A Multi-level Annotation Model for Fine-grained Opinion Detection in German Blog Comments

Bianka Trevisan
Textlinguistics/ Technical Communication
RWTH Aachen University
b.trevisan@tk.rwth-aachen.de

Melanie Neunerdt
Institute for Theoretical Information Technology
RWTH Aachen University
neunerdt@ti.rwth-aachen.de

Eva-Maria Jakobs
Textlinguistics/ Technical Communication
RWTH Aachen University
e.m.jakobs@tk.rwth-aachen.de

Abstract

Subject of this paper is a fine-grained multi-level annotation model to enhance opinion detection in German blog comments. Up to now, only little research deals with the fine-grained analysis of evaluative expressions in German blog comments. Therefore, we suggest a multi-level annotation model where different linguistic means as well as linguistic peculiarities of users' formulation and evaluation styles in blog comments are considered. The model is intended as a basic scheme for the annotation of evaluative expressions in blog data. This annotation provides suitable features for implementing methods to automatically detect user opinions in blog comments.

1 Introduction

Evaluations are complex linguistic acts that are realized through the use of linguistic means. In social media applications and blogs respectively, evaluations are, moreover, expressed by colloquial or grapho-stylistic means such as emoticons, multiple punctuations or Internet lingo (Schlobinski and Siever, 2005; Trevisan and Jakobs, 2010, 2012; Neunerdt et al., 2012). So far these text type-specific evaluative means are not considered in annotation models as most approaches focusing on strongly structured data such as newspaper texts where colloquial expressions are not common, generally. Therefore, we suggest a fine-grained multi-level annotation model, which consists of different annotation levels with various purposes of annotation, e.g., annotation of polarity vs. annotation of syntactical function. The model serves for the description of user-, context-, topic- and blog comment-related characteristics of opinion-indicating expressions in blog comments. The aim of the model is to determine how different users express themselves judgmental on a specific topic in blog comments, i.e., which linguistic means (typographic, rhetoric, syntactical etc.) they use in what combination and for what purpose, e.g., interaction signs at the end of a sentence to mark ironic expressions. We refer to this phenomenon as *evaluation pattern*. In this paper, the annotation model is presented using the example of emoticons.

The approach generates additional values for different disciplines. In computer sciences, for instance, the multi-level annotation model can serve as instrument for automatic opinion detection, where information on each annotation level serves as separate feature for classification. From the perspective of communication science, the fine-grained annotation model can be used for sentiment analysis, e.g., analysis of the grammatical function of an emoticon vs. its stylistic function. Up to now, the model is exemplarily used in acceptance research for the identification and analysis of opinion-indicating statements about *mobile communication systems (MCS)*.

The paper is structured as follows. Section 2 outlines characteristics of evaluative expressions and formulation styles in German blog comments. In section 3, related work on sentiment analysis and linguistic annotation are presented. Subject of Section 4 is the methodological approach and the developed multi-level annotation model; the different annotation levels and the related annotation schemes are thoroughly described. Results focusing on the validity and the identification of evaluation patterns are shown in

section 5. Finally, a summary and an outlook on future work are given in section 6.

2 Linguistic Means of Evaluation

In evaluating, writers use specific evaluative means that can be positive or negative connoted lexemes, specific syntactic structures (e.g. interrogative sentences), idioms, rhetorical means (e.g. metaphor) or non-verbal signs (e.g. emoticons) (Schlobinski and Siever, 2005). As we know from other contexts, which evaluative mean is chosen, depends on the function and purpose of the evaluation, group membership, the evaluating individual, the respective situation and the used text type (Sandig, 2003).

Moreover, the language in blogs has a low standardization degree, i.e. bloggers have specific modes of expression or formulation styles that are not characterized by recurring and recognizable structural features. This specific expression style can lead to processing errors, and, thus, impede the analysis accuracy of evaluative expressions in blog comments (Neunerdt et al., 2012).

Up to now, there is little research that deals with formulation and evaluation styles or patterns in blog comments. Rather, formulation styles of blog texts in general are described. First and foremost, it is evident that the language or vocabulary of bloggers is ordinary or colloquial (e.g. geil "cool") and, thereby, bears analogy to spoken language (Schlobinski and Siever, 2005; Thorleuchter et al., 2010). Typically, interaction words (e.g. IMO "in my opinion") or inflective expressions (e.g. seufz "sigh") are used as means of emphasis. Syntactic characteristics are simple word order and unclear sentence boundaries, e.g., because of missing sentence-terminating punctuations (Missen et al., 2009).

