Exploring the Use of Linguistic Features in Sentiment Analysis

Michel Généreux[1] and Marina Santini[1]

Abstract

In this paper we describe some explorations of the potential of genre-revealing features on automatic sentiment analysis. In particular, we use a small subset of the 'linguistic facets' employed in recent experiments on automatic genre identification in combination with more traditional sentiment-revealing features on two different single-genre corpora: a corpus of English blogs and a corpus of French reviews (relectures). Although still preliminary, results show that linguistic facets might have a positive influence on sentiment analysis because 6 out of 14 facets used in the experiments are among the first 22 most important discriminative features.

1 Introduction

In this article, we present experiments where we classify texts according to sentiment criteria. More specifically, we use supervised methods to discriminate among sentiments such as happiness or anger.

Automatic sentiment analysis is being extensively studied and applied. Sentiment analysis is applied for practical aims like opinion mining (Ghose et al., 2007; Esuli and Sebastiani, 2007; Blitzer et al., 2007; Devitt and Ahmad, 2007; Ku et al., 2007; Kim and Hovy, 2007; Kobayashi et al. 2007). Recent investigations on subjective language include Mihalcea et al. (2007), Medlock and Briscoe (2007), McDonald et al. (2007), and Read et al. (2007).

While most of the work in sentiment analysis is geared towards building a lexicon of emotions (e.g. cf. Riloff and Wiebe, 2003; Yang et al., 2007; Kaji and Kitsuregawa, 2007), in the experiments that we present in this paper we introduce a small set of linguistic features called *facets* (as interpreted in Santini, 2005) and normalisation methods that are original. In this way, we try to go beyond well-established approaches relying on the usual affect-bearing words and binary/frequency counts. Although still preliminary, our results allow us to gain some insights into the discriminative power of different kinds of features.

We have organised the article as follows: section 2 presents the learning algorithm, features and normalisation methods. The experiments are presented in section 3. We discuss and conclude in sections 4 and 5 respectively.

[1] NLTG, University of Brighton
e-mail: M.Genereux@brighton.ac.uk, MarinaSantini.MS@gmail.com

2 Methodology

In the experiments, we use a state-of-the-art classifier, the Support Vector Machine (SVM) (Joachims, 1997) as implemented in the Weka package[2] (Witten and Frank, 2005). SVM has proved successful in classifying opinion documents, including style (Diederich et al., 2000), and has the advantage of being able to accommodate a large number of features. In the training phase, the algorithm builds an hyperplan that separates maximally positive and negative examples. The task of classifying new examples amounts to finding on which side of the hyperplan they belong. Usually binary, the Weka implementation of SVM can of handle, instead, more than two classes.

2. 1 Features

Our sets of features can be divided in three groups: grammatical categories, such as adjectives, nouns, verbs and adverbs, linguistic facets, and terms with an emotional undertone (WordNet-Affect and Big-Six):

Group 1 - Adjectives, Nouns, Verbs and Adverbs. These grammatical categories have the capacity to express emotion or subjectivity (Turney, 2002).

Group 2 - Linguistic Facets. In research aiming at classifying documents according to their genre (Santini, 2007), these facets has proved very useful. The list of the facets used in this paper is provided online.

Group 3 - Terms with emotional undertone. These terms have been classified by others as having a particular affective orientation. WordNet-Affect (Strapparava and Valitutti, 2004) is an affective extension of WordNet[3]. These terms are divided as being positives, negatives and neutral. The Big-Six emotions are based on studies in psychology (Ekman, 1972) and reorganise WordNet-Affect according to the following six basic emotions: anger, joy, sadness, disgust, fear and surprise. The list of the terms used in this paper is available online.

Each term that belongs to one of the three groups and that qualifies as a feature is assigned a part-of-speech (POS) tag using Tree-Tagger[4]. To avoid counting negated terms ("You are not a nice person!"), all terms after the particle *not* (or *ne* in French) and the end of the phrase were not counted.

