3P1-IOS-2a-5

Beyond Conventional Recognition: Concept of a Conversational System Utilizing Metaphor Misunderstanding as a Source of Humor

Pawel Dybala[*1] Michal Ptaszynski[*2] Rafal Rzepka[*3] Kenji Araki[*3] Kohichi Sayama[*4]

[*1] JSPS Research Fellow / Otaru University of Commerce, Midori 3-5-21, 047-8501 Otaru, Japan
[*2] JSPS Research Fellow / Hokkai-Gakuen University, Toyohira, Asahimachi 4-1-40, 062-0911 Sapporo, Japan
[*3] Graduate School of Information Science and Technology, Hokkaido University, Kita 14 Nishi 9, Kita-ku, 060-0814 Sapporo, Japan
[*4] Otaru University of Commerce, Department of Information and Management Science, Midori 3-5-21, 047-8501 Otaru, Japan

Although we are still quite far from constructing a human-like conversational system, researchers all over the world keep investigating numerous factors that make conversations between humans. In this work we focus on two such factors: humor and metaphors.

Numerous research projects exist in the area of metaphor understanding and generation. We propose a unique approach to this subject, based on an observation that humans can not only properly understand and generate metaphors, but also make fun of their misunderstandings. For instance, an utterance "you have legs like a deer" can be understood as a compliment ("long and graceful"), as well as an insult ("very hairy"). If used properly, such misunderstanding can serve as source of humor in human-computer conversations.

In this paper we first briefly describe our previous research on humor-equipped conversational systems. We then summarize the state of the art in the metaphor processing research, and mention works showing that the salience imbalance theory, widely use in the field of metaphor understanding, can be slightly modified to be used for explaining humor understanding.

On this basis, we propose a research project aimed to construct a humorous metaphor misunderstandings generator, which will be implemented into a non-task oriented conversational system (chatterbot). The system, named HumMeR, takes as an input sentences possibly including metaphors. It will include procedures for processing known as well as novel metaphors, and, basing on salience imbalance degree changes, will try to find such common properties of source and target of metaphor that would allow to generate humorous misunderstandings. We describe the project's development steps, algorithms and procedures of the HumMeR system, as well as the overall algorithm outline of the chatterbot in which the system will be implemented.

1. Introduction

Turing Test [Turing, 1950], designed over 60 years ago, is probably one of the most widely discussed issues in computer science so far. Criticized by many, appraised by others, the test is still influential and attracts researchers attention. However, leaving appart its usability or appropriatness, it should not be a mistake to state that Turing Test's greatest contribution to modern science lies in triggering researchers all over the world to investigate the possibilities of constructing naturally talking machines. Even if algorithms and systems developed during these studies are not able to pass the test, and even if a machine able to do so is already constructed, the pursue of natural language should not be stopped at that point.

In recent years a tendency can be seen in computer science, especially in its language- and interaction-related areas, to focus not only on purely informative aspects of language, but also on those that make it sound natural, such as emotions, humor, metaphors, sarcasm or ironies. During an interaction with a human, a computer can simply say:

"-It is very hot today."

which, despite being gramatically correct, is somewhat plain and not very human-like. This computer's utterance, however, might as well sound like:

"-Damn, it's hot like hell today!"

which, in turn, sounds much more human-like, as it contains emotive expressions ("damn", "hell", exclamation mark) and a metaphor ("hot like hell"). In other words, the latter sentence is much more likely to be used by a human, and thus, since we are aiming at constructing human-like talking devices, we should investigate the possibilities to make them talk like this.

In our previous research we focused on two such human-like issues, which are humor and emotions. Below we summarize these projects (section 2). In this paper, however, we describe one of our current ventures, aimed at constructing a system that would be able to understand and generate metaphorical expressions. Here we particularly focus on a setup that would allow the system to uttilize humorous metaphorical misunderstandings, and on implementing this algorithm into a chatterbot (a non-task oriented conversational system, able to perform free talks with human users).

In section 3 we present some works in the field of cognitive science on humor and metaphor processing. We describe the salience imbalance mechanism (3.1) and lay special focus on a work that joins the topics of humor and metaphors, by describing

Contact: Pawel Dybala, Otaru University of Commerce, Midori 3-5-21, 047-8501 Otaru, Japan. e-mail: paweldybala@res.otaru-uc.ac.jp

how salience imbalance can be applied to humorous metaphors (3.2). In section 4 we briefly describe the current state of the art in the field of artificial metaphor generation and understanding, with special respect to Masui et al.'s MURASAKI system, able to generate word descriptors in Japanese [Masui et al., 2008]. Section 5 describes our research project, its development steps (5.1) and proposed algorithm of the HumMeR humorous metaphor misunderstanding generator, as well as a chatterbot in which it will be implemented (5.2).

2. Humor and emotions – our research so far

In this section we summarize our previous research in the field of humor and emotions processing. We developed PUNDA - a pun telling system for Japanese, and combined it with a chatterbot to obtain a pun-telling conversational system (see 2.1). To such a system we implemented ML-Ask - an emotiveness analysis system (2.2), which detects human users emotions from their utterances and on this basis decides, whether or not a joke should be told. As the final effect, we developed a pun-telling conversational system, which tells jokes accordingly to users emotions (see 2.3). Its performance was evaluated in numerous experiments. Some results are summarized in section 2.3.

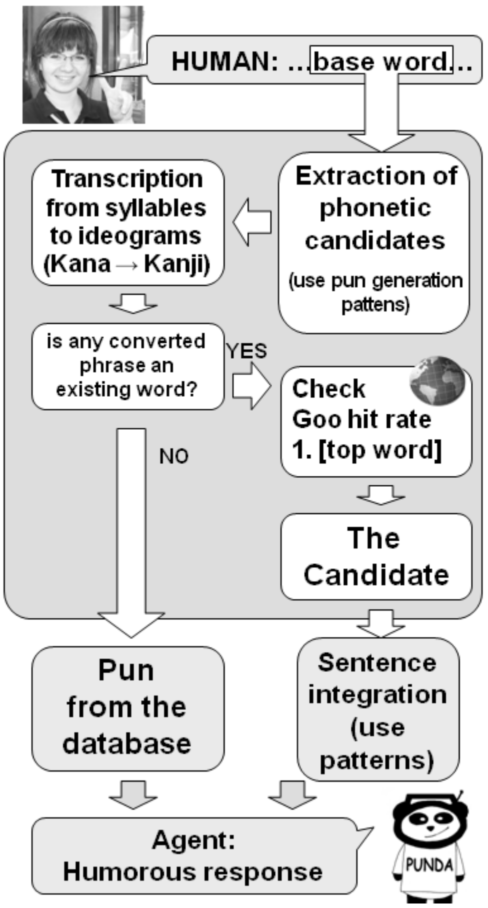

Figure 1: Pun generating system algorithm outline

2.1 Pun-telling conversational system

The PUNDA pun-telling conversational system for Japanese was developed by Dybala et al. [Dybala et al., 2010b; Dybala, 2011] by combining a pun generator [Dybala et al., 2008] and a chatterbot designed by Takahashi et al. [Takahashi, 2009]. Chatterbots, also known as non-task oriented conversational systems, are able to perform free, non-constrained conversations with humans, without any particular topic restriction.

