Automatic Emotion Annotation of Dream Diaries

Elena Frantova
Topological Media Lab and CLaC Lab
Computer Science and Software Engineering
Concordia University
Montreal, Canada
e_franto@cse.concordia.ca

Sabine Bergler
CLaC Lab
Computer Science and Software Engineering
Concordia University
Montreal, Canada
bergler@cse.concordia.ca

ABSTRACT
This paper explores automatic annotation of dream reports. We first compile "emotion dictionaries" from a thesaurus using Hall/Van de Castle emotion categories proposed for dream analysis. We capture the inherent ambiguity and polysemy of emotion words in *word profiles*, which give the fuzzy membership score of a word on all five emotion categories. We then use the derived dictionaries to assign emotion categories to texts in so-called *category profiles*. We analyze several parameters of the technique. A comparison with the manual annotation of texts from DreamBank indicate that this multi-faceted approach is promising.

1. INTRODUCTION
Attention in natural language processing (NLP) has recently shifted to include not only carefully edited text types using objective language, like newspaper articles, but also text types using subjective language, like commentary, reviews, journals, personal blogs. Not surprisingly, the form of the language differs in these genres (typos, slang, short forms), and the type of information transmitted is not mainly factual, as we find in newspapers or scientific writing, but highly opinionated, sentiment-laden, and emotional. And these outbursts of subjectivity cannot be ignored (as is usually done in "factual" text analysis), since often they are the message. Decoding the attitudes behind subjective language is a necessary first step in extracting the meaning of subjective texts.

This is the case for dream reports, studied, for example, by psychologists in sleep research. They are generally short, explicit descriptions of scenes and emotions (see Figure 1). Due to the nature of dreams, more than one emotion may be expressed in a single report. The goal of our work is to provide an automatic tool to categorize dream reports by the emotions expressed. This is currently done manually and presents a serious bottleneck for the comparison and evaluation of dream reports.

Using ideas that have proven effective in sentiment analysis, we present here an approach for building a dictionary of emotion laden terms, where each term is classified with its *word profile*, that is the degree of membership in the five emotion categories *happiness*, *sadness*, *anger*, *confusion*, and *apprehension* proposed by Hall and Van de Castle [10]. We then propose a multifaceted methodology to assign *category profiles* to dream reports and test the approach on DreamBank [6] data. To assess the dictionary's potential on other text types, we present preliminary results on using emotion terms for sentiment classification on blog sentences.

2. RELATED LITERATURE
While sentiment, opinion, and emotions are encoded mainly implicitly in "objective" text types traditional for NLP (such as newspaper articles), they are much more explicit in text types that are gaining importance in part through the increased use of social media (such as journal-like blogs). These categories of subjective language have, consequently, only recently found wider attention. Sentiment analysis, for instance, uses the "thumbs up/thumbs down" rating system of on-line movie reviews, extending the concept to product reviews, newspaper articles, and blogs using in-domain annotated texts for statistical approaches or lists of trigger words previously extracted and annotated for positive, negative, and neutral sentiment [3].

Opinion annotation adds to the positive/negative dimension a range of opinionated language (including argumentative and emotion indicating) [16]. [9] widen the attention to considering a variety of subjectivity indicators, combining their different triggers in a machine learning framework and report improved performance on the subjectivity sense labeling task (i.e. whether or not a word sense is subjective in context or not).

Emotions are a focal point for the scientific study of dreams. Hall and Van de Castle [10] developed a dream coding system, including classification of emotions. This has been widely used in the scientific study of dreams and in this paper we adopt their emotion categories *happiness*, *confusion*, *apprehension*, *anger*, and *sadness* to investigate whether automatic analysis can classify dream reports with sufficient accuracy to develop tools for dream researchers.

The task is similar to sentiment and subjectivity annotation, yet unlike for those tasks, no corpora have been prepared for evaluation of NLP tools, nor do reference word lists exist, in fact most research that does study emotions for automatic processing includes collecting and annotating data (see for instance [2, 15]).

