

Research Article

Affect and Metaphor Sensing in Virtual Drama

Li Zhang¹ and John Barnden²

¹ School of Computing, Teesside University, Middlesbrough TS1 3BA, UK

² School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK

Correspondence should be addressed to Li Zhang, l.zhang@tees.ac.uk

Received 24 March 2010; Revised 28 July 2010; Accepted 13 September 2010

Academic Editor: Abdenmour El Rhalibi

Copyright © 2010 L. Zhang and J. Barnden. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We report our developments on metaphor and affect sensing for several metaphorical language phenomena including affects as external entities metaphor, food metaphor, animal metaphor, size metaphor, and anger metaphor. The metaphor and affect sensing component has been embedded in a conversational intelligent agent interacting with human users under loose scenarios. Evaluation for the detection of several metaphorical language phenomena and affect is provided. Our paper contributes to the journal themes on believable virtual characters in real-time narrative environment, narrative in digital games and storytelling and educational gaming with social software.

1. Introduction

In our previous work, we have developed virtual drama improvisational software for young people age 14–16 to engage in role-playing situations under the improvisation of loose scenarios. The human users could be creative at their roleplays. A human director normally monitors the improvisation to ensure that the human actors have kept the general spirit of the scenarios. In order to reduce the burden of the human director, we have developed an affect detection component, EMMA (emotion, metaphor, and affect), on detecting simple and complex emotions, meta-emotions, value judgments, and so forth. This affect sensing component has been embedded in an intelligent agent, which interacts with human users and plays a minor role with the intention to stimulate the improvisation. In one session, up to 5 characters are involved in. The affect sensing component can detect 25 affective states in our previous development [1].

Metaphorical language has also been intensively used to convey emotions and feelings in the collected transcripts during the testing. The work presented here reports further developments on metaphor interpretation and affect detection for several particular metaphorical expressions with affect implication, which include affects as physical objects metaphor (“anger ran through me,” “fear drags me down”), food metaphor (“X is walking meat,” “Lisa has a pizza face”

and “you are a peach”), animal and size metaphor (“X is a fat big pig”, “shut ur big fat mouth”) and anger metaphor (“she exploded completely”, “he fired up straightaway”, and “she heated up just as fast”). Size metaphor also plays an important role in indicating affect intensities. We have detected these several metaphorical language phenomena using decision tree, naïve Bayes classifier, and support vector machine with the assistance of syntactic parsing and semantic analysis. WordNet and WordNet-affect domains have also been used to detect affect from the identified figurative language phenomena.

Also, several loose scenarios have been used in our study, including school bullying and Crohn’s disease. The animation engine adopts the detected affect implied in users’ text input to produce emotional gesture animation for the users’ avatars. The AI agent also provides appropriate responses based on the detected affect from users’ input in order to stimulate the improvisation.

In school bullying and Crohn’s disease scenarios, the AI agent plays a minor role in drama improvisation. For example, it plays a close friend of the bullied victim (the leading role) in school bullying scenario, who tries to stop the bullying and a close friend of the sick leading character in Crohn’s disease scenario who tries to give support to his friend with the decision on his friend’s life-changing operation.

We have also analysed affect detection performance based on the collected transcripts from user testing by calculating agreements via Cohen's Kappa between two human judges, human judge A/the AI agent and human judge B/the AI agent, respectively. A corpus extracted from the collected transcripts and other similar sources has also been used to evaluate the metaphorical phenomena recognition based on various machine learning approaches.

The content is arranged in the following way. We report relevant work in Section 2 and the new developments on metaphor, affect, and affect intensity detection for the processing of affect, food, animal, size, and anger metaphor in Section 3. Brief discussion on how the detected affects contribute to the emotional animation is provided in Section 4. Evaluation results of the metaphor and affect detection component are reported in Section 5. Finally, we summarize our work and point out future directions in Section 6.

2. Relevant Work

Textual affect sensing is a rising research branch for natural language processing. ConceptNet [2] is a toolkit to provide practical textual reasoning for affect sensing for six basic emotions, text summarization, and topic extraction. Shaikh et al. [3] provided sentence-level textual affect sensing to recognize evaluations (positive and negative). They adopted a rule-based domain-independent approach, but they have not made attempts to recognize different affective states from open-ended text input.

Although Façade [4] included shallow natural language processing for characters' open-ended utterances, the detection of major emotions, rudeness, and value judgements is not mentioned. Zhe and Boucouvalas [5] demonstrated an emotion extraction module embedded in an Internet chatting environment. It used a part-of-speech tagger and a syntactic chunker to detect the emotional words and to analyse emotion intensity for the first person (e.g., "I" or "we"). Unfortunately the emotion detection focused only on emotional adjectives and did not address deep issues such as figurative expression of emotion (discussed below). Also, it seems to have limited the system's functionalities on affect interpretation by focusing purely on first-person emotions. There has been relevant work on general linguistic cues that could be used in practice for affect detection (e.g., Craggs and Wood [6]).

There is also well-known research work on the development of emotional conversational agents. Egges et al. [7] have provided virtual characters with conversational emotional responsiveness. Elliott et al. [8] demonstrated tutoring systems that reason about users' emotions. They believe that motivation and emotion play very important roles in learning. Virtual tutors have been created in a way that not only having their own emotion appraisal and responsiveness, but also understanding users' emotional states according to their learning progress. Aylett et al. [9] also focused on the development of affective behaviour planning for the synthetic characters. Cavazza et al. [10] reported a conversational agent embodied in a wireless

robot to provide suggestions for users on a healthy living lifestyle. Hierarchical Task Networks (HTNs) planner and semantic interpretation were used in this work. The cognitive planner played an important role in assisting with dialogue management, for example, giving suggestions to the dialogue manager on what relevant questions should be raised to the user according to the healthy living plan currently generated. The user's response was also adopted by the cognitive planner to influence the change of the current plan. The limitation of such planning systems was that they normally worked reasonably well within the predefined domain knowledge, but their performance became worse when open-ended user input going beyond the planner's knowledge was used intensively during interaction. The system we present here intends to deal with such challenge.

Moreover, metaphorical language has drawn researchers' attention for a while since it has been widely used to provide effective vivid description. Fainsilber and Ortony [11] commented that "an important function of metaphorical language is to permit the expression of that which is difficult to express using literal language alone". Metaphorical language can be used to convey emotions implicitly and explicitly, which also inspires cognitive semanticists [12].