Based on a complex Japanese pun classification [Dybala, 2006], the system generates jokes-including answer using human interlocutor's utterance as an input, which makes it relevant to what the human said. Below we present an example of the system in action:

User: - <u>*Kaeru*</u> *daikirai!* (I hate frogs!)
System: -<u>*Kaeru*</u> *to ieba tsu*<u>*kaeru*</u> *no desu ne*. (Speaking of frogs, we could use that!)

The system algorithm is presented on Figure 1.

From the user utterance, the system first extracts a base word, which will be transformed into a pun. Next, it uses pun generation patterns, based on Dybala's classification [Dybala, 2006] to generate phonetic pun candidates towards the selected word. In the next step each candidate is converted to Japanese ideograms (Kanji characters). Then the system checks if any of converted phrases is an existing word. If no phrase was found to be an existing word, the system chooses a pun from a database. If yes, the system checks its hit rate on the Internet, and then chooses the one with the highest hit rate for the pun candidate, which is next incorporated into a sentence using Japanese pun sentence templates, prepared beforehand.

2.2 Emotiveness Analysis System

Another system developed in our previous research was Ptaszynski's et al. ML-Ask Emotiveness Analysis System [Ptaszynski et al., 2010; Ptaszynski, 2011], which detects emotions from the textual layer of speech. Its algorithm is presented on Figure 2.

The system first analyses the inputted sentence to check its emotiveness. This is done by checking if it contains so-called "emotive elements". For example, the sentence:

> *"Kono hon saa, sugee kowakatta yo. Maji kowasugi!"*
> (That book, ya know, 'twas a total killer. It was just too scary.),

is recognized as emotive, as it contains emotive elements: *saa* (emphasis), *sugee* (totally), *yo* (emphasis), *maji* (really), -*sugi* (too much) and an exclamation mark. If the sentence was recognized as emotive, the system next detects emotion types it contains. This is done by checking if the sentence contains any "emotive expressions", i.e. expressions that convey particular emotions. For example, in the sentence above, the agent found the emotive expression *kowai* (scary), which belongs to the group called *kyoufu* (fear). If no such expression is recognized, the system uses a web-mining technique to extract emotive associations from the Internet. As the result, we obtain an emotiveness analysis summary, such as one below:

Sentence: *Kono hon saa, sugee kowakatta yo. Maji kowasugi!"* (That book, ya know, 'twas a total killer. It was just too scary.)

Emotive elements: *saa* (emphasis), *sugee* (totally), *yo* (emphasis), *maji* (really), -*sugi* (too much), exclamation mark
Emotive value: 6 (above zero -> specify types of emotions)

Emotive expressions: *kowai* (frightening)
Emotions found: fear
Valence: negative

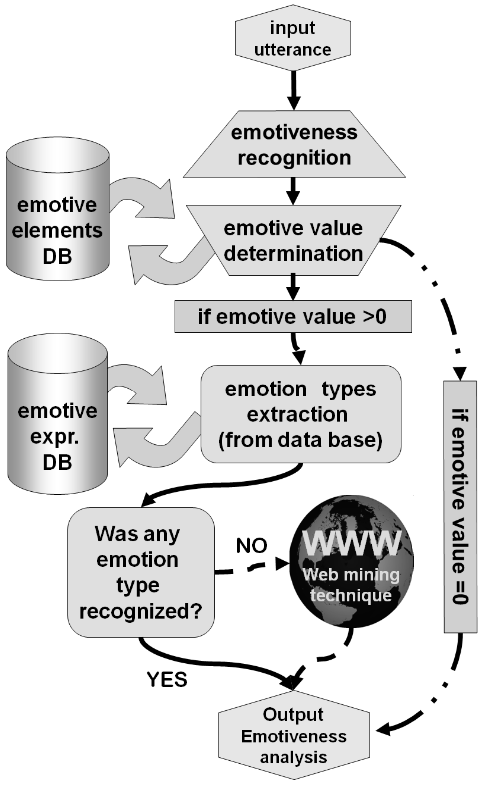

Figure 2: ML-Ask system algorithm outline

2.3 Emotion Aware Joking Conversational System

The two systems described in above sections were combined to construct an emotion aware joking conversational system. Its algorithm is presented on Figure 3.

The ML-Ask system was used to analyze users emotions and on this basis decide whether a joke should be told or not. Basing on existing research (summarized in [Dybala et al., 2010b; Dybala et al., 2012]), we decided that the system should tell jokes if users emotions were assessed as negative or neutral by the ML-Ask system. Thus, it can be said that the system makes effort to enhance humans moods by telling jokes (puns).

If the decision is that a joke should be told, the response to human utterance is generated by the humor-equipped chatterbot (the one described in 2.1). If the system decides otherwise, the response was generated by the baseline chatterbot.

The system's performance was evaluated in two main experiments: user-oriented and automatic, using methodology proposed by Dybala et al. [Dybala et al., 2010a]. In the first one, we asked human users to perform conversations with two systems: the emotion aware joking system and the baseline chatterbot, and to compare their performance. The results showed that most users evaluated the humor-equipped system as better, more friendly and making them fell better than the non-humorous one.

In the second experiment the chat logs from the first one were analyzed by the ML-Ask system to detect users emotive reactions toward both conversational systems. The results were consistent with those gained in the first experiment and showed that in most cases emotions triggered by the humor-equipped system were positive or changed to positive during the interactions.