Regarding grapho-stylistic means, emoticons are characteristic for blog texts. They perform the function of conversation particles, interjections, prosody, facial expressions and gestures and have, in conclusion, an expressive and evaluative function (Mishne, 2005). Asterisks play a similar role; they are used to mark emotive and evaluative speech acts, e.g., in ironic expressions (e.g. *grins* "*smile*"). Further typographical means of blog communication are word-internal capitalizations (e.g. CrazyChicks), iteration of letters (e.g. jaaaaa "yeeees") and italic (e.g. Das ist eine *große* Lüge. "That's a *big* lie."). Regarding stylistic means, most likely redemptions, e.g., of word endings occur (e.g. 'nen "einen/one") (Schlobinski and Siever, 2005). Other stylistic features are dialectic expressions (e.g. Mädl "girl"), foreign-language-switch phenomena (e.g. Killerapplikation "killer application"), increased use of rhetorical phrases and idioms (e.g. Alle Wege führen nach Rom. "All roads lead to Rome.") as well as ellipsis.

In total, on all linguistic levels transfer phenomena from written to spoken language are evident, which also impact the processing of evaluative expressions. Thus, an annotation scheme by which evaluative means and expressions can be analyzed holistically must consider the different evaluative means and relate relevant linguistic information to each other.

3 Related Work

The exploitation of opinions from unstructured content such as blog comments is a challenging task, because evaluative expressions are vague, have a high degree of semantic variability (Balahur and Montoyo, 2010), and, as shown in section 2, text type-specific formulation styles occur. How well users evaluations are captured, thus, depends crucially on the analysis approach. Up to date, there is little research done regarding the analysis of evaluative expressions in German blog comments. In German research, mainly annotation schemes for diachronic corpora such as forum and chat data are developed (Storrer, 2007; Luckhardt, 2009; Beißwenger, 2008, 2009, 2010; Beißwenger et al., 2012). The following section focuses established approaches in opinion detection and annotation of complex tasks.

3.1 Opinion Annotation

Depending on corpus, text type (written vs. spoken language), language, annotation task and annotation goal, the extent and kind of annotation or coding models varies. In our case, we annotate evaluative expressions in German blog comments, with the aim of identifying recurring evaluation patterns.

Annotation approaches and models for opinion detection and sentiment analysis can be distinguished according to their degree of granularity. Using *coarse-grained annotation schemes*, texts are annotated on the document

and sentence level with the aim of making general statements about the polarity, e.g., as it is of interest for product reviews. This annotation type is normally used in marketing research using shallow text-processing tools (e.g. PASW Modeler, evaluated in Trevisan and Jakobs, 2012). However, coarse-grained approaches possess the advantage that they provide a higher tagging or annotation accuracy compared to fine-grained annotations. Giesbrecht et al. (2009) have already shown this for part-of-speech (POS) tagging of German mixed corpora, Gimpel et al. (2011) reached a better annotator agreement for the tagging of English tweets.

So far, several coarse-grained approaches have been developed. Aman and Szpakowicz (2007) proposed a coarse-grained annotation scheme for English blog posts. According to the approach, emotional expressions are being identified in blog posts and assigned to the six basic emotions happiness, sadness, anger, disgust, surprise and fear by Ekman (1993). Every sentence, in which one or more related keyword occurs, is manually assigned to a basic emotion and classified by polarity. Strapparava and Mihalcea (2007) and Balahur and Montoyo (2010) developed a similar approach. However, according to Stoyanov and Cardie (2008) with a coarse-grained annotation, only a distinction between subjective and objective statements is possible, but they cannot be related to the object of evaluation.

In contrast, *fine-grained annotation schemes* serve for the annotation on the phrase and clause level or below. Fine-grained annotations are especially required for scientific purposes. In acceptance research, for instance, fine-grained annotations are essential for identifying evaluated components and properties of large-scale technologies such as *mobile communication systems (MCS)* (Trevisan and Jakobs, 2010, 2012).

In recent years, a number of fine-grained annotation approaches has been developed. Initialy, Wiebe et al. (2005) and Wilson (2008) provided fine-grained annotation approaches. In the further course, Stoyanov and Cardie (2008) pro-posed a topic annotation model that serves for the identification of opinion topics in text corpora. Annotated are six evaluation components: opinion expression, source (opinion holder), polarity, topic, topic span and target span. The inter-annotator results show that the fine-grained annotation system provides reliable results. Remus und Hänig (2011) present the *Polarity Composition Model (PCM)*, which is a two-level structure, where evaluations are analyzed on the word- and phrase-level. The authors draw the conclusion that evaluations are manifested through negations, positive and negative reinforces and the formal construct within the phrase; the world-level polarity analysis is carried out with recourse to the German-language polarity dictionary *SentiWS* (see also Remus et al., 2010). Lately, Fink et al. (2011) published their fine-grained annotation approach for sentiment analysis. Thereby, they identify sentiment-bearing sentences to spot sentiment targets and their valence.

