Computational Intelligence, Volume 59, Number 000, 2008

Crowdsourcing the Creation of a Word–Emotion Association Lexicon

Saif M. Mohammad and Peter D. Turney
Institute for Information Technology, National Research Council Canada.
Ottawa, Ontario, Canada, K1A 0R6
{saif.mohammad,peter.turney}@nrc-cnrc.gc.ca

Even though considerable attention has been given to semantic orientation of words and the creation of large polarity lexicons, research in emotion analysis has had to rely on limited and small emotion lexicons. In this paper we show how the combined strength and wisdom of the crowds can be used to generate a large, high-quality, word–emotion association lexicon quickly and inexpensively. We flesh out various challenges in emotion annotation in a crowdsourcing scenario and propose solutions to address them. Most notably, in addition to questions about emotions associated with terms, we show how the inclusion of a word choice question can discourage malicious data entry, help identify instances where the annotator may not be familiar with the target term (allowing us to reject such annotations), and help obtain annotations at sense level (rather than at word level). We perform an extensive analysis of the annotations to better understand the distribution of emotions evoked by terms of different parts of speech. We identify which emotions tend to be evoked simultaneously by the same term and show that certain emotions indeed go hand in hand. We also analyze the polarity of terms (positive and negative), as well as what colors are associated with words. We find that associations with colors is directly correlated with the order that colors terms first came into existence in language. Also, red and black are strongly associated with negative emotion terms, whereas white and green are strongly associated with positive terms. The lexicon with close to 10,000 entries (one entry for each word–sense pair) will be made freely available.

Key words: Emotions, semantic orientation, crowdsourcing, Mechanical Turk, lexicon, colors, word–emotion associations, word–color associations, sentiment analysis.

1. INTRODUCTION

Emotions are common and pervasive phenomena that have been studied by many, including psychologists, linguists, biologists, and neuro-biologists. Yet, it is easier to talk about emotions through examples—such as love, hate, joy, anger, fear, lust, and envy—than through a definition. Merriam Webster defines emotions as: *A mental reaction subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body.* Often, different emotions are expressed through different words. For example, *delightful* and *yummy* indicate the emotion of joy, *gloomy* and *cry* are indicative of sadness, *shout* and *boiling* are indicative of anger, and so on. In this paper, we describe an annotation project aimed at creating a large lexicon of such word–emotion associations.

Words may evoke different emotions in different contexts, and the emotion evoked by a phrase or a sentence is not simply the sum of emotions conveyed by the words in it, but the emotion lexicon will be a useful component for any sophisticated emotion detecting algorithm. When analyzing natural language text, or speech, automatically detecting the relevant entity's emotions such as joy, sadness, fear, anger, and surprise is useful for a number of purposes, including:

(1) Creating dialogue systems that respond appropriately to different emotional states of the user, for example, in customer relation models, intelligent tutoring systems, and emotion-aware games.
(2) Tracking sentiment towards politicians, movies, products.
(3) Determining emotional intelligence by analyzing documents written by a person.

(a) Guest editors: Diana Inkpen and Carlo Strapparava. (b) Name of conference or workshop of the initial publication: NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, June 5, 2010, Los Angeles, CA. (c) The title of the special issue: Computational Approaches to Analysis of Emotion in Text.

(4) Assisting in writing e-mails, documents, and other text to convey desired emotion (and avoiding misinterpretation).
(5) Depicting flow of emotions in novels and other books.
(6) Identifying what emotion a newspaper headline is trying to evoke.
(7) Re-ranking and categorizing information/answers in a web 2.0 world. For example, highly emotional responses may be ranked lower.
(8) Detecting how people use emotion-bearing-words and metaphors to persuade and coerce others.
(9) Developing more natural text-to-speech systems.
(10) Developing humanoid robots. For example, the robotics group in Carnegie Melon University is interested in building an emotion-aware physiotherapy coach robot.

The term–emotion association lexicon will also be useful for evaluating automatic methods that identify the emotions evoked by a word. Such algorithms may then be used to automatically generate emotion lexicons in languages where no such lexicons exist. As of now, high-quality high-coverage emotion lexicons do not exist for any language, although there are a few limited-coverage lexicons for a handful of languages, for example, the WordNet Affect Lexicon (WAL) (Strapparava and Valitutti, 2004) for six basic emotions and the General Inquirer (GI) (Stone *et al.*, 1966), which categorizes words into a number of categories, including positive and negative semantic orientation. The lack of emotion resources can be attributed to high cost and considerable manual effort required of the human annotators in a traditional hire-a-few-experts setting.

Crowdsourcing, as defined by Wikipedia, is *"the act of outsourcing tasks, traditionally performed by an employee or contractor, to a large group of people or community (a crowd), through an open call."* Some of the more well known crowdsourcing projects include Wikipedia, Netflix Prize, and Amazon's Mechanical Turk. Mechanical Turk is an online service that can be used to obtain a large amount of human input or annotation in an efficient and inexpensive manner (Snow *et al.*, 2008; Callison-Burch, 2009).[1] However, one must define the task carefully to obtain annotations of high quality. Several checks must be placed to ensure that random and erroneous annotations are discouraged, rejected, and re-annotated.

In this paper, we show how we compiled a large English emotion lexicon by manual annotation through Amazon's Mechanical Turk service. This dataset, which we will call **EmoLex**, is an order of magnitude larger than the only other known emotion lexicon, WordNet Affect Lexicon. We focus on the emotions of joy, sadness, anger, fear, trust, disgust, surprise, and anticipation—argued by many to be the basic and prototypical emotions (Plutchik, 1980). Complex emotions can be viewed as combinations of these basic emotions. The terms in EmoLex are carefully chosen to include some of the most frequent nouns, verbs, adjectives, and adverbs. Beyond unigrams, it has many commonly used bigrams as well. We also include words from the General Inquirer and the WordNet Affect Lexicon to allow comparison of annotations between the various resources.

We perform extensive analysis of the annotations to answer several questions, such as, how hard is it for humans to annotate words with the emotions they evoke? What percentage of commonly used terms, in each part of speech, evoke an emotion? Are emotions more commonly evoked by nouns, verbs, adjectives, or adverbs? How much do people agree with each other on emotions associated with words? Is there a correlation between the semantic orientation of a word and the emotion it evokes? Which emotions tend to go together; that is, which emotions are evoked simultaneously by the same term? What colors are associated with different emotions? What percentage of words have strong color associations? What colors are most commonly associated with words? And so on. Our lexicon now has more than 10,000 terms and work is on to make it even larger (about 40,000 terms).

2. RELATED WORK

WordNet Affect Lexicon Strapparava and Valitutti (2004) has a few hundred words annotated with the emotions they evoke.[2] It was created by manually identifying the emotions of a few seed words and then marking all their WordNet synonyms as having the same emotion. The General

[1] https://www.mturk.com/mturk/welcome
[2] http://wndomains.fbk.eu/wnaffect.html

Figure 1. An illustration (Figure 15) from Charles Darwin's The Expression of the Emotions in Man and Animals. It is captioned "FIG. 15. Cat terrified at a dog. From life, by Mr. Wood." (The image file is taken from Wikipedia Commons—a database of freely usable media files.)