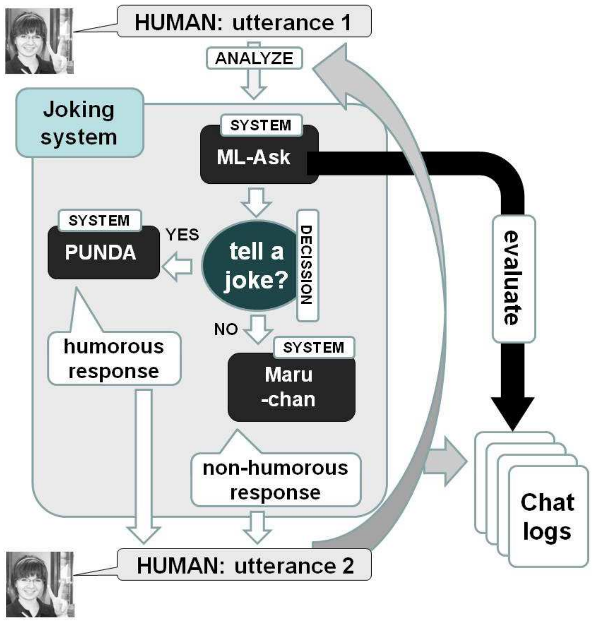

Figure 3: Emotion aware joking conversational system – algorithm outline

3. Humor and metaphors in cognitive science

Numerous publications exist in the field of cognitive science that focus separately on humor or metaphors. However, some works point out similarities between these two issues, stating that, despite some functional differences, they share some properties and tend to be processed using the same or similar mechanisms.

In this section we first describe the salience-imbalance theory, which is one of widely appreciated theories of metaphor understanding (see 3.1). Next, we mention the work of Shen and Engelmayer [Shen et al., 2012], which shows that the mechanism of salience-imbalance works also in humor understanding (3.2).

3.1 Salience imbalance in metaphors

One of the most influent theories explaining the mechanisms working in metaphors is Ortony's salience imbalance theory [Ortony, 1979]. Ortony explains it on two examples of comparisons: literal and metaphorical:

1) Billboards are like placards.
2) Billboards are like warts (they are ugly and stick out)

Both these sentences are constructed using a template:
X is like Y

where X is called the comparison's "target" and Y – its "source". The difference between the two above examples, states Ortony, lies in the relation between salience of properties shared by X and Y. Literal comparisons, like 1), highlight some highly salient properties of both X and Y (here: billboards and placards). Thus, it can be said that literal comparisons display what can be called "salience balance". On the other hand, in metaphorical comparisons, like 2), highly salient properties of source are matched with much less salient properties of target. In comparison 2), for instance, very salient properties of "warts", such as "ugliness" or "sticking out", are at the same time not very salient (albeit not completely implausible) properties of "billboards". In other words, in metaphorical comparisons certain properties of the target, which are normally perceived as not very salient (billboards are not commonly seen as ugly or sticking out) become more salient by comparing the common ground between the target and the source [Ortony, 1979]. Human perception of such comparisons can be summarized as: "indeed, ugliness and sticking out are not the most salient properties of billboards, but since they are compared to warts, these properties appear more salient".

3.2 Salience imbalance in humorous metaphors

The salience imbalance theory, described above, was proposed by Ortony to explain the process of understanding metaphors (metaphorical comparisons). However, it was also experimentally showed that (to some extent and after some modifications) a similar approach can be used when analyzing mechanisms that work in humor understanding [Shen et al., 2012].

One of the most influent theories in the field of humor understanding is the "incongruity theory", in which focus is laid on cognitive aspects of funniness. In this approach, a constituent factor of humor is the fact of bringing together two normally different and distant concepts, which surprises and amuses recipients of humorous acts [Ruch, 1998]. According to this theory, cognition process of humor includes first perception of facts in a more logical, serious and "normal" sense, and then sudden understanding of another sense, usually implausible, which results in amusement and laughter [Dybala, 2011].

Thus, it can be seen that the incongruity theory has much in common with the salience imbalance theory. Both of them lay focus on relationships between two concepts and the semantic distance between them (or their properties). If, for instance, we compare a non-humorous and humorous metaphors, such as:

1) A friend is like anchor – providing stability.
2) A friend is like anchor – sometimes you want to throw them out of the boat. [Shen et al., 2012]

we can see that the latter exhibits less salient properties of both target and source than the former (which is in fact the source of incongruity here). This is surprising and unexpected, and thus perceived as funny

On this basis, Shen and Engelmayer assumed that humorous effect in metaphors derives from a change in the pattern of the salience imbalance. The nature of this change lies in highlighting surprising commonalities between source and target [Shen et al., 2012]. In other words, the degree of salience imbalance (the difference between salience of target and salience of source) should be higher in humorous than in non-humorous metaphors.

To examine this assumption, Shen and Engelmayer conducted three experiments, in which they asked human participants to rate the appropriateness of word descriptions used in humorous and non-humorous metaphors (i.e. "friend: you want to throw it out"). The experiments are described in details in [Shen et al., 2012].

The results provided empirical proof that the assumptions were correct. Participants' rankings revealed greater difference in salience of target and source properties in humorous sentences than in non-humorous ones. This was observed for both explicit and implicit metaphorical comparisons.

Also, another experiment showed that in humorous metaphors a degree of "emotional connotation mismatch" was observed. The participants rated emotional valence of sentence components (negative-positive-neutral). Analysis of results showed that humorous metaphors tend to match concepts that are emotionally opposite (positive-negative or negative-positive) or at least not equal (positive-neutral, negative-neutral etc.). This mechanism works also in the above example 2), in which friends (an emotionally positive entity) is compared to an anchors in a rather negative manner (you want to throw them out of the boat). [Shen et al., 2012].

To summarize this section, it can be stated that the applicability of the salience imbalance theory was empirically proved. Humorous metaphors tend to exhibit higher degree of imbalance than non-humorous ones, which means that they can be processed and comprehended using the same mechanism. This is an important implication for our research project, described below (section 5).

4. Computing metaphors

Although some works exist in the field of computer science that focus separately on processing humor (including one described above) or metaphors (see below), to our best knowledge no research project exists that would join these two areas. Below we briefly summarize the state of the art in the field of automatic metaphor processing (4.1) and shortly describe the work of Masui et al. [Masui et al., 2008], which proposes a method of automatic generation of word sense description from Internet for Japanese (4.2). Some methods and results of this work can be utilized in our research (see 5).

4.1 State of the art

A good summary of current state of the art in metaphor processing in the field of Natural Language Processing is given by Shutova [Shutova, 2010]. Here we mention only some of works in this area.

Research projects in automatic metaphor processing can be divided into two groups: metaphor recognition and interpretation. Works of the former type deal with the problem of distinguishing metaphorical expressions from other types of texts. This includes detecting linguistic cues that indicate the presence of metaphors, such as "metaphorically speaking" or "so to speak", which were identified by Goatly [Goatly, 1997]. Such expressions themselves are not enough to identify metaphors, but they can be used in more complex recognition algorithms.

