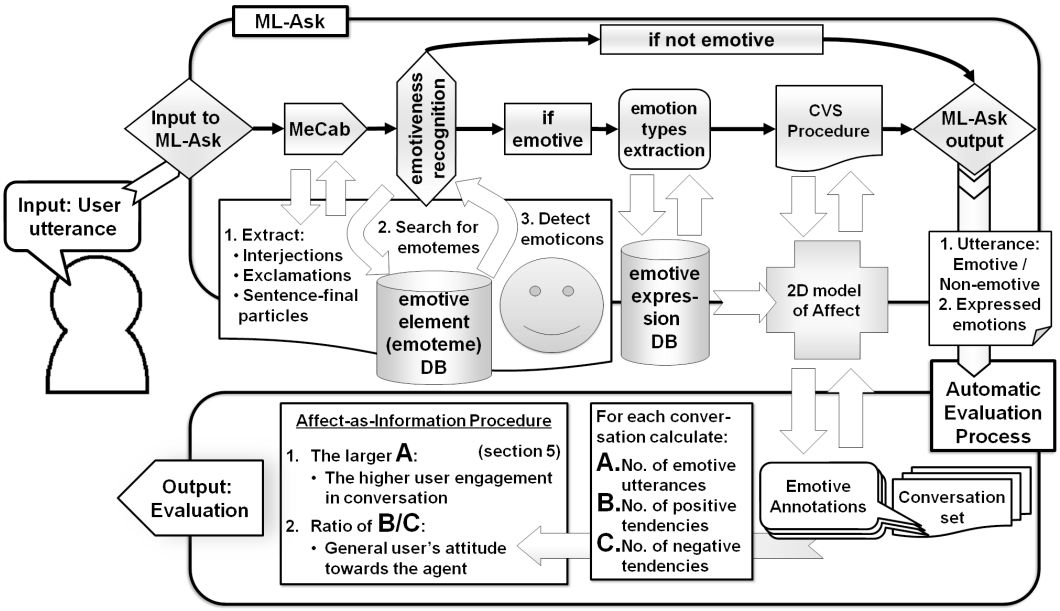


Figure 4.1:
Flow chart of the automatic evaluation procedure including: affect analysis system ML-Ask (upper part); further processing of information obtained by ML-Ask and the decision making process for the final evaluation (lower part).

presented in Figure 4.1.

Both types of the acquired information (the general engagement of the user in conversation and the attitude) provide an overview of the user's sentiment about the agent, and it is desirable for both types of information to harmonize rather than show dissonance. Such analysis, if accurate, realizes the first step of affective human factors design, which is to understand the user's affective needs [55].

4.1.3 Evaluation Experiment

To test the method, I performed an evaluation experiment of two non-task-oriented conversational agents. The first agent is a simple conversational agent which generates responses by 1) using Web-mining to gather associations to the content of user utterance; 2) making propositions by inputting the associations to the prepared templates; and 3) adding modality to the basic propositions to make the utterance more natural. The second agent, based on the first one, generates a humorous response to user utterance every third turn. The humorous response is a pun created by using user input as a seed to gather pun candidates from the Web and inputting the most frequent ones into pun templates (for more detailed description of the agents see below and references). The choice of the agent was deliberate. They differed only in one element - the humorous responses in the latter one. The assumption was that, as humor is an important factor in socialization [164], the joking agent should be evaluated higher by the users and this difference should be easily noticeable. If the automatic evaluation method then displayed the same tendencies, they should also be easily recognizable.

There were 13 participants in the experiment, 11 males and 2 females. All of them were university undergraduate students. The users were asked to perform a 10-turn conversation with both agents. No topic restrictions were made, so that the conversation could be as free and human-like as possible. The agents were first evaluated during the conversation using the proposed automatic evaluation method and the results were stored for further comparison with a subjective questionnaire. After the conversations, the users were asked to complete a questionnaire concerning their attitudes to the agents and their performance. The results of the automatic evaluation were

compared to the results of a subjective questionnaire filled in by the users in order to evaluate the two agents. Using these sets of results, I was looking for similarities between sentiment classification and the questionnaire.

Two Conversational Agents - Short Description

Modalin

Modalin is a non-task-oriented text-based conversational agent for Japanese. It automatically extracts from the Web sets of words related to a conversation topic set freely by a user in his utterance. The association words retrieved from the Web (with accuracy of over 80%) are then sorted by their co-occurrence on the Web, and the most frequent ones are selected to be used further in output generation. In the response generation, the extracted associations are put into one of the pre-prepared response templates. The choice of the template is random, but the agent keeps in its memory the last choice in order not to generate two similar sentence patterns in a row. Finally, the agent adds a modality pattern to the sentence and verifies its semantic reliability. The modality is added from a set of over 800 patterns extracted from a chat-room logs and evaluated. The naturalness of the final form of the response is then verified on the Web with a hit-rate threshold set arbitrary for 100 hits. The agent was developed by Higuchi and colleagues. For further details see [45].

Pundalin

Pundalin is a non-task-oriented conversational agent for Japanese, created on the base of Modalin combined with Dybala's Pun generating system PUNDA [30]. The PUNDA pun generator was developed by Dybala and

colleagues as a part of PUNDA research project, aiming to create a Japanese pun generating engine. The system works as follows. From the user's utterance, a base word is extracted and transformed using Japanese phonetic pun generation patterns, to create a phonetic candidate list. The candidate with the highest hit-rate in the Japanese search engine Goo[5] is chosen as the most common word that sounds similar to the base word. Next, the base word and the candidate are integrated into a sentence. The integration is done in two steps, one for each part of the sentence including the base word and the pun candidate, respectively. Firstly, the base phrase is put into one of several pre-prepared templates making up the first half of the sentence. The second half of the sentence is extracted from KWIC on WEB - on-line Keyword-in-context sentences database [165] as the shortest latter half of an emotive sentence including the candidate. Every third turn of the conversation, Modalin's output was replaced by a joke-including sentence, generated by the pun generator. Pundalin therefore is a humor-equipped conversational agent using puns to enhance communication with the user. Pundalin was developed by Dybala and colleagues as a conversational agent for use in experiments on the influence of humor on human-agent interaction [29].

Questionnaire - User's Evaluation

The questions the users were asked after the conversations with both agents were: A) Do you want to continue the dialogue?; B) Were the agent's utterances grammatically natural?; C) Were the agent's utterances semantically natural?; D) Was the agent's vocabulary rich?; E) Did you get an impression

[5]http://search.goo.ne.jp/

that the agent possesses any knowledge?; F) Did you get an impression that the agent was human-like?; G) Do you think the agent tried to make the dialogue more funny and interesting? and H) Did you find the agent's talk interesting and funny?. The answers for the questions were given in 5-point scale (1 - the lowest score; 5 - the highest score) with some explanations added. Each user filled two such questionnaires, one for each agent. The final, summarizing question was "Which agent do you think was better?"

Representation of Questionnaire in Sentiment Analysis

I made the following assumptions about how the questions the users were asked directly were represented in the results provided by the analysis. I assumed that the questions from A) to H) generally represent several kinds of information, such as: how highly did the users evaluate agents' talking abilities (questions A-D); how much the users were able to familiarize with the agents (questions E-F); and how much they were emotionally involved in the conversation (questions G-H). According to Dybala [32], in the evaluation of conversational agents there are two features that have to be evaluated. The first represents the agent's linguistic capabilities, and the second represents all features other than linguistic, such as subjective impression or ease of familiarization. In my assumption, the first set of questions (A-D) inquire about the linguistic features and the latter two sets of questions (E-F and G-H) represent the non-linguistic features. Furthermore, the general summarizing question represents the users' general attitude towards the agents, and therefore represents the second type of information obtained from the automatic analysis (for details see Section 4.1.2).

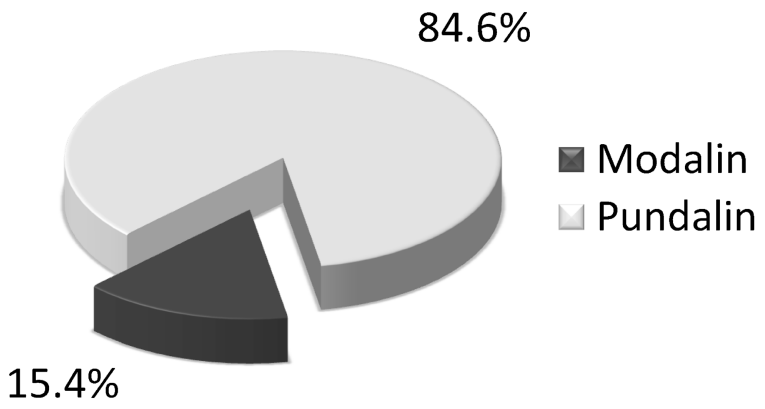


Figure 4.2:
Users' evaluation - results for the question "Which agent do you think was better?".

4.1.4 Results

The results of the evaluation are discussed below. First, the results of the questionnaire are discussed, then the results of the automatic evaluation method are summarized and compared to the users' direct opinions.

User Evaluation

Regarding the detailed questions, higher scores were given by the users to Pundalin (see Figures 4.3 Table 4.1 with its graphical representation in Figure 4.4). In all categories, overall results for both agents clearly showed that the performance of Pundalin was estimated as being more human-like and easier to familiarize with.

The questions about agents' conversational abilities (questions B-D) revealed that the humor-equipped agent was rated higher, although the differences were not as large as in other questions. The reason for this is that

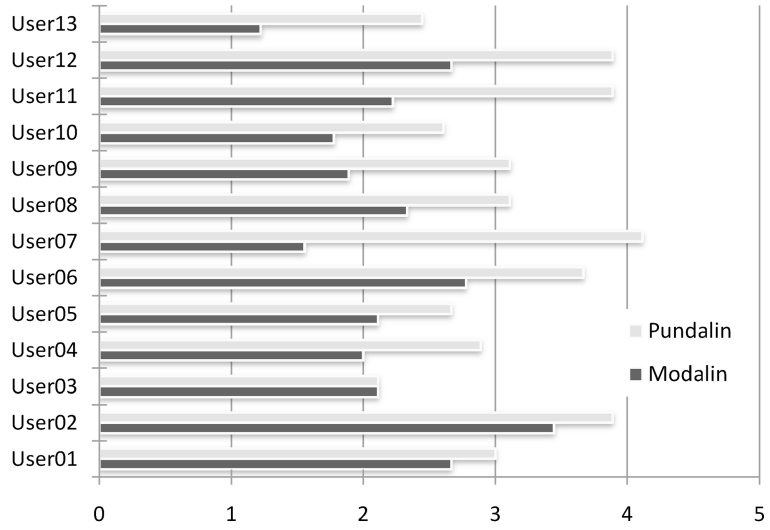


Figure 4.3:
Users' evaluation for Modalin and Pundalin, representing the approximated results of all detailed questions per user. Answers given in a 5-point scale.

Pundalin was based on Modalin and, with the exception of the humorous responses, all other responses were made in the same way as in Modalin. The questions inquiring how easily the users could familiarize with the agents (A and E-F) showed that Pundalin scored higher here as well. The most notable differences were seen in the questions investigating how much the users were emotionally involved in the conversation (questions A and G-H). Here, the joking agent was also evaluated higher. The results were summarized for all questions (with approximated values for users) in Table 4.1. All of the results were statistically significant at 5% level. The overall compared results of Modalin and Pundalin were extremely statistically significant, with P value = .0002. I also summarized the results for all users (with approximated values for questions), which also showed clearly that the users generally evaluated

Table 4.1:
Users' overall evaluation of Modalin and Pundalin for each detailed question. Answers given in a 5-point scale.

Questions	A	B	C	D	E	F	G	H
Modalin	2.62	2.15	1.85	2.08	2.15	2.38	1.92	2.46
Pundalin	3.38	2.92	2.69	3.00	2.85	3.31	4.15	4.08
P-values	.033	.040	.015	.014	.021	.006	.004	.006

the joking agent higher (see Figure 4.3). These results were also extremely statistically significant, with P value = .0002.

This corresponds to the results of the final question, in which the users were asked which agent was better in general. This question investigated the general attitude of the users towards each agent after the experiment. Eleven out of thirteen users (84.6%) evaluated Pundalin (humor-equipped agent) as better than Modalin (see Figure 4.2), which means the attitude was more positive towards the former agent.

After gathering the results of the questionnaire, I compared them to the automatic evaluation method. I assumed that if the tendencies were similar and the results were statistically significant, the method is applicable as an automatic evaluation method for non-task-oriented conversational agents.

Results of Sentiment Analysis

Evaluation based on sentiment analysis of the users' utterances showed tendencies similar to the questionnaire. The users were more emotionally involved in the conversations with Pundalin, which corresponds to the direct opinions about the agent - that it was more human-like, its utterances were more correct semantically, grammatically, etc. (see Figure 4.6) and therefore

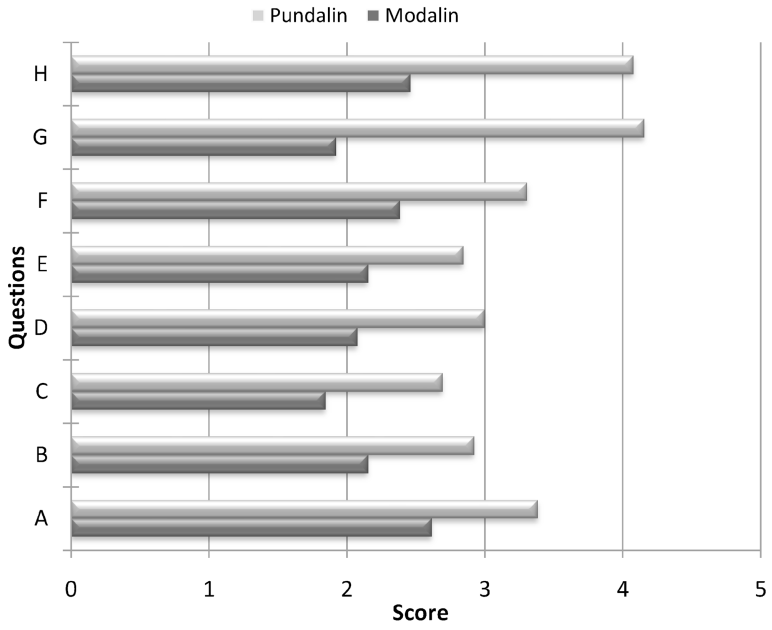


Figure 4.4:
Graphical representation of Table 4.1. Results for each detailed question per agent. Answers given in a 5-point scale.

the agent was easier to familiarize with. The results summarized for all users were very statistically significant (P value = .0053).

The analysis of specified emotion types conveyed by the users in conversations provided information clearly revealing the users' attitudes towards each agent. The users' general attitudes to Pundalin were mostly positive (67%), whereas to Modalin the attitudes of the users were mostly negative (75%). For details see Figure 4.5.

The results above indicate that the general attitude of a user towards an agent was better for Pundalin than for Modalin, which corresponds to the results of the questionnaire.