Inquirer (Stone *et al.*, 1966) has 11,788 words labeled with 182 categories of word tags, including positive and negative semantic orientation.[3] It also has certain other affect categories, such as pleasure, arousal, feeling, and pain but these have not been exploited to a significant degree by the natural language processing community.

Work in emotion detection can be roughly classified into that which looks for specific emotion denoting words (Elliott, 1992), that which determines tendency of terms to co-occur with seed words whose emotions are known (Read, 2004), that which uses hand-coded rules (Neviarouskaya *et al.*, 2009), and that which uses machine learning and a number of emotion features, including emotion denoting words (Alm *et al.*, 2005). Recent work by Bellegarda (2010) uses sophisticated dimension reduction techniques to automatically identify emotion terms and obtains marked improvements in classifying newspaper headlines into different emotion categories. Goyal *et al.* (2010) move away from classifying sentences as per the writer's perspective, towards attributing emotions to entities mentioned in the text. Their work deals with mental states such as positive and negative, but work on attributing emotions to entities mentioned in text is, similarly, a promising area of future work.

Much of this recent work focuses on six emotions studied by Ekman (1992) and Sautera *et al.* (2010). These emotions—joy, sadness, anger, fear, disgust, and surprise—are a subset of the eight proposed in Plutchik (1980). We focus on the Plutchik emotions because the emotions can be naturally paired into opposites—joy–sadness, anger–fear, trust–disgust, and anticipation–surprise. Natural symmetry apart, we believe that prior work on automatically computing word–pair antonymy (Lin *et al.*, 2003; Mohammad *et al.*, 2008; Lobanova *et al.*, 2010) can now be leveraged in automatic emotion detection. There is less work on non-basic emotions, for example, work by Pearl and Steyvers (2010) that focusses on politeness, rudeness, embarrassment, formality, persuasion, deception, confidence, and disbelief. They developed a game-based annotation project for these emotions.

3. EMOTIONS

Emotions are fairly pervasive and consistent in nature. It is well documented that even across cultures that have no contact with each other, facial expressions for basic human emotions are identical (Ekman and Friesen, 2003; Ekman, 2005; Sautera *et al.*, 2010). There is some contention on whether animal have emotions, but there are studies, especially for higher mammals, canines, felines, and even some fish, arguing in favor of the proposition (Masson, 1996; Guo *et al.*, 2007). (Some of the earliest work is by Charles Darwin in his book *The Expressions of the Emotions in Man and Animals* (Darwin, 1872). Figure 1 is an illustration from the book.) Studies by evolutionary biologists and psychologists show that emotions have evolved to improve the "reproductive fitness"

[3]http://www.wjh.harvard.edu/∼inquirer

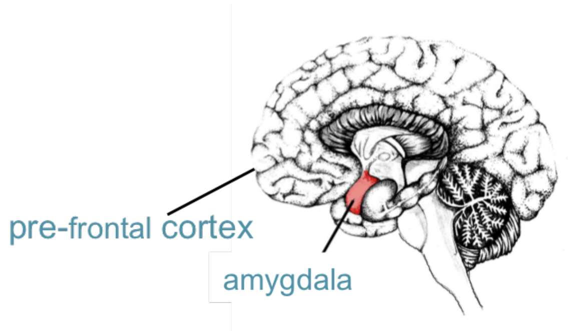

FIGURE 2. Illustration of the human brain showing the amygdala (associated with instinctual emotions) and the pre-frontal cortex (associated with cognitive emotions).

for a species, as they are triggers for behavior with high survival value. For example, fear inspires fight-or-flight response. However, more complex brains in primates and indeed humans are capable of experiencing not just the basic emotions such as fear and joy, but also more complex and nuanced emotions such as optimism and shame.

In this study, we are interested in human emotions associated with words. Psychologists have come up with a number of theories that classify the hundreds of emotions that humans can perceive into taxonomies. Emotions can be categorized into those that are **instinctual**, that is, emotions we sense and perceive, and those that are **cognitive**, that is, emotions we arrive at after some thinking and reasoning. Instinctual emotions have been associated with the amygdala and are found even in many of the species with a very primitive brain. Cognitive emotions are associated with the pre-frontal cortex and found in species with more developed brains. (Figure 2 shows the amygdala and pre-frontal cortex in a human brain.) Emotions can also be classified into basic (or pure) emotions and complex emotions that are mixtures or combinations of the basic emotions. There is a high correlation between the basic and instinctual emotions, as well as between complex and cognitive emotions. (Many of the basic emotions are also instinctual.) Emotions can also be categorized by duration into those that last for only a few seconds, for example surprise, and those that last for longer periods of time, for example love.

Since annotating for hundreds of emotions is expensive and also hard for annotators, we annotate the words with only the basic and instinctual emotions. Again a number of differing theories have been proposed on which emotions are basic. Ekman (1992) argues that there are six basic emotions: joy, sadness, anger, fear, disgust, and surprise. Plutchik (1962, 1980, 1994) proposes a theory with eight basic emotions. These include Ekman's six as well as trust and anticipation. Plutchik's 8 have a better balance of positive and negative emotions and they can be paired into opposites: joy–sadness, anger–fear, trust–disgust, and anticipation–surprise. Plutchik organizes the emotions in a wheel shown in Figure 3. Observe that similar emotions are placed next to each other, whereas opposite emotions are placed diametrically opposite to each other. The radius indicates intensity—the closer to the center, the higher is the intensity. The white spaces between the basic emotions are referred to as **primary dyads** and show complex emotions that are combinations of the two adjacent basic emotions. For example, optimism is a combination of joy and anticipation. **Secondary dyads** are emotions formed by combinations of basic emotions that are one hop away and **tertiary dyads** are emotions formed by combinations of basic emotions two hops away. These are shown in Figure 4.

We chose the Plutchik model of eight basic emotions for our annotations. However, it should be noted that there are many other models of basic emotions besides it and the one proposed by Ekman, such as that of Parrot (2001) and James (1884). See Ortony and Turner (1990) for a detailed review of many of these models.