Some attempts were made to detect metaphors using WordNet [Fellbaum, 1998] as the source of linguistic knowledge. Among these we can name the work of Peters and Peters [Peters et al., 2000], Mason [Mason, 2004] or Krishnakumaran and Zhu [Krishnakumaran et al., 2007]. Other works, such as the one of Gedigan et al. [Gedigian, 2006] use FrameNet [Fillmore et al., 2003] to obtain lexical data related to particular frames ("motion" and "cure"). Another popular source of linguistic data is large scale corpora, from which relations between source and target attributes can be extracted.

Perhaps the most common problem in metaphor detection comes from the fact that metaphors cannot be easily defined and distinguished from other linguistic phenomena, such as polysemy. Thus, in many cases the distinction stops at the "literal vs. non-literal" stage.

While some researchers focus on metaphor recognition, other attempt to deal with metaphor understanding. MIDAS system, developed by Martin [Martin, 1990] uses an existing metaphor database to search for metaphors similar to one inputted, and, if none such metaphor is found, it performs an ontological analysis to find analogical relations on a higher hierarchical level, comparing more general concepts. This work is worth mentioning also because it was integrated with the Unix Consultant system, which answers users' questions about Unix, which is one of the first and few attempts to incorporate metaphor understanding into a human-computer dialogue.

In other works authors developed metaphor-based reasoning frameworks, which use manually coded knowledge about domains and concepts. One such system was developed by Narayanan [Narayanan, 1997]. The system, named KARMA, takes parsed text as input and operates mostly within the source domain. The results are next projected on a target domain.

Another project worth mentioning is Veale and Hao's "Talking Points" [Veale et al., 2008]. The authors developed sets of characteristic concepts belonging to source and target domains. This knowledge was automatically extracted from the Internet (including WordNet). These sets are organized in a framework, in which operations like insertions or substitutions are performed to establish links between concepts. Unfortunately, it is still unclear, to what extent this approach is useful to interpret metaphors occurring in text.

4.2 MURASAKI Word Sense Description System

Above we mentioned some metaphor detecting and metaphor interpreting systems. However, since we are trying to construct a system able to extend concept salience calculation so that it covers also humorous metaphors, we need an algorithm that not only recognizes and interprets metaphors, but also can provide us with some sort of ranking of concepts properties or descriptions. As our previous research on humor-equipped chatterbots was done in Japanese, we decided to conduct this humorous metaphors related project also in this language (having in mind that, as we succeed, it should be possible to develop similar algorithms also for other languages). Thus, we would ideally need a system that generates word descriptions for Japanese, and ranks them according to their relation to the word or phrase.

A system that perform similar tasks is Masui et al.'s MURASAKI [Masui et al., 2008]. The system uses Internet query engines, such as Yahoo, to extract associations (descriptions) towards inputted word or phrase, and sorts generated associations from most to least plausible. This is done by checking cooccurrences of each description with the input. On this basis, score is calculated that reflects the description's appropriatness (i.e. degree of relevance to the inputted phrase). Thus, it can be stated that MURASAKI system calculates salience of concept properties. For instance, towards the word *ringo* (apple), the system would generate a list of descriptions like one below:

1. *kaori* (aroma) score: 0.147
2. *sawayaka* (invigorating) score: 0.075
3. *fruity* (fruity) score: 0.062
4. *sanmi* (sourness) score: 0.053

....

10. *hoo* (cheeks) score: 0.026

The details of score calculation can be found in [Masui et al., 2008].

5. Our research project

The goal of our current research project is to construct a system able to generate humorous metaphors (or humorous metaphor misunderstandings), and to implement it into a chatterbot, thus creating a conversational system able to use funny metaphors in proper way. We named the system HumMeR (abbreviation from HUMorous MEtaphor GeneratoR).

Below we summarize this project's development steps (see 5.1) and describe proposed system's algorithms (5.2).

5.1 Development steps

In this section we describe development steps of our research project, aimed at constructing a system able to generate humorous metaphor misunderstandings, which is next going to be implemented into a chatterbot.

In our research we initially focus on explicit metaphors, i.e. such that fit commonly used templates, like "X is like Y". After developing all necessary procedures for this type of metaphors, we then procede to the next phase of this research, in which we plan to create similar algorithms also for implicit metaphors.

1) Metaphor database and metaphor patterns database construction

In every language metaphors exist that are commonly known, such as, for instance, "as cool as cucumber". In many cases they gained idiomatical status, and thus can be immediately understood by anyone. Although we are aiming at constructing a system able to interpret metaphors automatically, in our oppinion a system like that should also need a database of most "classical" and commonly known metaphorical expressions. We are planning to construct such database by using existing metaphor dictionaries for Japanese (such as [Nakamura, 1995]), from which we will extract not only metaphors themselves, but also some structural patterns that are often used in Japanese metaphors, such as "*X no you na Y*" ("X such as Y"). Next, we will use the gathered metaphors and patterns to extend the database using the Internet and large scale text corpora, such as one created by Maciejewski et al. [Maciejewski et al., 2010].

2) Salience imbalance analysis in metaphors

Next we will analyze the salience imbalance degree that occurs in metaphors. The salience will be measured using scores calculated by the MURASAKI system. For each target and source descriptors will be generated from the Internet, and their relevance to the input will be calculated on the basis of their co-occurrence on te web. It might be necessary to modify some of the system's settings to narrow down the associations lists, to, for instance, avoid the situations when it contains only synonyms, which may require grouping of some of the descriptors. The results will be then analyzed to find regularities in salience imbalance, i.e. to extract some universal rules that would allow to assess the minimal and maximal level of salience imbalance for a pair of phrases to construct a metaphor. As a result, we will obtain a database of descriptors salience and a database of non-humorous metaphors salience imbalance thresholds (i.e. degrees of minimal and maximal salience imbalance necessary for two concepts to constitute a metaphor).

3) Salience imbalance analysis in humorous metaphors

It is quite problematic to acquire enough examples to construct a database of humorous metaphors. This is due to the fact that they often are quite novel and surprising, and thus not necessarily stored anywhere. Therefore, we are planning to employ the approach proposed by Shen and Engelmayer [Shen et al., 2012] and generate some possibly humorous metaphors first by simply increasing the degree of salience imbalance between the source and the target properties. Metaphors generated in this manner will then be evaluated for both aptness and funniness by human evaluators. The algorithm will be improved according to the results of this evaluation. As a result, we will obtain a database of humorous metaphors salience imbalance thresholds (i.e. degrees of minimal and maximal salience imbalance necessary for two concepts to constitute a metaphor).