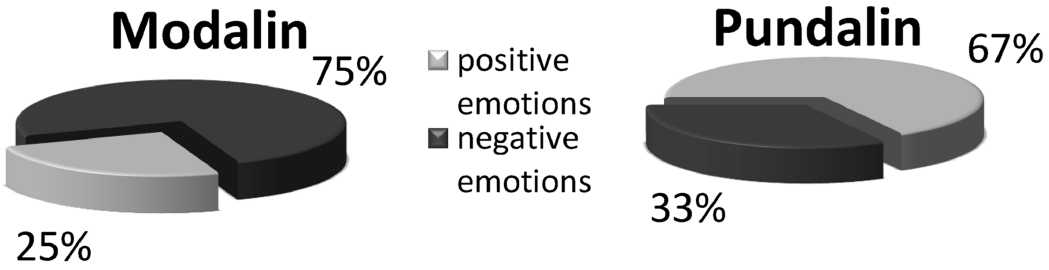


Figure 4.5:
The total ratio of all emotions positive to all negative conveyed in the utterances of users with Modalin and Pundalin.

Correlations Between Automatic Evaluation and Questionnaire

In order to check which questions were correlated best with the results of automatic evaluation, I calculated the correlation coefficient. As the base for calculations I used Spearman's rank correlation coefficient (Spearman's ρ [rho]). The usual Pearson's correlation coefficient represents a linear dependence between data, which is not the issue in subjective evaluation. Therefore Spearman's rank correlation coefficient used to calculate any monotonic dependence is more appropriate for my task. The results are presented in Table 4.2.

In this research I aim to propose a method to substitute the usual subjective evaluation questionnaire, and thus the most important particular question sets for me were those representing non-linguistic features (especially G and H). The correlation test revealed accordingly that the strongest correlation was between sentiment analysis and questions G (Did the agent try to be interesting?) and H (Was the agent interesting?). Therefore, I can say that the automatic evaluation method is applicable in subjective evaluation of non-linguistic features, especially those related to entertaining the user.

The test revealed also other correlations. A medium-strong correlation was found in the results of question E (Did the agent posses any knowledge?). This can be interpreted to signify that people usually become more involved in conversation with intelligent interlocutors. Medium correlation was also found with questions A (Continuing the dialog) and B (Grammatical naturalness). The first can be interpreted as a natural consequence of the results for question E - stronger involvement in the conversation with an intelligent partner logically makes one more obligated to continue the dialog. The cause of respectively high correlation of B is not visible at first glance, but when set together with questions C (Semantic naturalness) and D (Vocabulary richness) becomes more understandable. The correlation of these questions with the automatic evaluation declines along with an increase in possibilities of interpretation. This is presumably also the reason for question F (Human-likeness) to be the least correlated, since, as noted also by Dybala and colleagues [33], the concept of human-likeness in machines is still vague and undefined.

It is also possible that changing the formulation of the questions and improving the method itself will enhance the correlation as well. Moreover, there already exist automatic evaluation methods for only linguistic abilities of conversational agents [50], although their accuracy is not high. However, combining them with my method might show improvement of the overall evaluation.

4.1.5 Discussion

In the primary evaluation experiment of this method, performed on two conversational agents, previously I showed [100] on five user-testers, that there

Table 4.2:
The results of Spearman's rank correlation test between sentiment analysis and each question.

Question	A	B	C	D	E	F	G	H
Correlation (ρ)	**.333**	**.350**	.202	.164	**.480**	.035	**.559**	**.597**

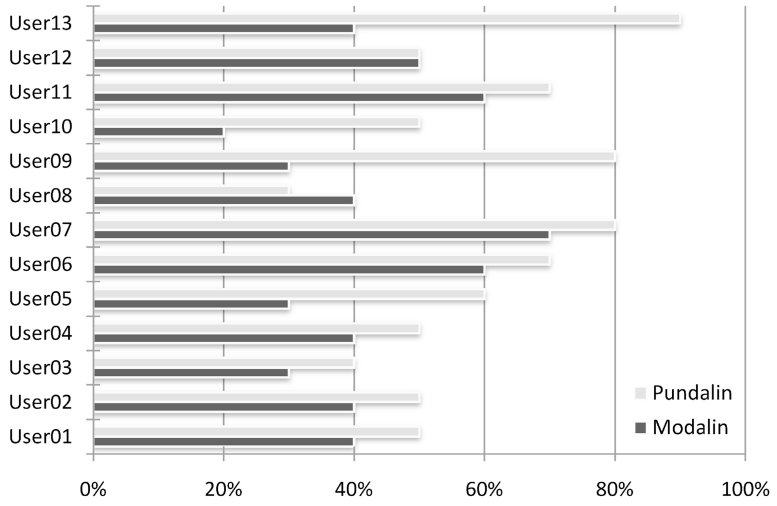


Figure 4.6:
Average appearance of emotively engaged utterances for all 13 users in conversations with both agents ("90%" means that in 10-turn conversation there were 9 emotive utterances).

were similar tendencies in the results acquired by the method and the results of the questionnaire. The results presented here, although the number of evaluators was nearly three times larger (13 participants), show that the tendencies remained the same. Users showed higher emotive engagement and positive attitudes in conversations with the agent which used jokes. This proves that the method is applicable as a means of evaluation for conversational agents.

The differences between results of the questionnaire and the method were not in a one-to-one ratio, however, it should be remembered that both evaluations, although aiming to provide answers to similar questions, were based on different assumptions. In the questionnaire the users are aware of the points they deliberately assign, whereas in the automatic evaluation method the users did not know that what they say will be used as a material for evaluation of the agent. Compared to traditional subjective questionnaires, this makes the proposed method less invasive and therefore provides objective information on the users' sentiments about the machine interlocutor.

The automatic evaluation correlated strongest with the questions about non-linguistic features. As there has not previously been a method for automatic evaluation of such features, this is probably one of the most significant achievements of this method. The questions about linguistic abilities also correlated, although in a weaker manner. However, it can be predicted that improving the method, either by improving the intermediary procedures or by combining it with other automatic evaluation methods, will improve the overall quality of evaluation. Moreover, the representation of sentiment analysis results in the questions was set arbitrarily and it is possible that there could be a set of questions which represent the information obtained by automatic evaluation in a more straightforward manner. However, for the experiment presented in this section, the attention should rather be focused on the similarities in tendencies that appeared in general comparison of the two agents and on the fact that all compared results were statistically significant.

Approximate time of processing one utterance is below 0.15 s, which makes the method applicable in providing actual information on changes in the users' attitudes towards the machine interlocutor in real time. This does

not only provide fast and up-to-date information on users' sentiments, but also, appropriately utilized, can provide hints about potential undesirable changes in the users' attitudes and the need for appropriate counteractions, during everyday use.

4.1.6 Conclusions

In this section I presented an automatic method of evaluation for conversational agents. The method is based on analyzing affective states conveyed by a user in a conversation with an agent. Borrowing the notion of affect-as-information [124], the results of affect analysis performed by a system created by Ptaszynski and colleagues [102, 107] provide information about the user's emotional involvement in a conversation, the user's psychological distance to the agent, and ease of familiarization with the machine. This corresponds to direct questions about the agent's performance. Next, analysis of specified emotion types conveyed by the user in the whole conversation, and their classification by applying the two-dimensional model of emotions [114] provides information on the polarity of the users' attitudes towards the machine interlocutor during the conversation.

By applying the proposed method in evaluation of conversational agents, the evaluating information is acquired during the user-testers' conversations with the agents. Therefore as means of evaluation, the method saves time, effort and funds spent each time on preparing and performing laborious questionnaires. It is desirable for the proposed method to be accepted widely in the field as a full equivalent or at least a strong supportive means to objectivize the results of traditional questionnaires.

4.2 Method of Verifying Contextual Appropriateness of Emotion

"Public opinion is a second conscience."

- William R. Alger

"Conscience is, in most men, an anticipation of the opinions of others."

- Henry Taylor

Recent introduction in our lives of communication technologies based on ubiquitous networks, such as 3G or wireless LAN, made the Internet, a social phenomenon only a few years ago [82], an indispensable everyday article. There was a rapid increase of fields using Internet resources as an object of research in information or opinion retrieval [128]. The Web is becoming a determinant of human commonsense [116]. Within the last few years there have been several attempts to retrieve information concerning human emotions and attitudes from the Internet [1]. Some of them aimed at supporting emotion recognition by using statistical approximation of numerous data gathered form the Web [101, 145].

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 [101, 145]. However, in these methods, although the results of experiments appear to be positive, two extremely different approaches are mixed, the language-syntax based and the Web mining based one. 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 at a certain time. The latter one 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 a complementary mean for the language based approach, although achieving reasonable results, means not fully exploiting the great potential lying in the Web [118].

In this chapter I present a novel method utilizing these two approaches in a more effective way. The method I propose is capable to analyze affect with regard to a context and estimate whether an emotion conveyed in a conversation is appropriate for the particular situation. In the method I used previously developed systems for affect analysis of utterances (see chapter 3). Next, I used a method for gathering emotive associations from the Web developed by Shi et al. [125].

Furthermore, I checked several versions of the method to optimize its procedures. Firstly, I checked two versions of ML-Ask, with and without Contextual Valence Shifters. Secondly, I checked two versions of the Web mining technique, one performing search on the whole Internet and the second one searching only through blogs.

A problem with the Web mining technique was that it was gathering too much noise using the whole Internet. In research such as the one by Abbasi et al. [1] it was proved that public Internet services consisting of social networks, such as forums, or blogs are a good material for affect analysis because of their richness in evaluative and emotive information. Restricting the query scope of the Web mining technique was thus reasonable and I assumed it should improve the method. Since blogs provide specifically information on

user's emotions, evaluations and attitudes, such restriction is also assumed to be more accurate for the described task, than the Web content in general. Therefore, in the second version of the technique I restricted the Web mining process to the contents of *Yahoo! Japan - Blogs*[6], a robust blog service, one of the most popular in Japan.

In this work I focus on enhancing the human-computer interaction by processing emotions in more sophisticated manner than the present affect analysis methods. In most of the popular present approaches the emotions are divided into several classes [145], or sometimes simplified to only two - corresponding to positive and negative valence [141]. The method proposed here determines not only the valence and the specific type of emotions, but also verifies whether the expressed emotion is appropriate for the context. Moreover, the results of experiments drove me into even further conclusions.

4.2.1 Blogs as Generalized Human Conscience

The development of widely accessible wireless Internet brought to the light different kinds of activities unavailable till then. Today statistically every family has at least one personal computer and an access to the Internet [23]. Socialization of the Web as personal and social space caused a development of services to connect people from all over the world. One of that kind of services are blogs, open diaries in which people encapsulate their own experiences, opinions and feelings to be read and commented by other people. In this shape blogs have become an indicator of general human commonsense and conscience, and therefore came into the focus of scientific fields such as opinion mining, or sentiment and affect analysis [2]. As I assumed, the

[6]blogs.yahoo.co.jp

information hidden in blogs can be used for one other purpose, namely to help distinguish whether an emotion conveyed by a user during the interaction with an agent is appropriate for the situational context it is used in. As the blog service used for mining the desirable information I used *Yahoo! Japan - Blogs* Internet service as one of the most popular and robust blog services in Japan.

4.2.2 Methods

Affect Analysis

As the first step, of the method for verification of contextual appropriateness of emotions, I used the two affect analysis systems described in chapter 3. The affect analysis provides information on whether an utterance was emotive or not, and what type of emotion was expressed in the utterance. However, differently to the automatic evaluation method described in section 4.1, here I do not focus on all summarized emotion scores from the conversation. On the contrary, since it is desirable to spot any undesirable and inappropriate user behavior, the emotion appropriateness verification is performed for every emotionally emphasized utterance. In a conversation between a user and an agent, the affect analysis is performed on each utterance in user-agent conversation. For every emotive utterance with specified emotion type a Web mining technique is used as a verifier of emotion appropriateness.

Web Mining Technique

To verify the appropriateness of the speaker's affective states I applied Shi's et al. [125] Web mining technique for extracting emotive associations from the

Web. Ptaszynski et al. [103] 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 gathers from the Internet large number of examples and derives from this data an approximated reasoning about what emotion types usually associate with the 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) extracting context phrases from an utterance; 2) adding causality morphemes to the context phrases; 3) cross-referencing the modified phrases on the Web with the emotive lexicon and extracting emotion associations for each context phrase.

Phrase Extraction Procedure An utterance is first processed by MeCab, a tool for part-of-speech analysis of Japanese [63]. Every element separated by MeCab is treated as a unigram. All the unigrams are grouped into larger groups of n-grams 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. N-grams ending with particles are excluded, since they gave too many ambiguous results in pre-test phase. An example of phrase extraction

Table 4.3:
Example of context n-gram phrases separation from an utterance.

Original utterance	*Aa, pasokon ga kowarete shimatta...*					
English Translation	(Oh no, the PC has broken...)					
longest n-gram	(1) *Aa*	*pasokon*	*ga*	*koware*	*te*	*shimau*
(here: hexagram)	[interjection]	[N]	[SUBJ]	[V]	[GER]	[PRF]
pentagram	(2) *pasokon*	*ga*	*koware*	*te*	*shimau*	
tetragram	(3) *Aa*	*pasokon*	*ga*	*kowareru*		
trigrams	(4) *pasokon*	*ga*	*kowareru*	(5) *koware*	*te*	*shimau*

is presented in table 4.3.

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 [162] that Japanese people tend to convey emotive meaning after causality morphemes. This was independently confirmed experimentally by Shi et al. [125]. They distinguished eleven emotively stigmatized 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 the Web mining they decided to use only those five ones, namely: *-te*, *-to*, *-node*, *-kara* and *-tara* (see Table 4.4). An example of morpheme modification is presented on Table 4.5.

Table 4.4:
Hit-rate results for each of the eleven morphemes with the ones used in the Web mining technique in bold font.

morpheme	***-te***	***-node***	***-tara***	*-nara*	*-kotoga*	*-nowa*
result	**41.97%**	**7.20%**	**5.94%**	1.17%	0.35%	2.30%
morpheme	***-to***	***-kara***	**subtotal**	*-ba*	*-noga*	*-kotowa*
result	**31.97%**	**6.32%**	**93.4%**	3.19%	2.15%	0.30%

Table 4.5: Examples of n-gram modifications for Web mining.

Original n-gram	pasokon ga koware te shimau	/causality morpheme/
n-gram	pasokon ga koware te shimat-	/te/
phrase	pasokon ga koware te shimau	/to/
adjusting	pasokon ga koware te shimau	/node/
(morpheme	pasokon ga koware te shimau	/kara/
modification)	⋮	⋮

Emotion Type Extraction Procedure In this step the modified n-gram phrases are used as a query in Google search engine and 100 snippets for one morpheme modification per query phrase is extracted. This way a maximum of 500 snippets for each queried phrase is extracted. These are 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 a list of emotions associated with the queried sentence is obtained. It is the approximated emotive commonsense used further as an appropriateness indicator (an example is shown in Table 4.6).