3. CODING DREAMS

Dream reports, accompanied by psycho-physiological data, such as electrical activity of the brain or muscular activity, are studied by psychologists for dream research. Dream reports may be collected in a sleep laboratory during experimental awakening, logged by subjects in dream diaries in their normal domestic setting for a given period of time, or, recently, submitted through the Internet. Dream reports allow researchers to study connections between dreaming and certain socio-demographic, physiological factors or medical conditions[1].

The analysis of dream reports in this setting has to be carefully normed if the results are to be compared to those of other researchers. Hall and Van de Castle developed a very detailed coding scheme in [10][2]. As part of this scheme, they propose five emotion categories, *happiness*, *confusion*, *apprehension*, *anger*, and *sadness* with sample trigger words that are meant to be indicative of the emotion category.

An online database of dream reports, DreamBank[3] [6], launched in 1999, currently contains a collection of 16000 dream reports, in part manually annotated with the Hall/Van de Castle coding scheme. DreamBank is presented through a web interface that enables different types of searches, clearly indicating the need for automatic tools to navigate the database of dreams. The fact that only some dreams are annotated is undoubtedly due to the time intensive process of manually annotating dreams. An automatic emotion annotation would enhance the search options over the entire corpus and allow ad hoc annotation of additional reports. It also serves human annotators as a first approximation of the annotation to be corrected, akin to computer-aided translation tools [11], possibly enhancing consistency between annotators.

I was in biology lab and had a bag of cookies and a bag of after dinner mints. I put them on the desk and said everyone was welcome to help themselves. When I went to get some, they were all gone, so I went around to each person in the lab and made them put some back. I was perplexed because they had taken advantage of my generosity and disgusted because some of them had made such pigs of themselves. (81 words)

Figure 1: DreamBank report #0005 annotated with categories CO (confusion) and AN (anger)

In dream research, emphasis is put on the annotated categories being expressed explicitly, to minimize the bias the annotator brings to the task. Thus a sad story expressed ironically with all overtly positive words will not be annotated for *sadness* in the Hall/Van de Castle scoring scheme. In Figure 1, the annotated emotions are CO (for confusion) and AN (for anger). Note that the annotation is given as a global summary and the trigger words are not indicated. We can only guess that the label CO stems from *perplexed* and the label AN from *disgusted*.

Compared to traditional texts analyzed automatically, dream reports are unpredictable, inconsistent, and abundant in dramatic narrative and emotional changes, requiring adding emotion analysis in addition to the subjectivity categories currently investigated in other research. A number of emotion classifications exist, there is, however, no generally accepted model of emotions. Ekman presents a set of six basic emotions [8, 7] - *anger, disgust, fear, happiness/joy, sadness, surprise*, which have been used in NLP studies [1, 13]. A different approach to emotion classification has been proposed by Watson and Tellegen with the Circumplex Theory of Affect [17] applied by [15], who compiled a manually annotated corpus for the task. The inter-annotator agreement in these studies is reportedly low [15, 12, 2], thus we prefer to use the Hall/Van de Castle coding here because of the detailed annotation instructions and because the DreamBank corpus has been annotated independently for dream research.

4. EMOTION DICTIONARY

The first step to automatically annotate dream reports with emotion categories is to compile an emotion dictionary. Our goal is to compile a more comprehensive list of emotion words than can be found in the literature while exploring the influence of different parameters on

[1]For instance, Nielsen suggests that emotion patterns from the waking life reappear in dreams in disguised form [14].
[2]The schema codes dream reports into various categories: characters, social interactions, activities, success and failure, misfortune and good fortune, emotions, settings, objects and descriptive elements. Each category consists of a number of coded subcategories and rules for their application are given.
[3]www.dreambank.net

satisfaction
1 *(he derived great satisfaction from his work)* contentment, pleasure, gratification, fulfillment, enjoyment, happiness, pride; self-satisfaction, smugness, complacency.
2 *(the satisfaction of consumer needs)* fulfillment, gratification; appeasement, assuaging.
3 *(investors turned to the courts for satisfaction)* compensation, recompense, redress, reparation, restitution, repayment, payment, settlement, reimbursement, indemnification, indemnity.