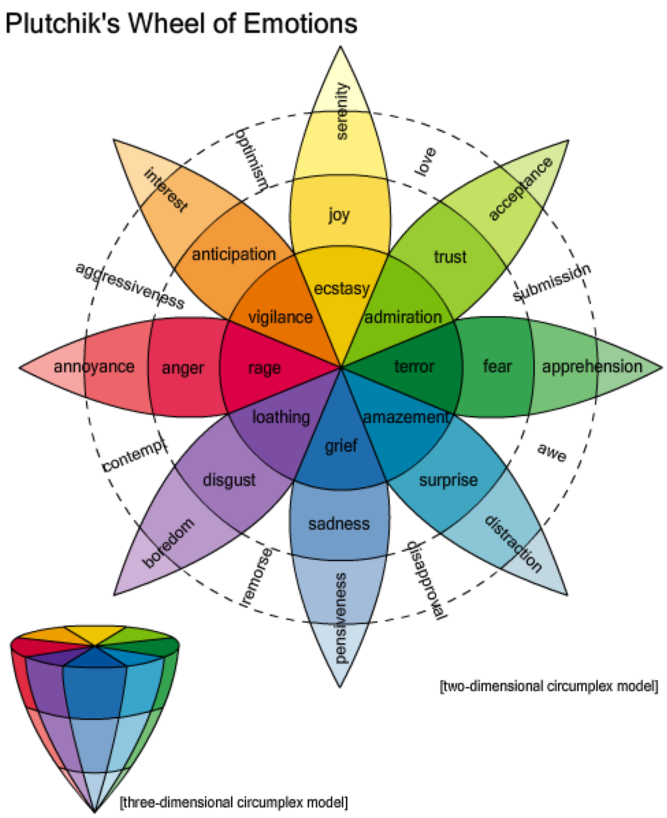

Figure 3. Plutchik's wheel of emotions. Similar emotions are placed next to each other. Contrasting emotions are placed diametrically opposite to each other. Radius indicates intensity. White spaces in between the basic emotions represent primary dyads—complex emotions that are combinations of adjacent basic emotions. (The image file is taken from Wikimedia Commons.)

4. TARGET TERMS

In order to generate a word–emotion association lexicon, we first identify a list of words and phrases for which we want human annotations. We chose the *Macquarie Thesaurus* as our source pool for unigrams and bigrams. The categories in the thesaurus act as coarse senses of the words. (A word listed in two categories is taken to have two senses.) Any other published dictionary would have worked well too. However, apart from over 57,000 commonly used English word types, the *Macquarie Thesaurus* also has entries for more than 40,000 commonly used phrases. From this list we chose those terms that occurred frequently in the Google n-gram corpus (Brants and Franz, 2006). Specifically we chose the 200 most frequent unigrams and 200 most frequent bigrams from four parts of speech: nouns, verbs, adverbs, and adjectives. When selecting these sets, we ignored terms that occurred in more than one *Macquarie Thesaurus* category. (There were only 187 adverb bigrams that matched these criteria. All other sets had 200 terms each.) We chose all words from each of the six emotion categories in the WordNet Affect Lexicon that had at most two senses (terms listed in at most two thesaurus categories)—640 word–sense pairs in all. We included all terms in the General Inquirer that were not too ambiguous (had at most 3 senses)—8132 word–sense pairs in all.[4] Some of these terms occur in more than one set. The union of the three sets (Google n-gram terms, WAL terms, and GI terms) has 10,170 term–sense pairs.

[4]We started the annotation on monosemous terms, and gradually included more and more ambiguous terms as we became more and more confident that the quality of annotations we were getting was good.

FIGURE 4. Robert Plutchik's set of primary dyads, secondary dyads, tertiary dyads, and opposites. Primary Dyads are formed by combinations of adjacent basic emotions. Secondary Dyads are formed by combinations of basic emotions that are one hop away on the emotion wheel. Tertiary Dyads are formed by combinations of basic emotions that are two hops away on the emotion wheel.

5. MECHANICAL TURK

We used Amazon's Mechanical Turk service as a platform to obtain large-scale emotion annotations. An entity submitting a task to Mechanical Turk is called the **requester**. The requester breaks the task into small independently solvable units called **HITs (Human Intelligence Tasks)** and uploads them on the Mechanical Turk website. The requester specifies (1) some key words relevant to the task to help interested people find the HITs on Amazon's website, (2) the compensation that will be paid for solving each HIT, and (3) the number of different annotators that are to solve each HIT. The people who provide responses to these HITs are called **Turkers**. Turkers usually search for tasks by entering key words representative of the tasks they are interested in and often also by specifying the minimum compensation per HIT they are willing to work for. The annotation provided by a Turker for a HIT is called an **assignment**.

We created Mechanical Turk HITs for each of the terms specified in Section 4. Each HIT has a set of questions, all of which are to be answered by the same person. (A complete example HIT with directions and all questions is shown in Section 7 ahead.) We requested five different assignments for each HIT (each HIT is to be annotated by five different Turkers). Different HITS may be attempted by different Turkers, and a Turker may attempt as many HITs as they wish.

6. CHALLENGES

In the subsections below, we describe some of the main challenges in using a crowdsourcing platform such as Mechanical Turk, and challenges associated specifically with emotion annotation.

6.1. Key challenges in using Mechanical Turk

Even though there are a number of benefits to crowdsourcing, such as low cost, less organizational overhead, and quick turn around time, there are also some inherent challenges. First and foremost is quality control. The task and compensation may attract cheaters (who may input garbage information) and even malicious annotators (who may deliberately enter incorrect information). We have no control over the educational background of a Turker, and we cannot expect the average Turker to read and follow complex and detailed directions, that one can some times get away with in a more traditional annotation setting. However, this may not necessarily be a disadvantage of crowdsourcing. We believe that good annotations often follow clear, brief, and simple instructions. Another challenge in crowdsourcing is finding enough Turkers interested in doing the task. If the task does not require any special skills, then more Turkers will do the task. The number of Turkers and the number of annotations they provide is also dependent on how interesting they find the task and how high the compensation is. A high compensation is not always the best approach though, since it attracts cheaters. With the right amount of compensation one can attract Turkers who are doing the task not just for the money but also because they find the task interesting.

6.2. Key challenges in emotion annotation

Native and fluent speakers of a language are good at identifying emotion associations of words. So for our task we do not require the annotators to have any special skills other than that they be native or fluent speakers of English. However, emotion annotation, especially in a crowdsource setting, has some important challenges.

Words used in different senses can evoke different emotions. For example, the word *shout* evokes a different emotion when used in the context of admonishment, than when used in "*Give me a shout if you need any help.*" Getting human annotations on word senses is made complicated by decisions about which sense-inventory to use and what level of granularity the senses must have. On the one hand, we do not want to choose a fine-grained sense-inventory because then the number of word–sense combinations will become too large and difficult to easily distinguish, and on the other hand we do not want to work only at the word level because when used in different senses a word may often evoke different emotions.

Yet another challenge is how best to convey a word sense to the annotator. Long definitions will take time to read and limit the number of annotations we can obtain for the same amount of resources. Further, we want the users to annotate a word only if they are already familiar with it and know its meanings. Definitions are good at conveying the core meaning of a word but not so effective in conveying the subtle emotional connotations. Therefore, we wanted to discourage a Turker from annotating for a word they are not familiar with. And lastly, we must ensure that malicious and erroneous annotations are discarded.