2.2 The facets

Facets, as intended in (Santini, 2007), are macro-features that can be "functionally-interpreted". For instance, the first person facet includes first person pronouns, singular and plural. The first person facet indicates that the communication context is

---

[2] We used the algorithm SMO with the following parameters: -C 1.0 -E 1.0 -G 0.01 -A 250007 -L 0.0010 -P 1.0E-12 -N 0 -M -V -1 -W 1.
[3] Available at < http://wordnet.princeton.edu/>.
[4] Available at <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>.

related to the text producer, i.e. the speaker or the writer. A high frequency of first person facets in a text signals an impressionistic or subjective stance of the text producer. While in most text classification tasks, features are used individually without any further interpretation, with facets the aim is to interpret a particular stance in communication. For example, a high frequency of first person facets is assumed to be found in ARGUMENTATIVE texts, like COMMENTS and OPINIONS. (Santini, 2007) used 100 facets, divided in several subtypes (e.g. functional cues, syntactic patterns or HTML facets). In the experiment reported in this paper we used only fourteen facets. While many of the 100 facets used in (Santini, 2007) were grammatical facets, for the experiments described in this paper we selected a small subset of facets as a preliminary exploration of their potential in sentiment classification.

The use of facets introduces two innovations in sentiment classification. The first is related to the grammatical nature of facets. While most of the research on affect (opinion classification, sentiment analysis, affect detection, and so on) relies on affect-bearing words (Hatzivassiloglou and Wiebe, 2000; Riloff and Wiebe, 2003), i.e. mainly adjectives and adverbs, facets focus on the use of grammatical cues. For instance, we noticed that there was a variation in the use of personal pronouns across the different types of French reviews. For instance, the reviews of video games often refer directly to the players, using second person pronouns, as in “Vous incarnez un guerrier [...]”, while parliamentary debates often make use of first person pronouns to emphasise the view expressed by the speaker, as in “Nous avons passé des dizaines d’heures en juillet et en août [...] à analyser ce projet. Mon sentiment est qu’il répond bel et bien à l’évolution du monde [...]”. Similarly, activity verbs[5] appear more common in the reviews of video games, while paper reviews seem more characterised by communication verbs and mental verbs. Similar observations are valid also for English blogs. For example, blogs with happy mood seems to be more characterized by activity verbs than blogs in angry mood. We also hypothesised that the frequencies of occurrence of nominals and predicates (usually revealing the difference between written and spoken texts) could vary across the different sentiments.

The second innovation refer to the composite nature of facets. In other words, facets are macro-features, i.e. each facet is made of a number of individual features that share a similar semantic and textual interpretation. We defined facets as “functionally-interpreted” features because they help interpret and reconstruct the context of communication through linguistic cues. The use of macro-features has a practical benefit. In fact, facets reduce the risk of over-fitting, a phenomenon that usually occurs when a statistical model has too many attributes.

2.3 Methods used to normalise feature count

A textual document is modelled as a vector of feature counts. The simplest approach in counting features is the **binary** approach, where each feature is given the value zero if it does not appear in the document, or one if it does appear at least once. Another basic counting approach is based on **frequency**, where each feature is given a value representing the exact number of times it appears in a document, often normalised to a document of fixed length (in our case 1000 words). In our experiments, we have also

[5] The semantic classification of verbs into seven categories is taken from (Biber et al., 1999).

considered additional ways of normalising the counts, by factoring counts with at least one of the following elements:

*** idf -** (*Inverse Document Frequency*). This factor evaluates the importance of a term *i*, the assumption being that the importance of a term lowers as it appears in a growing proportion of documents in the corpus. The exact formula to compute IDF is:

$$idf_i = log\frac{D}{d_i}$$

where D is the total number of documents in the corpus and d_i is the number of documents in which the term *i* appears.