Figure 2: Thesaurus entry for satisfaction

the task.

4.1 Thesaurus

We adopt an approach that has been successfully used for sentiment annotation [4]: we automatically compile a dictionary using an existing linguistic resource. We use an electronic version of the Oxford American Writers Thesaurus[4] using different seed lists derived from [10].

As Figure 2 illustrates, a thesaurus doesn't provide definitions for words, but groups synonymous words, usually under the heading of the most general synonymous term. The algorithm starts from a small seed list for one emotion category[5]. For each seed word, all synonyms listed under each emotion word sense are added to form an expanded seed list. This expanded seed list is used as seed list in a next iteration. With each iteration, the strength of the membership degree of the added expansion words in the emotion category is decreased.

For seed lists that start from a single term ({*happiness*}, for instance), the algorithm starts from an initial seed list that contains all the terms of the synonym group for this single seed term. For *happiness*, *satisfaction* is an immediate synonym and receives a membership degree of 1.

4.2 Word Profiles

The fact that there are several emotion categorization schemes proposed and used without convergence on a core set is, in our opinion, due to the multifaceted character of emotions not lending itself to rigid categorization models. Ambiguity in emotion words is illustrated by the word *shocked*: it can indicate a pleasant surprise or serious trauma. But beyond simple ambiguity, with any emotion model there are always borderline cases, words that do not fit into the classification. One example for both Hall/Van de Castle and Ekman models is *guilt*, which, while unambiguous as an emotion, does not fit any of their categories. Hall/Van de Castle suggest to classify *guilt* as *apprehension* in the coding guidelines. *Guilt* feels actually more like the inverse of *apprehension* to us, and *sadness* is as good an option, illustrating bias in the classification task.

We address this potential overlap of emotion words with several categories using a multidimensional membership degree approach. In similar research, fuzzy membership has been used successfully to model centrality of words in the sentiment categories *positive*, *negative*, and *neutral* [3], expressing likelihood in the exclusive categories. Here, we rely more strongly on the underlying assumption that any linguistic variable (here: word) has a certain fuzzy membership degree in all 5 (overlapping) categories [18]. This means that the fuzzy membership in all five categories in a *word profile* here positions the word over the five non-exclusive meaning dimensions as well as indicating its likelihood of membership in a given emotion category.

4.3 Dictionary Compilation

We compile separate dictionaries for each of the five emotion categories and for each part of speech considered (adjective, noun, verb). Compilation of a particular dictionary proceeds iteratively from a seed list of terms which describe the category. All synonyms listed in a thesaurus category headed by a seed word are added to the list as long as the word sense is determined also by a second term in the seed list. Word sense drift can occur when a seed word has several word senses. This redundancy test ensures that only word senses for which at least two synonyms are in the seed list are added to the dictionary, to keep word sense drift to a minimum.

The algorithm is repeated for a fixed number of iterations (*total_iter.*). With fewer iterations we have fewer words in the dictionary and the collected words are closer in meaning. More iterations yield more words, but with each iteration the new words are less specific to the category, that is they tend to have membership in more emotion categories, increasing “fuzziness”.

We compared resulting dictionaries derived in different numbers of iterations. The empirically determined optimum is five iterations for seed lists starting from a single term, four for larger seed lists.

For each term in the dictionary, we record the iteration in which it was added (*recruiting_iter.*) as a measure of distance from the original seed list. The membership degree $\mu_{C_i}(w)$ in emotion category C_i is computed using (1):

[4] 1st Edition, built into the OSX Dictionary application and accessed via the Dictionary Service Library of the Apple Developer Libraries.
[5] The seed list may contain only a single word or a list. The resulting dictionaries differ. A comprehensive evaluation for different seed lists was beyond this paper. We use as initial seeds the terms suggested by [10], see Section 4.4.

$$\mu_{C_i}(w) = WSR(w) * Weight(w) \quad (1)$$

$$Weight(w) = \frac{total\_iter. - (recruiting\_iter.(w) - 1)}{total\_iter.} \quad (2)$$

WSR is the word sense ratio of a term given in (3) as the number of emotion word senses that have been validated by at least two synonyms divided by the total number of word senses.

$$WSR(w) = \frac{emotion\_senses(w)}{word\_senses(w)} \quad (3)$$

The WSR for *satisfaction* is 2/3. This allows us to devalue abstract words like *thought* with multiple meanings, few of which fall into emotion categories.