4) Humorous metaphors misunderstanding generator construction

Having gathered all the necessary data and resources, we will construct HumMeR - a system able to generate humorous misunderstandings of metaphors. Its algorithm is showed on Figures 4, 5 and 6 and described in section 5.2.

5) Evaluation I

The HumMeR system will be evaluated in a third-person oriented evaluation experiment (see [Dybala et al., 2010a] for details). We will again use some of the existing metaphors from the database (different than those used in step 3) as an input and have the system generate humorous misunderstandings towards them. The input-output pairs will then be assessed by human evaluators for aptness and funniness.

As described in section 5,1, we are also planning to develop a procedure to generate misunderstandings also towards novel metaphors, i.e. such that cannot be found in the database. This procedure will be evaluated in a first-person oriented experiment, in which we will ask human participants to use the system and input some novel metaphors, towards which the system will try to generate humorous misunderstandings. The output will be evaluated by the participants for aptness and funniness.

6) Implementation into a chatterbot

The HumMeR system developed in previous stages will be implemented into the same chatterbot that was used in our previous research (see section 2). By doing this, we will construct a chatterbot able to use two types of humor: puns and humorous metaphor misunderstandings, accordingly to users' emotions (detected by ML-Ask, as described in 2.2). As for the timing of using these two types of humor, we are first going to give priority to metaphors, as possibilities to generate misunderstandings occur less often than to generate puns. This, however, will be only an initial setup, which can be changed under the influence of numerous factors. It is also possible that current emotion-based timing rule (use humor if user emotions are negative or neutral) should be different for metaphor misunderstandings than for puns. These settings will be tested empirically and adjusted according to the results.

7) Evaluation II

The chatterbot with implemented metaphor misunderstanding generation system will be evaluated in first person oriented experiments, in which human participants will interact with the system and evaluate its performance. We are also planning to conduct an automatic evaluation experiment, in which the chat logs from the first person oriented one will be analysed by the ML-Ask system in order to investigates users emotive reactions towards the system (using the standards proposed by Dybala et al. [Dybala et al., 2010a]).

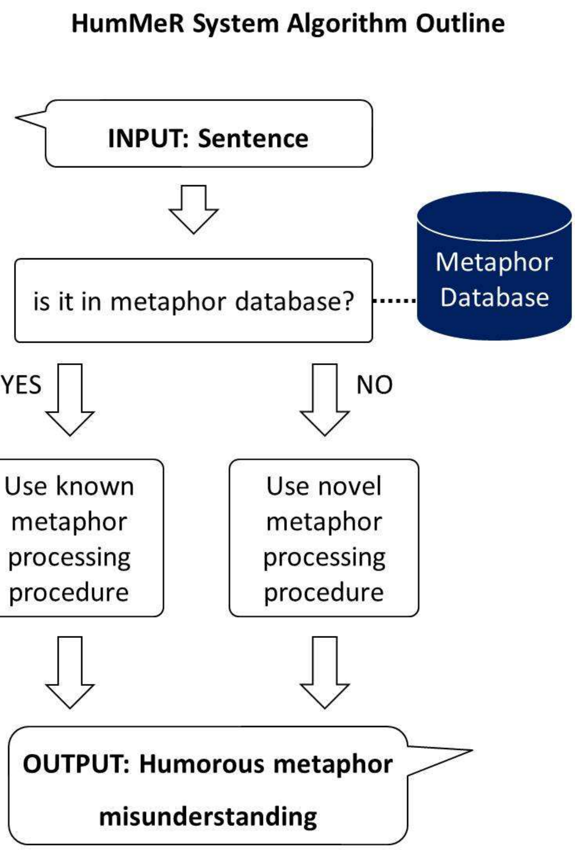

Figure 4: HumMeR System – alhorithm outline

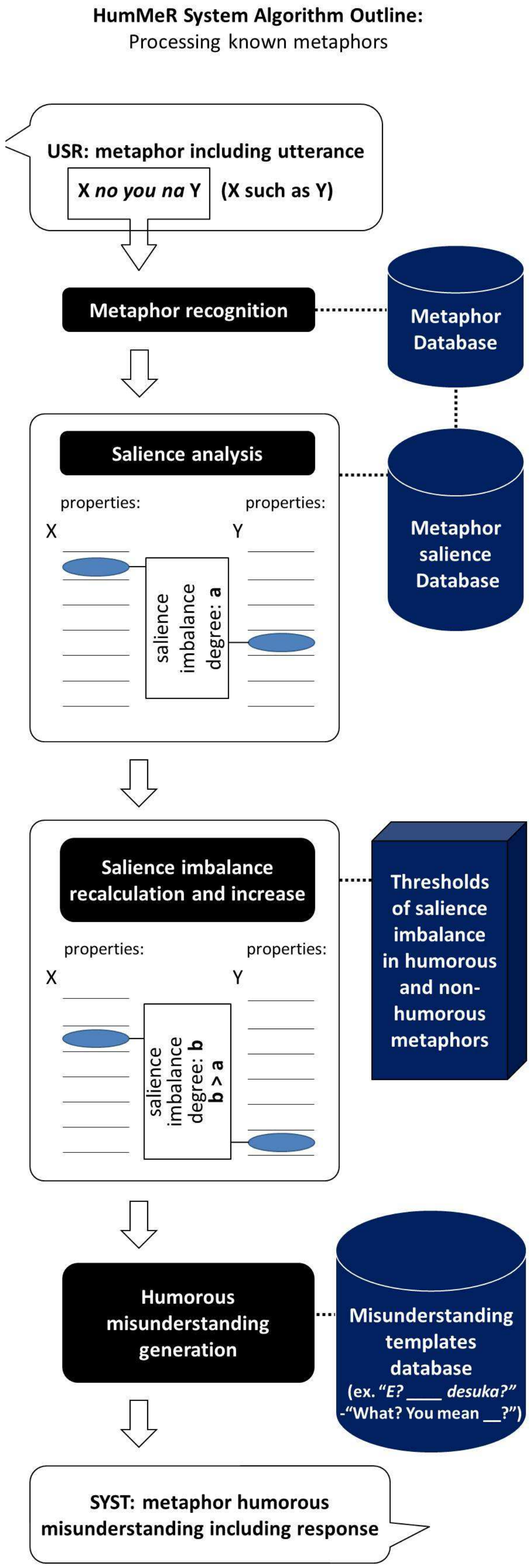

Figure 5: HumMeR System – known metaphor processing procedure outline

5.2 System algorithm

In this section we explain the HumMeR system'algorithm, including the procedure used to process known metaphors (5.2.1), novel metaphors (5.2.2), and the general flow of chatterbot algorithm after implementation of the humorous misunderstandings generation system (5.2.3).