Blog Mining The baseline Web mining method, using Google to search through the whole Web, was gathering a large amount of noise. To solve

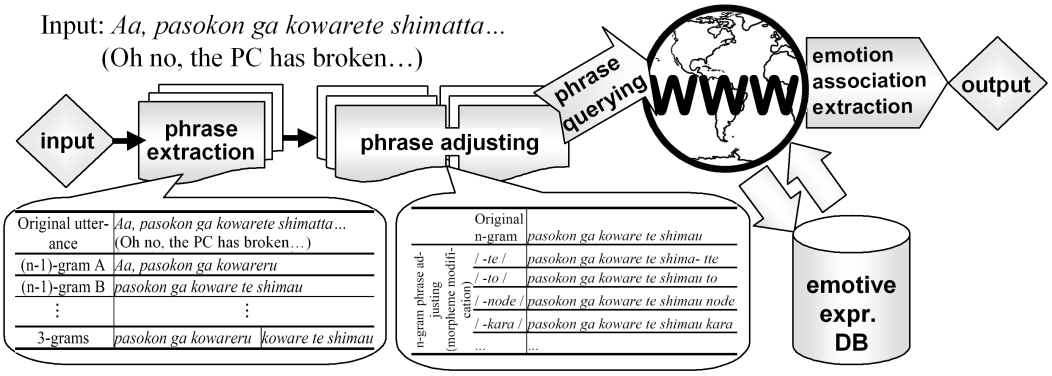


Figure 4.7: Flow chart of the Web mining technique.

this problem I made two modifications. Firstly, I 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, I 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 4.6 .

4.2.3 Contextual Appropriateness of Emotion Verification (CAEV) Procedure

The data obtained by affect analysis systems and the Web mining technique described above are combined as follows: when ML-Ask discovers an emotive

Table 4.6:
Example of emotion association extraction from the Web and its improvement by blog mining procedure.

Sentence: *Konpyūta 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)
surprize	30/284(0.105)	excitement	1/610 (0.001) [rejected
excitement	30/284(0.105)	fear	1/610 (0.001) as
fear	29/284(0.102)	relief	1/610 (0.001) noise]
⋮	⋮	⋮	⋮

utterance and the emotion types are successfully specified, 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 dialog 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!

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.

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 manage his/her emotions. Two examples are shown below.

Positive-inappropriate emotion:

User: I'd be happy if that bastard was hit by a car!

`ML-Ask: joy; Web mining: fear, sadness`

Agent: Are you sure that is what you are really feeling?

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!

Two-dimensional Model of Affect in CAEV Procedure

According to Solomon [131], people sometimes misunderstand the specific emotion types, but they rarely misunderstand their valence. One could, e.g., confuse anger with irritation, but not admiration with detestation. Therefore, I checked if at least the general features matched even when specific emotion types did not match perfectly with the emotive associations. By general features I refer to those proposed by Russell [114] in the theory of the two-dimensional model of affect (valence and activation). Using the mapping of Nakamura's emotion types on Russell's model I 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.

4.2.4 Evaluation Experiment

To evaluate the method I performed an experiment. In the experiment I used the chatlogs from the evaluation experiment of Modalin and Pundalin (see section 4.1.3). All 26 conversations were analyzed by ML-Ask. Six 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. I 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) both improvements, affect analysis supported with 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 five conversations, where differences in results appeared for the four compared versions of the procedure, 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 I obtained a total number of 250 different evaluations performed by different people.

For every conversation set I calculated how many of the human evaluators confirmed the system's results. The evaluated items were: A) specific emotion types determination; and B) general valence determination accuracies of affect analysis systems (ML-Ask and CAO); and the accuracy of the method as a whole (affect analysis verified by Web mining) to determine the contextual appropriateness of C) specific emotion types and D) valence.

Evaluation Criteria

A common problem in emotion processing research, is the number of evaluators employed in the evaluation process. In a third person evaluation, it is desirable to engage in the process of evaluation as many evaluators as possible to get a wide view on the results, calculate an overall agreement and rectify potential errors. Unfortunately there has been no standard for a desirable number of evaluators. In many research the evaluation is limited to, e.g., five people [147], three people [35] or even one [145]. The evaluation

with only one person, like in [145] assumes that if at least one person agrees with the system, the system has performed the evaluated task on a human level. The evaluation criteria where three people are employed [35] usually assume that at least two people from the group of three must agree about the evaluated object. The problem becomes complicated with a larger number of evaluators. In the evaluation performed by Tsuchiya et al. [147] there were five people employed in the evaluation. According to their explanations, a) if four or five people agreed with the system, the results were positive; b) if three or two people agreed with the system, the results were acceptable; c) if only one person or no person agreed with the system, the results were negative.

It can be easily noticed, that although the number of evaluators grows with the introduced approach, the general idea is that at least one person needs to agree with the system for the results to be positive (first approach), or the results are negative only in the case when one person or less from a larger group agree with the system (two latter approaches). These somewhat lenient conditions in evaluation of emotion processing-related systems comes from the fact that it is difficult to obtain a perfect agreement between people about emotion-related topics, since the cognition of emotions in people is highly subjective and context related.

Therefore I decided to look on the results from a more analytical point of view. In my research, apart from the 13 people who took part in conversations with the agents, I looked to evaluate every questionnaire set 10 times. Then I checked how many people agreed with the results given by the system. Since every questionnaire set was evaluated 10 times, a number of agreements for each evaluated item (A-D) in all twenty evaluated cases could be from 0 to 10.

Table 4.7: Criteria conditions and naming of the level of agreements.

Criteria conditions	**No. of people who had to agree**
ideal	all 10
rigorous	at least 9
grand majority	at least 8
fair	at least 7
weak majority	at least 6
medium	at least 5
optimistic	at least 4
easy	at least 3
lenient	at least 2
negligible	at least 1
no agreement	0

When comparing the four versions of the method mentioned in section 4.2.4, I assumed the better version of the system is the one which achieved more agreements with a larger number of evaluators. I took into the consideration all types of agreement conditions, from ideal (all 10 people agreed) to the smallest one (negligible), applied in the research described above (at least one person had to agree with the system). To visualize the results I named every level of agreement like in the table 4.7. Each level of agreement assumed that the results are positive if at least the specified number of evaluators agreed with the system. To verify whether the agreements are statistically significant I independently calculated a multi-rater kappa [111] for all sets of the results.

4.2.5 Results and Discussion

I analyzed three aspects of the results. Firstly, I focused on evaluation of the affect analysis procedure. Although the two affect analysis systems were

evaluated separately previously in chapter 3, there was no overall evaluation. This corresponds to answers to the questions A) and B) from the questionnaire. Secondly, I summarized the results of the CAEV procedure for all of the results considered together. This corresponds to questions C) and D) from th questionnaire. Finally, I analyzed the results separately for the two agents to check whether there were any differences between verifying the emotion appropriateness in a usual conversational agent (Modalin) and the joking agent (Pundalin).

Evaluation of Affect Analysis Procedure

The first part of the evaluation process consisted in evaluation of affect analysis procedure. The results were as follows. For all possible agreements of the system with 10 evaluators about the 20 evaluation sets (200 possible cases of agreement) baseline version of affect analysis obtained 110 (55%) agreements about determining emotion type and 126 (63%) agreements about determining valence. Statistical strength of agreements in this setting was κ=0.66 and κ=0.68, respectively, which indicates that both sets of agreements were statistically significant. As for the affect analysis procedure upgraded with CVS (ML-Ask last part of the procedure; see section 3.1.4 for details), there were 120 agreements (60%) for emotion types and 138 (68%) for valence determination. Statistical strength of these sets of agreements was κ=0.66 and κ=0.68, respectively. As for the distribution of the agreements, most of the results for emotions types (over 50% of all actual agreements) were enclosed in a group where at least 8 people agreed with the system (grand majority conditions). Similarly, most of the results for valence were enclosed in a group where at least 9 people agreed with the system (nearly ideal, rigorous condi-

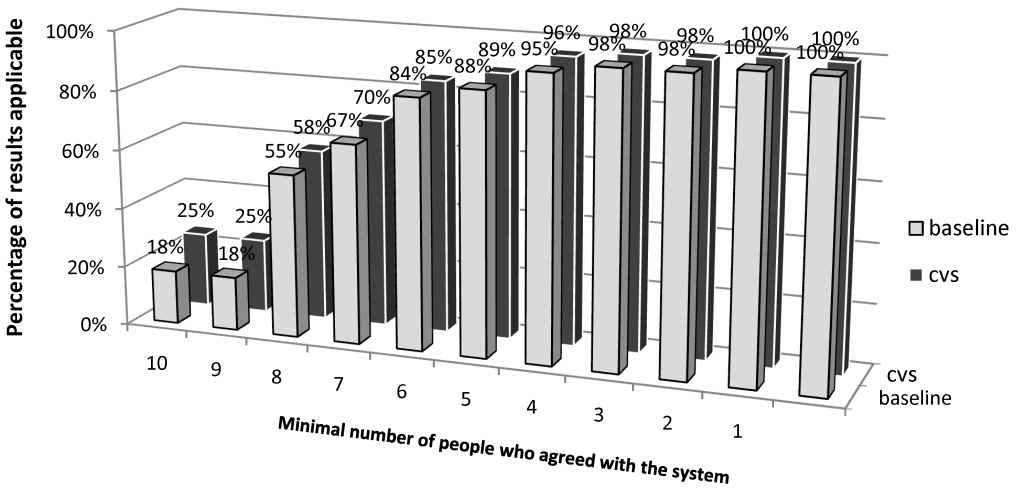


Figure 4.8:
Visualization of the distribution of agreements for both versions of affect analysis procedure in determining about emotion types. Figure corresponds to the upper part of Table 4.8.

tions). For the negligible condition ("at least one person"), often applied in other research, the results enclosed 100% of cases. However, ideal conditions (all agree) appeared from 18% to 29% of cases, which shows that the negligible condition is far from objectivity. However, the grand majority of the results (over 80%) were enclosed in a group where at least 6 people agreed with the system. The conditions including medium (at least 5 people) and more relaxed conditions enclosed from nearly 90% and above. The results are represented in Table 4.8. Visualization of the distribution of agreements for both versions of affect analysis procedure is represented on Figure 4.8 (for emotion type determination) and Figure 4.9 (for valence determination).

Table 4.8:
Results for evaluation of affect analysis procedure. Upper part of the table: results for specifying emotion **types**; Lower part: results for specifying **valence**. The table presents numbers of people who agreed with the system. Distribution of numbers shows how many there were agreements with how many people; **% of all**: shows percentage of this group of agreements within all agreements; **% sums**: shows percentage of results applicable when the condition for agreement was set as "at least this group of agreements (or higher)"; **agr.ratio**: overall number of agreements divided by ideal number of agreements and ratio; **kappa**: statistical strength of agreements in this setting.

A)	**Number of people who agreed with the system**											**agr.ratio**
TYPES	10	9	8	7	6	5	4	3	2	1	0	**(kappa)**
baseline	2	0	5	2	3	1	2	1	0	2	2	110/200
% of all	18%	0%	36%	13%	16%	5%	7%	3%	0%	2%	0%	55%
% sums	18%	18%	55%	67%	84%	88%	95%	98%	98%	100%	100%	(κ= 0.66)
cvs	3	0	5	2	3	1	2	1	0	2	1	120/200
% of all	25%	0%	33%	12%	15%	4%	7%	3%	0%	2%	0%	60%
% sums	25%	25%	58%	70%	85%	89%	96%	98%	98%	100%	100%	(κ= 0.66)

B)	**Number of people who agreed with the system**											**agr.ratio**
VALENCE	10	9	8	7	6	5	4	3	2	1	0	**(kappa)**
baseline	3	4	2	1	2	3	0	2	2	0	1	126/200
% of all	24%	29%	13%	6%	10%	12%	0%	5%	3%	0%	0%	63%
% sums	24%	52%	65%	71%	80%	92%	92%	97%	100%	100%	100%	(κ= 0.68)
cvs	4	4	2	1	2	3	0	2	2	0	0	136/200
% of all	29%	26%	12%	5%	9%	11%	0%	4%	3%	0%	0%	68%
% sums	29%	56%	68%	73%	82%	93%	93%	97%	100%	100%	100%	(κ= 0.68)

Evaluation of CAEV Procedure: General

Secondly I checked the results for the determination of emotion appropriateness by the CAEV procedure. The results were as follows. For all possible agreements of the system with 10 evaluators about the 20 evaluation sets (200 possible cases of agreement) baseline version of CAEV procedure obtained 69 (35%) agreements about determining emotion type and 85 (43%) agreements about determining valence. Statistical strength of agreements in this setting was κ=0.652 and κ=0.677, respectively, which indicates that both

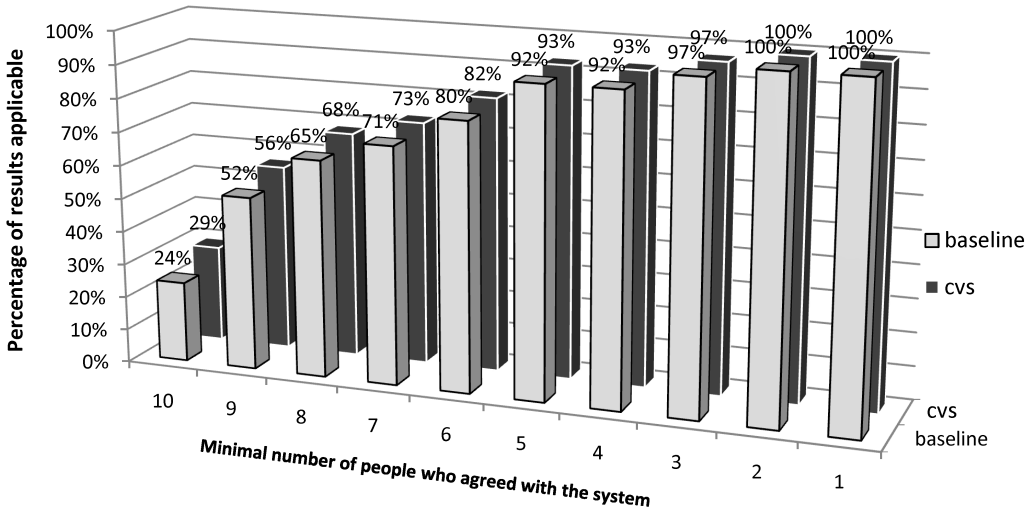


Figure 4.9:
Visualization of the distribution of agreements for both versions of affect analysis procedure in determining about valence of emotions. Figure corresponds to the lower part of Table 4.8.

sets of agreements were statistically significant. As for the version of CAEV procedure with affect analysis upgraded with CVS, there were 78 agreements (39%) for emotion types and 94 (47%) for valence determination. Statistical strength of these sets of agreements was κ=0.642 and κ=0.667, respectively. As for the version of CAEV procedure with Web mining restricted to blogs, there were 81 agreements (41%) for emotion types and 95 (48%) for valence determination. Statistical strength of these sets of agreements was κ=0.667 and κ=0.643, respectively. Finally, for the version of CAEV procedure with both improvements (ML-Ask upgraded with CVS and Web mining restricted to blogs), there were 90 agreements (45%) for emotion types and 104 (52%) for valence determination. Statistical strength of these sets of agreements was κ=0.643 and κ=0.633, respectively.

As for the distribution of the agreements, the majority of the results (over 50% of all actual agreements) for determining about appropriateness of emotion types were enclosed in a group where at least 5 people agreed with the system (medium conditions). For determining about valence appropriateness, most of the results were enclosed in a group where at least 8 people agreed (grand majority conditions). The results which enclosed at least 80% of agreements oscillated for emotion types verification around groups where at least three (easy) to four (optimistic) people agreed. For valence verification the groups enclosing at least 80% of agreements oscillated from optimistic (at least 4) to medium (at least 5) condition group. Although there were no cases with ideal conditions, the best version of the system (both improvements) encapsulated with the use of grand majority condition (at least 8 people) 48% of results for emotion types and 64% for valence.