A third measure used is each entry's *connectivity*, that is how many other entries in the category reference this term as its synonym. Connectivity is thus a global measure that is determined after the dictionary has been acquired. Connectivity is a measure of centrality in the category, we surmise that an entry with high connectivity is also more specific to the category than one with low connectivity and use it between the iterations, for a transitional mid-iteration. After a main iteration is complete, terms acquired on the previous iteration with connectivity greater then 2 are revisited again in a mid-iteration[6]. Connectivity above 2 signifies that the term has been referenced in one of its senses by other terms during the iteration. Some of its word senses that did not qualify during the main iteration may qualify now. For this step, any term acquired in previous iterations of the given emotion category can qualify the word sense as valid. Table 1 illustrates these measures on the example of the term *satisfaction* using the SingleSeed dictionary (see below).

4.4 Seed Lists

The choice of seed word(s) leads to different emotion dictionaries and has to be taken into account, when designing an automated system. We experimented with different seeds to assess two different strategies: growing dictionaries from a single, central seed or using a wider list of manually determined seed words, such as the complete list of coding examples for each category and part of speech described in [10].

For any seed list, we compile emotion category dictionaries that have three parts. *Part 1* contains adjectives,

[6]During mid-iterations the *recruiting-iter.* value is increased by 0.5.

Table 1: *Satisfaction* statistics in each emotion category

Category	Weight	WSR	μ	Conn.
anger	0.125	2/3	0.083	2
apprehension	0.125	2/3	0.083	2
confusion	0	2/3	0	0
happiness	1	2/3	0.666	12
sadness	0	2/3	0	0

Part 2 contains nouns, and *Part 3* contains verbs. We compile these dictionaries separately not only because our technique might confound different part of speech and different word senses more easily, but also because language uses them differently. Adjectives are the most explicit indicators of emotions in text, they are predominantly used to express a first person perspective in the dream diaries. Nouns associated with emotion categories are found more likely in official-sounding context, often for reporting (the familiar third person perspective). By sorting the dictionaries by part of speech and recognizing their different shades of meaning, we set the stage for exploiting these insights later, while not capitalizing on them in this context.

The smallest seed list[7] starts from exactly one seed word for each emotion category, for adjectives these are *angry, apprehensive, confused, happy*, and *sad*. The initial seed list used in the iterative algorithm in this case consists of the thesaurus synonyms of these five terms. Running the algorithm for five iterations leads to Part 1 of the respective dictionaries. Part 2 is derived from the nouns *anger, apprehension, confusion, happiness, sadness.*

Verb seeds are manually derived from the seed list for adjectives. The resulting Part 3 of the category dictionaries is, however, not as comprehensive as Parts 1 and 2. We therefore include also those verbs which have identical forms as adjectives or nouns and which were derived for Parts 1 or 2, such as *hurt* (adjective) and *anger* (noun). In our experiments, adding verb dictionaries increased results by only an insignificant amount, possibly because the most frequent verbal form encountered is the participle, which is already included in the adjective lists.[8]

We also compiled dictionaries from larger adjective seed lists taken from the Hall/Van de Castle coding instruc-

[7]We refer to dictionaries derived from these single seeds as SingleSeed dictionaries.

[8]It is interesting to note that the members of the adjective dictionaries are mainly derived from nouns or verbs. Deverbal adjectives convey the implicit presence of a causer, while nouns convey neutrality or a self-inflicted, more permanent emotional state.