7. OUR APPROACH

In order to overcome the challenges described above, before asking the annotators questions about what emotions are associated with a target term, we first present them with a word choice problem. They are provided with four different words and asked which word is closest in meaning to the target. This single question serves many purposes. Through this question we convey the word sense for which annotations are to be provided, without actually providing annotators with long definitions. Further, if an annotator is not familiar with the target word and still attempts to answer questions pertaining to the target, or is randomly clicking options in our questionnaire, then there is a 75% chance that they will get the answer to this question wrong, and we can discard all responses pertaining to this target term by the annotator (that is, we also discard answers to the emotion questions provided by the annotator for this target term).

We generated these word choice problems automatically using the *Macquarie Thesaurus* (Bernard, 1986). As mentioned before, published thesauri, such as *Roget's* and *Macquarie*, divide the vocabulary into about a thousand categories, which may be interpreted as coarse senses. Each category also has a head word that best captures the meaning of the category. The word choice question for a target term is automatically generated by selecting the following four alternatives (choices): the head word of the thesaurus category pertaining to the target term (the correct answer); and three other head words of randomly selected categories (the distractors). The four alternatives are presented to the annotator in random order. We generated a separate HIT (and a separate word choice question) for every sense of the target.

We created Mechanical Turk HITs for each of the terms (n-gram–sense pairs) specified in Section 4 earlier. The first and second column of Table 1 list the various sets of target terms as well as the number of terms in each set for which annotations were requested. $\mathbf{EmoLex_{Uni}}$ stands for all the unigram–sense pairs taken from the thesaurus. **EmoLex$_{\mathbf{Bi}}$** refers to all the bigram–sense pairs. **EmoLex$_{\mathbf{GI}}$** are all the word–sense pairs taken from the General Inquirer. **EmoLex$_{\mathbf{WAL}}$** are all the word–sense pairs taken from the WordNet Affect Lexicon.

Each HIT has a set of questions, all of which are to be answered by the same person. As mentioned before, we requested five different assignments for each HIT. Below is a complete example HIT for the target word *startle*. All questions were multiple choice questions, and the Turkers could select exactly one option for each question.

Title: Emotions associated with words
Keywords: emotion, English, sentiment, word association, word meaning
Reward per HIT: $0.04

Directions:

(1) This survey will be used to better understand emotions. Your input is much appreciated.
(2) If any of the questions in a HIT are unanswered, then the assignment is no longer useful to us and we will be unable to pay for the assignment.
(3) Please return/skip HIT if you do not know the meaning of the word.
(4) Attempt HITS only if you are a native speaker of English, or very fluent in English.
(5) Certain "check questions" will be used to make sure the annotation is responsible and reasonable. Assignments that fail these tests will be rejected. If an annotator fails too many of these check questions, then it will be assumed that the annotator is not following instructions 3 and/or 4 above, and ALL of the annotator's assignments will be rejected.
(6) We will approve HITs about once a week. Expected date all the assignments will be approved: April 14, 2010.
(7) Confidentiality notice: Your responses are confidential. Any publications based on these responses will not include your specific responses, but rather aggregate information from many individuals. We will not ask any information that can be used to identify who you are.
(8) Word meanings: Some words have more than one meaning, and the different meanings may be associated with different emotions. For each HIT, Question 1 will guide you to the intended meaning. You may encounter multiple HITs for the same target term, but they will correspond to different meanings of the target word, and they will have different guiding questions.
(9) Note that the order of options in Q13 changes with every HIT.

Prompt word: **startle**

1. Which word is closest in meaning (most related) to *startle*?

- *automobile*
- *shake*
- *honesty*
- *entertain*

2. How positive (good, praising) is the word startle?

- *startle* is not positive
- *startle* is weakly positive
- *startle* is moderately positive
- *startle* is strongly positive

3. How negative (bad, criticizing) is the word startle?

- *startle* is not negative
- *startle* is weakly negative
- *startle* is moderately negative
- *startle* is strongly negative

4. How much is *startle* associated with the emotion joy? (For example, *happy* and *fun* are strongly associated with joy.)

- *startle* is not associated with joy
- *startle* is weakly associated with joy
- *startle* is moderately associated with joy
- *startle* is strongly associated with joy

5. How much is *startle* associated with the emotion sadness? (For example, *failure* and *heart-break* are strongly associated with sadness.)

- *startle* is not associated with sadness
- *startle* is weakly associated with sadness
- *startle* is moderately associated with sadness
- *startle* is strongly associated with sadness

6. How much is *startle* associated with the emotion fear? (For example, *horror* and *scary* are strongly associated with fear.)

- Similar choices as in 4 and 5 above

7. How much is *startle* associated with the emotion anger? (For example, *rage* and *shouting* are strongly associated with anger.)

- Similar choices as in 4 and 5 above

8. How much is *startle* associated with the emotion trust? (For example, *faith* and *integrity* are strongly associated with trust.)

- Similar choices as in 4 and 5 above

9. How much is *startle* associated with the emotion disgust? (For example, *gross* and *cruelty* are strongly associated with disgust.)

- Similar choices as in 4 and 5 above

10. How much is *startle* associated with the emotion surprise? (For example, *startle* and *sudden* are strongly associated with surprise.)

- Similar choices as in 4 and 5 above

11. How much is *startle* associated with the emotion anticipation? (For example, *expect* and *eager* are strongly associated with anticipation.)

- Similar choices as in 4 and 5 above

12. Is *startle* an emotion? (For example: *love* is an emotion; *shark* is associated with fear (an emotion), but *shark* is not an emotion.)

- No, *startle* is not an emotion
- Yes, *startle* is an emotion

13. What color is associated with *startle*?

- white
- black
- red
- green
- yellow
- blue
- brown
- pink
- purple
- orange
- grey

For each term, the color options in question 13 were presented in random order. Before going live, the survey was approved by the ethics committee at the National Research Council Canada.

The phrasing of questions in any survey can have a huge impact on the results. With our questions we hoped to be clear and brief, so that different annotators do not misinterpret what was being asked of them. However, before arriving at the above way of phrasing the questions, we performed a couple of pilot experiments on a smaller set of 2100 terms. One, where we asked questions exactly as described above, and another independent set of annotations where questions 5 through 11 were phrased slightly differently. Instead of asking if a word is *associated* with a certain emotion, we asked if a word *evokes* a certain emotion. We found that the annotators agreed with each other much more in the *associated* case than in the *evokes* case. (Details in Section 9.3 ahead.) So all subsequent annotations were done with *associated*. All results, except those presented in Section 9.3, are for the *associated* annotations.

8. ANNOTATION SATISTICS AND POST-PROCESSING

We conducted annotations over two batches, starting first with a pilot set of about 2100 terms, which was annotated in about a week. The last batch of about 8000 terms (HITs) was annotated in about two weeks. Notice that the amount of time take is not linearly proportional to the number of HITs. As one builds a history of tasks and payment, more and more Turkers do subsequent tasks. Also, if there are a large number of HITs, then more people find it worth the effort to understand and become comfortable at doing the task. Each HIT had a compensation of \$0.04 (4 cents) and the Turkers spent about a minute on average to answer the questions in a HIT. This resulted in an hourly pay of about \$2.4.