Indeed, the metaphorical description of emotional states is common and has been extensively studied (Fussell and Moss [13]), for example, "he nearly exploded" and "joy ran through me," where anger and joy are being viewed in vivid physical terms. Such examples describe emotional states in a relatively explicit if metaphorical way. But affect is also often conveyed more implicitly via metaphor, as in "his room is a cess-pit"; affect (such as "disgust") associated with a source item (cess-pit) gets carried over to the corresponding target item (the room). There is also other work conducting theoretical research on metaphor in general (see, e.g., Barnden et al. [14]; Barnden [15]), which could be beneficial to our application as a useful source of theoretical inspiration.

Our work is distinctive in the following aspects: (1) metaphor and affect detection in figurative expressions; (2) real-time affect sensing for basic and complex affects, meta-emotions, value judgments and so forth, (including 25 affective states) from improvisational open-ended user input; (3) expressive animation driven by the detected affective states from users' input.

3. Metaphor and Affect Sensing

Before we introduce the new developments on affect, food, animal, size, and anger metaphor, we briefly introduce our previous work on affect detection and responding strategy development for the AI agent. As mentioned earlier, our original system has been developed for age 14–16 secondary school students to engage in role-play situations in virtual social environments. Without predefined constrained scripts, the human users could be creative in their role play within the highly emotionally charged scenarios. The AI agent could be activated to interact with human actors by playing a minor bit-part character in the two scenarios.

We have used responding regimes for the conversational AI agent in order to stir up the discussion and stimulate the improvisation. For example, the responses can be activated when its confidence on interpreting affect from users' input is high.

The language used by the secondary school students during their roleplay is highly diverse with various online chatting features. Thus, before affect detection processing, we have implemented preprocessing procedures including spelling checking, abbreviation checking, and Metaphone algorithm dealing with letter repetitions in interjections and onomatopoeia in order to recover the standard user input. The recovered user input is sent to the Rasp parser to obtain syntactic information. We have particularly focused on users' input with potential emotional implication, such as diverse imperatives ("Lisa, go away", "you leave me alone", "Dave bring me the menu", and "do it or I will kill u") and statements with a structure of "first-person + present-tense verb" ("I like it", "I hate u", and "I enjoy the meal"). In addition, the approach followed for the detection of affect intensity was limited to checking punctuation (e.g., repeated exclamation marks) and capitalization in users' input.

Overall, we have adopted rule-based reasoning, robust parsing, pattern matching, and semantic and sentimental profiles (e.g., WordNet and a semantic profile [16]) in our approach. Jess, the rule engine for Java platform, has been used to implement the rule-based reasoning while Java has been used to implement other algorithms and processing with the integration of the off-the-shelf language processing tools, such as Rasp and WordNet.

In this study, we have made further developments on affect detection especially from several different types of metaphorical expressions—size, affect, food, animal, and anger metaphor. Affect intensity has been further explored on size metaphor, size adjectives, and degree adverbs. Especially, these several metaphorical language phenomena will also be detected by several machine learning approaches (classifiers) based on our previous rule-based development.

The machine learning approaches have also been trained by 400 extracted examples of these several metaphorical phenomena and literal expressions which are represented by the identified extracted semantic and syntactic structures. The implementation detail is presented in the following.

3.1. Size Metaphor and Affect Intensity. In our study, size adjectives are often used to emphasize the affect conveyed in the users' literal and metaphorical input ("shut ur big fat mouth", "u r a big bully"). As degree adverbs, they could be used to measure intensity of the affect conveyed. In our previous work, affect intensity is simply judged by punctuations and repeated letters, syllables in interjections and ordinary words, and so forth. We now employ size adjectives and degree adverbs to reason about intensity. In order to facilitate our study, we have created our own semantic dictionary. It contains not only size adjectives and degree adverbs with their corresponding semantic tags but also emotional and affective terms, food terms, animal names, and so forth. The semantic annotations used in our

TABLE 1: Size adjectives and degree adverbs and their corresponding semantic tags.

Size adjectives and degree adverbs	Semantic tag
Maximizer adjectives (e.g., huge)	n3.2++
Booster adjectives (e.g., "big", "massive", and "fat")	n3.2+
Diminisher adjectives (e.g., "little", "small", and "tiny")	n3.2-
Maximizer adverbs (e.g., "completely")	a13.2
Booster adverbs (e.g., "greatly")	a13.3
Approximator adverbs (e.g., "almost")	a13.4
Compromiser adverbs (e.g., "pretty")	a13.5
Diminisher adverbs (e.g., "slightly")	a13.6
Minimizer adverbs (e.g., "hardly")	a13.7
Frequency minimizer adverbs (e.g., "rarely").	n6-

semantic dictionary have been borrowed from Wmatrix [17], which facilitates users to obtain corpus annotation with semantic and part-of-speech tags to compose dictionary. For example, for size adjectives, since "n3.2" represents measurement, size according to Wmatrix, a semantic tag "n3.2++" is used to label maximizer adjectives such as "huge", with "n3.2+", used to indicate booster adjectives such as "big", "massive", and "fat", and "n3.2-" used to mark diminisher adjectives such as "little", and "small", "tiny". For degree adverbs, Wmatrix uses "a13" to represent degree generally with "a13.x" to indicate a particular type of such adverbs. "N6-" is used to indicate frequency minimizer adverbs (e.g., "rarely"). After the metaphorical phenomena and affect detection using various methods reported in the following with the assistance of sentence types information obtained from Rasp, the system checks for these intensity indicators (size adjectives and degree adverbs) to reason about affect intensities. Table 1 lists all types of size adjectives and degree adverbs that have been considered in our paper.

First of all, at the beginning of metaphor and affect detection, Rasp is used to obtain the sentence type information from user input. It also reports part-of-speech information for each word in the user input. Then after affect is detected from user input, all the adjectives and adverbs (indicated by their part-of-speech tags) from the user input are attached with their corresponding semantic tags provided by the semantic dictionary mentioned above. If maximizer and booster size adjectives (e.g., "huge", "big", "fat" etc.) and degree adverbs (e.g., "completely", "greatly", and "extremely") are detected, then we conclude that the affect intensity is strong (e.g., "u r completely a big idiot" and "keep your big mouth shut"). If approximator and compromiser adjectives and adverbs (e.g., "almost", "nearly", "rather", "quite", and "pretty") are presented, then we believe that the user input implies affect with medium intensity ("u r quite cool", "fear nearly kills me"). Otherwise, if diminisher size adjectives (e.g., "little", "small", and "tiny") and degree adverbs (e.g., "slightly") are found, the system believes that the intensity of the affect expressed in user input is weak

(or minor) (e.g., “u r just a little idiot”). Finally, if minimizer degree and frequency adverbs (e.g., “hardly”, “rarely”, and “seldom”) are detected, the affect detected from the user input is discounted (e.g., “Lisa hardly is a pizza/freak”, “fear rarely controls me”). Although the intensity processing could be fooled by user input with complex syntactic structures, our current processing is effective enough in dealing with intensity detection in sentence-level conversational interaction.