Category	Seed Words
Anger	annoyed, irritated, mad, provoked, furious, enraged, belligerent, incensed, indignant
Apprehension	terrified, horrified, frightened, scared, worried, nervous, concerned, panicky, alarmed, uneasy, upset, remorseful, sorry, apologetic, regretful, ashamed
Sadness	disappointed, distressed, hurt, depressed, lonely, lost, miserable, hopeless, crushed, heartbroken
Confusion	surprised, astonished, amazed, awestruck, mystified, puzzled, perplexed, strange, bewildered, doubtful, conflicted, undecided, uncertain
Happiness	contented, pleased, relieved, amused, cheerful, glad, relaxed, gratified, gay, wonderful, elated, joyful, exhilarated

Table 2: HVDC adjective seeds

tions illustrated in Table 2 for comparison. We call these HVDC dictionaries. When starting from a larger seed list, we reduce the number of total iterations to four.

The resulting emotion category dictionaries differ significantly, as illustrated by the example of *guilt* in Table 3. In the HVDC dictionary, its highest membership is in the category *apprehension*, followed by *sadness*. This conforms with Hall and Van de Castle's recommendation to classify *guilt* under *apprehension*. This conformity shows consistency in the bias of Hall/Van de Castle classification terms, since the SingleSeed dictionary attributes the highest membership degree for *guilt* in the category *sadness* with very low membership in *apprehension*. The SingleSeed dictionary demonstrates the semantic field of emotion terms as conceived by the thesaurus lexicographers only, whereas the HVDC dictionary adds a user bias. This demonstrates the potential danger of overpowering the lexical resource with seed lists that are too big or too skewed. Future research has to take this effect into account.

Another difference between the SingleSeed dictionaries and the HVDC dictionaries stems from the strong bias of HVDC for explicit markers of emotion. The SingleSeed dictionary is bigger than the HVDC dictionary, because in the first iteration all thesaurus synonyms of the category seeds are part of the seed list, that includes indirect emotion markers such as *grinning*. No indirect category indicators are part of the HVDC seeds and these indirect markers occur less frequently in the HVDC dictionary.

Comparing Part 1 of the SingleSeed and HVDC dictionaries, we observe the following: the SingleSeed dictionary contains 7594 words, the HVDC dictionary 6640. Of these, 6451 are in common. Interestingly, most of the words that are exclusive to only one dictionary were added during the last iteration of the algorithm, indicating that the number of iterations is a parameter that controls specificity of the dictionary. Terms occurring only in the SingleSeed dictionary generally have lower category membership degrees and are mostly indirect markers of their category, but also include a few strong emotion markers such as *scary* and *homesick*.

This demonstrates again that the underlying bias of the seed lists interacts with the implicit structure of the lexicographic resource used. For every application task, the appropriateness of seed lists and lexicographic resources has to be carefully assessed. The implicit structure of a lexicographic resource can be made explicit by deriving category dictionaries from seed lists and studying them for their usefulness. We presented here the notions of connectivity and word sense ratio as interesting tools for such an endeavor. While we can only present few insights from our analysis here, it is our strong conviction that this type of lexicographic analysis is necessary also for statistical techniques and that it is useful to identify which features should become standard for emotion annotation.

5. EMOTION ANNOTATION OF TEXTS

Figure 1 illustrates the annotation format of the DreamBank data. 483 dream reports in the DreamBank corpus are annotated for the Hall/Van de Castle categories for each dream. The annotation might indicate that the dream contains two markers of *happiness* and one marker of *sadness*, but unfortunately the markers themselves are not identified. In order to use DreamBank data as a gold standard, we convert our five-dimensional word profiles for each emotion word into a summary emotion category ranking for the text. To this end, for each emotion category, we add the membership degree $\mu_{C_i}(w)$ of each identified emotion word w in that cate-

Table 3: Emotion category membership for *guilt* for different seed lists

Category	SingleSeed	HVDC
Anger	0.0625	0.0625
Happiness	0	0
Sadness	0.3125	0.1875
Apprehension	0.0625	0.25
Confusion	0	0

gory. To normalize for text length, we divide this sum by the number EW of emotion words in the text. We compare the five sums for each dream report and consider their relative ranking to be the *category profile* for that dream report.