The results are represented in Table 4.9. Visualization of the distribution of agreements for all four versions of CAEV procedure is represented on Figure 4.11 (for emotion type determination) and Figure 4.13 (for valence determination). The visualization of percentage of results encapsulated for each condition (from "at least 9" to "at least 1") is presented on Figure 4.10 (for emotion type determination) and Figure 4.12 (for valence determination). Some of the successful examples are represented in Table 4.13.

Evaluation of CAEV Procedure: Agents Separately

Finally, I checked the results for the verification of emotion appropriateness by the CAEV procedure separately for each agent, Modalin and Pundalin. This was done to find out whether verification of emotion appropriateness

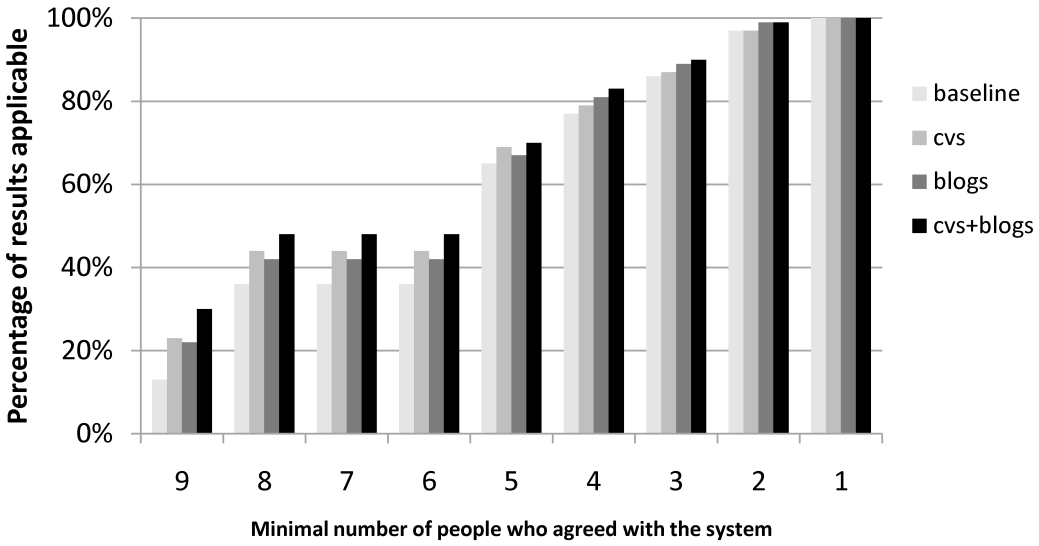


Figure 4.10:
Visualization of percentage of results encapsulated for each condition, from "at least 9" to "at least 1" (for emotion type determination).

differs for different types of conversations. Modalin is a non-task oriented keyword-based conversational agent, which uses modality to enhance dialog propositions extracted from the Web. Apart from this, the agent has no distinctive features. Modalin was designed by Higuchi et al. [45]. For detailed description see section 4.1.3. Pundalin is also a non-task oriented conversational agent. It was created by adding to Modalin a pun generating system developed by Dybala et al. [30]. Therefore the only distinctive feature of Pundalin with comparison to Modalin was using puns. I compared the results achieved by the agents separately to check whether the presence of jokes (puns) helps or interrupts the process of verification of emotion appropriateness.

The results were as follows. The overall number of agreements with hu-

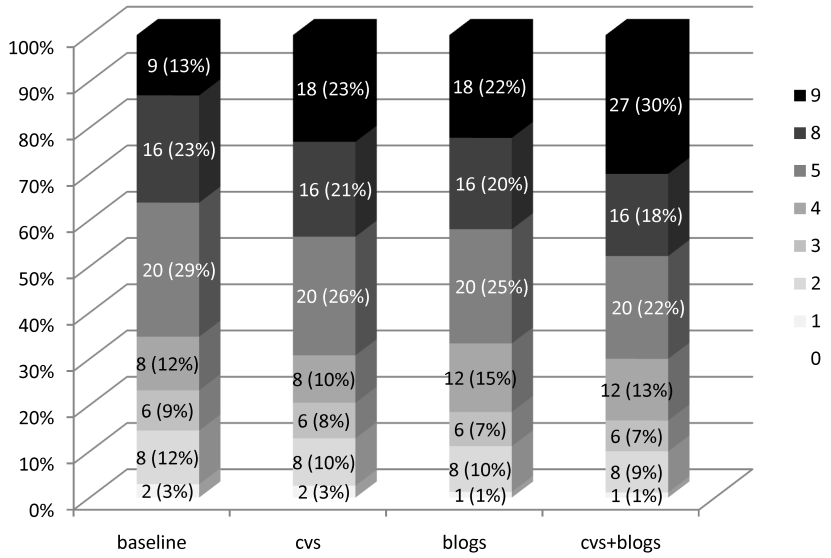


Figure 4.11:
Visualization of the distribution of agreements for all four versions of CAEV procedure (for emotion type determination).

man evaluators about verification of contextual appropriateness was better for Modalin (from 41% to 55%) than for Pundalin (from 28% to 49%). This was true for both, emotion types and valence. Also the number of agreements encapsulated for different conditions showed similar tendency. The conditions which encapsulated at least 50% of agreements were, for Pundalin/emotion types, from medium (at least 5 people agreed) to rigorous (at least 9 people agreed). For Pundalin/valence, the results were from medium to grand majority (at least 8 people agreed). The same results for Modalin were approximately higher. For emotion types the condition that encapsulated most of the agreements (over 70%) was a stable medium condition, and for valence it was a stable condition of grand majority (at least 8 peo-

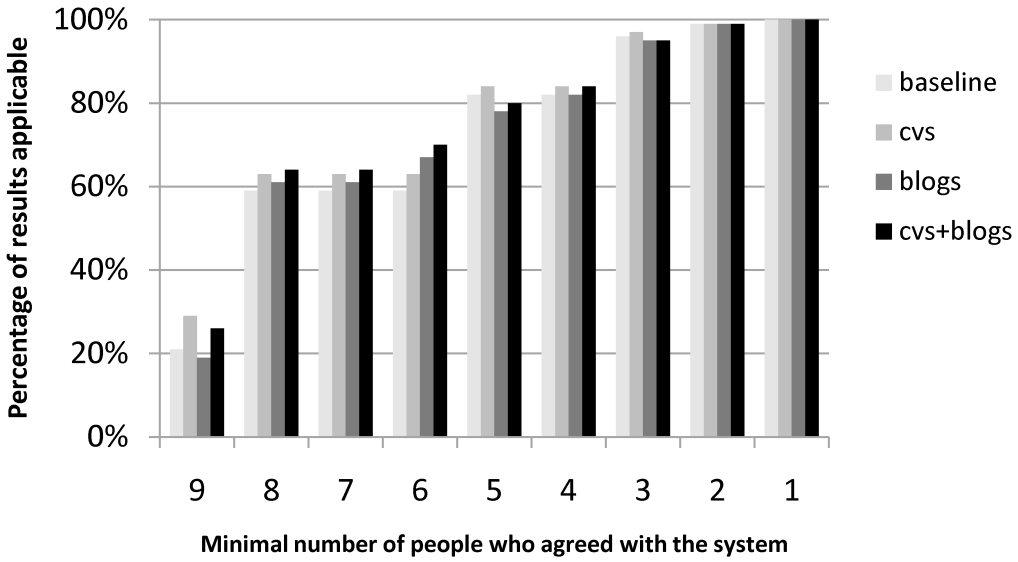


Figure 4.12:
Visualization of percentage of results encapsulated for each condition, from "at least 9" to "at least 1" (for valence determination).

ple agreed). The conditions that enclosed over 80% of the results oscillated around optimistic and medium conditions. The statistical strength of agreements was considerably high with kappa oscillating from 0.573 to 0.697 across all separate results. The results are summarized for Modalin in Table 4.10 and for Pundalin in Table 4.11.

These results indicate that contextual appropriateness was more difficult to verify in the conversations with pun-telling agent. It is reasonable, since humor is said to be one of the most creative and therefore difficult tasks in Artificial Intelligence [12].

Both improvements, the one with CVS procedure and the one limiting the query scope in the Web mining procedure to search only through blog

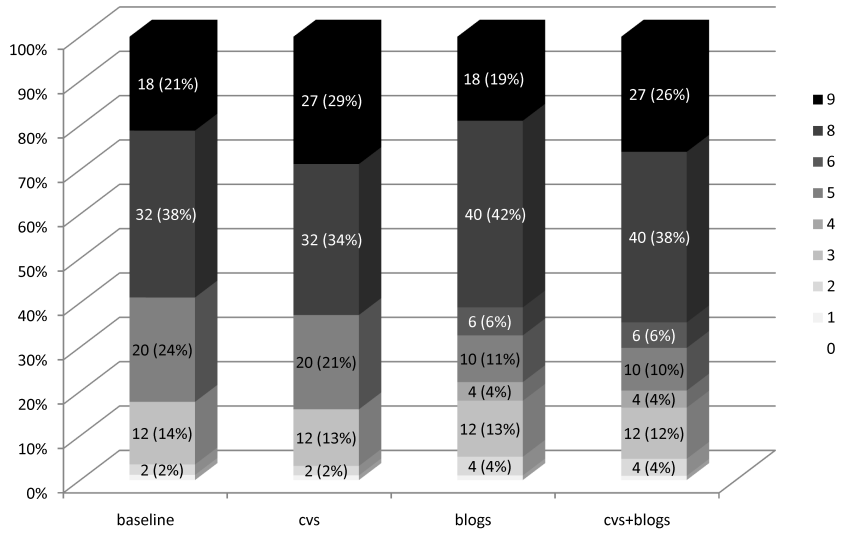


Figure 4.13:
Visualization of the distribution of agreements for all four versions of CAEV procedure (for valence determination).

contents, positively influenced the performance of the Contextual Appropriateness of Emotion Verification procedure, in all of the cases for both of the agents. The improvement was noticeable both on the level of specific emotion types and of valence, and also for the result of both agents taken together as well as separately.

The most effective version of the method was the one with both improvements applied, by which the system's performance (number of agreements with evaluators) was generally improved for all considered cases. For example, for the grand majority condition (at least 8 people agreed) the results were improved from 36% to 48% (emotion types) and from 59% to 64% (valence), with the highest score achieved by Modalin (75%).

In almost all cases the results which changed after the improvement were

statistically significant on a 5% level (see Table 4.12). 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 method and with CVS, were statistically significant (P value = 0.0274) and, what is the most important, the results of the version fully improved were the most significant of all (P value = 0.0119).

Although the method for verification of contextual appropriateness of emotion, presented here, is still not ideal, the increase in results after implementation of different improvements to the intermediary systems (ML-Ask in affect analysis procedure and Web mining) indicate the method is easily improvable. Considering the further enhancements that are already in plans, I am expecting a high improvement of this method in the near future.

4.2.6 Emotion Appropriateness as "Conscience Calculus": Implications Towards Computational Conscience.

As mentioned above, expressing and understanding emotions is one of the most important cognitive human behaviours present in everyday communication. In particular, Salovey and Mayer [121] showed that emotions are a vital part of human intelligence, and Schwarz [124] showed, that emotional states influence the decision making process in humans. If the process of decision making is defined 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 devel-

opment of human conscience was proved by Thompson and colleagues [143] 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. [116], who defined 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. [119], 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 [153], Human Aspects in Ambient Intelligence [146], and in Artificial Intelligence in general. In my research I performed 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. I understand, that the idea of conscience is far more sophisticated, however, when defined narrowly as the ability to distinguish between what is appropriate and what is inappropriate, my method for verifying contextual appropriateness of emotions could be applied to obtain simplified conscience calculus for machines. I 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., [120].

Table 4.9:
Results for evaluation of Contextual Appropriateness of Emotion Verification (CAEV) Procedure. Upper part of the table: results for specifying emotion **types**; Lower part: results for specifying **valence**. The table presents numbers of people who agreed with the system. Distribution of numbers shows how many there were agreements with how many people; **% of all**: shows percentage of this group of agreements within all agreements; **% sums**: shows percentage of results considered when the condition for agreement was set as "at least this group of agreements (or higher)"; **agr.ratio**: overall number of agreements divided by ideal number of agreements and ratio; **kappa**: statistical strength of agreements in this setting.

C) TYPES	**Number of people who agreed with the system** 9	8	7	6	5	4	3	2	1	0	**agr.ratio (kappa)**
BASELINE	1	2	0	0	4	2	2	4	2	3	69/200
% of all	13%	23%	0%	0%	29%	12%	9%	12%	3%	0%	35%
% sums	13%	36%	36%	36%	65%	77%	86%	97%	100%	100%	(κ= 0.652)
CVS	2	2	0	0	4	2	2	4	2	2	78/200
% of all	23%	21%	0%	0%	26%	10%	8%	10%	3%	0%	39%
% sums	23%	44%	44%	44%	69%	79%	87%	97%	100%	100%	(κ= 0.642)
BLOGS	2	2	0%	0%	4	3	2	4	1	2	81/200
% of all	22%	20%	0%	0%	25%	15%	7%	10%	1%	0%	41%
% sums	22%	42%	42%	42%	67%	81%	89%	99%	100%	100%	(κ= 0.667)
CVS+BLOGS	3	2	0%	0%	4	3	2	4	1	1	90/200
% of all	30%	18%	0%	0%	22%	13%	7%	9%	1%	0%	45%
% sums	30%	48%	48%	48%	70%	83%	90%	99%	100%	100%	(κ= 0.643)

D) VALENCE	**Number of people who agreed with the system** 9	8	7	6	5	4	3	2	1	0	**agr.ratio (kappa)**
BASELINE	2	4	0	0	4	0	4	1	1	4	85/200
% of all	21%	38%	0%	0%	24%	0%	14%	2%	1%	0%	43%
% sums	21%	59%	59%	59%	82%	82%	96%	99%	100%	100%	(κ= 0.677)
CVS	3	4	0	0	4	0	4	1	1	3	94/200
% of all	29%	34%	0%	0%	21%	0%	13%	2%	1%	0%	47%
% sums	29%	63%	63%	63%	84%	84%	97%	99%	100%	100%	(κ= 0.667)
BLOGS	2	5	0	1	2	1	4	2	1	2	95/200
% of all	19%	42%	0%	6%	11%	4%	13%	4%	1%	0%	48%
% sums	19%	61%	61%	67%	78%	82%	95%	99%	100%	100%	(κ= 0.643)
CVS+BLOGS	3	5	0	1	2	1	4	2	1	1	104/200
% of all	26%	38%	0%	6%	10%	4%	12%	4%	1%	0%	52%
% sums	26%	64%	64%	70%	80%	84%	95%	99%	100%	100%	(κ= 0.633)

Table 4.10:
Results for evaluation of Contextual Appropriateness of Emotion Verification (CAEV) Procedure for Modalin. Description of table contents like in Table 4.9.