From the collected transcripts, there is also one particular phenomenon of theoretical and practical interest, that is, physical size is often metaphorically used to emphasize evaluations, as in “you are a big bully”, “you’re a big idiot”, and “you’re just a little bully.” The bigness is sometimes literal as well. Sharoff [18] indicates that “big bully” expresses strong disapproval and “little bully” can express contempt, although “little” can also convey sympathy or be used as an endearment.

3.2. Affect Metaphor Interpretation. Affect terms have been used intensively during online interaction. Besides they have been used literally to convey users’ emotional states (e.g., “I am angry”, “I get bored”), affect terms have been mentioned in affective metaphorical language. One category of such metaphorical expression is “Ideas/Emotions as Physical Objects” [12, 19], for example, “joy ran through me”, “my anger returns in a rush”, “fear is killing me”, and so forth. In these examples, emotions and feelings have been regarded as external entities. The external entities are often, or usually, physical objects or events. Therefore, affects could be treated as physical objects outside the agent in such examples, which could be active in other ways [19]. Implementation has been carried out to provide the affect detection component with the ability to deal with such affect metaphor.

In order to effectively detect such metaphorical expressions, their general semantic and syntactic structures have to be identified, so that these metaphorical expressions could be converted into these structures (to train the classifiers for future recognition). Thus, Rasp has been used to detect statements with a structure of “a singular common noun subject + present-tense/past-tense lexical verb phrase” or “a singular common noun subject + present-tense copular form + -ing form of lexical verb phrase”. A syntactic annotation for each word in the user input has also been provided by Rasp.

Various user inputs could possess such syntactic forms, for example, “the girl is crying”, “the big bully runs through the grass”, and so forth. Our special semantic dictionary has been employed to recover corresponding semantic tags for the singular common noun subjects. As mentioned earlier, the semantic dictionary created consists mainly of emotion and affect terms, food terms, animal names, measurable adjectives (such as size), special verbs (e.g., explode, fire, heat) and so forth, with their corresponding semantic tags due to the fact they have the potential to convey affect and feelings. For example, if the main subject is an affective term (“joy”), then its corresponding semantic tag (“e4.1+”) will be recovered. If it is not recorded in the semantic dictionary (“girl”), then the syntactic part-of-speech tag obtained from Rasp for the main subject is retained (“nn1”).

Thus, with the assistance of the semantic and syntactic analysis, the user input with affect terms as main subjects will be converted into the following structure: “the semantic tag for the main subject + the part-of-speech tag (obtained from Rasp) for the lexical main verb + the part-of-speech tag for the object”. The step-by-step analysis is listed in the following for the user input “anger runs through me”:

- (1) Rasp recognizes the input with a structure of “nn1 (a singular common noun subject: anger) + vvz (present-tense lexical verb phrase: runs) + ppio1 (object: me)”;
- (2) the subject noun term, “anger”, has been sent to the semantic dictionary;
- (3) then the input is interpreted as a semantic syntactic structure of “e3- (semantic tag: anger) + vvz (runs) + ppio1 (me)”.

From such an expression, the system realizes that an emotional state has been used as a subject which carries out an activity indicated by the verb phrase(s). It has been noticed that this extracted structure could be at least extended to other similar expressions belonging to “affects as external entities” metaphor such as “a terrible rage began to seize hold of me”, “blind waves of panic swept over and over him”, and “joy runs through me”. Thus, we need to train the system to regard any new input with such a semantic and syntactic structure as affective metaphor belonging to the category of “affects as entities”. Therefore, we have gathered 80 examples of such metaphorical expressions not only from the collected transcripts from the previous testing but also from an online metaphor databank [19] and represented these examples in the above semantic syntactic structures. Since quantifiers (e.g., “completely”, “almost”, and “hardly”) play an important role in the interpretation of the affect conveyed in the user input as mentioned earlier, they have also been incorporated in the extracted structure represented by their semantic annotations. Moreover, sentence types sometimes also may become affect indicators. For example, imperatives may contain potential emotional implication, especially without softeners such as “please”. Thus, sentence types have been taken into our study.

Thus, these examples have been served as (part of) the training data for several chosen classifiers including decision tree, naïve Bayes classifier, and support vector machine, in order to provide our system with the ability of detecting such a metaphorical language phenomenon effectively and distinguishing these figurative expressions from other types of metaphors and literal expressions. The chosen classifiers have also been trained with examples of other metaphorical phenomena (such as food, animal, and anger metaphor) and literal expressions. Some training examples for affect metaphor are presented in Table 2.

In our processing, we allow the quantifiers to be present at any position within the sentence. For any new input, Rasp informs the system of any input with a structure of “a singular common noun subject + present-tense/past-tense lexical verb phrase” or “a singular common noun subject + present-tense copular form + -ing form of lexical verb

TABLE 2: Affect metaphor training examples (s represents a statement sentence).

Sentence type	Subject	Quantifiers	Verb	Object	Category
s	E3-(“anger”)	null	VVZ (“runs through”)	PPIO1 (“me”)	Affect metaphor
s	E3-(“anger”)	A13.4 (“almost”)	VVZ (“hits”)	PPIO1 (“me”)	Affect metaphor
s	E5-(“panic”)	A13.2 (“completely”)	VVZ (“sweeps over”)	PPHS1 (“him”)	Affect metaphor

phrase”. With the assistance of the semantic dictionary, the semantic annotation of the singular common noun subject is derived as discussed above. The semantic syntactic structure is sent to the classifiers, which are trained by examples from several different language phenomena. Generally the classifiers perform reasonably well for the detection of such affect metaphor although they could be challenged by its variations (e.g., “I stare at her, fighting a rising tide of disbelief”). The evaluation detail is presented in Section 5.