Fine-grained and, especially, coarse-grained annotations schemes are used for qualitative and automatic text analysis methods. However, these annotation approaches do not allow the investigation of specific linguistic analysis purposes and furthermore do not provide a well-separated feature space for automatic opinion detection methods as multi-level annotation schemes do.

3.2 Annotation Models

Each token fulfills different grammatical functions, which are also relevant for the constitution of evaluations. In usual annotation schemes, these different information are not separated from each other. Queries and studies related to one type of linguistic mean are therefore difficult to perform. In multi-level annotations systems, several information such as different linguistic or evaluative means can be assigned independently to a single token or sequence of tokens, e.g., the morpho-syntactic function of a token vs. its semantic meaning (Lüdeling et al., 2005). Such multi-level annotations systems are commonly used for the annotation of learner corpora and transcripts (Lüdeling, 2011). Regarding the annotation of evaluative expressions in blog comments, multi-level annotation systems provide several advantages compared to flat or tabular annotation models, such as in TreeTagger used (Lüdeling et al., 2005; Dipper, 2005):

- Level-specific annotation standards can be applied.
- The number and categories of annotation levels is expandable and modifiable dur-

ing the editing process, e.g., adding or deleting of an annotation level.

- Token columns are mergeable and separable again, e.g., in case of multi-word expressions.
- Characteristic features of the text data can be considered, e.g., in case of written texts providing a linear mapping.

Adapting this approach, evaluative expressions in blog comments can be analyzed holistically. However, standards for processing evaluative expressions in German blog comments are mainly missing, such as a multi-level annotation scheme and text type-specific annotation guidelines for multi-level architectures[1] (Balahur and Montoyo, 2010; Lüdeling, 2011; Clematide et al., 2012), which are part of the proposed model.

4 Approach

The aim is to develop a multi-level annotation scheme for the identification of evaluative expressions in blog comments. In the following, the process of model development is described.

4.1 Corpus

As an exemplary corpus, a topic-specific German blog dealing with mobile communication systems (MCS) is selected and blog comments from two years semi-automatically collected (t=2008, 2009) (Trevisan and Jakobs, 2012). In total, the corpus contains 12,888,453 tokens and 160,034 blog comments[2]. The collected blog comment corpus was bowdlerized: Enclosed website elements, e.g., menu and navigation elements (*anchor texts*), which do not belong to the blog comment content, have been deleted. Blog comments and their contextual metadata, such as the bloggers name and the date and time of publication, are stored in a database for further analysis.

4.2 Tokenization

Tokenization is a fundamental initializing step to divide the input text into coherent units (tokens), in which each token is either a word or something else, e.g., a punctuation mark. The resulting tokens serve as basic information for designing the annotation scheme. Annotation labels must be determined according to all feasible tokens. Blog comments pose a special challenge to the task of tokenization due to the usage of non-standard language, including emoticons (e.g. :-), ;)) and multiple punctuation (e.g. ???, ?!!). Such irregularities are not considered in standard tokenizers, provided with POS taggers, e.g. TreeTagger and Standford Tagger, which are developed on well-structured data such as newspaper texts. Hence, tokenization and POS tagging results based on blog comments suffer from high error rates (Neunerdt et al., 2012).

POS tagging is the initial level in multi-level annotation; therefore we develop a tokenizer particularly referring to this annotation level. A rule-based tokenizer is developed, which is adapted to the language in blog comments. By means of regular expressions text type-specific expressions, e.g., URLs, multiple punctuations and emoticons are detected as coherent tokens. Furthermore, the tokenizer treats text type-specific writing styles, like short forms, e.g., *geht's "it works"*, *gibts "there's"*, filenames, e.g. *test.jpg*, interaction words, e.g., **lol**, numbers, e.g., *3. November* and so on. We design the tokenization rules with respect to the desired annotation scheme, e.g., *geht's* is seperated into two tokens *geht* and *'s*. After successful tokenization the text can be annotated on all annotation levels.

4.3 POS Tagging

Initially, evaluation patterns emerge on the POS level, e.g., word orders of noun phrases [ADJA NN – *useful application*] or comparatives [ADJD KOKOM – *better than*][3]. Therefore, blog comments of the selected corpus are automatically labeled with POS tags by means of the TreeTagger. Instead of the provided tokenizer, we use the developed blog comment tokenizer.