MODALIN								
C) TYPES	**Number of people who agreed with the system** 8	5	4	3	2	1	0	**agr.ratio (kappa)**
BASELINE	2	3	1	1	1	1	1	41/100
% of all	39%	37%	10%	7%	5%	2%	0%	41%
% sums	39%	76%	85%	93%	98%	100%	100%	(κ= 0.606666)
CVS	2	3	1	1	1	1	1	41/100
% of all	39%	37%	10%	7%	5%	2%	0%	41%
% sums	39%	76%	85%	93%	98%	100%	100%	(κ= 0.606666)
BLOGS	2	3	2	1	1	0	1	44/100
% of all	36%	34%	18%	7%	5%	0%	0%	44%
% sums	36%	70%	89%	95%	100%	100%	100%	(κ= 0.573333)
CVS+BLOGS	2	3	2	1	1	0	1	44/100
% of all	36%	34%	18%	7%	5%	0%	0%	44%
% sums	36%	70%	89%	95%	100%	100%	100%	(κ= 0.573333)
D) VALENCE	**Number of people who agreed with the system** 9	8	5	4	3	2	0	**agr.ratio (kappa)**
BASELINE	1	4	1	0	2	0	2	52/100
% of all	17%	62%	10%	0%	12%	0%	0%	52%
% sums	17%	79%	88%	88%	100%	100%	100%	(κ= 0.688888)
CVS	1	4	1	0	2	0	2	52/100
% of all	17%	62%	10%	0%	12%	0%	0%	52%
% sums	17%	79%	88%	88%	100%	100%	100%	(κ= 0.688888)
BLOGS	1	4	1	0	3	0%	1	55/100
% of all	16%	58%	9%	0%	16%	0%	0%	55%
% sums	16%	75%	84%	84%	100%	100%	100%	(κ= 0.642222)
CVS+BLOGS	1	4	1	0	3	0%	1	55/100
% of all	16%	58%	9%	0%	16%	0%	0%	55%
% sums	16%	75%	84%	84%	100%	100%	100%	(κ= 0.642222)

Table 4.11:
Results for evaluation of Contextual Appropriateness of Emotion Verification (CAEV) Procedure for Pundalin. Description of table contents like in Table 4.9.

PUNDALIN										
C) TYPES	**Number of people who agreed with the system** 9	8	6	5	4	3	2	1	0	**agr.ratio (kappa)**
BASELINE	1	0	0	1	1	1	3	1	2	28/100
% of all	32%	32%	32%	18%	14%	11%	21%	4%	0%	28%
% sums	32%	32%	32%	50%	64%	75%	96%	100%	100%	(κ= 0.698)
CVS	2	0	0	1	1	1	3	1	1	37/100
% of all	49%	49%	49%	14%	11%	8%	16%	3%	0%	37%
% sums	49%	49%	49%	62%	73%	81%	97%	100%	100%	(κ= 0.678)
BLOGS	2	0	0	1	1	1	3	1	1	37/100
% of all	49%	49%	49%	14%	11%	8%	16%	3%	0%	37%
% sums	49%	49%	49%	62%	73%	81%	97%	100%	100%	(κ= 0.678)
CVS+BLOGS	3	0	0	1	1	1	3	1	0	46/100
% of all	59%	59%	59%	11%	9%	7%	13%	2%	0%	46%
% sums	59%	59%	59%	70%	78%	85%	98%	100%	100%	(κ= 0.658)
D) VALENCE	**Number of people who agreed with the system** 9	8	6	5	4	3	2	1	0	**agr.ratio (kappa)**
BASELINE	1	0	0	3	0	2	1	1	2	33/100
% of all	27%	0%	0%	45%	0%	18%	6%	3%	0%	33%
% sums	27%	27%	27%	73%	73%	91%	97%	100%	100%	(κ= 0.664)
CVS	2	0	0	3	0	2	1	1	1	42/100
% of all	43%	0%	0%	36%	0%	14%	5%	2%	0%	42%
% sums	43%	43%	43%	79%	79%	93%	98%	100%	100%	(κ= 0.644)
BLOGS	1	1	1	1	1	1	2	1	1	40/100
% of all	23%	20%	15%	13%	10%	8%	10%	3%	0%	40%
% sums	23%	43%	58%	70%	80%	88%	98%	100%	100%	(κ= 0.644)
CVS+BLOGS	2	1	1	1	1	1	2	1	0	49/100
% of all	37%	16%	12%	10%	8%	6%	8%	2%	0%	49%
% sums	37%	53%	65%	76%	84%	90%	98%	100%	100%	(κ= 0.624)

Table 4.12:
Statistical significance of differences between the results for different versions of the system.

	Versions of the methods			
	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 4.13:
Three examples of the results provided by the emotion appropriateness verification procedure (CAVP) with a separate display of the examples showing the improvement of the procedure after applying CVS.

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

Chapter 5

Concluding Remarks and Further Work

In this dissertation I presented my work on developing affect analysis systems and applying them to enhance human-computer interaction. I developed two systems for affect analysis and two methods for enhancement of human-computer interaction making use of those systems.

The first system for affect analysis I developed, was ML-Ask, a system for affect analysis of textual input utterance in Japanese.

I first performed a study on the emotive function of language in Japanese. Basing on that study I created ML-Ask. The system was developed for automatic annotation of corpora with emotive information. A need for such system is expressed in many research on emotion processing. To verify the system's utility I performed a series of experiments. I verified how much the discrimination between emotive and neutral utterances differs between linguistic and laypeople viewpoint. I found out that, linguistic approach is in a strong agreement with the layperson one, and the linguistic examples can be successfully utilized in creating systems like mine for other languages.

After the positive verification of the system's utility I performed an annotation of a large corpus containing discussions from a popular Japanese forum *2channel*. In comparison with manual annotation of the same coprus I found

out that the system was sufficiently (and significantly) successful in providing information about tendencies of emotive utterances in conversation, number of extracted tokens and rank setting for the most frequently conveyed emotions - crucial tasks in, monitoring of Internet forums for security reasons, gaining on importance by the day.

ML-Ask system answers the main problem present in NLP methods for affect recognition - it can determine if an utterance is emotive or not. To do that it extracts emotemes - indicators of the emotive function of language and provides a detailed description of the structure of an emotive utterance.

ML-Ask achieved a high accuracy result of 90% in recognizing the general emotiveness of an utterance. A speaker-specific evaluation of affect analysis procedure showed that the system recognizes specific types of emotions conveyed in utterances on a fair but improvable level of 0.47 balanced F-score. This level was also confirmed the observer-specific evaluation (0.45).

The system is applicable as an affect recognition system, however its accuracy in this matter is not ideal, since to specify the particular emotion types it uses a dictionary of emotive expressions - a source reliable, but already out of date. However, this can be improved in several ways, such as **1)** updating the emotive expressions lexicon, which is a simple but laborious task, or **2)** statistical disambiguation of emotemes by attaching to them potential emotive affiliations (e.g. an exclamation mark would be used with "excitement", rather than with "gloom"). There was also a problem with inability of the system to process emoticons. I address this problem in the next section. The system can be used in real-time applications, since the approximate time for processing one utterance is less than 0.15 s.

Finally, since I verified that ML-Ask can be used as a tool for automatic

annotation of corpora with emotive information, in the future I plan to continue experiments with the system and perform a large scale annotation of another corpora with a particular focus on those containing natural dialogs to find out how emotive utterances function within context.

The second system for affect analysis I developed, was CAO *a system for emotiCon Analysis and decOding of affective information.* CAO is a prototype system for automatic affect analysis of Eastern type emoticons. The system was created using a database of emoticons containing over ten thousand of unique emoticons collected from the Internet. These emoticons were automatically distributed into emotion type databases with the use of the previously developed ML-Ask. Finally, the emoticons were automatically divided into semantic areas, such as mouths or eyes and their emotion affiliations were calculated based on occurrence statistics. The division of emoticons into semantic areas was based on Birdwhistell's [10, 11] idea of kinemes as minimal meaningful elements in body language. The database applied in CAO contains over ten thousand raw emoticons and several thousands of elements for each unique semantic area (mouths, eyes, etc.). The evaluation on both the training set and the test set showed that the system outperforms previous methods, achieving results close to ideal, and has other capabilities not present in its predecessors: detecting emoticons in input with very strong agreement coefficient (kappa = 0.95); and extracting emoticons from input and dividing them into semantic areas, which, calculated using balanced F-score, reached over 97%. The system estimated emotions of separate emoticons with an accuracy of 93.5% for the specific emotion types and 97.3% for groups of emotions belonging to the same two dimensional affect space [114].

At present CAO is the most accurate and reliable system for emoticon analysis known to the author. In the near future I plan to apply it to numerous tasks. Beginning with contribution to computer-mediated communication, I plan to make CAO a support tool for e-mail reader software. Although emoticons are used widely in online communication, there is still a wide spectrum of users (often elderly), who do not understand the emoticon expressions. Such users, when reading a message including emoticons, often get confused which causes future misunderstandings with other people. CAO could help such users interpret the emoticons appearing in e-mails. As processing time in CAO is very short (processing of both training and test sets took no more than a few seconds), this application could be also extended to instant messaging services to help interlocutors understand each other in the text based communication. As a support system for Affect and Sentiment Analysis systems, such as [107], CAO could also contribute to preserving online security, which has been an urgent problem for several years [1]. To standardize emoticon interpretation I plan to contribute to the Smiley Ontology Project [110]. Finally, I plan to annotate large corpora of online communication, like Yacis Corpus, to contribute to linguistic research on emotions in language.

Those two systems are used in an affect analysis procedure applied in two methods for enhancing human-computer interaction.

As the first application of the two affect analysis systems (ML-Ask and CAO), I presented a method of automatic evaluation for conversational agents. The method is based on analyzing affective states conveyed by a user in a conversation with an agent. Borrowing the notion of Affect-as-Information [124], the results of affect analysis performed by ML-Ask and CAO provide

information about the user's emotional involvement in a conversation, his/her psychological distance, and ease of familiarization with the machine. This corresponds to direct questions about the agent's performance. Next, analysis of specified emotion types conveyed by the user in the whole conversation and their classification by applying the two-dimensional model of emotions [114] provides information on the polarity of the users' attitudes towards the machine interlocutor during the conversation.

By applying the proposed method in evaluation of conversational agents, the evaluative information is acquired during the conversations of user-testers with the agents. Therefore as means of evaluation, the method saves time, effort and funds spent each time on preparing and performing laborious questionnaires. It is desirable for the proposed method to be accepted widely in the field as a full equivalent or at least a strong supportive means to objectivize the results of traditional questionnaires.

The method, although proven to be effective, still has still some deficiencies which I aim to rectify in the near future. The imperfections of the sub-systems used in the method influence its accuracy. The slight deficiency in the emotion types extraction procedure in ML-Ask limit the information about affective states conveyed by users in conversation. However, it can be predicted that applying the two-dimensional model of emotions into assigning emotional affiliations of emotive elements will disambiguate the emotional affiliations of emotive elements, thus improving the performance of ML-Ask. Some ideas about the ways to improve the system were already proposed by Ptaszynski and colleagues [104]. I plan to implement them in the near future.

The method should be also tested on other agents than the two presented here. Dybala and colleagues, after adding some improvements mentioned

above, have already used this method to evaluate two different conversational agents [33]. However, the differences between their agents were similar to the two agents compared in this chapter - one was a simple conversational agent (HMM based) and the second one used jokes, although the appropriate timing for joke generation was not set arbitrarily every third turn, like here, but was based on analysis of the emotional states of the users. Therefore, it is desirable to verify the usability of this automatic evaluation method also on conversational agents which differ in features other than the generation of humorous responses.

The notion of affect-as-information, although with a firm scientific background in psychology and social psychology [16, 17], is not a common notion in the fields I referred to in this chapter, Agent Development, Evaluation Methods, Affect Analysis, or Artificial Intelligence in general. The mapping of questions on the results of affect analysis, although supported with strong theory, is still rather intuitive. Therefore, I will aim to make the mapping more precise in future by looking for the questions that correlate strongest with the automatic evaluation. However, in this experiment I tried to prove that affective states do influence judgments and attitudes towards agents and, properly analyzed, reveal similar tendencies to usual evaluation questionnaires, providing valuable and important information in evaluation - a significant part of the product design process.

In the last part of the research described in this dissertation I presented a novel method for estimating contextual appropriateness of emotions. The method is composed of two parts, a language based affect analysis procedure utilizing two affect analysis systems developed previously (used as an emotion detector), and a Web mining technique for extracting from the Internet

lists of emotional associations considered as a generalized emotive common sense (used as an emotion verifier). I checked the performance of four versions of the method. The affect analysis procedure is compared with and without Contextual Valence Shifters. As for the Web mining technique, two versions are compared: one, using all the Internet resources and a second one improved by restricting the search scope to the contents of blog documents. The improvements positively influenced the results and were statistically significant. I observed that emotion appropriateness was difficult to determine in conversations containing puns.

The method provides the conversational agent with computable means to determine whether emotions expressed by a user are appropriate for the context they appear in. A conversational agent equipped with this method could be provided with hints about what communication strategy would be the most desirable at a certain moment. For example, a conversational agent could choose to either sympathize with the user or take precautions and help the user manage his/her feelings. By proposing computational means for verification of contextual appropriateness of emotional states in conversation, this research defines a new set of goals for Affective Computing. Enhancing a conversational agent with this ability is a step toward implementation of the full scope of Emotional Intelligence in machines.

The theory of Emotional Intelligence, to which I referred in this dissertation, is a quickly developing field of research. It frequently delivers new discoveries about the structures and functions of emotions and therefore should be in focus of researchers attempting to develop means for computing human intelligence. To create a machine capable to communicate with user on a human level, there is a need to equip it with an Emotional Intelligence

Framework [121]. Implementation of the full scope of Emotional Intelligence Framework is the key task on the way to full implementation of Emotional Intelligence in machines and therefore is a valuable research in the field of Artificial Intelligence in general.

I showed that computing emotions in context is a feasible task. Although the system as a whole is still not perfect and its components (ML-Ask and the Web mining technique) need further improvement, there have been seen a significant improvement by restricting Web mining to the contents of *Yahoo! Japan - Blogs*. As for the future work, I plan to focus on deepening the understanding of emotions by bootstrapping the context phrases. For example, in a sentence "I'm so depressed since my girlfriend left me..." the context phrase would be "girlfriend left". The Web mining procedure provides for such phrases a list of appropriate emotions. However, using similar Web mining procedure I plan to go further and find out the reason for an emotion object to happen. For example, to find out "why girls leave their boyfriends?". An answer for this question, found in the Internet, could be, e.g., "because boys are not sporty enough", or "because boys have no money". Next asked question could be, e.g., "why boys have no money?", etc. Sufficient accuracy in such bootstrapping method would provide a deeper knowledge about the causality of experiences. When applied in a companion agent this would help providing hints about probable undesirable consequences of user activities.

References

[1] Ahmed Abbasi and Hsinchun Chen. Affect Intensity Analysis of Dark Web Forums. *Intelligence and Security Informatics 2007*, pp. 282-288, 2007.

[2] Shuya Abe, Moe Eguchi, Asuka Sumida, Azusa Ohsaki, Kentaro Inui. *Minna no keiken: Burogu kara chūshutsu shita ibento oyobi senchimento no DB-ka* [Everyone's experiences: Creating a Database of Events and Sentiments Extracted from Blogs] (in Japanese), In *Proceedings of NLP-2009*, pp. 296-299, 2009.

[3] Cecilia Ovesdotter Alm, Dan Roth and Richard Sproat. Emotions from text: machine learning for text based emotion prediction. In *Proc. of HLT/EMNLP*, pp. 579-586, 2005.

[4] Elisabeth André, Matthias Rehm,Wolfgang Minker, and Dirk Bühler. Endowing Spoken Language Dialogue Systems with Emotional Intelligence. *LNCS*, Vol. 3068, pp. 178-187, 2004.