Further processing has also been conducted to sense affect from the identified affective metaphorical expressions. Although the semantic annotation for the main subject has suggested that the subject belongs to emotional states, and the overall user input has been recognized as affect (as entities) metaphor, further processing is needed in order to recover the appropriate affect conveyed in the user input.

WordNet-affect domain (part of WordNet-domain 3.2) [20] has been used in our application. It provides an additional hierarchy of “affective domain labels”, with which the synsets representing affective concepts are further annotated. Thus the singular common noun subject is sent to WordNet-affect in order to obtain the hierarchical affect information. For example, if the subject is the affective term “panic”, then the hierarchical affect information obtained from WordNet-affect is “negative-fear → negative-emotion → emotion → affective-state → mental-state”. A further processing based on the hierarchical affect result leads to the exact affective state conveyed in user’s input—fear (negative emotion). If such an input has a first-person object, “me” (e.g., “panic is dragging me down”), then it indicates the user currently experiences fear. Otherwise if such input has a third-person object, “him/her” (e.g., “panic is sweeping over and over him”), it implies that it is not the user who currently experiences “fear”, but another character. The step-by-step analysis is listed in the following for the new input “panic is dragging me down”, which is not included in the training examples for the classifiers.

- (1) With the assistance of Rasp and the semantic dictionary, the input becomes “e5- (emotional state: panic) + vvg (-ing form of lexical verb: dragging) + ppio1 (me)” (by the procedure described as in the above training example).
- (2) The classifiers deduce and conclude that the input = an affects as entities metaphor.
- (3) The main subject (panic) → WordNet-affect.
- (4) WordNet-affect → the hierarchical affect information → panic: fear (negative emotion).

Obviously the quantifiers may influence the detected affect from the user input as mentioned above. For example,

“hardly” may dismiss the detected affect from the input although it could be recognized as a metaphorical expression (e.g., “anger hardly touches him”); “completely” may emphasize the affect conveyed in the input (e.g., “sorrow completely hits him”).

Moreover, if the user input is literal (e.g., “Lisa hits me”, “the boy sweeps the floor”), the classifiers will not regard it as a metaphorical expression but literal. Thus, other suitable processing methods (e.g., checking syntactic information and affect indicators etc) are adopted to extract affect. On the whole, such processing on metaphor interpretation is indeed at a very initial stage. However, it provides a useful way to recognize affect and affect metaphor in which emotions are used as external entities.

3.3. Food and Animal Metaphor Interpretation. Food has been used extensively as metaphor for social position, group identity, and so forth. For example, food could be used as a metaphor for national identity. British have been called “roastbeefs” by the French, while French have been referred to as “frogs” by the British. It has also been used to indicate social hierarchy. For example, in certain Andean countries, potatoes have been used to represent poor rural farmers of native American descent and white flour and bread have been used mainly to refer to wealthy European descent. In our school bullying scenario, the big bully has called the bullied victim (Lisa) names, such as “u r a pizza”, “Lisa has a pizza face” to exaggerate that fact that the victim has acne. Another most commonly used food metaphor is to use food to refer to a specific shape. For example, body shape could be described as “banana”, “pear”, and “apple”. In our application, “Lisa has a pizza face” could also be interpreted as Lisa has a “round (shape)” face. Therefore, insults could be conveyed in such food metaphorical expression. We especially focus on the statements of “second-person/a singular proper noun + present-tense copular form + food term”.

In our application, Rasp informs the system of the user input with the following structure: “second-person/a singular proper noun + present-tense copular form + noun phrases” (e.g., “Lisa is a pizza”, “u r a hard working man”, and “u r a peach”). The noun phrases are examined in order to recover the main noun term. Then its corresponding semantic tag is derived from the composed semantic dictionary if it is a food term, or an animal-name and so forth. Syntactic annotations for the user input are also obtained from Rasp. For example, “u r a peach” has been regarded as “ppy (second-person) + vbr (present-tense copular form) + f1-l3 (semantic tag for food terms but also plants)”, while the input “Lisa is a pizza” is described as “np1 (singular proper noun) + vbz (present-tense copular form) + f1 (semantic tag for food terms)”.

TABLE 3: Food and animal metaphor training examples.

Sentence type	Subject	Quantifiers	Verb	Object	Category
s	PPHS1 (“she”)	null	VBZ (“is”)	F1-L3 (“peach”)	Food metaphor
s	NP1 (“Lisa”)	A13.2 (“completely”)	VBZ (“is”)	F1 (“pizza”)	Food metaphor
s	NP1 (“Lisa”)	null	VBZ (“is”)	L2 (“pig”)	Animal metaphor
s	PPY (“you”)	N3.2- (“little”)	VBR (“are”)	L2y (“bunny”)	Animal metaphor

TABLE 4: Anger metaphor training examples.

Sentence type	Subject	Quantifiers	Verb	Object	Conjunction	Category
s	PPHS1 (“she”)	A13.4 (“nearly”)	E3- (“exploded”)	null	null	Anger metaphor
s	PPHS1 (“she”)	A13.2 (“completely”)	O4.6+/a2.1 (“heated up”)	null	CS (“when”)	anger metaphor
s	NP1 (“Lisa”)	A13.7 (“hardly”)	O4.6+/a2.1 (“fires up”)	null	null	Anger metaphor

Also, in common sense, calling someone a baby animal name may indicate affection, while calling someone an adult animal name may convey insults. Thus, we have put animal names in our own newly created semantic dictionary with the semantic tag—“l2” to indicate living creatures generally for adult animal names [17] and the semantic tag—“l2y” to indicate young living creatures for baby animal names (e.g., “puppy”, “bunny”, “kitten” etc). According to the processing described above, for example, the user input “Lisa is a pig” is also converted into “np1 (singular proper noun) + vbz (present-tense copular form) + l2 (semantic tag for adult animal names). Thus, we collected 80 training examples for each of the two metaphorical language phenomena from the collected transcripts and converted them into the above identified semantic and syntactic structures to train the classifiers. Examples are listed in Table 3.

The classifiers are trained on these examples and deduce if any new user input belongs to these two metaphorical phenomena. Once an input has been recognized as an animal metaphor, then it carries affectionate if the semantic tag for the object indicates a young animal name. Otherwise the input conveys insulting if the object implies an adult animal name.