*** so-pmi-ir -** From Semantic Orientation - Pointwise Mutual Information - Information Retrieval. This strategy allows us to compute the semantic orientation (SO) of terms or full texts by computing their degree of association (A) with a list of positive or negative words (P and N). This approach was used by (Turney, 2002) to classify terms according to their *sentimentality*, which can be more or less positive or negative. This method, called SO-A, can be expressed in the following formal terms:

$$\sum_{p}^{P} A(term, p) - \sum_{n}^{N} A(term, n)$$

Note that the quantity of positive terms P must be equal to the quantity of negative terms N. To compute SO-A, (Turney, 2002) uses the notion of PMI-IR. PMI (Church and Hanks, 1989) between two terms is defined as:

$$\log_2 \frac{prob(term_1\ is\ around\ term_2)}{prob(term_1) * prob(term_2)}$$

PMI is positive when two terms tend to co-occur and negative otherwise. PMI-IR comes from Information Retrieval (IR), where multiple occurrences of a single term in a document is counted as one single occurrence; according to (Turney, 2002), this appears to give a measure more resistant to noise. By computing probabilities using the number of documents (*nd*) extracted as in IR, this yields, for PMI-IR:

$$\log_n \frac{D * (nd(term_1\ AROUND\ term_2) + 1/D)}{(nd(term_1) + 1) * (nd(term_2) + 1)}$$

where D is the total number of documents in the corpus. The positive paradigm words P employed *were good, nice, excellent, positive, fortunate, correct, superior,* and the negative ones N were bad *nasty, poor, negative, unfortunate, wrong, inferior*. Smoothing values (1/D and 1) have been selected so that PMI-IR is zero for terms not in the corpus, a term is *around* another term if it is no more than twenty words apart and $\log_2$ has been replaced by $\log_n$, since the natural logarithm is more common in the literature and this does make any difference for the algorithm. We have used the (English) Waterloo[6] corpus to compute the probabilities, this corpus has approximately forty-six millions pages. For the experiment on French documents, each term was translated and searched in the Waterloo corpus; terms not in the dictionary has been assigned a neutral value **so-pmi-ir** of zero

*** so-sim -** This time we use a measure of similarity between two terms obtained throughWordNet to compute SO-A. This approach is similar to (Kamps and Marx, 2002), where similarity is computed by using a count of the number of edges between two terms in WordNet, a technique similar the computation made to know the genetic relation between two persons through their common ancestors (Budanitsky and Hirst, 2001). Only nouns, verbs, adjectives and adverbs can have a semantic relation in WordNet. The positive paradigm words (nouns and verbs) P used are: *good* and *better*, *win* and *win*, *excellence* and *excel*, *superiority* and *surpass*, while the negative terms N are *bad* and *worsen*, *lost* and *loose*, *poverty* and *impoverish*, *negativity* and *negate*. For the French corpus (see section 3.1), each term which does not appear in the corpus has been assigned a neutral value **so-sim** of zero. The PERL package WordNet::Similarity[7] was used to compute so-sim.

*** sen -** (Esuli and Sebastiani, 2006) provides a valuable resource in the form of *SentiWordNet* in which each *synset s* is given three numerical values describing the degree of objectivity and subjectivity (positive or negative). The three values must add up to one, which means that each term can possess, to a different degree, more than one property at the same time. A unique measure of subjectivity can therefore be obtained for each term listed in SentiWordNet. The approach used to develop SentiWordNet is based on a quantitative analysis of the glosses associated with each *synset* by training a committee of classifiers for the three classes (objective, positive and negative) (Esuli and Sebastiani, 2005). The value attributed to each class corresponds to the proportion of classifiers that have selected this class in particular. SentiWordNet has been used favourably on the *General Inquirer* (Stone et al., 1966). For the French corpus, each term which does not appear in the dictionary has been assigned a neutral value **sen** of zero.

*** hum -** A list of terms annotated manually as being positive or negative (Turney, 2002). The list of the terms used in this paper are available online. For the French corpus, each term which does not appear in the dictionary has been assigned a neutral value **hum** of zero.

*** binf -** Hybrid normalisation, this allows a distinction to be made between group 1 (normalised as *binary*) and the groups 2 and 3 (normalised as *frequency*).