The general outline of the HumMeR system is showed on Figure 4. The input is a sentence, possibly including a metaphor. First the system checks if the input can be found in the metaphor database. If yes, it uses the known metaphor processing procedure. If no, it uses the novel metaphor procedure, which first checks if the input is a metaphor (see below).

5.2.1 Known metaphors processing procedure

Figure 5 presents an outline of procedure that will be used to process known metaphors (i.e. such that can be found in the database) and to generate humorous metaphorical understandings toward them.

The system first analyzes the input to check if it contains a metaphor that can be found in metaphor database. This will recquire some grammatical transformations of the input to query in in all possible forms (i.e. various tenses, aspects etc.). Next, if the metaphor is found in the database, the system checks its salience imbalance, using the data from database. The calculated salience imbalance "a" is next used as a baseline to extract another pair of descriptors with salience imbalance "b", which should be higher than "a" (as explained above, humorous metaphors have higher salience imbalance than non humorous ones). This is compared with thresholds of salience imbalance in humorous and non-humorous metaphors, which will be defined in previous stage of this project (see 5.1). If, for instance, recalculated salience imbalance "b" exceeds the maximal level defined for humorous metaphors, there is a possibility that the components of metaphor (new properties of the target and source) are too distant and thus constitute rather abstract than humorous sentence.

Finally, the system uses the newly extracted description (property) of inputted metaphor's source and target to generate a humorous misunderstanding including response. To do this, it will use a database of misunderstanding patterns, prepared manually beforehand. The database will contain templates such as "*E? {extracted description} desuka*?" ("What? You mean {extracted description}?). As an output, we will obtain a humorous misunderstanding of the inputted metaphor.

If, for instance, in the input the system founds the metaphor like "*ookami no you na otoko*" ("A man like a wolf"), it also checks extract the common description (property) of "man" and "wolf" used in this metaphor, here being "sly, cunning", along with its salience. Next, it checks other descriptors of these two components, in the search of other common properties, which match the humorous metaphors salience imbalance threshold. This, for instance, can be the property of "being hairy", which can next be inserted into a template to generate humorous metaphor misunderstanding such as "*E? Kebukai desuka?*" ("What? You mean hairy?).

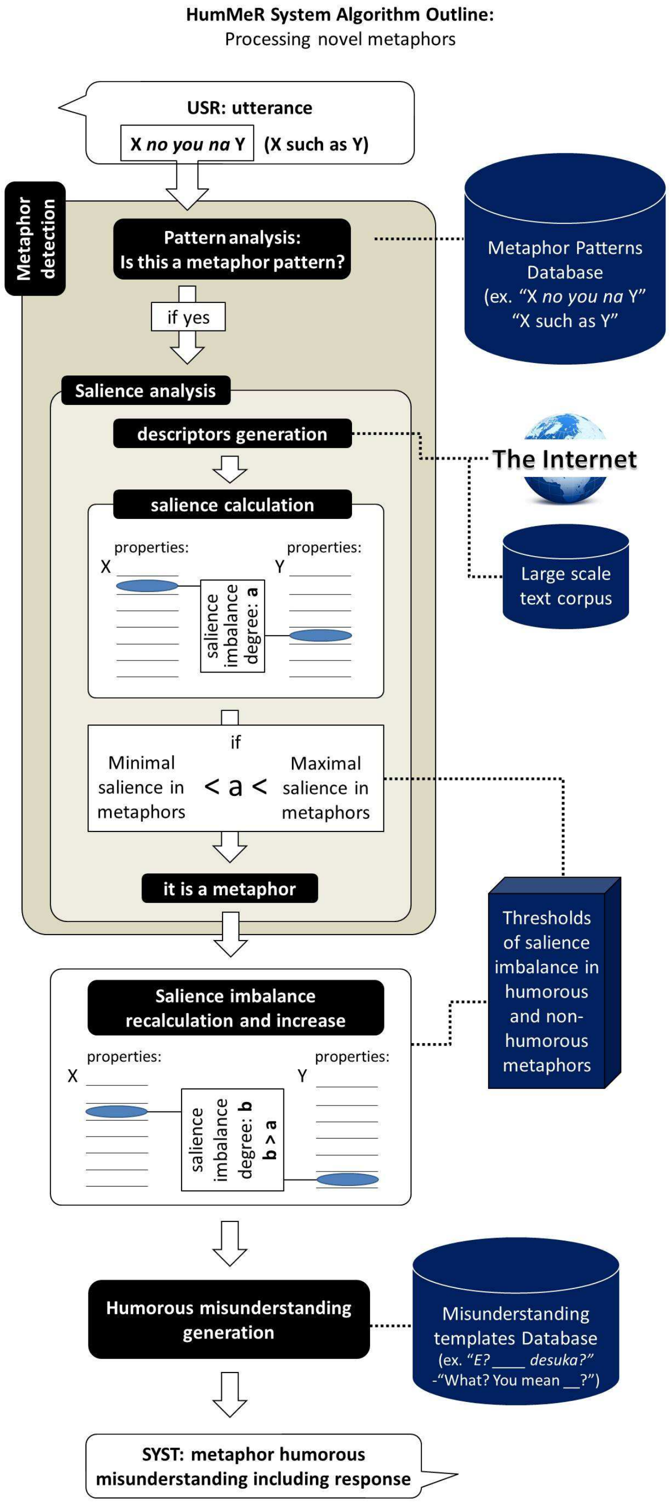

Figure 6: HumMer System – novel metaphor processing procedure outline

5.2.2 Novel metaphors processing procedure

Figure 6 presents an outline of procedure that will be used to process novel metaphors (i.e. such that cannot be found in the database) and to generate humorous metaphorical understandings toward them.