If the input is identified as a food metaphor, WordNet has been employed in order to get the synset of the food term. If among the synset, the food term has been explained as a certain type of human being, such as “beauty” and “sweetheart”. Then another small slang-semantic dictionary collected in our previous study containing terms for special person types (such as “freak”, “angle”) and their corresponding evaluation values (negative or positive) has been adopted in order to obtain the evaluation values of such synonyms of the food term. If the synonyms are positive (e.g., “beauty”), then we conclude that the input is affectionate with a food metaphor (e.g., “u r a peach” → deduced as a food metaphor → “peach” is sent to WordNet → synonyms of peach from WordNet: beauty and sweetheart → from slang-semantic profiles: beauty is positive → the input is affectionate).

3.4. Anger Metaphor Interpretation. There are several anger metaphors that have also been widely used, such as “anger is the heat of a fluid in a container” [12] (e.g., “she nearly exploded”, “he fired up straightaway”, and “she heated up just as fast”) and “anger is giving birth” [21] (e.g., “he had a baby when he heard what happened” and “don’t have a cow! It’s no big deal”). They all depict an attack of anger. As Kövecses [12] suggested, the above examples of the first anger metaphor imply the conceptualization of the “release of pressure”. The second anger metaphor indicates the similarity of the behaviour during labour to the behaviour expressed during an attack of anger. We are particularly interested in the metaphor “anger is the heat of a fluid in a container” in this study.

In order to train the classifiers, Rasp is used to detect user input with a structure of “a singular proper noun/second person/a third-person singular pronoun + past-tense or present-tense lexicon verb phrase” (e.g., “Lisa hits me”, “Peter needs the operation”, “she exploded completely”, and “Lisa likes school”) and provide part-of-speech tags for the user input. Then the adverb and the base form of the main verb (identified by their part-of-speech) are sent to the semantic dictionary. If the verb is among those verbs with strong affective implication (e.g., “fire”, “heat”, “explode”, “toast”, “steam” etc), then its semantic annotation is derived. Otherwise, its part-of-speech tag remains.

For example, Rasp converts the input “she nearly exploded” into “pphs1 (she) + rr (nearly) + vvd (exploded)”. With the assistance of the semantic dictionary, the input becomes “pphs11 (she) + a13.4 (semantic tag for quantifiers: nearly) + e3- (semantic tag for verbs indicating emotional states—Calm/Violent/Angry: explode)”. Examples such as “she heated up completely when she heard the news” are interpreted as “pphs1 (she) + a13.2 (semantic tag for quantifiers: completely) + o4.6+/a2.1 (semantic tag for verbs indicating Temperature/Affect: heat) + rp (up) + cs (when)”. The example training data of this language phenomenon are presented in Table 4.

Conjunctions are considered since our study shows that they tend to be used in such affective metaphorical expressions. The classifiers have also been trained by 80 examples of such a language phenomenon represented in the above structures in order to equip themselves with the ability to detect such expressions effectively and improve future performance.

When dealing with any new user input, Rasp informs the system about any input with the desired structure: “singular proper noun/second person/third-person singular + past-tense or present-tense lexicon verb phrase”. After the semantic annotations for the adverb and the main verb derived from the semantic dictionary, the derived semantic and syntactic representation of the input is sent to the classifiers to identify if it is an anger metaphorical expression or not. If it is, then the quantifier is employed to measure the affect intensity. For example, for the new input (not included in the training samples), “she fired up completely when she heard the news”, the processing detail is provided in the following:

- (1) Rasp recognizes the input with a syntactic structure: “pphs1 (she) + vvd (fired) + rp (up) + rr (completely) + cs (when)” and recognizes the input as a statement sentence;
- (2) since “completely” has a syntactic tag, “RR”, indicating it is a general adverb, it is sent to the semantic dictionary to recover its corresponding semantic tag, “a13.2”. In a similar way, the base form of the verb, “fire”, is sent to the semantic dictionary to derive its corresponding semantic tag, “o4.6+/a2.1”;
- (3) the user input has been interpreted as “pphs1 (she) + o4.6+/a2.1 (main verb: fire) + a13.2 (quantifier: completely) + cs (conjunction: when)”;
- (4) the above derived structure has been sent to the classifiers. It has been recognized as an anger metaphor (“anger is the heat of a fluid in a container”);
- (5) the quantifier, a13.2 (completely), indicates the input implying “anger” with a strong intensity.

In general, we have made some initial attempts on the interpretation and detection of several metaphorical language phenomena and the implied affect in such expressions. The overall component on affect and metaphor sensing is integrated with the conversational intelligent agent. The component is implemented in Java, integrated with Rasp, WordNet-affect, WordNet, and the APIs for the classifiers (decision trees, naïve Bayes, and support vector machine) embedded in Weka, which is a well-known data mining tool and incorporates many machine learning approaches together for various classification and clustering tasks.

The approaches we have taken provided some flexibility in the processing and recognition of metaphorical expressions. However, our processing could also be challenged by various variations of these several language phenomena (e.g., “I couldn’t bear to touch the memories”, “he felt his anger rising step by step”, “he was red with anger. I could see the smoke coming out of his ears”, “she was filled



FIGURE 1: Emotional animation.

with joy” etc), and there are also other figurative language phenomena such as irony, humor, and simile that we have not even touched. But our work points out a potential positive direction for affect and metaphor interpretation and detection for figurative language processing.

4. Emotional Animation

The detected affective states from users’ open-ended text input play an important role in producing emotional animation of human players’ avatars. The emotional animation mainly includes emotional gesture and social attention (such as eye gazing). The expressive animation engine, Demeanour (Gillies and Ballin [22]), makes it possible for our characters to express the affective states detected by EMMA. When EMMA detects an affective state in a user’s text input, this is passed to the Demeanour system attached to this user’s character and a suitable emotional animation is produced.

The Demeanour system has also used character profiles, particularly including personality traits and relationships with other characters, to provide expressive animation for other avatars when the “speaking” avatar experiences affect. For example, Figure 1 shows a screen shot of user interaction at the beginning of Crohn’s disease scenario. Briefly, in Crohn’s disease scenario, Peter has had Crohn’s disease since the age of 15. Crohn’s disease attacks the wall of the intestines and makes it very difficult to digest food properly. The character has the option to undergo surgery (ileostomy) which will have a major impact on his life. The task of the role-play is to discuss the pros and cons with friends and family and decide whether he should have the operation. The other characters are Mum, who wants Peter to have the operation, Matthew (younger brother), who is against the operation, Dad, who is not able to face the situation, and David (the best friend), who mediates the discussion.