TreeTagger is a statistical tool for the annotation of text data with POS tags, according to the Stuttgart-Tübingen Tagset (STTS), and lemma information, according to a special lexicon

[1] This point requires special consideration to ensure sustainability and reusability of annotated data.

[2] While there is currently no firm legislation for the use and disclosure of Internet-based corpora, we can not share the corpus to the scientific community. We are working to solve this problem.

[3] The speech tags are taken from the STTS-Tagset, http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html.

(Schmid, 1995; Schiller et al., 1999). In total, the tagset consists of 54 tags for part of speech, morpho-syntactic functions and non-speech. TreeTagger is trained on newspaper articles, which are grammatically well structured. Hence, the annotation of blog comments results in a high number of unknown words, thus annotation errors. Therefore we manually correct the annotation in a second step. Resulting data serve as training data for later development of an automatic POS tagger for blog comments.

Basically, syntactic information is given on the POS annotation level. Hence, text type-specific expressions such as emoticons and interaction words (*netspeak jargon*) as well as topic-specific terms such as URLs and file names are annotated according to their syntactical function, which has considerable advantages. First, there is no need to extend the existing STTS-tagset for blog comment annotation. Second, existing tools developed for texts with given STTS-annotation can still be applied.

For instance, emoticons are tagged according to their position as sentence-internal or sentence-final token. Much more difficult is the morpho-syntactic annotation of interaction words. Taken literally, interaction words are acronyms of multi-word expressions, e.g., *lol* for *Laughing out loud*. Thus, they cannot be classified as a part of speech. Rather, interaction words are similar to interjections, which are defined as single words or fixed phrases, which are invariable in their form and, moreover, they are seen as syntactically unconnected sentence-valent expressions. Table 1 shows which STTS-tags are assigned for exemplarily text type-specific expressions in blog comments.

Tag	Description	Example
$.	Emoticons	:-) (*_*) o.O
NE	File names, Internet address	test.jpg www.rwth-aachen.de
ITJ	Interaction words, inflectives	lol seufz
$(	Special characters	# * @ ^

Table 1: Morpho-syntactic annotation of text type-specific expressions.

As a result, we receive POS annotated blog comments with lemmas, which provide additional information for higher annotation levels.

4.4 Multi-level Annotation Model

In our understanding, an evaluation consists of different components, which are the *evaluated topic*, e.g., MCS, the *source of evaluation* (author or blogger), the *expressed evaluation* itself and the *textual context,* in which the evaluation is embedded. Therefore, in our model four types of levels are distinguished: *User-related levels*, *context-related levels, topic-related levels* and *blog comment-related levels*.

User-related levels. Blogger use in blogs typically nicknames such as *Andy2002* or *Triamos81*. The information contains suppositions about the gender of the blogger, his age as well as the accession date of his membership in the blog. The annotation of this data is most useful in identifying user profiles and user types due to the distribution by gender and age.

Context-related levels. For each blog comment, contextual metadata is supplied that are bloggers name, comment title, date and time of comment submission. The contextual metadata provides information about the user's blogging behavior, e.g., periods in which the user post comments vs. frequency of commentary.

Topic-related levels. According to the discussion topic, different terminologies are important. Particularly in the case of MCS, many topic-specific terms occur in blog comments, e.g., Sendemast "transmitter mast" vs. Nokia 808 PureView. These terms are typically not part of common lexicons respectively taggers are not trained on these terms which at worst leads to tagging errors. Thus, topic-specific terms must be detected in the corpus and classified (noun vs. proper noun, topic-specific vs. non-specific). The collected terms become lexical entries and are used for the development of the blog comment tagger. Moreover, the use and distribution of topic-specific terms in the corpus can also draw conclusions about topic preferences, e.g., system-related topics vs. device-related topics.