---

[6] Available at <http://canolal.uwateloo.ca/>.
[7] Avalilable at <http://www.d.umn.edu/~tpedersen/similarity.html>.

3 Experiments

3.1 Experiment 1: Exploring Sets of Features and Normalisation Methods

In this experiment we looked at different combinations of features and normalisation methods and evaluated classification accuracy on two different corpora, an corpus of English blogs and a corpus of French reviews (relectures).

The English corpus contains 8000 self-annotated blog posts[8] into one of the four following classes: **Class 1** Obstructive, such as *tense* and *angry*; **Class 2** Low Power/Control, such *worried* and *lonely*; **Class 3** Conducive, such as *calm* and *confident*; and **Class 4** High Power/Control, such as *happy* and *aroused*. These classes represent the four ubiquitous affective classes stemming from research in psychology. We have tried to avoid ambiguous moods such as *confident*, but we have kept moods such as *impressed* belonging to the grey zone. The grey zone is an area where there is no clear-cut between the four classes. Figure 1 shows the class distribution based on Osgood semantic differential (Osgood *et al.*, 1957) as reported in (Généreux and Evans, 2006).

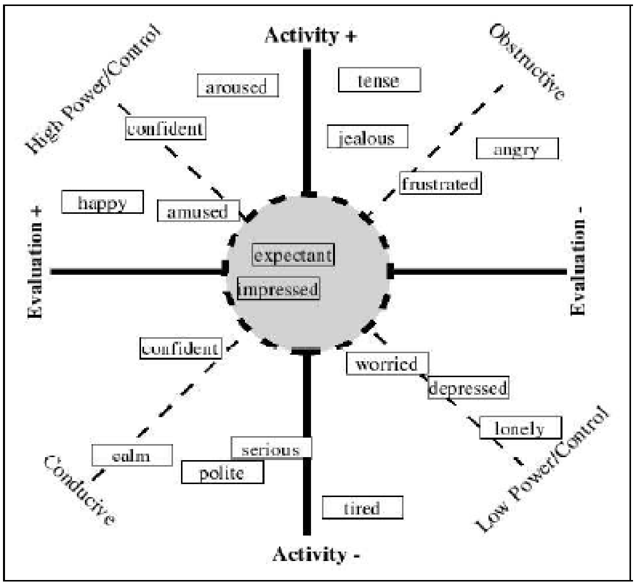

Figure 1: A two-factor structure of affect

In this experiment, we used only 800 blogs equally divided among the four classes as training corpus.

The French corpus is a collection of reviews (relectures) used in DEFT07[9] by (Généreux and Santini, 2007). This corpus is made up of 227 negative texts, 278 neutral texts and 376 positive texts. Because resources for French are not as extensive as for English, a bilingual dictionary was created so that part of the resources available for English could be mapped to French. This dictionary is composed of 1244 terms translated manually from groups 2 and 3.

[8] As collected during Oct-Nov 2005 from <http://www.livejournal.com/>.
[9] See <http://deft07.limsi.fr>.

We have devised a set of seventeen different feature combinations: *1, 2, 3, 4, 5, 6, 7, A, B, C, D, E, F, G, H, I* and *J*, as shown in Table 1. For example, combination *J* is made up of the 500 most frequent adjectives and adverbs as counted in the training data, as well as the three sub-groups of WordNet-Affect features.

Features/Normalisation	binary	frequency	idf	so-pmi-ir	so-sim	sen	hum	binf
Group 1 Adjectives	1ABHI	2 F	3G	4	5	6	7	J
Group 1 Nouns	1 B	2	3	4	5	6	7	
Group 1 Verbs	1 B I	2	3	4	5	6	7	
Group 1 Adverbs	HI							J
Group 2 Facets (14)	1	2C	3	4	5	6	7	
Group 3 WN-Affect (3)	1	2 D	3	4	5	6	7	J
Group 3 Big-Six (6)	1	2 E	3	4	5	6	7	

Table 1: 17 feature and normalisation combinations

We have used the seventeen combinations to classify both corpora (blogs and relectures), each time using cross-validation (ten-fold). The results are shown in Table 2, where they have been sorted in descending order by the average of accuracy in both corpora. The results for both corpora are positively correlated (Pearson coefficient: 0.12), i.e. they both vary in the same direction: when one goes up/down, the other tends to go up/down too. The baseline is 25 percent for blogs (4 classes) and 33 percent for relectures (3 classes).