In Figure 1, from left to right, the characters are Peter (with Crohn’s disease and needs to go through another life-changing operation), Janet (Peter’s mum who approves the idea of operation), Matthew (Peter’s younger brother who is against the operation idea since Peter could be bullied because of the side effects of the operation) and Dave (Peter’s best friend who tries to mediate the discussion). The character suffering from the disease (Peter) tends to



FIGURE 2: Emotional animation and the AI character.

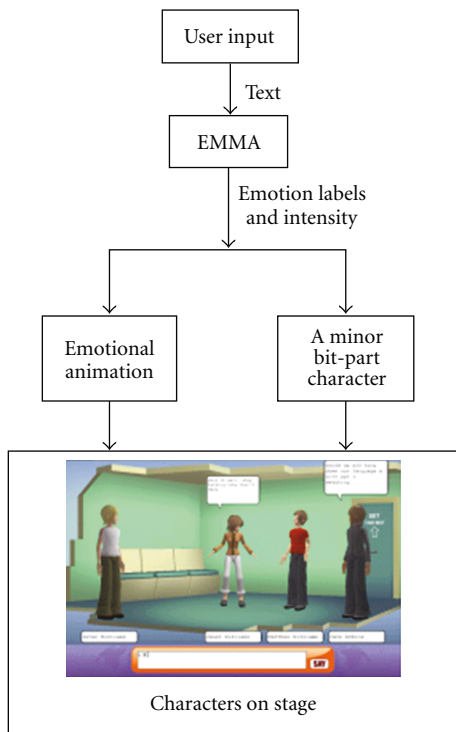


FIGURE 3: Affect detection and the control of characters.

feel uncomfortable and sad. Demeanour makes this type of personality trait expressed by a default, low-level emotional state (see Peter in Figure 1).

Figure 2 shows an interface of emotional interaction and animation with the same order of the characters as shown in Figure 1 in Crohn’s disease scenario. The mum character, Janet, expressed her anger towards Matthew (her other son, Peter’s younger brother) by saying “shut it matt, stop talking like that 2 Dave”. Since Matthew and Janet had a positive relationship (mother and son), Matthew showed a mild emotional response of acceptance of Janet’s suggestion by gesture. Dave, played by EMMA, had also provided a conversational response, “Could we all tone down our language a bit? ppl (people) r (are) watching...”, in order to mediate the discussion. All the characters shared

social attention as well by looking at the angry “speaking” character, Janet.

Figure 3 gives an overview of the control of the expressive characters. Users’ text input is analyzed by EMMA in order to detect affect in the text. The output is an emotion label with intensity derived from the text. This is then used in two ways. Firstly, it is used by the minor bit-part character (played by EMMA) to generate a response. Secondly, the label and the intensity are sent to the emotional animation system (via an XML stream) where it is used to generate animation.

As discussed earlier, the main intention of Dave’s (the AI agent) responses is to stimulate the improvisation. For example, Dave may tune down the argument, or he may also stir up the discussion by mentioning emotionally charged sensitive topics such as “Arnold, u (you) r (are) family to Peter and he needs ur (your) support”, “Arnold, Peter is ur (your) son and you can’t just ignore it”, when Dave detected that Arnold (the Dad character) was “embarrassed” to talk about Peter’s disease in public (“Peter we know about it. Stop talking about it”). Since in our previous pilot user testing, EMMA was commented by the testing subjects (14–16 years old school children) that its responses were different from theirs, we have also used abbreviations, acronyms and slang language in the construction of EMMA’s responses in order to simulate the language style used by school children. More discussion on user testing and EMMA’s performance on affect sensing and drama improvisation is provided in the following section.

5. Evaluation

We carried out user testing with 220 secondary school students from Birmingham schools and Education Village in Darlington for the improvisation of school bullying (SB) and Crohn’s disease scenarios. Briefly, the methodology of the testing is that we had each testing subject have an experience of both scenarios, one including the AI minor character only and the other including the human-controlled minor character only. Such arrangement could not only enable us to measure any statistically significant difference to users’ engagement and enjoyment due to the involvement of the AI minor character but also provide us the opportunity to compare the performance of the AI minor character with that of the human-controlled one. After the testing sessions, we obtained users’ feedback via questionnaires and group debriefings. Improvisational transcripts were automatically recorded during the testing so that it allows further evaluation of the performance of affect detection component.

Moreover, in order to identify the following several particular metaphors—*affect*, *food*, *animal*, and *anger*, we used 400 different language phenomena examples (80 samples for each metaphorical phenomenon and 80 for literal expressions) for the training of the classifiers. We have collected a small test set to evaluate the performance of the classifiers, with 50 examples for each category (most from the collected transcripts, which were produced by the testing subjects and automatically recorded during the testing, and

TABLE 5: Detailed recognition results for metaphorical and literal test expressions using decision tree learning.

Metaphorical and literal expressions	Results for decision tree learning		
	Precision	Recall	F-Measure
Affect metaphor	1	1	1
Food metaphor	0.929	1	0.963
Animal metaphor	1	0.813	0.897
Angry Metaphor	1	1	1
Literal expression	0.667	0.667	0.667

TABLE 6: Detailed recognition results for metaphorical and literal test expressions using naïve Bayes.

Metaphorical and literal expressions	Results for Naïve Bayes		
	Precision	Recall	F-Measure
Affect metaphor	1	1	1
Food metaphor	0.929	1	0.963
Animal metaphor	0.933	1	0.966
Angry Metaphor	1	0.917	0.957
Literal expression	0.667	0.667	0.667

a small portion from an online chatting transcripts database on travel (<http://akayoglu.s.web.ibu.edu.tr/webheads.htm>). All the chosen classifiers obtained reasonably good results for the metaphor sensing. Decision tree, naïve Bayes, and support vector machine achieved more than 90% of the accuracy rates for the recognition of these four types of metaphorical expressions. Since the training samples for the literal expressions are dramatically less than those for metaphorical expressions, the recognition results of such expressions are generally worse than those of the figurative phenomena, with the *F-Measure* 0.667 for both decision tree and naïve Bayes, and 0.333 for the support vector machine. Detailed evaluation results including *Precision*, *Recall*, and *F-Measure* obtained from Weka have been presented in Tables 5, 6, and 7 for the recognition of literal and metaphorical expressions using the three approaches, respectively. (The *True Positive (TP)* rate provided by Weka is the proportion of examples which were classified as class x , among all examples which truly have class x , that is, how much part of the class was captured. It is equivalent to *Recall*, while the *Precision* is the proportion of the examples which truly have class x among all those which were classified as class x . According to Weka, the *F-Measure* is produced in the following way: $2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$, that is, a combined measure for precision and recall. Overall, although there is room for further improvements, the evaluation results for the three approaches are generally promising.