$$\text{category profile } = (C_1, C_2, C_3, C_4, C_5), \\ \text{where } C_1 \geq C_2 \geq C_3 \geq C_4 \geq C_5 \quad (4)$$

$$C_i = \frac{1}{EW} \sum_{\forall w \in \text{emotion dictionary}} \mu_{C_i}(w) \quad (5)$$

The gold standard annotations are also ranked by the number of annotations in each emotion category. So if a text has 2 annotations for *anger*, 3 for *apprehension* and 1 for *confusion*, the gold standard category profile for this text will be *(apprehension, anger, confusion)*. The gold standard category profiles are compared to the automatically derived emotion category profiles using the notion of *position correctness* (see Equation 6). DPR is necessarily an imprecise mapping. Our automatic annotation assigns categories to implicit emotion indicators (such as *grinning*) and has an average of around 20 emotion markers per dream, compared to an average 3 recorded explicit emotion markers in the gold standard. Interestingly, while the total number of emotion markers differs drastically, the relative proportions seem stable and we use these relative emotion profiles for evaluation.

For comparison, we also calculate a winner take all strategy, where each dictionary entry is assigned only the strongest emotion category, not the profile over all five categories.

5.1 DreamBank Evaluation

Our goal is not to duplicate DreamBank annotation automatically, but to propose an automatic emotion dictionary acquisition method and to index texts for their emotion category profiles. Because texts usually contain more than one emotion, we propose a way to accumulate and combine fuzzy membership degrees in all five emotion categories of words into a category profile for a text by essentially adding and normalizing the fuzzy membership degrees for all words in each emotion category into a 5-tuple, the category profile. The usefulness of these techniques we are proposing cannot be evaluated straightforwardly, since no standard tasks and annotated corpora exist. DreamBank is a very close annotated corpus and serves us for evaluation purposes here, but important differences exist, most importantly we do not limit our annotation to explicit emotion markers, but consider also implicit markers. Our approach also considers words 4-5 (synonym-) links removed from the seeds, which introduces considerable fuzziness.

	AN	**SA**	**AP**	**CO**	**HA**
# dreams	85	146	215	124	113
	Thesaurus-based seed dictionary				
WTA	.48	.43	.59	.30	.77
WWTA	.65	.61	.54	.37	.75
5CAT	.77	.94	.87	.74	.93
	HVDC dictionary				
WTA	.40	.67	.56	.31	.74
WWTA	.51	.74	.69	.57	.81
5CAT	.71	.94	.91	.78	.91

Table 4: Comparing winner take all, weighted WTA and word category quintuples (5CAT)

Still, we do not prioritize recall over precision here. Given that the system is to benefit a psychologist in retrieving from or analyzing a comparatively small set of documents, comprehensiveness in retrieval (recall) should possibly be more important than keeping the retrieved set accurate (precision). But since our interest here is to assess our dictionary acquisition and category assignment methodology, precision is at least as important as recall.

The task then is to compare two ranked lists of potentially different length, *gp* (gold standard profile) and *cp* (our category profile). For each of the N dream reports and for each of the k (up to five) emotion categories that the gold standard identifies for that dream report, the *DreamBank positional rating* (DPR) sums *pos._cor.*, the correctness of the result for that profile position.

$$DPR = \frac{1}{kN} \sum_{i=1}^{N} \sum_{j=1}^{k} pos.\_cor.(ij) \\ \text{where } 1 \leq k \leq 5, \\ i \text{ is a dream report and } j \text{ is a profile position} \quad (6)$$

$$pos.\_cor.(ij) = \begin{cases} 1 & \text{if } gp_i(j) = cp_i(j) \\ 0.5 & \text{if } \mid gp_i(j) - cp_i(j) \mid \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