COMB->	J	A	H	3	F	I	1	2	4	B	E	C	G	6	7	D	5
Blogs	35	35	33	37	37	36	36	37	36	34	35	36	34	34	34	31	31
Relectures	50	49	51	45	45	46	46	44	43	44	43	41	43	42	42	43	42
Average	43	42	42	41	41	41	41	41	40	39	39	39	39	38	38	37	37

Table 2: Classification accuracy (percentage) of both corpora using the different combinations

3.2 Experiment 2: Classifying Blog Posts

In the second experiment, we classified the whole blog corpus (8000 English blog posts, i.e. two thousands per class) into one of the four classes (Obstructive, Low Power/Control, Conducive, and High Power/Control).

We used Information Gain (IG) to explore the discriminative power of each feature. Table 3 shows the ranked features with IG *not equal to zero*. This list includes fifty-two features. In this experiment we used only these fifty-two features to classify the 8000 blog posts in four classes. Interestingly, six features from group 2, and only six features from group 3 are included in the list.

The classification results are presented in Table 4. The results are sorted in descending order of accuracy for each combination of groups of features. The best accuracy is obtained when all three groups (i.e. the fifty-two features in Table 3) are employed. The normalisation method was kept constant to *binf*. Each line of the confusion matrix represents the distribution of blogs among the four classes. For example, using combination G1, the two thousands blog posts from class one (obstructive) have been classified as follows: 1290 as class one, 186 as class two (low power/control), 117 as class three (conducive) and 407 as class four (high power/control).

Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature
1	joy_emotion	14	activity_verbs	27	want	40	pictures
2	positive_emotion	15	first_person	28	interesting	41	friday
3	anger_emotion	16	cute	29	favorite	42	sick
4	negative_emotion	17	depressed	30	fear_emotion	43	alone
5	past_facet	18	shit	31	show	44	miss
6	great	19	beautiful	32	issue	45	vanilla
7	nominal_facet	20	wrong	33	boy	46	knowledge
8	happy	21	sleep	34	really	47	amused
9	fuck	22	passive_facet	35	never	48	supplies
10	sadness_emotion	23	cool	36	shitty	49	racist
11	mental_verbs	24	pisses	37	frustrated	50	shine
12	rant	25	something	38	just	51	fuckers
13	met	26	movie	39	shopping	52	suspected

Table 3: Feature selection: the fifty-two top-features with Information Gain not equal to zero

Combination	Cross-Validation 5-fold				Confusion Matrix				Class
	Accuracy	Precision	Recall	F-score	1	2	3	4	
G1+G2+G3	34.9%	0.357	0.579	0.441	1157	198	123	522	1
		0.333	0.155	0.211	848	309	169	674	2
		0.350	0.133	0.193	664	256	266	814	3
		0.345	0.531	0.418	574	164	201	1061	4
G1+G3	34.4%	0.350	0.598	0.441	1196	193	121	490	1
		0.331	0.151	0.207	876	302	177	645	2
		0.374	0.135	0.194	705	254	270	771	3
		0.340	0.492	0.402	644	164	209	983	4
G1	32.8%	0.321	0.645	0.428	1290	186	117	407	1
		0.333	0.154	0.210	1000	307	169	524	2
		0.336	0.119	0.175	873	265	237	625	3
		0.338	0.397	0.365	860	164	183	793	4
G2+G3	32.4%	0.315	0.698	0.434	1396	55	76	473	1
		0.320	0.036	0.065	1192	72	102	634	2
		0.349	0.072	0.119	960	52	144	844	3
		0.334	0.489	0.397	885	46	91	978	4
G1+G2	32.2%	0.329	0.596	0.424	1192	198	131	479	1
		0.330	0.154	0.210	918	307	174	601	2
		0.327	0.121	0.176	781	263	241	715	3
		0.338	0.458	0.389	731	162	191	916	4
G3	31.1%	0.308	0.719	0.431	1437	53	15	495	1
		0.286	0.033	0.059	1228	66	34	672	2
		0.395	0.025	0.046	1007	64	49	880	3
		0.314	0.468	0.376	990	48	26	936	4
G2	26.4%	0.257	0.924	0.402	1847	19	71	63	1
		0.247	0.009	0.017	1794	18	81	107	2
		0.327	0.057	0.096	1773	22	113	92	3
		0.332	0.065	0.109	1775	14	81	130	4