Once the assignments were collected, we used automatic scripts to validate the annotations. Some assignments were discarded because they failed certain tests (described below). A subset of the discarded assignments were officially **rejected** (the Turkers were not paid for these assignments) because instructions were not followed. About 2,666 of the 50,850 (1,0170 $\times$ 5) assignments included at least one unanswered question. These assignments were discarded and rejected. Even though distractors for Q1 were chosen at random, every now and then a distractor may come too close to the meaning of the target term, resulting in a bad word choice question. For 1045 terms, 3 or more annotators gave an answer different than the one automatically taken from the thesaurus. These questions were marked as bad questions and shelved. All corresponding assignments (5,225 in total) were discarded. Turkers were paid in full for these assignments regardless of their answer to Q1.

More than 95% of the remaining assignments had the correct answer for the word choice question This was a welcome result showing that, largely, the annotations were done in a responsible manner. We discarded all assignments that had the wrong answer for the word choice question. If an annotator obtained an overall score that is less than 66.67% on the word choice questions (that is, got more than one out of three wrong), then we assumed that, contrary to instructions, HITs for words not familiar to the annotator were attempted. We discarded and rejected *all* assignments by such annotators (not just the assignments for which they got the word choice question wrong).

For many of the terms, the emotion questions (Q5 through Q11) may not have a "right answer". (Even though many may have a clear wrong answer. For example, *morose* is definitely not associated with joy.) Therefore, for each of the annotators, we calculated the maximum likelihood probability with which the annotator agrees with the majority on the emotion questions. We calculated the mean of these probabilities and the standard deviation. Consistent with standard practices in identifying outliers, we discarded annotations by Turkers who were more than two standard deviations away from the mean (annotations by 111 Turkers).

Finally, there remained three or more valid assignments for 8,883 of the 10,170 target terms. We will refer to this set of assignments as the **master set**. We created the word–emotion association lexicon from this master set containing 38,726 assignments from about 2,216 Turkers who attempted 1 to 2,000 assignments each. About 300 of them provided 20 or more assignments each (more than 33,000 assignments in all). The master set has, on average, about 4.4 assignments for each of the 8,883 target terms. (See Table 1 for more details.) The total cost of the annotation was about US\$2,100. This includes fees that Amazon charges (about 13% of the amount paid to the Turkers) as well as the cost for the dual annotation of the pilot set with both *evokes* and *associated*. [5]

9. ANALYSIS OF EMOTION ANNOTATIONS

The different emotion annotations for a target term were consolidated by determining the **majority class** of emotion intensities. For a given term–emotion pair, the majority class is that intensity level that is chosen most often by the Turkers to represent the degree of emotion evoked by the word. Ties are broken by choosing the stronger intensity level. Table 2 lists the percent of 8,883 target terms assigned a majority class of no, weak, moderate, and strong emotion. For example, it tells us that 5% of the target terms strongly evoke joy. The table also presents an average of the numbers in each column (micro average). The last row lists the percent of target terms that evoke

[5] HITs of discarded assignments will be uploaded on Mechanical Turk for a second round of annotations.

Table 1. Break down of target terms into various categories. Initial refers to terms chosen for annotation. Master refers to terms for which three or more valid assignments were obtained using Mechanical Turk.

EmoLex	**# of terms Initial**	**Master**	**Annotations per word**
EmoLex$_{\mathbf{Uni}}$:			
adjectives	200	190	4.4
adverbs	200	187	4.5
nouns	200	178	4.5
verbs	200	195	4.4
EmoLex$_{\mathbf{Bi}}$:			
adjectives	200	162	4.4
adverbs	187	171	4.3
nouns	200	185	4.5
verbs	200	178	4.4
EmoLex$_{\mathbf{GI}}$:			
negatives in GI	2119	1837	4.4
neutrals in GI	4226	3653	4.4
positives in GI	1787	1541	4.4
EmoLex$_{\mathbf{WAL}}$:			
anger terms in WAL	165	160	4.5
disgust terms in WAL	37	34	4.4
fear terms in WAL	100	89	4.4
joy terms in WAL	165	149	4.5
sadness terms in WAL	120	112	4.5
surprise terms in WAL	53	51	4.4
Union	**10170**	**8883**	**4.45**

Table 2. Percent of terms assigned a majority class of no, weak, moderate, and strong emotion.

Emotion	Intensity no	weak	moderate	strong
anger	81.6	8.5	5.1	4.5
anticipation	84.2	8.9	4.2	2.4
disgust	84.6	8.3	3.8	3.1
fear	79.6	10.3	5.6	4.3
joy	79.5	8.9	6.4	5.0
sadness	80.9	10.0	4.8	4.2
surprise	89.5	6.6	2.2	1.4
trust	81.9	7.9	5.9	4.1
micro average	**82.7**	**8.7**	**4.8**	**3.6**
any emotion	**35.6**	**21.2**	**20.5**	**22.5**

some emotion (any of the eight) at the various intensity levels. We calculated this using the intensity level of the strongest emotion expressed by each target. Observe that 22.5% of the target terms strongly evoke at least one of the eight basic emotions.

Even though we asked Turkers to annotate emotions at four levels of intensity, practical NLP applications often require only two levels—associated with a particular emotion (**emotion emotive**) or not (**emotion non-emotive**). For each target term–emotion pair, we convert the four-level annotations into two-level annotations by placing all no- and weak-intensity assignments in the non-emotive bin, all moderate- and strong-intensity assignments in the emotive bin, and then choosing the bin with the majority assignments. Table 3 shows how many terms overall, and within each category, are emotive of the different emotions.

Analysis of question 12 revealed that 9.3% of the 8,883 target terms (826 terms) were considered not just associated with emotions, but represented emotions themselves.

TABLE 3. Percent of terms, in each target set, that are emotive. Highest individual emotion scores for EmoLex$_{\text{WAL}}$ are shown bold. Observe that WAL fear terms are marked most as fear emotive, joy terms as joy emotive, and so on.

	anger	anticipn.	disgust	fear	joy	sadness	surprise	trust	any
EmoLex	13	12	10	14	16	12	6	16	54
EmoLex$_{\text{Uni}}$:									
adjectives	14	14	10	13	29	14	10	15	68
adverbs	13	20	8	10	23	11	7	23	67
nouns	7	18	3	7	16	6	3	24	46
verbs	11	21	5	16	14	11	7	15	52
EmoLex$_{\text{Bi}}$:									
adjectives	12	25	8	14	30	15	8	16	66
adverbs	6	23	1	7	19	3	9	29	54
nouns	9	23	6	14	20	9	7	29	58
verbs	8	25	5	7	21	6	3	27	60
EmoLex$_{\text{GI}}$:									
negatives in GI	36	4	29	34	0	33	8	2	67
neutrals in GI	4	11	3	8	10	4	5	13	36
positives in GI	1	13	0	2	40	1	4	33	62
EmoLex$_{\text{WAL}}$:									
anger terms in WAL	**83**	1	53	18	0	16	0	0	90
disgust terms in WAL	44	0	**94**	14	0	2	0	0	94
fear terms in WAL	17	17	19	**74**	1	20	15	3	89
joy terms in WAL	2	14	0	2	**78**	2	7	28	91
sadness terms in WAL	9	0	13	13	0	**94**	0	0	96
surprise terms in WAL	2	6	4	8	42	6	66	6	**88**

9.1. Discussion

Table 3 shows that a sizable percent of nouns, verbs, adjectives, and adverbs are emotive. Adjectives and adverbs are some of the most emotion inspiring terms and this is not surprising considering that they are used to qualify nouns and verbs, respectively. Trust, and joy come through as the most common emotions associated with terms. Nouns are more commonly associated with trust, whereas adjectives are more commonly associated with joy.