We also noticed that some of the testing metaphorical examples collected from our recorded transcripts showed much resemblance to some of the training data although they have been produced by different testing subjects in different testing sessions. Therefore, we need to adopt a bigger size sample in order to evaluate the classifiers fully and choose the most effective approach for further development.

Also, we provided Cohen's Kappa in order to evaluate the efficiency of the affect detection processing for the detection

of 25 affective states. The following formula was used for the interagreement calculation: $\text{Kappa} = (\text{the number of actual agreed annotation} - \text{the number of agreed annotation by chance}) / (\text{the total number of annotation} - \text{the number of agreed annotation by chance})$.

As indicated in the formula, we have removed the effects of producing agreement between annotators purely by chance in interagreement calculation. Two human judges (not involved in development) have been employed to annotate part of the recorded transcripts of the SB scenario (72 turn-taking user input) filed from the testing. The interagreement between human judge A and B is 0.896. The agreements for human judge A/the AI agent and human judge B/the AI agent are, respectively, 0.662 and 0.729. Although improvement is needed, the AI agent's affect detection performance is acceptable and could achieve satisfactory level in good cases (e.g., in good cases, the interagreement between the human judge B and the AI agent is close to that between two human judges).

Inspection of the transcripts collected indicates that the AI agent usefully pushed the improvisation forward on various occasions. Box 1 shows an example about how the AI actor contributed to the drama improvisation in Crohn's disease scenario. In it, Dave was played by the AI actor, which successfully led the improvisation on the desirable track. In another scenario (school bullying) used for the testing, example transcripts have also shown that the AI actor has helped to push the improvisation forward.

Other evaluation results were also provided for the performance of the AI character. Generally the results indicated that the involvement of the AI character has not made any statistically significant difference to users' engagement and enjoyment with the emphasis of users' notice of the AI character's contribution throughout. Figure 4 also shows some evaluation results from a "within-subjects" analysis looking at the difference made PER SUBJECT by having

TABLE 7: Detailed recognition results for metaphorical and literal test expressions using support vector machine.

Metaphorical and literal expressions	Results for support vector machine		
	Precision	Recall	F-Measure
Affect metaphor	1	0.857	0.923
Food metaphor	1	1	1
Animal metaphor	0.833	0.938	0.882
Angry Metaphor	0.923	1	0.96
Literal expression	0.333	0.333	0.333

Peter Williams: I'M THE ILL ONE
 Dave Atkins: Arnold, Peter needs ur support and u can't just ignore it.
 Janet Williams: nobody cares peter
 Arnold Williams: we know peter now stop talking about it
 Peter Williams: yes I need your support
 Dave Atkins: Hey guys, lets all calm down a bit.
 Arnold Williams: we will discuss it later
 Janet Williams: well you have got time
 Peter Williams: help me daddy
 Arnold Williams: not now son

Box 1: Example transcripts; Crohn's disease scenario; Dave: the AI character. Part of an example transcript produced in the user testing.

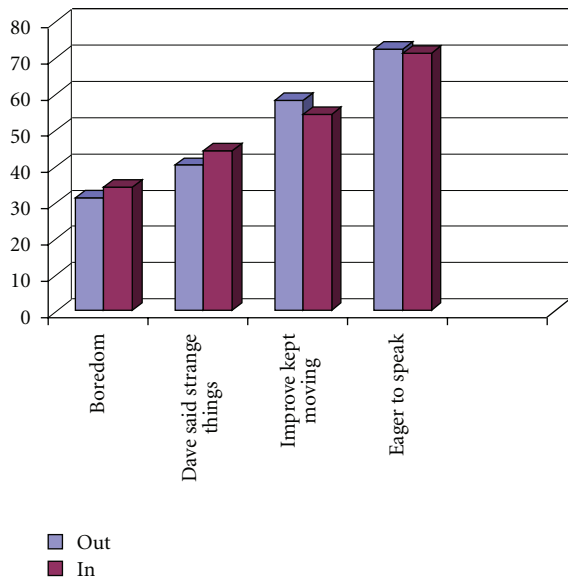


FIGURE 4: Statistical results for “boredom,” “Dave said strange things,” “improvisation kept moving,” and “eager to make own character speak” when EMMA is OUT of or IN an improvisation.

EMMA IN (= playing Dave, in either scenario) or OUT. When EMMA is out, the overall boredom is 31%. When EMMA is in, it changes to 34%. The results of “human Dave and EMMA Dave said strange things”, respectively are 40% and 44%. When EMMA changes from in to out of an improvisation, the results of “improvisation kept moving” are, respectively, 54% to 58% and the results of “the eagerness

to make own character speak” are, respectively, 71% to 72%. Although the measures were “worsened” by having EMMA in, in all cases the worsening was numerically fairly small and not statistically significant.

The preliminary results from statistical analysis also indicate that when the AI actor is involved in the improvisation, users’ abilities to concentrate on the improvisation are somewhat higher in Crohn’s disease scenario than school bullying scenario. When the AI actor is not involved in the improvisation, users’ abilities to concentrate on the improvisation are a lot higher in school bullying than Crohn’s disease. This seems very interesting, as it seems to be showing that the AI actor can make a real positive difference to an aspect of user engagement when the improvisation is comparatively uninteresting.

Moreover, as we mentioned earlier, the AI agent’s responses are mainly directed based on the detected affect from users’ input and at the beginning of the testing, we also concealed the fact that one character was computer controlled in order to get some fair results for the testing of the AI agent. In the debriefing sessions, it surprised us that no testing subject realized that sometimes one character was computer controlled. Generally, our statistical results gathered from the analysis of the questionnaires indicated that our AI agent performed as good as another 14–16-year-old school pupil. Analysis results also indicated that improvement is needed for negative affect detection (e.g., using context information). In our future development, we intend to employ context-based emotional modeling (e.g., using hidden Markov models) and psychological and linguistic contextual indicators to deduce affect conveyed in the input with the assistance of user profiles.