If the input is not found in the metaphor database, the system analyses it to check if it contains any of the metapor patterns (such as, for instance, "X *no you na* Y" – "X such as Y"). To do that, it uses the metaphor patterns database. If such pattern is found, the system enters the next phase, in which it calculates salience of the components found in the input. The system first generates descriptors (properties) for each component (possible source and target), using similar procedures that those used in MURASAKI [Masui et al., 2008]. As the resource, the system will use the Internet, although we are also planning to use a large scale text corpus [Maciejewski et al., 2010], or a combination of these two. Next, the system calculates salience of each descriptor and salience imbalance between them ("a"). This value is then compared to thresholds acquired in earlier steps of this research (see 5.1). If it is higher than the minimal and lower then maximal value of salience inbalance set for non-humorous metaphors, the input is classified as metaphor.

Next, the system uses the salience imbalance "a" as a baseline to extract another pair of descriptors with salience imbalance "b" higher than "a". This is compared with thresholds of salience imbalance in humorous and non-humorous metaphors in a procedure similar to that used in the known metaphor processing procedure.Finally, the system uses the newly extracted description (property) of inputted metaphor's source and target to generate a humorous misunderstanding including response (using the misunderstanding templates database).

5.2.3 Chatterbot implementation

The HumMeR system, including its both procedures (for known and novel metaphors processing) will be implemented into a chatterbot in order to place metaphor misunderstanding generation in its natural environment, which is a dialogue. In this phase we are planning to use the system that was developed in our earlier research (see section 2), in which we already implemented the pun telling procedure (PUNDA System). As the baseline chatterbot, at this moment we are intending to use Takahashi's Maru-chan system [Takahashi, 2009], this, however, can be easily changed afterwards.

The outline of the chatterbot algorithm after the implementation of the HumMeR system is showed on Figure 7.

User utterance will first be analyzed by ML-Ask to detect his / her emotional states. On this basis, the system will decide if humor should be used or not to make the user feel better (currently: use humor if emotions are negative or neutral). If the answer is yes, the system next moves to the humor generation procedure. In the initial setup, we decided to give priority to the HumMeR system, as possibilities of generating metaphor misunderstandings seem to occur less often in daily dialogue than those of generating simple puns (this setup, however, can be changed in further phases of the HumMeR's development). Thus, the system first checks if the input (utterance) can be found in the metaphor database. If yes, it will use the known metaphor processing procedure to check if a misunderstanding can be generated. If the system cannot do that, humorous (punning) response is generated by the PUNDA system.

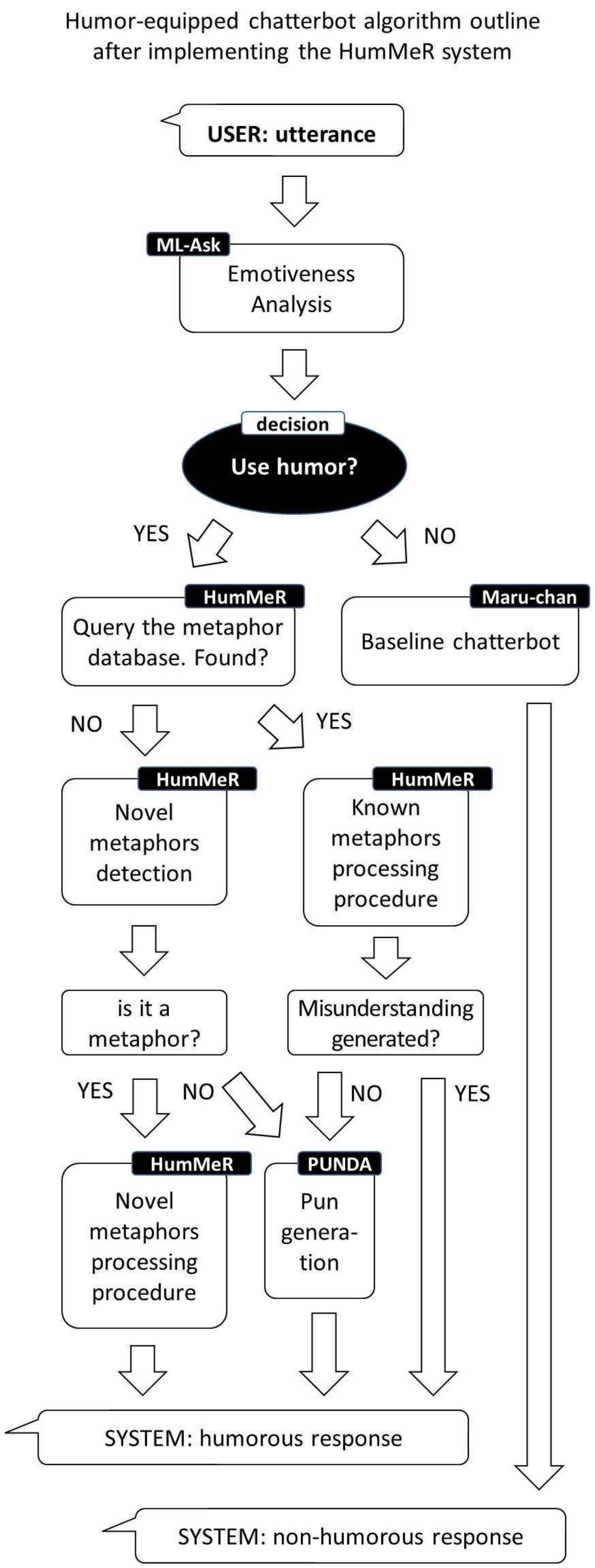

Figure 7: Pun-tellinge emotion aware chatterbot after implementation of the HumMeR system – algorithm outline

If the system does not find any known metaphor in the input, it will use the novel metaphor processing procedure, which first checks if the utterance can contain a metaphor. If yes, the system attempts to generate a humorous misunderstanding. If no, the response towards user utterance is generated by the PUNDA system, which is also used if the HumMeR system fails to generate a misunderstanding (i.e. no common properties will be found that would match the humorous salience imbalance patterns).

As a result, we will obtain a humorous (if the system's decision was to use humor) or non-humorous response to user's utterance.

6. Conclusion and further ideas

In above sections we presented the outline of our research project, aimed at creating a system able to generate humorous metaphorical misunderstandings, which will be implemented into a chatterbot. The system, named HumMeR, is currently under construction, and thus we still need to consider some possible improvements and issues in the project development.

The plan described above assumes working on explicit metaphors, i.e. such created using some typical formal templates, like "A is like B". Such metaphors are easier to test all the mechanisms and algorithms, which then can be used to process also implicit metaphors. This will recquire some corrections or creating additional databases, but implicit metaphors still remain in our focus of interest.