The **EmoLex$_{\text{WAL}}$** rows are particularly interesting because they serve to determine how much the Turker annotations match annotations in the Wordnet Affect Lexicon (WAL). The most common Turker-determined emotion for each of these rows is marked in bold. Observe that WAL anger terms are mostly marked as anger emotive, joy terms as joy emotive, and so on. Here is the complete list of terms that are marked as anger terms in WAL, but were NOT marked as anger terms by the Turkers: *baffled, exacerbate, gravel, pesky, and pestering*. One can see that indeed many of these terms are not truly anger emotive. We also observed that the Turkers marked some terms as being associated with both anger and joy. The complete list includes: *adjourn, credit card, find out, gloat, spontaneously, and surprised*. One can see how many of these words are indeed associated with both anger and joy.

The **EmoLex$_{\text{WAL}}$** rows also indicate which emotions tend to be jointly associated to a term. Observe that anger terms tend also to be emotive of disgust. Similarly, many joy terms are also emotive of trust. The surprise terms in WAL are largely also associated with joy.

The **EmoLex$_{\text{GI}}$** rows rightly show that words marked as negative in the General Inquirer are mostly associated with negative emotions (anger, fear, disgust, and sadness). Observe that the percentages for trust and joy are much lower. On the other hand, positive words are associated with anticipation, joy, and trust.

9.2. Agreement

In order to analyze how often the annotators agreed with each other, for each term–emotion pair, we calculated the percentage of times the majority class has size 5 (all Turkers agree), size 4 (all but one agree), size 3, and size 2. Table 4 presents these agreement values. Observe that for almost 60% of the terms, at least four annotators agree with each other. Since many NLP systems may rely only on two intensity values (emotive or non-emotive), we also calculate agreement at that level (Table 5).

Table 4. Agreement at four intensity levels for emotion (no, weak, moderate, and strong): Percent of terms for which the majority class size was 2, 3, 4, and 5.

	Majority class size			
Emotion	two	three	four	five
anger	13.7	21.7	25.7	38.7
anticipation	19.2	31.7	28.3	20.7
disgust	13.8	20.7	23.8	41.5
fear	16.7	27.7	25.6	29.9
joy	16.1	24.3	21.9	37.5
sadness	14.3	23.8	25.9	35.7
surprise	11.8	25.3	32.2	30.6
trust	18.8	27.4	27.7	25.9
micro average	**15.6**	**25.3**	**26.4**	**32.6**

Table 5. Agreement at two intensity levels for emotion (emotive and non-emotive): Percent of terms for which the majority class size was 3, 4, and 5.

	Majority class size		
Emotion	three	four	five
anger	13.2	19.4	67.2
anticipation	18.8	32.6	48.4
disgust	13.4	18.4	68.1
fear	15.3	24.8	59.7
joy	16.2	22.6	61.0
sadness	12.8	20.2	66.9
surprise	10.9	22.8	66.2
trust	20.3	28.8	50.7
micro average	**15.1**	**23.7**	**61.0**

Table 6. Agreement at two intensity levels for emotion (emotive and non-emotive) for *evokes* and *associated*: Percent of terms in the pilot set for which the majority class size was 5.

	Majority class size five	
Emotion	evokes	associated
anger	61.6	**68.2**
anticipation	34.8	**49.6**
disgust	65.4	**66.4**
fear	**62.0**	59.4
joy	54.6	**62.3**
sadness	**66.7**	65.3
surprise	54.0	**67.3**
trust	47.3	**49.8**
micro average	55.8	**61.0**

Observe that for more than 60% of the terms, all five annotators agree with each other, and for more than 80% of the terms, at least four annotators agree. This shows that even though emotions are inherently subjective, there is a reasonable degree of agreement on word–emotion associations. This, despite no control over the educational background and qualifications of the annotators.

9.3. Associated versus Evokes

As mentioned earlier, we performed two separate sets of annotations on the pilot set. One where we asked if a word *evokes* a certain emotion, and another where we asked if a word is *associated* with a certain emotion. Table 6 lists the the percent of times all five annotators agreed with each other on the classification of a term as emotive of a particular emotion or not, for the two scenarios.

TABLE 7. Percent of terms assigned a majority class of no, weak, moderate, and strong polarity.

	Intensity			
Polarity	no	weak	moderate	strong
negative	64.3	9.1	10.8	15.6
positive	61.9	9.8	13.7	14.4
polarity average	**63.1**	**9.5**	**12.3**	**15.0**
any polarity	**29.9**	**15.4**	**24.3**	**30.1**

Observe that the agreement numbers are markedly higher with *associated* than with *evokes* for anger, anticipation, joy, and surprise. In case of fear and sadness, the agreement is only slightly better with *evokes*, whereas for trust and disgust the agreement is slightly better with *associated*. Overall, *associated* leads to a jump in agreement by more than 5 percentage points over *evokes*. Therefore, all subsequent annotations were performed with *associated* only. (All results shown in this paper, except for those in Table 6, are for *associated*.)

We speculate that to answer which emotions are evoked by a term, people sometimes bring in their own varied personal experiences, and so we see relatively more disagreement than when we ask what emotions are associated with a term. In the latter case, people may be answering what is more widely accepted rather than their own personal perspective. Further investigation on the differences between *evoke* and *associated* and why there is a marked difference in agreements for some emotions and not so much for others, is left as future work.

10. ANALYSIS OF SEMANTIC ORIENTATION ANNOTATIONS

We consolidate the semantic orientation (polarity) annotations in the same manner as for emotion annotations. Table 7 lists the percent of 8,883 target terms assigned a majority class of no, weak, moderate, and strong semantic orientation. It states, for example, that 15.6% of the target terms are strongly negative. The last row in the table lists the percent of target terms that have some semantic orientation (positive or negative) at the various intensity levels. Observe that 30.1% of the target terms are either strongly positive or strongly negative.

Just as in the case for emotions, practical NLP applications often require only two levels of semantic orientation—having particular semantic orientation (**evaluative**) or not (**non-evaluative**). For each target term–emotion pair, we convert the four-level semantic orientation annotations into two-level ones, just as we did for the emotions. Table 8 shows how many terms overall and within each category are positively and negatively evaluative.