[5] Richard E. Aquila. Emotions, Objects and Causal Relations, *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 26, No. 3/4, pp. 279-285, 1974.

[6] Junko Baba. 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, 2003.

[7] Adel Salem Bahameed. Hindrances in Arabic-English Intercultural Translation. *Translation Journal*, Vol. 12, No. 1, 2008.

[8] Monika Bednarek. *Emotion Talk Across Corpora*. Palgrave Macmillan, 2008.

[9] Fabian Beijer. The syntax and pragmatics of exclamations and other expressive/emotional utterances. *Working Papers in Linguistics 2*, The Dept. of English in Lund, 2002.

[10] Ray L. Birdwhistell. *Introduction to kinesics: an annotation system for analysis of body motion and gesture*, University of Kentucky Press, 1952.

[11] Ray L. Birdwhistell, *Kinesics and Context*, University of Pennsylvania Press, Philadelphia, 1970.

[12] Margaret A. Boden. Creativity and artificial intelligence. *Artificial Intelligence*, Vol. 103, No. 1-2, pp. 347-356, 1998.

[13] Steven J. Breckler and Elizabeth C. Wiggins. On defining attitude and attitude theory: Once more with feeling. In A. R. Pratkanis, S. J. Breckler, and A. C. Greenwald (Eds.). *Attitude structure and function*. Hillsdale, NJ: Erlbaum. pp. 407-427, 1992.

[14] Karl Bühler. *Theory of Language. Representational Function of Language*. John Benjamins Publ. 1990. (reprint from Karl Bühler. *Sprachtheorie. Die Darstellungsfunktion der Sprache*, Ullstein, Frankfurt a. M., Berlin, Wien, 1934.)

[15] Kou-Chan Chiu. Explorations in the Effect of Emoticon on Negotiation

Process from the Aspect of Communication, Master's Thesis, Department Information Management, National Sun Yatsen University, 2007.

[16] Gerald L. Clore, Karen Gasper and Erika Garvin. Affect as information. In Joseph P. Forgas. *Handbook of affect and social cognition*, Routledge, 2001.

[17] Gerald L. Clore and Justin Storbeck. Affect as information about liking, efficacy, and importance. *Affect in Social Thinking and Behavior*, 2006.

[18] Justine Coupland. Small Talk: Social Functions. *Research on Language & Social Interaction*, Vol. 36, No. 1, pp. 1-6, 2003.

[19] David Crystal. *The Cambridge Encyclopedia of Language*. Cambridge University Press, 1989.

[20] Antonio Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*, New York: G.P. Putnam's Sons, 1994.

[21] Antonio Damasio. *The Feeling of what Happens: Body and Emotion in the Making of Consciousness*, Harcourt Brace and Co., 1999.

[22] Charles R. Darwin. *The Expression of the Emotions in Man and Animals*. John Murray, London. 1872.

[23] Tetsuo Sakurai, Jun Daienoki, Satoshi Kitayama. *Dejitaru Nettowāku Shakai - Nyūmon kōza* [Digital Network Society - Introduction] (In Japanese), Heibonsha, 2005.

[24] Daniel C. Dennett. *The Intentional Stance*, The MIT Press, 1989.

[25] Daantje Derks, Arjan E.R. Bos, Jasper von Grumbkow. Emoticons and social interaction on the Internet: the importance of social context, *Computers in Human Behavior*, Vol. 23, pp. 842-849, 2007.

[26] Rene Descartes. *The Passions of the Soul*, Hackett Publishing, 1989 (reprint from 1649).

[27] Alan J. Dix, Janet E. Finlay, Gregory D. Abowd, Rusel Beale. *Human-Computer Interaction*. Prentice Hall. 2004.

[28] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, J-C. Burgelman. Scenarios for Ambient Intelligence in 2010. *ISTAG Report*, European Commission. 2001.

[29] Pawel Dybala, Michal Ptaszynski, Shinsuke Higuchi, Rafal Rzepka and Kenji Araki. Humor Prevails! - Implementing a Joke Generator into a Conversational System. *LNAI*, Vol. 5360, pp. 214-225, 2008.

[30] Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki. Extracting *Dajare* Candidates from the Web - Japanese Puns Generating System as a Part of Humor Processing Research, In *The Proceedings of the First International Workshop on Laughter in Interaction and Body Movement (LIBM'08)*, pp. 46-51, Asahikawa, Japan, June 2008.

[31] Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki. Humoroids - Conversational Agents That Induce Positive Emotions with Humor. In *Proceedings of AAMAS 2009*, pp. 1171-1172, 2009.

[32] Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki. Subjective, But Not Worthless - Non-linguistic Features of Chatterbot Eval-

uations, The 6th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems, in *Working Notes of Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*, pp. 87-92, 2009.

[33] Pawel Dybala, Michal Ptaszynski, Jacek Maciejewski, Mizuki Takahashi, Rafal Rzepka, and Kenji Araki. Multiagent system for joke generation: Humor and emotions combined in human-agent conversation, *The IOS Journal on Ambient Intelligence and Smart Environments*, Thematic Issue on Computational Modeling of Human-Oriented Knowledge, pp. 31-48, 2010.

[34] Phoebe C. Ellsworth. William James and Emotion: Is a Century of Fame Worth a Century of Misunderstanding?, *Psychological Review*, Vol. 101, No. 2, pp. 222-229, 1994.

[35] Daisuke Endo, Manami Saito and Kazuhide Yamamoto. *Kakariuke kankei wo riyo shita kanjoseikihyogen no chushutsu.* (Extracting expressions evoking emotions using dependency structure) [in Japanese], *Proceedings of The Twelve Annual Meeting of The Association for Natural Language Processing (NLP2006)*, 2006.

[36] Linda E. Francis. Laughter, the Best Mediation: Humor as Emotion Management in Interaction. *Symbolic Interaction*, Vol. 17, No. 2, pp. 147-163, 1994.

[37] Nico H. Frijda. *The Emotions.* Cambridge University Press, 1987.

[38] Susan R. Fussell. *The Verbal Communication of Emotions: Interdisciplinary Perspectives.* Lawrence Erlbaum Associates, 2002.

[39] Georg Geiser. *Mensch-Maschine Kommunikation*. Oldenbourg. 1990.

[40] Carlos Gershenson. Contextuality: A Philosophical Paradigm, with Applications to Philosophy of Cognitive Science, *POCS Technical Report*, COGS, University of Sussex, 2002.

[41] Gregory Grefenstette, Yan Qu, James G. Shanahan and David A. Evans. Coupling Niche Browsers and Affect Analysis for an Opinion Mining. In *Proceedings of RIAO-04*, pp. 186-194, 2004.

[42] Joseph C. Hager, Paul Ekman, Wallace V. Friesen. *Facial action coding system*. Salt Lake City, UT: A Human Face, 2002.

[43] Masao Hase, Shiori Kenta and Junichi Hoshino. *Hatsuwa wo okonau kagu ni yoru nichijōteki entāteinmento (taiwa)* [The Everyday Entertainment by Talking Furniture] (in Japanese). *IPSJ SIG Technical Report 2007-NL-181*, Vol. 2007(94), pp. 41-46, 2007.

[44] James Curtis Hepburn, *A Japanese-English and English-Japanese Dictionary*, Shanghai, American Presbyterian Mission Press, 1886.

[45] Shinsuke Higuchi, Rafal Rzepka and Kenji Araki. A Casual Conversation System Using Modality and Word Associations Retrieved from the Web. In *Proceedings of the EMNLP 2008*, pp. 382-390, 2008.

[46] William Huitt. The affective system. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University, 2003.

[47] Edmund Husserl. *Ideas: General Introduction to Pure Phenomenology*. Collier Books, 1962.

[48] Takashi Inui, Manabu Okumura. A Survey of Sentiment Analysis. *Journal of Natural Language Processing*. Vol. 13, No. 3, pp. 201-241, 2006.

[49] Amy Ip. The Impact of Emoticons on Affect Interpretation in Instant Messaging, 2002. http://amysmile.com/pastprj/emoticon_paper.pdf

[50] Naoki Isomura, Fujio Toriumi, Ken'ichiro Ishii. Evaluation method of Non-task-oriented dialogue system by HMM, *IEICE Technical Report*, Vol. 106, No. 300, pp. 57-62, 2006.

[51] Carroll E. Izard (ed.). *Human Emotions*. Plenum Press, 1977.

[52] Roman Jakobson. Closing Statement: Linguistics and Poetics. *Style in Language*, pp.350-377, The MIT Press, 1960.

[53] William James. *The Letters of William James*, ed. Henry James, Boston: Little Brown, 1926.

[54] William James. What is emotion?, *Mind*, Vol. ix, No. 189, 1884.

[55] Jianxin (Roger) Jiao, Qianli Xu and Jun Du. Affective Human Factors Design with Ambient Intelligence. In *Proceedings of The First International Workshop on Human Aspects in Ambient Intelligence*, pp. 45-58, 2007.

[56] Takashi Kamei, Rokuro Kouno and Eiichi Chino (eds.). *The Sanseido Encyclopedia of Linguistics*, Vol. VI, Sanseido, 1996.

[57] Bong-Seok Kang, Chul-Hee Han, Sang-Tae Lee, Dae-Hee Youn and Chungyong Lee. Speaker dependent emotion recognition using speech signals. In *Proc. ICSLP*, pp. 383-386, 2000.

[58] Masahiro Kawakami. The database of 31 Japanese emoticon with their emotions and emphases, *The human science research bulletin of Osaka Shoin Women's University*, Vol. 7, pp.67-82, 2008.

[59] Alistair Kennedy and Diane Inkpen. 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)*, 2005.

[60] Antony Kenny. *Action, Emotion and Will*, Routledge, 2003 (reprint from 1963).

[61] Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto. Opinion Mining from Web Documents: Extraction and Structurization. *Transactions of the Japanese Society for Artificial Intelligence*, Vol. 22, No. 2, pp. 227-238, 2007.

[62] Hidekazu Kubota, Kouji Yamashita, Tomohiro Fukuhara, Toyoaki Nihsida. POC caster: Broadcasting Agent Using Conversational Representation for Internet Community [in Japanese], *Transactions of JSAI*, Vol. AI-17, pp. 313-321, 2002.

[63] Taku Kudo. MeCab: Yet Another Part-of-Speech and Morphological Analyzer, 2001. http://mecab.sourceforge.net/

[64] Taku Kudo and Yuji Matsumoto. Chunking with support vector machines, In *Proceedings of Second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL 2001)*, pp. 192-199, 2001.

[65] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*. Vol. 33, pp. 159-174, 1977.

[66] Richard S. Lazarus, A. D. Kanner, S. Folkman. Emotions: A cognitive-phenomenological analysis, *Emotion, Theory, Research*, Academic Press, 1980.

[67] Richard S. Lazarus. Cognition and Motivation in Emotion, *American Psychologist*, Vol. 46, pp. 362-367, 1991.

[68] Joseph E. LeDoux. Cognitive-emotional interactions in the brain, *Cognition & Emotion*, Vol. 3, Issue 4, pp. 267-289, 1989.

[69] Jennifer S. Lerner and Dacher Keltner. Beyond valence: Toward a model of emotion-specific influences on judgment and choice. *Cognition & Emotion*, Vol. 14, No. 4, pp. 473-493, 2000.

[70] Michael Lewis, Jeannette M. Haviland-Jones, Lisa Feldman Barret (eds.). *Handbook of Emotions*. The Guilford Press, Second Edition, 2000.

[71] Diane J. Litman, Shimei Pan, Marilyn A. Walker. Evaluating response strategies in a Web-based spoken dialogue agent. In *Proceedings of the 17th International Conference on Computational linguistics*, pp. 780-786, 1998.

[72] Gorge Loewenstein and Jennifer S. Lerner. The Role of Affect in Decision Making. *Hanbook of Affective Sciences*, pp. 619-642, 2003.

[73] Karl F. MacDorman and Hiroshi Ishiguro. The uncanny advantage of using androids in social and cognitive science research, *Interaction Studies*, Vol. 7, No. 3, pp. 297-337, 2006.

[74] David Mandel. Counterfactuals, emotions, and context. *Cognition & Emotion*, Vol. 17, No. 1, pp. 139-159, 2003.

[75] Jack M. Maness. A Linguistic Analysis of Chat Reference Conversations with 18-24 Year-Old College Students, *The Journal of Academic Librarianship*, Vol. 34, No. 1, pp. 31-38, January 2008.

[76] Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, Kazuma Takaoka, and Masayuki Asahara. Japanese Morphological Analysis System ChaSen version 2.2.1, 2000.

[77] Naohiro Matsumura, Asako Miura, Yasufumi Shibanai, Yukio Ohsawa and Mitsuru Ishizuka. The Dynamism of 2channel. *Transactions of Information Processing Society of Japan*, Vol. 45, No. 3, pp. 1053-1061, 2004.

[78] Bernard Comrie, Martin Haspelmath, Balthasar Bickel. *Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*, Developed jointly by: Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology and Department of Linguistics of the University of Leipzig, 2004. http://www.eva.mpg.de/lingua/resources/glossing-rules.php

[79] John D. Mayer and Peter Salovey. What is emotional intelligence?. In Peter Salovey and David Sluyter, Eds. *Emotional Development and Emotional Intelligence*, pp. 3-31, New York, Basic Books, 1997.

[80] Junko Minato, David B. Bracewell, Fuji Ren and Shingo Kuroiwa. Statistical Analysis of a Japanese Emotion Corpus for Natural Language Processing. *LNCS* 4114, pp. 924-929, 2006.

[81] Tetsuya Miyoshi and Yu Nakagami. Sentiment classification of customer reviews on electric products, *IEEE International Conference on Systems, Man and Cybernetics (SMC-2007)*, pp. 2028-2033, 2007.

[82] Merrill Morris and Christine Ogan. The Internet as Mass Medium. *Journal of Computer-Mediated Communication*, Vol. 1, No. 4, 1996.

[83] Kevin Mulligan. Intentionality, Knowledge and Formal Objects, *Electronic Festschrift for W. Rabinowicz*, In: T. Ronnow-Rasmussen et al. (eds.), 2007. www.fil.lu.se/HommageaWlodek/site/papper/MulliganKevin.pdf

[84] Jinpei Nakamura, Takeshi Ikeda, Nobuo Inui and Yoshiyuki Kotani. Learning Face Mark for Natural Language Dialogue System, *IEEE NLP-KE 2003*, pp. 180-185, 2003.

[85] Akira Nakamura. *Kanjō hyōgen jiten* [Dictionary of Emotive Expressions] (in Japanese), Tokyodo Publishing, 1993.

[86] Hitori Nakano. *Densha otoko* [Train man] (in Japanese). Tokyo, Shinchosha, 2005.

[87] Norio Nakayama, Koji Eguchi and Noriko Kando. Proposal for Extraction of Emotional Expression. *IEICE Technical Report*, Vol. 104, No. 416, pp. 13-18. 2004.

[88] Nasir Naqvi, Baba Shiv, Antoine Bechara. The Role of Emotion in Decision Making - A Cognitive Neuroscience Perspective, *Current Directions in Psychological Science*, Vol. 15, No. 5, pp. 260-264, 2006.

[89] Nel Noddings. Conversation as Moral Education. *Journal of Moral Education*, Vol. 23, No. 2, pp. 107-118, 1994.

[90] Hajime Ono. *An emphatic particle DA and exclamatory sentences in Japanese*. University of California, Irvine, 2002.