An interesting issue in humor generation in general is the fact that some topics seem to have greater funniness potential than other. This should work also in generating humorous metaphor misunderstandings. For instance, someone's appearance might be easier to make fun of than his / her beliefs or preferences. Thus, we also consider giving priority in salience calculation to properties from certain domains, in order to increase perceived funniness of the output.

Acknowledgements

This work was supported by KAKENHI (Project Number: 23-01348)

References

[Dybala, 2006] Pawel Dybala, *Dajare - Nihongo ni okeru do'on igi ni motozuku gengo yugi* (Dajare - Japanese puns based on homophony), Krakow, Poland: Jagiellonian University, 2006.

[Dybala et al., 2008] Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, "Humor Prevails! - Implementing a Joke Generator into a Conversational System", *LNAI, Vol. 5360*, 214-225, Springer Berlin & Heidelberg, 2008.

[Dybala et al., 2010a] Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, "Evaluating subjective aspects of HCI on an example of a non-task oriented conversational system", *Int. J. Artif. Intell. T., Special Issue on AI Tools for HCI Modeling 19*(6), 819-856, 2010.

[Dybala et al., 2010b] Pawel Dybala, Michal Ptaszynski, Jacek Maciejewski, Mizuki Takahashi, Rafal Rzepka, Kenji Araki, "Multiagent system for joke generation: Humor and emotions combined in human-agent conversation", *Journal of Ambient Intelligence and Smart Environments 2,* 31-48, 2010.

[Dybala, 2011] Pawel Dybala, *Humor to Facilitate HCI: Implementing a Japanese Pun Generator into a Non-task Oriented Conversational System*, Lambert Academic Publishing, 2011.

[Dybala et al., 2011] Pawel Dybala, Michal Ptaszynski, and Kohichi Sayama, "Reducing Excessive Amounts of Data: Multiple Web Queries for Generation of Pun Candidates," *Advances in Artificial Intelligence, vol. 2011*, Article ID 107310, 12 pages, 2011. doi:10.1155/2011/107310

[Dybala et al., 2012] Pawel Dybala, Michal Ptaszynski, and Kohichi Sayama, "A Step Towards Joking AI: Multiagent Humor-Equipped Conversational System", Chapter in *Agent-Based Approaches to Ambient Intelligence*, IOS Press (to appear in 2012).

[Fellbaum, 1998] Christiane Fellbaum, editor, *WordNet: An Electronic Lexical Database* (ISBN: 0-262-06197-X). MIT Press, first edition, 1998.

[Fillmore et al., 2003] Charles J. Fillmore, Christopher R. Johnson, Miriam R.L. Petruck, 2003, "Background to FrameNet", *International Journal of Lexicography, 16*(3):235–250, 2003.

[Gedigian, 2006] Matt Gedigian, John Bryant, Srini Narayanan, Branimir Ciric, "Catching metaphors", In In Proceedings of the 3rd Workshop on Scalable Natural Language Understanding, pages 41–48, New York, 2006.

[Goatly, 1997] Andrew Goatly, *The Language of Metaphors*, Routledge, London, 1997.

[Krishnakumaran et al., 2007] Saisuresh Krishnakumaran, Xiaojin Zhu, "Hunting elusive metaphors using lexical resources", In Proceedings of the Workshop on Computational Approaches to Figurative Language, pages 13–20, Rochester, NY, 2007.

[Maciejewski et al., 2010] Jacek Maciejewski, Michal Ptaszynski, Pawel Dybala, "Developing a Large-Scale Corpus for Natural Language Processing and Emotion Processing Research in Japanese", In Proceedings of the International Workshop on Modern Science and Technology (IWMST), Kitami, Japan, pp. 192-195, 2010.

[Martin, 1990] James H. Martin, *A Computational Model of Metaphor Interpretation*, Academic Press Professional, Inc., San Diego, CA, USA, 1990.

[Mason, 2004] Zachary J. Mason, "Cormet: a computational, corpus-based conventional metaphor extraction system", *Computational Linguistics, 30*(1):23–44, 2004.

[Masui et al., 2008] Fumito Masui, Yosifumi Kawamura, Jun'ichi Fukumoto, Naoki Isu, "MURASAKI: Web-based Word Sense Description System", The 23rd International Technical Conference on Circuits/Systems, Computers and Communications, pp. 1285-1288, July 2008.

[Nakamura, 1995] Akira Nakamura, *Hiyu hyougen jiten* (Dictionary of metaphorical expressions), Kadokawa Shoten, 1995.

[Narayanan, 1997] Srini Narayanan, *Knowledge-based action representations for metaphor and aspect (KARMA)*, PhD thesis, University of California at Berkeley, 1997.

[Ortony, 1979] Andrew Ortony, "Beyond literal similarity", *Psychological Review, 86*(1), 161-180, 1979.

[Peters et al., 2000] Wim Peters and Ivonne Peters, "Lexicalised systematic polysemy in wordnet", In Proceedings of LREC 2000, Athens, 2000.

[Ptaszynski et al., 2010] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka, 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, 73-89, 2010.

[Ptaszynski, 2011] Michal Ptaszynski, *Emotion Awareness in Dialog Agents: Affect Analysis of Textual Input Utterance and its Application in Human-Computer Interaction*. Lambert Academic Publishing, 2011.

[Ruch, 1998] Willibald Ruch, Ed., *The Sense of Humor: Explorations of a Personality Characteristic*, Berlin/New York: Mouton de Gruyter, 1998.

[Shutova, 2010] Ekaterina Shutova, "Models of Metaphor in NLP", Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 688–697, Uppsala, Sweden, July 2010

[Takahashi, 2009] Mizuki Takahashi, *Web ni yoru kyoukihindo to n-gram moderu wo mochiita hatsuwabunnseiseishuhou* (Utterance Generation Method Using Web Search and N-grams). Bachelor diss., Hokkaido University, Sapporo, Japan, 2009.

[Turing, 1950] Alan Turing, "Computing Machinery and Intelligence", *Mind LIX* (236): 433–460, October 1950.

[Veale et al., 2008] Tony Veale, Yanfen Hao, "A fluid knowledge representation for understanding and generating creative metaphors", In Proceedings of COLING 2008, pages 945–952, Manchester, UK, 2008.

[Shen et al., 2012] Yeshayahu Shen and Galia Engelmayer, **"**A friend is like an anchor - sometimes you want to throw them out of the boat: What makes a metaphorical metaphorical comparison humorous?", Submitted for publication in Metaphor and Symbol, 2012 (available at: http://www.tau.ac.il/~yshen/publications/A%20friend%20is%20like%20an%20anchor.pdf)