Table 4: Classification results on eight thousands blog posts (baseline = 25 percent)

4 Discussion

According to the averages shown in Table 2, the best accuracies are returned by traditional features and traditional normalization, i.e. unigrams, adjectives and adverbs with binary normalisation. A natural approach to consider for improvement is the use of bigrams.

In general, accuracies on English texts is about 10 percent above the baseline, while accuracies on French texts is about 15 percent above the baseline. One explanation to this state of affairs is that blog posts are by nature extremely noisy and use a language which is very diverse and difficult to modelise. Additionaly, we did not count emoticons (cf. Yang et al. 2007), thus subtracting some affective content from classification.

Classification on French texts is higher because English texts are classified using a set of four classes which is still not fully validated (Généreux and Evans, 2006), while the French relectures are classified along a more “natural” set of classes: positive, negative and neutral. However, a small positive correlation between the two corpora gives some evidence that the combinations of features and normalising factors have, to some limited extent, similar impacts across language and corpora.

Interestingly, 6 out of 14 facets are among 22 most important discriminative features (Table 2). They then appear potentially interesting for sentiment discrimination. We defer to future research a more comprehensive assessment of the impact of these features as sentiment-revealing attributes.

5 Conclusion

In this paper we have explored the discriminative power of some features for sentiment analysis. Our approach was to see if a well-known supervised learning method (SVM) could be more accurate when using features other than unigrams taken from relevant grammatical categories (group 1): those additional features were defined as linguistic facets, group 2, and affect-bearing words, group 3. We have also experimented with a few normalising factors other than binary or frequency count.
Apparently, the most productive approach seems to be based on binary count for adjectives and adverbs and frequency count of the WordNet-Affect terms. However, although still preliminary, results also show that linguistic facets might have a positive influence on sentiment analysis because 6 out of 14 facets used in experiment 2 are among the first 22 most important discriminative features.

In conclusion, the overall feeling is that there is still a lot to explore to gain insights into the classification of sentiment categories, especially for web genres like blogs, where difficulties are caused by innovative language, noisy format (e.g. typos and abbreviations), and graphical ways of expressing emotions (e.g. emoticons).

References

Biber D., Johansson S., Leech G., Conrad S. and Finegan E. (1999). *Longman Grammar of Spoken and Written English*. Longman, Harlow.

Blitzer J., Dredze M. and Pereira F. (2007). Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. *Proceedings* *of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague.

Budanitsky A. and Hirst G. (2001). Semantic distance in wordnet: an experimental, application-oriented evaluation of five measures. In *NAACL 2001 Workshop on WordNet and Other Lexical Resources*, Pittsburgh.

Church K. W. and Hanks P. (1989). Word association norms, mutual information and lexicography. In *27th annual Conf. of the ACL*, p. 76–83: New Brunswick, NJ:ACL.

Devitt A. and Ahmad K. (2007). Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics,* Prague.

Diederich J., Kindermann J., Leopold E. and Paass G. (2000). Authorship attribution with support vector machines.

Ekman P. (1972). Universal and cultural differences in facial expression of emotion. In J. Cole, Ed., *Nebraska Symposium on Motivation*, p. 207–282, Lincoln: University of Nebraska Press.