10.1. Discussion

Observe in Table 8 that, across the board, a sizable number of terms are evaluative with respect to some semantic orientation. Unigram nouns have a markedly lower proportion of negative terms, and a much higher proportion of positive terms. It may be argued that the default semantic orientation of noun concepts is neutral or positive, and that usually it takes a negative adjective to make the phrase negative.

The **EmoLex**$_{\mathbf{GI}}$ rows in the two tables show that words marked as having a negative semantic orientation in the General Inquirer are mostly marked as negative by the Turkers. And similarly, the positives in GI are annotated as positive. Again, this is confirmation that the quality of annotation obtained is high. Observe that the Turkers mark 12% of the GI neutral terms as negative and 30% of the GI neutral terms as positive. This may be because the boundary between positive and neutral terms is more fuzzy than between negative and neutral terms. The **EmoLex**$_{\mathbf{WAL}}$ rows show that anger, disgust, fear, and sadness terms tend not to have a positive semantic orientation and are mostly negative. In contrast, and expectedly, the joy terms are positive. The surprise terms are more than twice as likely to be positive than negative.

TABLE 8. Percent of terms, in each target set, that are evaluative. The highest scores for EmoLex$_{GI}$ positives and negatives are shown bold. Observe that the positive GI terms are marked mostly as positively evaluative and the negative terms are marked mostly as negatively evaluative.

	negative	positive	any
EmoLex	30	35	65
EmoLex$_{Uni}$:			
adjectives	32	48	79
adverbs	26	55	80
nouns	8	39	46
verbs	26	37	63
EmoLex$_{Bi}$:			
adjectives	30	47	77
adverbs	11	42	52
nouns	14	45	57
verbs	14	48	60
EmoLex$_{GI}$:			
negatives in GI	**83**	1	85
neutrals in GI	12	30	41
positives in GI	2	82	**84**
EmoLex$_{WAL}$:			
anger terms in WAL	96	1	97
disgust terms in WAL	97	0	97
fear terms in WAL	85	1	86
joy terms in WAL	4	93	97
sadness terms in WAL	91	4	95
surprise terms in WAL	26	57	80

TABLE 9. Agreement at four intensity levels for polarity (no, weak, moderate, and strong): Percent of terms for which the majority class size was 2, 3, 4, and 5.

	Majority class size			
Polarity	two	three	four	five
negative	12.8	27.3	27.2	32.5
positive	23.5	28.5	18.0	29.8
micro average	**18.2**	**27.9**	**22.6**	**31.2**

TABLE 10. Agreement at two intensity levels for polarity (evaluative and non-evaluative): Percent of terms for which the majority class size was 3, 4, and 5.

	Majority class size		
Polarity	three	four	five
negative	11.5	22.3	66.1
positive	24.2	26.3	49.3
micro average	**17.9**	**24.3**	**57.7**

10.2. Agreement

In order to analyze how often the annotators agreed with each other, for each term–polarity pair, we calculated the percentage of times the majority class has size 5 (all Turkers agree), size 4 (all but one agree), size 3, and size 2. Table 9 presents these agreement values. Observe that for more than 50% of the terms, at least four annotators agree with each other. Table 10 gives agreement values at the two-intensity level. Observe that for more than 55% of the terms, all five annotators agree with each other, and for more than 80% of the terms, at least four annotators agree.

TABLE 11. Percent of terms marked as being associated with each color.

	Colors										
	white	black	red	green	yellow	blue	brown	pink	purple	orange	grey
raw	11.9	12.2	11.7	12.0	11.0	9.4	9.6	8.6	4.2	4.2	4.6
voted	22.7	18.4	13.4	12.1	10.0	6.4	6.3	5.3	2.1	1.5	1.3

TABLE 12. Color signature of emotive terms: percent of emotive terms associated with each color. For example, 32.4% of the anger terms are associated with the color red.

	Colors										
	white	black	red	green	yellow	blue	brown	pink	purple	orange	grey
anger	2.1	**30.7**	**32.4**	5.0	5.0	2.4	6.6	0.5	2.3	2.5	9.9
anticipation	**16.2**	7.5	11.5	**16.2**	10.7	9.5	5.7	5.9	3.1	4.9	8.4
disgust	2.0	**33.7**	**24.9**	4.8	5.5	1.9	9.7	1.1	1.8	3.5	10.5
fear	4.5	**31.8**	**25.0**	3.5	6.9	3.0	6.1	1.3	2.3	3.3	11.8
joy	**21.8**	2.2	7.4	**14.1**	13.4	11.3	3.1	11.1	6.3	5.8	2.8
sadness	3.0	**36.0**	**18.6**	3.4	5.4	5.8	7.1	0.5	1.4	2.1	16.1
surprise	11.0	**13.4**	**21.0**	8.3	**13.5**	5.2	3.4	5.2	4.1	5.6	8.8
trust	**22.0**	6.3	8.4	**14.2**	8.3	**14.4**	5.9	5.5	4.9	3.8	5.8

11. ANALYSIS OF COLOR ANNOTATIONS

Table 11 has the percentage of times different colors were chosen as being associated with the target words. The first row shows percentages of raw counts obtained from all the annotators on all the terms. For each term, the color that gets the majority votes (from the 3 to 5 annotators for that HIT) is chosen as the true color associated with the target term, and the "voted" row of Table 11 shows the percentage of terms associated with the different colors. Observe that even though the color options were presented in random order to the annotators, the order of the most frequently associated colors is identical to the order in which colors emerged in human language (first white, then black, then red, then green, and so on) (Berlin and Kay, 1969).

From the annotated associations between words and emotions together with associations between words and colors, we can determine the color signature of different emotions—the rows in Table 12. The top two most frequently associated colors with each of the eight emotions are shown in bold. The "anger" row shows the percentage of anger terms associated with different colors. Observe that red and black are the most associated colors with anger terms (red slightly more). Red and black are also most frequently associated with other negative emotions disgust, fear, and sadness. However, for these emotions black is much more dominant. Grey comes third for all four of the negative emotions. Among the positive emotions: anticipation is most frequently associated with white and green; joy with white, green, and yellow; and trust with white, blue, and green. Surprise, which can be positive or negative, is associated with red, yellow, and black.

Table 13 shows the color signature for all terms marked positive and all terms marked negative (these include positive and negative terms that may or may not be associated with any of the 8 basic emotions). As expected from the results in Table 12, the negative terms are strongly associated with black and red, whereas the positive terms are strongly associated with white and green.

11.1. Agreement

Table 14 shows how often the majority class in color associations is 1, 2, 3, 4, and 5 respectively. Since the annotators are given 11 different color options to choose from, the chance that none of the 5 annotators agrees with each other (majority class size of 1) is $1 \times 10/11 \times 9/11 \times 8/11 \times 7/11 = 0.344$. Thus, if there was no real correlation among any of the words and colors, then 34.4% of the time none of the annotators would have agreed with each other. However, this happens only 15.1% of the time in the annotations. For about 2.1% of the words, all 5 annotators agree on the associated

TABLE 13. Color signature of positive and negative terms: percent of polar terms associated with each color. For example, 28.3% of the negative terms are associated with black.