6. Conclusions

Metaphorical affective expressions have been employed to provide powerful vivid descriptions when literal expressions seem weak and unlikely to describe a feeling effectively. Such metaphorical expressions also challenge any natural language processing system if accurate semantic and sentiment interpretations are exploited. In our study, we have made a step towards automatic metaphor and affect sensing from several metaphorical figurative phenomena, including size, affect, food, animal, and anger metaphor. Although our system mainly focused on the interpretation of a few variations of the above metaphors, the study has been used as a test application and shows inspiration for theoretical metaphor studies and research. However, there is still a long way to go in order to successfully process the rich diverse variations of metaphorical language and other figurative expressions, such as humor, lies, and irony. Also, context information sometimes is very crucial for textual affect detection. These indicate that our strength needs to lie in the future development. We also intend to make the AI agent capable of recognizing and generating metaphor using metaphor ontologies to stimulate the improvisation and conduct autonomous learning of new concepts.

Overall, our work provides automatic improvisational agents for virtual drama improvisation situations. It makes a contribution to the issue of what types of automation should be included in human-robots interaction, and as part of that the issue of what types of affect should be detected and how. It also provides an opportunity for researchers to explore how emotional issues embedded in the scenarios, characters, and open-ended metaphorical expressions can be represented visually without detracting users from the learning situation. Finally, the automated conversational AI agent and the emotional animation may contribute to improving the perceived quality of social interaction.

We envisage that there is great potential for the use of our system in education in areas such as citizenship, PHSE, and drama. Beyond the classroom, our system can be easily customised for use in professional training, where face-to-face training can be difficult or expensive, such as customer services training and e-learning in the workplace.

References

- [1] L. Zhang, M. Gillies, K. Dhaliwal, A. Gower, D. Robertson, and B. Crabtree, "E-drama: facilitating online role-play using an AI actor and emotionally expressive characters," *International Journal of Artificial Intelligence in Education*, vol. 19, no. 1, pp. 5–38, 2009.
- [2] H. Liu and P. Singh, "ConceptNet: a practical commonsense reasoning toolkit," to appear in *BT Technology Journal*.
- [3] M. A. M. Shaikh, H. Prendinger, and I. Mitsuru, "Assessing sentiment of text by semantic dependency and contextual valence analysis," in *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction (ACII '07)*, vol. 4738 of *Lecture Notes in Computer Science*, pp. 191–202, Lisbon, Portugal, 2007.
- [4] M. Mateas, *Interactive drama, art and artificial intelligence*, Ph.D. thesis, School of Computer Science, Carnegie Mellon University, 2002.
- [5] X. Zhe and A. C. Boucouvalas, "Text-to-emotion engine for real time internet communication," in *Proceedings of International Symposium on Communication Systems, Networks and DSPs*, pp. 164–168, Staffordshire University, Stafford, UK, July 2002.
- [6] R. Craggs and M. Wood, "A two dimensional annotation scheme for emotion in dialogue," in *Proceedings of AAAI Spring Symposium: Exploring Attitude and Affect in Text*, 2004.
- [7] A. Egges, S. Kshirsagar, and N. Magnenat-Thalmann, "A model for personality and emotion simulation," in *Proceedings of Knowledge-Based Intelligent Information and Engineering Systems (KES '03)*, vol. 2773 of *Lecture Notes in Computer Science*, pp. 453–461, Springer, Berlin, Germany, 2003.
- [8] C. Elliott, J. Rickel, and J. Lester, "Integrating affective computing into animated tutoring agents," in *Proceedings of Workshop on Intelligent Interface Agents (IJCAI '97)*, pp. 113–121, 1997.
- [9] R. Aylett, S. Louchart, J. Dias et al., "Unscripted narrative for affectively driven characters," *IEEE Computer Graphics and Applications*, vol. 26, no. 3, pp. 42–52, 2006.
- [10] M. Cavazza, C. Smith, D. Charlton, L. Zhang, M. Turunen, and J. Hakulinen, "A 'companion' ECA with planning and activity modelling," in *Proceedings of the 7th International Conference on Autonomous Agents and Multi-Agent Systems*, pp. 1281–1284, Estoril, Portugal, 2008.
- [11] L. Fainsilber and A. Ortony, "Metaphorical uses of language in the expression of emotions," *Metaphor and Symbolic Activity*, vol. 2, no. 4, pp. 239–250, 1987.
- [12] Z. Kövecses, "Are there any emotion-specific metaphors?" in *Speaking of Emotions: Conceptualization and Expression*, A. Athanasiadou and E. Tabakowska, Eds., pp. 127–151, Mouton de Gruyter, New York, NY, USA, 1998.
- [13] S. Fussell and M. Moss, "Figurative language in descriptions of emotional states," in *Social and Cognitive Approaches to Interpersonal Communication*, S. R. Fussell and R. J. Kreuz, Eds., Lawrence Erlbaum, 1998.
- [14] J. Barnden, S. Glasbey, M. Lee, and A. Wallington, "Varieties and directions of interdomain influence in metaphor," *Metaphor and Symbol*, vol. 19, pp. 1–30, 2004.
- [15] J. A. Barnden, "Metaphor, semantic preferences and context-sensitivity," in *Words and Intelligence II: Essays in Honor of Yorick Wilks*, K. Ahmad, C. Brewster, and M. Stevenson, Eds., pp. 39–62, Springer, Dordrecht, The Netherlands, 2007.
- [16] A. Esuli and F. Sebastiani, "Determining term subjectivity and term orientation for opinion mining," in *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL '06)*, pp. 193–200, Trento, Italy, 2006.
- [17] P. Rayson, *Matrix: a statistical method and software tool for linguistic analysis through corpus comparison*, Ph.D. thesis, Lancaster University, 2003.
- [18] S. Sharoff, "How to handle lexical semantics in SFL: a corpus study of purposes for using size adjectives," in *Systemic Linguistics and Corpus*, Continuum, London, UK, 2005.
- [19] "ATT-Meta Databank: Examples of Usage of Metaphors of Mind," July 2008, <http://www.cs.bham.ac.uk/~jab/ATT-Meta/Databank/>.
- [20] C. Strapparava and A. Valitutti, "WordNet-affect: an affective extension of WordNet," in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC '04)*, pp. 1083–1086, Lisbon, Portugal, 2004.

- [21] D. Glynn, "Love and anger: the grammatical structure of conceptual metaphors," *Style*, vol. 36, no. 3, pp. 541–574, 2002.
- [22] M. Gillies and D. Ballin, "Integrating autonomous behavior and user control for believable agents," in *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '04)*, pp. 336–343, Columbia University, New York, NY, USA, July 2004.