DPR tries to assess recall and precision in a single measure. To assess the influence of our decision to model emotion markers as quintuples of fuzzy membership degrees in the five emotion categories, we present a winner take all (WTA) strategy. Here words are classified only under one category, the one with highest membership degree, that means four positions in the word profile quintuple are zero. Two variations are presented in Table 4, *WTA*, where all emotion markers have the same strength, and weighted WTA(*WWTA*), where the category membership degree of the highest category is retained as a weight. The evaluation procedure is otherwise unchanged. Table 4 shows highest values for DPR, confirming that the secondary category memberships are informative and the multidimensional quintuple representation is effective for words. WWTA outperforms WTA, highlighting that even if the secondary category memberships are not available, the degree of membership in the highest category is informative, which lends support to our dictionary compilation strategy.

To establish that all parts of speech contribute to the results, we ran a baseline experiment that mimics a very naive emotion annotation. Using as a word list only the adjectives suggested in the Hall/Van de Castle coding instructions, we simply count the occurrence of each category marker and assign the category that received the highest count to the text. This baseline procedure has an overall score of .55, which is high given the relatively short list of adjectives it used. This confirms findings in the literature that adjectives are high-quality markers, but that coverage can be improved by considering also nouns and verbs.

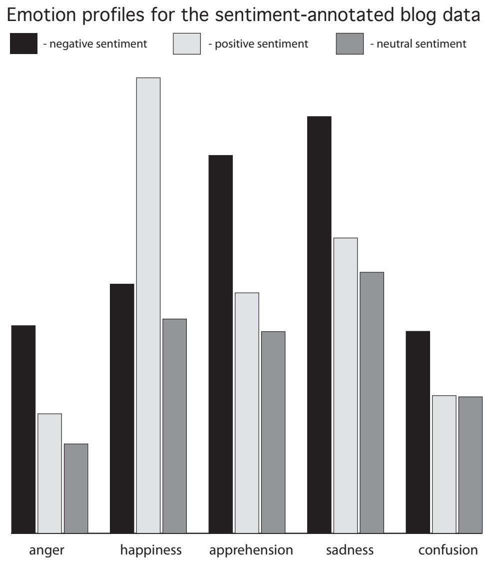

Figure 3: Emotion profiles for different sentiment categories

6. BLOG SENTENCE SENTIMENT

As a last experiment we looked at the emotion category distribution for blog sentences which were annotated with their sentiment: positive, negative, or neutral [5]. Figure 3 illustrates the cumulative emotion profiles for sentences of negative, positive, and neutral sentiment.

There is an indication that the three sentiments have clearly distinct emotion profiles. Positive sentiment sentences are highest in happiness; neutral sentences are muted in emotion in general; negative sentences have highest scores in sadness followed by apprehension. Emotion levels in general are markedly higher for negative sentences.

[5] describe that sentiment annotation schemes do not perform as well on short texts, in particular sentences show too little redundancy for most approaches. The interesting difference in emotion profiles for blog sentences of different sentiment suggests that a secondary annotation with emotion categories might add additional consistent features and help improve sentiment annotation.

7. CONCLUSION

We present here a technique to acquire emotion dictionaries from freely available lexicographic resources and a methodology to assign emotion profiles to texts based on these dictionaries. We chose to follow the emotion categorization proposed by Hall and Van de Castle and evaluated the usefulness of our techniques on dream reports that are available annotated from dreambank.net. We propose a way to bridge the difference in our emotion annotations and the gold standard and report the influence of different parameters. In particular, we suggest that *word profile* quintuples, encoding the membership degree of a word in all five emotion categories, are a particularly useful notion. Emotion words are often polysemous or indirect in isolation, but show reliable patterns in context. Thus our dictionaries, which include implicit emotion markers such as *grinning*, achieve good performance in summary dream report annotation. We also compile emotion dictionaries for nouns and verbs, not only adjectives and show that this leads to better results.