	Colors										
	white	black	red	green	yellow	blue	brown	pink	purple	orange	grey
negative	2.9	**28.3**	**21.6**	4.7	6.9	4.1	9.4	1.2	2.5	3.8	14.1
positive	**20.1**	3.9	8.0	**15.5**	10.8	12.0	4.8	7.8	5.7	5.4	5.7

TABLE 14. Percent of terms in a majority color class of one, two, three, four, and five.

majority class size				
one	two	three	four	five
15.1	52.9	22.4	7.3	2.1

colors. One can argue that even the terms for which the majority class was 3 (22.4% of the terms) and 4 (7.3% of the terms) have strong color associations.

12. CONCLUSIONS

Emotion detection and generation have a number of practical applications. Even though there is work in speech and facial expressions, it is only recently that we see growing interest in analyzing text for emotions. However, only a small number of limited-coverage emotion resources exist, and that too only for English. In this paper we show how the combined strength and wisdom of the crowds can be used to generate a large word–emotion association lexicon quickly and inexpensively. This lexicon, EmoLex, has entries for more than 10,000 word–sense pairs. Each entry lists the association of the a word–sense pair with 8 basic emotions. We used Amazon's Mechanical Turk as the crowdsourcing platform. Lexicons can be created, in a similar manner, for other languages too as long as there are enough Turkers who speak the target language.

We fleshed out various challenges associated with crowdsourcing the creation of an emotion lexicon (many of which apply to other language annotation tasks too), and presented various solutions to address those challenges. Notably, we used automatically generated word choice questions to detect and reject erroneous annotations and to reject all annotations by unqualified Turkers and those who indulge in malicious data entry. The word choice question is also an effective and intuitive way of conveying the sense for which emotion annotations are being requested.

We compared a subset of our lexicon with existing gold standard data to show that the annotations obtained are indeed of high quality. A detailed analysis of the lexicon revealed insights into how prevalent emotion bearing terms are among common unigrams and bigrams. We also identified which emotions tend to be evoked simultaneously by the same term. As part of the annotation we also compiled a list of 826 terms that are not just associated with emotions, but refer directly to emotions. All of the 10,170 terms in the lexicon are also annotated with whether they have a positive, negative, or neutral semantic orientation. We find that many of the terms in a language are strongly associated with certain colors. Also, the frequency of association of a color with terms is in exact correlation with the order in which color terms first appeared in language. From the various word–emotion and word–color annotations, we also generate color signatures for different emotions.

Our future work includes expanding the coverage of the lexicon even further, creating similar lexicons in other languages such as German, identifying cross-cultural and cross-language differences in emotion and color associations, and last but not least to use the lexicon in various emotion detection applications such as those listed in Section 1. The lexicon will be made available for free download.[6]

[6]http://www.purl.org/net/emolex

Acknowledgments

This research was funded by the National research Council Canada (NRC). Thanks to Diana Inkpen and Diman Ghazi for early discussions on emotion. Thanks to Joel Martin for encouragement and support. Thanks to Norm Vinson and the Ethics Committee at NRC for examining, guiding, and approving the survey. And last but not least, thanks to the more than 2000 anonymous annotators who answered the emotion survey with diligence and care.

REFERENCES

Alm, C. O., Roth, D., and Sproat, R. (2005). Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of the Joint Conference on Human Language Technology / Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, pages 579–586, Vancouver, Canada.

Bellegarda, J. (2010). Emotion analysis using latent affective folding and embedding. In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, California.

Berlin, B. and Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

Bernard, J., editor (1986). *The Macquarie Thesaurus*. Macquarie Library, Sydney, Australia.

Brants, T. and Franz, A. (2006). Web 1t 5-gram version 1. *Linguistic Data Consortium*.

Callison-Burch, C. (2009). Fast, cheap and creative: Evaluating translation quality using amazon's mechanical turk. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, pages 286–295, Singapore.

Darwin, C. (1872). *The Expressions of the Emotions in Man and Animals*. John Murray.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, **6**(3), 169–200.

Ekman, P. (2005). *Emotion in the Human Face*. Oxford University Press.

Ekman, P. and Friesen, W. V. (2003). *Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions*. Malor Books.

Elliott, C. (1992). *The affective reasoner: A process model of emotions in a multi-agent system*. Ph.D. thesis, Institute for the Learning Sciences, Northwestern University.

Goyal, A., Riloff, E., Daume III, H., and Gilbert, N. (2010). Toward plot units: Automatic affect state analysis. In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, California.

Guo, K., Hall, C., Hall, S., K., M., and Mills, D. (2007). Left gaze bias in human infants, rhesus monkeys, and domestic dogs. *Animal Cognition*, **12**, 409–418.

James, W. (1884). What is an emotion? *Mind*, **9**, 188–205.

Lin, D., Zhao, S., Qin, L., and Zhou, M. (2003). Identifying synonyms among distributionally similar words. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03)*, pages 1492–1493, Acapulco, Mexico.

Lobanova, A., van der Kleij, T., and Spenader, J. (2010). Defining antonymy: A corpus-based study of opposites by lexico-syntactic patterns. *International Journal of Lexicography (in press)*, **23**, 19–53.

Masson, J. M. (1996). *When Elephants Weep: The Emotional Lives of Animals*. Delta.

Mohammad, S., Dorr, B., and Dunn, C. (2008). Computing word-pair antonymy. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, pages 982–991, Waikiki, Hawaii.

Neviarouskaya, A., Prendinger, H., and Ishizuka, M. (2009). Compositionality principle in recognition of fine-grained emotions from text. In *Proceedings of the Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM-09)*, pages 278–281, San Jose, California.

Ortony, A. and Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, **97**, 315–331.

Parrot, W. (2001). *Emotions in Social Psychology*. Psychology Press.

Pearl, L. and Steyvers, M. (2010). Identifying emotions, intentions, and attitudes in text using a game with a purpose. In *Proceedings of the NAACL-HLT 2010 Workshop on Computational*

Approaches to Analysis and Generation of Emotion in Text, Los Angeles, California.

Plutchik, R. (1962). *The Emotions*. New York: Random House.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Emotion: Theory, research, and experience*, **1**(3), 3–33.

Plutchik, R. (1994). *The psychology and biology of emotion*. New York: Harper Collins.

Read, J. (2004). *Recognising affect in text using pointwise-mutual information*. Ph.D. thesis, Department of Informatics, University of Sussex.

Sautera, D. A., Eisner, F., Ekman, P., and Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, **107**(6), 2408–2412.

Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. (2008). Cheap and fast - but is it good? Evaluating nonexpert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, pages 254–263, Waikiki, Hawaii.

Stone, P., Dunphy, D. C., Smith, M. S., Ogilvie, D. M., and associates (1966). *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.

Strapparava, C. and Valitutti, A. (2004). Wordnet-Affect: An affective extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004)*, pages 1083–1086, Lisbon, Portugal.