Esuli A. and Sebastiani F. (2005). Determining the semantic orientation of terms through gloss analysis. In *Proceedings of CIKM-05, 14th ACM International Conference on Information and Knowledge Management*, Bremen.

Esuli A. and Sebastiani F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining.

Esuli A. and Sebastiani F. (2007). PageRanking WordNet Synsets: An Application to Opinion Mining. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague.

Génèreux M. and Evans R. (2006). Towards a validated model for affective classification of texts. In *Sentiment and Subjectivity in Text*, p. 55–62, Sydney, Australia.

Génèreux M. and Santini M. (2007). Classification de textes français subjectifs. In *Proceedings of DEFT'07*.

Ghose A., Ipeirotis P. and Sundararajan A. (2007). Opinion Mining using Econometrics: A Case Study on Reputation Systems. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague.

Hatzivassiloglou V. and Wiebe J. (2000). Effects of adjective orientation and gradability on sentence subjectivity. In *International Conference on Computational Linguistics*.

Joachims T. (1997). *Text Categorization With Support Vector Machines: Learning With Many Relevant Features.* Rapport Interne Ls8-Report 23, Universität Dortmund. Ls Viii-Report.

Kaji N. and Kitsuregawa M. (2007). Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (EMNLP-CoNLL), Prague.

Kamps J. and Marx M. (2002). Words With Attitude. In *1st International Conference On Global Wordnet, Mysore, India*.

Kim S.-M. and Hovy E. (2007). Crystal: Analyzing Predictive Opinions on the Web. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (EMNLP-CoNLL), Prague.

Kobayashi N., Inui K. and Matsumoto Y. (2007). Extracting Aspect-Evaluation and Aspect-Of Relations in Opinion Mining. *Proceedings of the 2007 Joint*

Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Prague.

Ku L.-W., Lo Y.-S. and Che H.-H. (2007). Test Collection Selection and Gold Standard Generation for a Multiply-Annotated Opinion Corpus. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, Prague.

McDonald R., Hannan K., Neylon T., Wells M. and Reynar J. (2007). Structured Models for Fine-to-Coarse Sentiment Analysis. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague.

Medlock B. and Briscoe T. (2007). Weakly Supervised Learning for Hedge Classification in Scientific Literature. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague.

Mihalcea R., Banea C. and Wiebe J. (2007) Learning Multilingual Subjective Language via Cross-Lingual Projections. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague.

Osgood C. E., Suci G. J. and Tannenbaum P. H. (1957). *The Measurement Of Meaning*. University Of Illinois.

Read J., Hope D. and Carroll J. (2007), Annotating Expressions of Appraisal in English. *Proceedings of the Linguistic Annotation Workshop*, held in conjunction with the 45th Annual Meeting of the Association of Computational Linguistics, Prague.

Riloff E. and Wiebe J. (2003). Learning Extraction Patterns For Subjective Expressions. In *Conference On Empirical Methods In Natural Language Processing (Emnlp-03). Acl Sigdat.*

Santini M. (2005). *Linguistic Facets For Genre and Text Type Identification: A Description Of Linguistically-Motivated Features*. Technical Report ITRI-05-02, University of Brighton.

Santini M. (2007). *Automatic Identification Of Genre In Web Pages*. Phd Thesis, University Of Brighton.

Stone P. J., Dunphy D. C., Smith M. S. and Ogilvie D. M. (1966). *The General Inquirer: A Computerapproach To Content Analysis*. Mit Press. Mit Press.

Strapparava C. and Valitutti A. (2004). Wordnet-Affect: An Affective Extension Of Wordnet. In *The 4th International Conference On Language Resources and Evaluation (Lrec 2004)*, P. 1083–86, Lisbon.

Turney P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *40th Annual Meeting of the Association for Computational Linguistics, Philadelphia*.

Witten I. and Frank E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, Amsterdam (The Netherlands). 2nd edition.

Yang C., Lin K. H.-Y. and Chen H.-H. (2007). Building Emotion Lexicon from Weblog Corpora. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, Prague.