[91] Ooso Mieko. Nagoya University Conversational Corpus, 2003. http://tell.fll.purdue.edu/chakoshi/public.html

[92] Andrew Ortony, Gerald L. Clore and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press. 1988.

[93] Yuriko Oshima-Takane, Brian MacWhinney (Ed.), Hidetosi Shirai, Susanne Miyata and Norio Naka (Rev.). *CHILDES Manual for Japanese*, McGill University, The JCHAT Project, 1995-1998.

[94] Bo Pang and Lillian Lee. Opinion Mining and Sentiment Analysis. In *Foundations and Trends in Information Retrieval*, Vol. 2, No. 1-2, pp. 1-135, 2008.

[95] Rosalind W. Picard, Elias Vyzas, Jennifer Healey. Toward Machine Emotional Intelligence: Analysis of Affective Physiological State, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 10, pp. 1175-1191, 2001.

[96] Rosalind W. Picard. *Affective Computing*, MIT Press, 1997.

[97] Livia Polanyi, Annie Zaenen. Contextual valence shifters, *Computing Attitude and Affect in Text: Theory and Applications*, Springer-Verlag, 2006.

[98] Christopher Potts and Shigeto Kawahara. Japanese honorifics as emotive definite descriptions. In *Proceedings of Semantics and Linguistic Theory 14*, pp. 235-254, 2004.

[99] Michal Ptaszynski. 2006. Moeru gengo: Intānetto keijiban no ue no nihongo kaiwa ni okeru kanjōhyōgen no kōzō to kigōrontekikinō no bunseki – "2channeru" denshikeijiban o rei toshite –, *[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, 2006.

[100] Michal Ptaszynski, Pawel Dybala, Shinsuke Higuchi, Rafal Rzepka and Kenji Araki. Affect-as-Information Approach to a Sentiment Analysis Based Evaluation of Conversational Agents, In *Proceedings of the 2008 International Conference on Intelligent Agents, Web Technologies & Internet Commerce (IAWTIC'08)*, pp. 896-901, 2008.

[101] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki. Disentangling emotions from the Web. Internet in the service of affect analysis. In *Proceedings of KEAS'08*, pp. 51-56, 2008.

[102] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki. Effective Analysis of Emotiveness in Utterances Based on Features of Lexical and Non-Lexical Layers of Speech. In *Proceedings of The 14th Annual Meeting of The Association for NLP*, pp. 171-174, 2008.

[103] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki. 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), 2009.

[104] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki. Ideas for Using Large-Scale Corpora to Improve Verification of Emotion Appropriateness in Japanese, In *Proceedings of The Joint GCOE Symposium for Cultivating Young Researchers*, 2009.

[105] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki. Towards Context Aware Emotional Intelligence in Machines: Computing Contextual Appropriateness of Affective States', In *Proceedings of Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*, pp. 1469-1474, 2009.

[106] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki. Contextual Valence Shifters Supporting Affect Analysis of Utterances in Japanese, In *Proceedings of The Fifteenth Annual Meeting of The Association for Natural Language Processing (NLP-2009)*, pp. 825-828, 2009.

[107] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki. Affecting Corpora: Experiments with Automatic Affect Annotation System - A Case Study of the 2channel Forum -", In *Proceedings of The Conference of the Pacific Association for Computational Linguistics 2009 (PACLING-09)*, pp. 223-228, 2009.

[108] Michal Ptaszynski, Pawel Dybala,Wenhan Shi, Rafal Rzepka and Kenji Araki. Contextual Affect Analysis: A System for Verification of Emotion

Appropriateness Supported with Contextual Valence Shifters, *International Journal of Biometrics*, Vol. 2, No. 2, pp. 134-154, 2010.

[109] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki. Towards Fully Automatic Emoticon Analysis System (^o^), In *Proceedings of The Sixteenth Annual Meeting of The Association for Natural Language Processing (NLP2010)*, March 8 -11, pp. 583-586, 2010.

[110] Filip Radulovic and Nikola Milikic. Smiley Ontology, In *Proceedings of Social Network Interoperability Workshop* held within Asian Semantic Web Conference, 2009.

[111] Justus J. Randolph. Online Kappa Calculator. 10.07.2010, http://justusrandolph.net/kappa/

[112] Jonathon Read. Recognizing Affect in Text using Pointwise-Mutual Information. M. Sc. Dissertation, University of Sussex. 2004.

[113] Jonathon Read. Using Emoticons to reduce Dependency in Machine Learning Techniques for Sentiment Classification, In *Proceedings of the Student Research Workshop ACL-05*, pp. 43-48, 2005.

[114] James A. Russell. A circumplex model of affect. *J. of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161-1178, 1980.

[115] Rafal Rzepka and Kenji Araki. What Statistics Could Do for Ethics?– The Idea of Common Sense Processing Based Safety Valve. *AAAI Fall Symposium Technical Report FS-05-06*, pp. 85-87, 2005.

[116] Rafal Rzepka, Yali Ge and Kenji Araki. 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, 2006.

[117] Rafal Rzepka and Kenji Araki. What About Tests In Smart Environments? On Possible Problems With Common Sense In Ambient Intelligence. In *Proceedings of 2nd Workshop on Artificial Intelligence Techniques for Ambient Intelligence*, IJCAI'07, pp. 92-96, 2007.

[118] Rafal Rzepka and Kenji Araki. Consciousness of Crowds - The Internet As a Knowledge Source of Human's Conscious Behavior and Machine Self-Understanding, In *Proceedings of AAAI Fall Symposium on AI and Consciousness: Theoretical Foundations and Current Approaches*, Technical Report, pp. 127-128, 2007.

[119] Rafal Rzepka, Shinsuke Higuchi, Michal Ptaszynski and Kenji Araki. Straight Thinking Straight from the Net - On the Web-based Intelligent Talking Toy Development, In *Proceedings of SMC-08*, pp. pp. 2172-2176, 2008.

[120] Rafal Rzepka, Fumito Masui and Kenji Araki. 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.

[121] Peter Salovey and John D. Mayer. Emotional intelligence, *Imagination, Cognition, and Personality*, Vol. 9, pp. 185-211, 1990.

[122] Kaori Sasai. 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, 2006.

[123] Harold Schlosberg. The description of facial expressions in terms of two dimensions, *Journal of Experimental Psychology*, Vol. 44, No. 4, pp. 229-237, 1952.

[124] Norbert Schwarz. Emotion, cognition, and decision making. *Cognition & Emotion*, Vol. 14, No. 4, pp. 433-440, 2000.

[125] Wenhan Shi, Rafal Rzepka and Kenji Araki. Emotive Information Discovery from User Textual Input Using Causal Associations from the Internet [In Japanese]. *FIT2008*, pp. 267-268, 2008.

[126] Abdullah Shunnaq. Lexical Incongruence in Arabic-English Translation due to Emotiveness in Arabic. *Turjuman*. Vol. 2, No. 2, pp. 37-63, 1993.

[127] John Simpson and Edmund Weiner (eds). *Oxford English Dictionary*, Additions Series. 1993. OED Online. Oxford University Press. 17 Aug. 2008, http://www.oed.com/

[128] Amit Singhal. Modern Information Retrieval: A Brief Overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, Vol. 24, No. 4, pp. 35-43, 2001.

[129] Jonas Sjöbergh. Vulgarities are fucking funny, or at least make things a little bit funnier. *Technical Report of KTH*, Stockholm, 2006.

[130] Jonas Sjöbergh and Kenji Araki. 2008. A Multi-Lingual Dictionary of Dirty Words, *LREC*, 2008.

[131] Robert C. Solomon. *The Passions: Emotions and the Meaning of Life*, Hackett Publishing, 1993.

[132] Robert C. Solomon. *Not Passion's Slave*, Oxford University Press, 2003.

[133] Robert C. Solomon. *True To Our Feelings: What Our Emotions Are Really Telling Us*, Oxford University Press, 2007.

[134] Bas R. Steunebrink, Mehdi Dastani, John-Jules Ch. Meyer. Emotions as Heuristics in Multi-Agent Systems, In *Proceedings of the 1st Workshop on Emotion and Computing*, pp. 15-18, 2006.

[135] Charles L. Stevenson. *Facts and Values-Studies in Ethical Analysis.* Yale University Press, 1963.

[136] Nobuo Suzuki and Kazuhiko Tsuda. Automatic emoticon generation method for web community, *IADIS International Conference on Web Based Communities 2006 (WBC2006)*, pp. 331-334, 2006.

[137] Nobuo Suzuki and Kazuhiko Tsuda. Express Emoticons Choice Method for Smooth Communication of e-Business, *KES 2006*, Part II, LNAI 4252, pp. 296-302, 2006.

[138] Toshiaki Takahashi, Hiroshi Watanabe, Takashi Sunda, Hirofumi Inoue, Ken'ichi Tanaka and Masao Sakata. Technologies for enhancement of operation efficiency in 2003i IT Cockpit. *Nissan Technical Review*, Vol. 53, pp. 61-64, 2003.

[139] Kazumasa Takami, Ryo Yamashita, Kenji Tani, Yoshikazu Honma, Shinichiro Goto. Deducing a User's State of Mind from Analysis of the Pictographic Characters and Emoticons used in Mobile Phone Emails for Personal Content Delivery Services, *International Journal On Advances in Telecommunications*, Vol. 2, No, 1, pp. 37-46, 2009.

[140] Yuki Tanaka, Hiroya Takamura, Manabu Okumura. Extraction and Classification of Facemarks with Kernel Methods, *IUI'05*, January 9-12, 2005, San Diego, California, USA, 2005.

[141] Jorge Teixeira, Vasco Vinhas, Eugenio Oliveira, Luis Paulo Reis. A New Approach to Emotion Assessment Based on Biometric Data, 2nd International Workshop on Human Aspects in Ambient Intelligence (HAI'08), In *Proceedings of the WI-IAT'08*, pp. 459-500, 2008.

[142] Fabrice Teroni. Emotions and Formal Objects, *dialectica*, pp. 395-415, 2007.

[143] Ross A. Thompson, Deborah J. Laible and Lenna L. Ontai. Early understandings of emotion, morality, and self: Developing a working model, *Advances in child development and behavior*, Vol. 31, pp. 137-171, 2003.

[144] Tin Lay Nwe, Foo Say Wei, and Liyanage C. De Silva, Speech Emotion Recognition Using Hidden Markov Models, In *Elsevier Speech Communications Journal*, Vol. 41, Issue 4, pp. 603-623, 2003.

[145] Ryoko Tokuhisa, Kentaro Inui, Yuji Matsumoto. Emotion Classification Using Massive Examples Extracted from the Web. In *Proc. of Coling 2008*, pp. 881-888, 2008.

[146] Jan Treur. On Human Aspects in Ambient Intelligence, In *Proceedings of The First International Workshop on Human Aspects in Ambient Intelligence*, pp. 5-10, 2007.

[147] Seiji Tsuchiya, Eriko Yoshimura, Hirokazu Watabe and Tsukasa Kawaoka. The Method of the Emotion Judgement Based on an Associ-

ation Mechanism. *Journal of Natural Language Processing*, Vol. 14, No. 3, pp. 219-238, 2007.

[148] Naoko Tsuchiya. *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*, 1999.

[149] Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of ACL'02*, pp. 417-424, 2002.

[150] Mayumi Usami (Ed.). *BTS ni yoru nihongo hanshikotoba kopasu 1 (hatsu tamen, yūjin; zatsudan, tōron, sasoi)* [Conversation corpus of spoken Japanese using the Basic Transcription System (first meeting, friend's conversation, small talk, discussion, invitation)] (in Japanese), Tokyo University of Foreign Studies, Tokyo, Japan, 2007.

[151] Gary VandenBos. *APA Dictionary of Psychology.* Washington, DC: American Psychological Association, 2006.

[152] Marjorie Fink Vargas. Louder than Words: An Introduction to Nonverbal Communication, Ames: Iowa State UP, 1986.

[153] Gianmarco Veruggio, Fiorella Operto. Roboethics: a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics, *International Review of Information Ethics, Ethics in Robotics*, Vol. 6, No. 12, 2006.

[154] Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, Alicia Abella. PARADISE: a framework for evaluating spoken dialogue agents, In *Pro-*

ceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics, pp. 271-280, 1997.

[155] Edda Weigand (ed.). *Emotion in Dialogic Interaction: Advances in the Complex*. John Benjamins, 2002.

[156] Anna Wierzbicka. *Emotions Across Languages and Cultures: Diversity and Universals*. Cambridge University Press, 1999.

[157] Theresa Wilson and Janyce Wiebe. Annotating Attributions and Private States. *Proceedings of the ACL Workshop on Frontiers in Corpus Annotation II*, pp. 53-60, 2005.

[158] John R. S. Wilson. *Emotion and Object*, Cambridge University Press, 1972.

[159] Alecia Wolf. Emotional Expression Online: Gender Differences in Emoticon Use, *CyberPsychology & Behavior*, Vol. 3, No. 5, pp. 827-833, 2004.

[160] Chung-Hsien Wu, Ze-Jing Chuang and Yu-Chung Lin. Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models. *ACM Transactions on Asian Language Information Processing*, Vol. 5, No. 2, pp. 165-183, 2006.

[161] Taichi Yamada, Seiji Tsuchiya, Shingo Kuroiwa and Fuji Ren. Classification of Facemarks Using N-gram, *International Conference on NLP and Knowledge Engineering*, pp. 322-327, 2007.

[162] Yoshitaka Yamashita. Kara, Node, Te-Conjunctions which express

cause or reason in Japanese (in Japanese), *Journal of the International Student Center*, Hokkaido University, Vol. 3, 1999.

[163] Changhua Yang, Kevin Hsin-Yih Lin, Hsin-Hsi Chen. Building Emotion Lexicon from Weblog Corpora, In *Proceedings of the ACL 2007 Demo and Poster Sessions*, pp. 133-136, 2007.

[164] Jeremy A. Yip and Rod A. Martin. Sense of humor, emotional intelligence, and social competence, *Journal of Research in Personality*, Vol. 40, No. 6, pp. 1202-1208, 2006.

[165] Kenji Yoshihira, Yoshiyuki Takeda, Satoshi Sekine. KWIC system for Web Documents (in Japanese). In *Proceedings of The 10th Annual Meeting of the Japanese Association for NLP*, pp. 137-139, 2004.

[166] Polly Young. The Effects of Mood and Emotional State on Decision Making. *The Journal of Behavioral Decision Making*, Special Issue, 2006.

[167] Chen Yu, Paul M. Aoki, Allison Woodruff. Detecting User Engagement in Everyday Conversations. In *Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP)*, Vol. 2, pp. 1329-1332. Jeju Island, Republic of Korea, 2004.

Research Achievements

First Author Publications

Journals

1. Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"CAO: A Fully Automatic Emoticon Analysis System Based on Theory of Kinesics"**, *IEEE Transactions on Affective Computing*, 19 July 2010. IEEE Computer Society Digital Library, IEEE Computer Society.

2. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, **"Contextual Affect Analysis: A System for Verification of Emotion Appropriateness Supported with Contextual Valence Shifters"**, *International Journal of Biometrics*, Vol. 2, No. 2, 2010, pp. 134-154, Inderscience Enterprises.

3. Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"An Automatic Evaluation Method for Conversational Agents Based on Affect-as-Information Theory"**, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Special Issue on Emotions, Vol. 22, No. 1 (February), 2010, pp. 73-89.

4. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, **"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), 2009, pp. 30-49 (194-213).

Books

1. Michal Ptaszynski, Pawel Dybala, Shinsuke Higuhi, Wenhan Shi, Rafal Rzepka and Kenji Araki, **"Towards Socialized Machines: Emotions and Sense of Humour in Conversational Agents"**, Chapter in *Web Intelligence and Intelligent Agents*, In-Tech, Vienna, Austria, 2010, ISBN 978-953-7619-85-5, pp. 173-206.

International Conferences

1. Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"CAO: A Fully Automatic Emoticon Analysis System"**, In *Proceedings of The Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, pp. 1026-1032, July 11 - 15, 2010, Atlanta, Georgia, USA.

2. Michal Ptaszynski, Pawel Dybala, Tatsuaki Matsuba, Fumito Masui, Rafal Rzepka and Kenji Araki, **"Machine Learning and Affect Analysis Against Cyber-Bullying"**, In *Proceedings of The Thirty Sixth Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB'10)*, Symposium on Linguistic and Cognitive Approaches to Dialog Agents (LaCATODA'2010), March 29 - April 1, 2010, Leicester, UK, pp. 7-16.

3. Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"For-**

getful and Emotional: Recent Progress in Development of Dynamic Memory Management System for Conversational Agents", In *Proceedings of The Thirty Sixth Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB'10)*, Symposium on Linguistic and Cognitive Approaches to Dialog Agents (LaCATODA'2010), March 29 - April 1, 2010, Leicester, UK, pp. 32-38.

4. Michal Ptaszynski, Pawel Dybala, Radoslaw Komuda, Rafal Rzepka and Kenji Araki, **"Development of Emoticon Database for Affect Analysis in Japanese"**, In *Proceedings of The 2010 International Symposium on Global COE Program of Center for Next-Generation Information Technology Based on Knowledge Discovery and Knowledge Federation (GCOE-NGIT 2010)*, Sapporo, Japan, January 18-20, 2010, pp. 203-204.

5. Michal Ptaszynski, Rafal Rzepka and Kenji Araki, **"A Pragmatic Approach to Implementation of Emotional Intelligence in Machines"**, In *Proceedings of The AAAI 2009 Fall Symposium on Biologically Inspired Cognitive Architectures (BICA-09)*, Washington, D.C., USA, November 5 - 7, 2009, pp. 101-102.

6. Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"Affecting Corpora:Experiments with Automatic Affect Annotation System - A Case Study of the 2channel Forum -"**, In *Proceedings of The Conference of the Pacific Association for Computational Linguistics (PACLING-09)*, Japan, September 1-4, 2009, pp. 223-228.

7. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, **"Conscience of Blogs: Verifying Contextual Appropriateness of Emotions Basing on Blog Contents"**, In *Proceedings of The Fourth International Conference on Computational Intelligence (CI 2009)*, Honolulu, Hawaii, USA, August 17 - 19, 2009, pp. 1-6.

8. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, **"Shifting Valence Helps Verify Contextual Appropriateness of Emotions"**, In *Working Notes of Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*, Workshop on Automated Reasoning about Context and Ontology Evolution (ARCOE-09), Pasadena, California, USA, 2009, pp. 19-21.

9. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, **"Towards Context Aware Emotional Intelligence in Machines: Computing Contextual Appropriateness of Affective States"**, In *Proceedings of Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*, Pasadena, CA, USA, 2009, pp.1469-1474.

10. Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"Development of Corpus for Affect Analysis in Japanese"**, In *Proceedings of The 2nd International Symposium on Global COE Program of Center for Next-Generation Information Technology Based on Knowledge Discovery and Knowledge Federation (GCOE-NGIT 2009)*, Sapporo, Japan, January 20-21, 2009.

11. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji

Araki, **"How to find love in the Internet? Applying Web mining to affect recognition from textual input"**, In *Proceedings of the 2008 Empirical Methods for Asian Languages Processing Workshop (EMALP'08)* at The Tenth Pacific Rim International Conference on Artificial Intelligence (PRICAI'08), Hanoi, Vietnam, December 2008, pp. 67-79.

12. Michal Ptaszynski, Pawel Dybala, Michal Ptaszynski, Shinsuke Higuchi, Rafal Rzepka and Kenji Araki, **"Affect-as-Information Approach to a Sentiment Analysis Based Evaluation of Conversational Agents"**, In *Proceedings of the 2008 International Conference on Intelligent Agents, Web Technologies & Internet Commerce (IAWTIC'08)*, Vienna, Austria, December 2008, pp. 896-901.

13. Michal Ptaszynski, Pawel Dybala, Shinsuke Higuchi, Rafal Rzepka and Kenji Araki, **"Affect as Information about Users' Attitudes to Conversational Agents"**, In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT'08)*, Second International Workshop on Human Aspects in Ambient Intelligence (HAI'08), Sydney, Australia, December 2008, pp. 459-500.

14. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, **"Disentangling emotions from the Web. Internet in the service of affect analysis"**, In *Proceedings of the Second International Conference on Kansei Engineering & Affective Systems (KEAS'08)*, Nagaoka, Japan, November 2008, pp 51-56. [BEST STUDENT PAPER AWARD]

National Conferences and Meetings

1. Michal Ptaszynski, Rafal Rzepka and Kenji Araki, **"On the Need for Context Processing in Affective Computing"**, In *Proceedings of Fuzzy System Symposium (FSS2010)*, Organized Session on Emotions, September 13-15, 2010 (to appear).

2. Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"Towards Fully Automatic Emoticon Analysis System (^o^)"**, In *Proceedings of The Sixteenth Annual Meeting of The Association for Natural Language Processing (NLP2010)*, March 8 -11, 2010, pp. 583-586.

3. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, **"Ideas for Using Large-Scale Corpora to Improve Verification of Emotion Appropriateness in Japanese"**, In *Proceedings of The Joint GCOE Symposium for Cultivating Young Researchers*, Sapporo, Japan, September 30 - October 1, 2009.

4. Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"Processing the Contextual Appropriateness of Emotions"**, In *Proceedings of The Joint GCOE Seminar for Cultivating Young Researchers*, April 25-26, 2009, Jozankei, Japan.

5. Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"Contextual Valence Shifters Supporting Affect Analysis of Utterances in Japanese"**, In *Proceedings of The Fifteenth Annual Meeting of The Association for Natural Language Processing (NLP-2009)*, March 2-5, pp.825-828.

6. Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"Double Standpoint Evaluation Method for Affect Analysis System"**, In *Proceedings of The 22nd Annual Conference of The Japanese Society for Artificial Intelligence (JSAI2008)*, June 2008, CD-ROM Proceedings, Paper No. 2P2-07.

7. Michal Ptaszynski, Pawel Dybala, Rafal Rzepka and Kenji Araki, **"Effective Analysis of Emotiveness in Utterances based on Features of Lexical and Non-Lexical Layer of Speech"**, In *Proceedings of The Fourteenth Annual Meeting of The Association for Natural Language Processing (NLP-2008)*, March 17-21, 2008, pp 171-174.

8. Michal Ptaszynski, Koichi Sayama, Rafal Rzepka, Kenji Araki, **"A dynamic memory management system based on forgetting and recalling."**, In *2007 Joint Convention Record. The Hokkaido Chapters of the Institutes of Electrical and Information Engineers*, Japan, October, 2007, pp. 295-296.

9. Michal Ptaszynski, Pawel Dybala, Wen Han Shi, Rafal Rzepka, Kenji Araki, **"Lexical Analysis of Emotiveness in Utterances for Automatic Joke Generation"**, *ITE Technical Report*, Vol. 31, No.47, ME2007-204, October, 2007, pp.39-42.

10. Michal Ptaszynski, Koichi Sayama, **"The idea of dynamic memory management system based on a forgetting-recalling algorithm with emotive analysis."**, *Language Acquisition and Understanding Research Group (LAU) Technical Reports*, Sapporo, August 2007, pp. 12-15.

Co-Authored Publications

Journals

1. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki, **"Evaluating Subjective Aspects of HCI on an Example of a Non-task Oriented Conversational System"**, *International Journal of Artificial Intelligence Tools*, Special Issue on AI Tools for HCI Modeling, 2010. (to appear)

2. Rafal Rzepka, Shinsuke Higuchi, Michal Ptaszynski, Pawel Dybala and Kenji Araki, **"When Your Users Are Not Serious - Using Web-based Associations, Affect and Humor for Generating Appropriate Utterances for Inappropriate Input"**, *Transactions of the Japanese Society for Artificial Intelligence*, Vol. 25, No. 1, 2010, pp.114-121.

3. Pawel Dybala, Michal Ptaszynski, Jacek Maciejewski, Mizuki Takahashi, Rafal Rzepka and Kenji Araki, **"Multiagent system for joke generation: Humor and emotions combined in human-agent conversation"**, *Journal of Ambient Intelligence and Smart Environments*, Vol. 2, Thematic Issue on Computational Modeling of Human-Oriented Knowledge within Ambient Intelligence, 2010, pp. 31-48.

4. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki, **"Activating Humans with Humor - A Dialogue System that Users Want to Interact With"**, *IEICE Transactions on Information and Systems*, Special Issue on "Natural Language Processing and its Applications", Vol. E92-D, No. 12 (December), 2009, pp.2394-2401.

5. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, **"Humorized Computational Intelligence: Towards User-Adapted Systems with a Sense of Humor"**, In *Lecture Notes in Computer Science (LNCS)*, Vol.5484, Springer-Verlag Berlin Heidelberg, 2009, pp.452-461.

6. Pawel Dybala, Michal Ptaszynski, Shinsuke Higuchi, Rafal Rzepka and Kenji Araki, **"Humor Prevails! - Implementing a Joke Generator into a Conversational System"**, In *Lecture Notes in Artificial Intelligence (LNAI)*, Vol. 5360, Springer-Verlag Berlin Heidelberg, 2008, pp. 214-225.

International Conferences

1. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, **"Multi-humoroid: Joking System That Reacts With Humor To Humans' Bad Moods"**, In *Proceedings of The Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, Toronto, Canada, May 10-14, 2010, pp. 1433-1434.

2. Radoslaw Komuda, Michal Ptaszynski, Rafal Rzepka and Kenji Araki, **"An idea of a web-crowd based moral reasoning agent"**, In *Proceedings of The Thirty Sixth Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB'10)*, Symposium on Linguistic and Cognitive Approaches to Dialog Agents, March 29 - April 1, 2010, Leicester, UK, pp. 17-19.

3. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, **"Chain**

of Events: Multi-stage Approach to Humor and Emotions in HCI", In *Proceedings of The Thirty Sixth Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB'10)*, Symposium on Linguistic and Cognitive Approaches to Dialog Agents, March 29 - April 1, 2010, Leicester, UK, pp. 53-58.

4. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, **"Computers with personalized sense of humor"**, In *Proceedings of The 2010 International Symposium on Global COE Program of Center for Next-Generation Information Technology Based on Knowledge Discovery and Knowledge Federation (GCOE-NGIT 2010)*, Hokkaido University, Sapporo, Japan, January 18-20, 2010, pp. 205-206.

5. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki, **"Crossing Word Borders - Towards Phrasal Pun Generation Engine"**, In *Proceedings of The Conference of the Pacific Association for Computational Linguistics (PACLING-09)*, Hokkaido University, Sapporo, Japan, September 1-4, 2009, pp. 242-247.

6. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki, **"Getting Closer to Human Level - Human Likeness and its Relation to Humor in Non-task Oriented Dialogue Systems"**, In *Proceedings of The Fourth International Conference on Computational Intelligence (CI 2009)*, August 17-19, 2009, Honolulu, Hawaii, USA, pp.7-14.

7. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki, **"Sub-**

jective, But Not Worthless - Non-linguistic Features of Chatterbot Evaluations", In *Working Notes of Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*, The 6th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems, Pasadena, California, USA, 2009, pp. 87-92.

8. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, **"Humoroids - Conversational Agents That Induce Positive Emotions with Humor"**, In *Proceedings of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, May, 10-15, 2009, Budapest, Hungary, pp. 1171-1172.

9. Rafal Rzepka, Pawel Dybala, Wenhan Shi, Shinsuke Higuchi, Michal Ptaszynski and Kenji Araki, **"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, 2009, pp. 487-488.

10. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, **"Conversational Agents Can Joke"**, In *Proceedings of The 2nd International Symposium on Global COE Program of Center for Next-Generation Information Technology Based on Knowledge Discovery and Knowledge Federation (GCOE-NGIT 2009)*, Hokkaido University, Sapporo, Japan, January, 2009, 20-21.

11. Rafal Rzepka, Shinsuke Higuchi, Michal Ptaszynski, Kenji Araki, **"Straight Thinking Straight from the Net - On the Web-based Intelli-**

gent Talking Toy Development", In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC'08)*, Singapore, October 2008, pp. 2172-2176.

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

National Conferences and Meetings

1. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, **"PUNDA Numbears: Proposal of Goroawase Generating System for Japanese"**, In *Proceedings of The Sixteenth Annual Meeting of The Association for Natural Language Processing (NLP-2010)*, March 8-11, 2010, pp. 345-348.

2. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka and Kenji Araki, **"Do humans like it human like?"**, In *Proceedings of The Joint Global COE Symposium for Cultivating Young Researchers*, Sapporo, Japan, October 2009, pp. 169-170.

3. Pawel Dyabla, Michal Ptaszynski, Rafal Rzepka, Kenji Araka, **"Joking Computers - Why, When and How?"**, In *Proceedings of Joint Global COE Symposium for Cultivating Young Researchers*, Jozankei, Japan, April, 2009.

4. Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, **"When Should Computers Joke? - Concept of Emotiveness Analysis Based Timing Algorithm for Humor-Equipped Conversational Systems"**, In *Proceedings of The Fifteenth Annual Meeting of The Association for Natural Language Processing (NLP-2009)*, March 2-5, 2009,pp.542-545.

5. Shi Wenhan, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, **"Emotive Information Discovery from User Textual Input Using Emotion Expression Element and Web Mining"** (In Japanese), *IEICE Technical Report*, Thought and language, Vol. 108, No. 353, 2008, pp.31-34.

Awards

1. **BEST STUDENT PAPER AWARD for:**
   Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, **"Disentangling emotions from the Web. Internet in the service of affect analysis"**, In *Proceedings of the Second International Conference on Kansei Engineering & Affective Systems (KEAS'08)*, Nagaoka, Japan, November 2008, pp 51-56.

Acknowledgements

I would like to express my deepest gratitude to Professor Kenji Araki for his invaluable help and guidance during the course of my Doctor of Philosophy Degree. I would like to thank him for his advice and kind support throughout my research, and his understanding and openness to different cultures. This gratitude extends of course, to all the professors in the Division of Media and Network Technologies.

I would also like to thank Professor Rafal Rzepka for his invaluable insight into the world of Artificial Intelligence, his precious inputs to my research and most of all for his help into easing me into the Japanese way of life, work and study - you made me feel home a thousand miles away.

My studies in Hokkaido University would not have been possible without the support of the Japanese Ministry of Education which awarded me this scholarship. Therefore, I would like to express my gratitude to the Ministry of Education and the Embassy of Japan in Poland for giving me this chance.

Finally, I would like to thank my family for having supported me through all these years, and for making me who I am today. You are always in my heart.

Lastly, I want to thank all my friends, close and distant, for all the help they have provided.