There are differences between our emotion annotation approach and the DreamBank gold standard annotation. For the gold standard, only very explicit emotion indicators are marked, the same emotion indicator for the same subject mentioned consecutively does not count (Hall/Van de Castle recommendations for annotators). Our system, in contrast, compiles a very open, large dictionary (17240 entries for SingleSeed, 16496 for HVDC-seed dictionaries) that include non-explicit and borderline emotion words.

However, the system obtains good results when we compute fuzzy *category profiles.* Fuzziness seems to be an inherent feature of emotions but the observed relative ordering and strength encoded in the category profiles seems to be stable, even on a very different text type and task, blog sentence sentiment annotation.

Emotions are an integral part of many text types and form a central role in the emerging social media, which are focused largely on sharing experiences and ideas.

The automatic analysis of texts for their emotion content is desirable for many purposes, but the exploratory research to date has not settled on standard notions. We claim that in this pioneering phase it is most important to carefully analyze data and system parameters and present our work as a step in this direction.

8. REFERENCES

[1] C. Alm, D. Roth, and R. Sproat. Emotions from text: Machine learning for text-based emotion prediction. In *Human Language Technology and Empirical Methods in Natural Language Processing*, 2005.

[2] S. Aman and S. Szpakowicz. Identifying expressions of emotion in text. In *Speech and Dialogue 10th International Conference, TSD 2007, Pilsen, Czech Republic, September 3-7, 2007*. Springer Verlag, LNCS 4629, 2007.

[3] A. Andreevskaia and S. Bergler. Senses and sentiments: Sentiment tagging of adjectives at the meaning level. In *19th Canadian Conference on Artificial Intelligence, AI'06*, LNAI 4013, 2006.

[4] A. Andreevskaia and S. Bergler. Clac and CLaC-NB: Knowledge-based and corpus-based approaches to sentiment tagging. In *SemEval-2007*, Prague, 2007.

[5] A. Andreevskaia, S. Bergler, and M. Urseanu. All blogs are not made equal: Exploring genre differences in sentiment tagging of blogs. In *International Conference on Weblogs and Social Media (ICWSM-2007)*, Boulder, CO, 2007.

[6] G. W. Domhoff and A. Schneider. Studying dream content using the archive and search engine on DreamBank.net. *Consciousness and Cognition*, 17:1238–1247, 2008.

[7] P. Ekman. Are there basic emotions? *Psychological Review*, 99:550–553, 1992.

[8] P. Ekman. An argument for basic emotions. *Cognition and Emotion*, 6:169–200, 1992.

[9] Y. Gyamfi, J. Wiebe, R. Mihalcea, and C. Akkaya. Integrating knowledge for subjectivity sense labeling. In *NAACL-HLT 2009*, 2009.

[10] C. Hall and R. Van De Castle. *The Content Analysis of Dreams*. Appleton-Century-Croft, New York, 1966.

[11] J. Hutchins. The development and use of machine translation systems and computer-based translation tools. In *International Symposium on Machine Translation and Computer Language Information Processing*, Beijing, China, 1999.

[12] D. Nadeau, C. Sabourin, J. D. Koninck, S. Matwin, and P. Turney. Automatic dream sentiment analysis. In *Workshop on Computational Aesthetics at the 21st AAAI*, Boston, MA, 2006.

[13] V. Nastase, M. Sokolova, and J. Shirabad. Do happy words sound happy? A study of the relation between form and meaning for English words expressing emotions. In *Recent Advances in Natural Language Processing (RANLP 2007)*, 2007.

[14] T. Nielsen and P. Stenstrom. What are the memory sources of dreaming? *Nature*, 437:1286–1289, 2005.

[15] V. Rubin, J. Stanton, and E. Liddy. Discerning emotions in texts. In *Exploring Attitude and Affect in Text: Theories and Applications*. AAAI Report SS-04-07, 2004.

[16] S. Somasundaran, J. Wiebe, and J. Ruppenhofer. Discourse level opinion interpretation. In *22nd Coling*, Manchester, UK, August 2008.

[17] D. Watson and A. Tellegen. Toward a consensual structure of mood. *Psychological Bulletin*, 98:219–235, 1985.

[18] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.