Blog comment-related levels. Annotations on these levels provide information about linguistic means, which form an evaluative expression. Actually, there are five different kinds of levels distinguished according to the grammatical fields: the graphemic level, the morphological level, the syntactic level, semantic level and the pragmatic level. At the *graphemic level*, expressions at the text surface as well as grapho-

TOK	Die	haben	es	doch	begriffen	,	die	liefern	einfach	immer	weniger	:)
TRANS	*They*	*have*	*it*	*yet*	*realized*	*,*	*they*	*provide*	*simply*	*increasingly*	*less*	*:)*
POS	PIS	VAFIN	PPER	ADV	VVPP	$,	PDS	VVFIN	ADV	ADV	PIS	$.
LEM	d	haben	es	doch	begreifen	,	d	liefern	einfach	immer	weniger	:)
TYPO												EMO
POL										^	%	
ESA	IRONIZE											

Table 2: Example of a multi-level annotation: comment token (TOK) part of speech (POS), lemma (LEM), typography (TYPO), polarity (POL), evaluative speech act (ESA). The line translation (TRANS) is actually not part of the annotation system, but has been added here for reasons of comprehensibility.

stilistic features that show special notational styles, e.g., emoticons. The annotation at the *morphological level* focuses on word formation, inflection and formation of abbreviated words, e.g., topic-specific new word creations. The designs of the tag sets at both levels are inspired by Gimpel et al. (2011). At the *syntactical level*, syntactic structures are assigned as pointed out in Stoyanov and Cardie (2008). At the *semantic level*, semantic characteristics at the word (lexical semantics) and sentence level (sentence semantics) are recorded, e.g., polarities; tags are partially taken from Clematide et al. (2012). Finally at the *pramatic level*, information is given about the evaluative substance of a speech act, e.g., someone has the intention to BLAME, PRAISE, CRITICIZE something (Austin, 1972; Sandig, 2006).

Thus, each token and each evaluation-indicating token sequence is, in addition to the automatic annotation by the POS tagger, enriched with information regarding its various grammatical functions. Table 2 shows an example of a multi-level annotated text passage, which shows that annotations can cover one or more tokens, depending on the annotation level. Reading the annotations vertically allows for recognizing tag sequences of evaluative expressions in blog comments. These evaluation patterns can be useful for the purpose of automatic opinion detection.

5 Model Application

The aim of the model application was to validate the provided annotation model in terms of its reliability, exemplarily, and, to demonstrate its ability for pattern recognition. The annotation is performed in the score editor EXMARaLDA[4].

1st Level	2nd Level	Tag	Description
Graphemic	Typography	AKR	interaction words
		EMO	emoticons
		ITER	iterations
		MAJ	capital letters
		MARK	highlighting
		MAT	maths symbols
Semantic	Polarity	+	positive
		-	negative
		~	shifter
		^	intensifier
		%	diminisher
Pragmatic	Evaluative speech act	ANGER	be upset about sth.
		BLAME	express disapproval
		CLAIM	accuse so., insist on sth.
		COMPARE	oppose sth. to each other
		CONCLUDE	sum up judgments
		CRITICIZE	judge by standards
		ESTIMATE	speak valuing about sth.
		IRONIZE	draw sth. ridiculous
		NEGATE	consider as non existant
		OVERSTATE	pose sth. overly positive
		PRAISE	express appreciation
		SUSPECT	raise concerns about sth.
		UNDERSTATE	pose sth. overly negative

Table 3: Excerpt of the level-specific tagsets.

5.1 Inter-annotator Agreement Study

To determine the reliability of the model, a two-stage inter-annotator agreement study is carried out. The test corpus contains comments of bloggers with 20 posts in the respective blog over two years. In total, the corpus comprises 50 comments and 5,362 token.

[4] EXMARaLDA is a freely available tool and can be downloaded under http://www.exmaralda.org/downloads.html.

		Typography						Polarity				
	Tag	AKR	MARK	ITER	EMO	MAJ	MAT	+	-	~	^	%
EI	A1	23	31	20	5	6	3	298	172	120	31	10
	A2	22	34	2	4	4	4	295	185	124	50	23
	D	1	3	18	1	1	1	3	13	4	19	13
	%	4.31	8.8	90	20	33.3	23.1	0.99	7.4	2.9	37.9	56.5
	T	20.7						50.7				
EII	A1	29	27	3	5	3	3	190	80	121	24	35
	A2	41	29	4	5	3	3	215	99	111	25	36
	D	25	7	1	0	0	3	25	19	10	1	1
	%	29.1	6.5	23.1	0	0	0	23.7	19.4	8.3	1.4	2.9
	T	17.4						7.4				

Table 4: Results of inter-annotator agreement study: first evaluation (EI), second evaluation (EII), annotator 1 (A1), annotator 2 (A2), allocation difference (D) in percent (%), allocation difference in total/percent per evaluation (T).

The model is tested on three blog comment-related sub-levels, exemplarily: (i) grapho-stylistic means, (ii) polarity and (iii) evaluative speech act. The annotation process is guided by a stylebook that contains information about (i) the level-specific tagsets and (ii) the annotation guidelines. Before each of the evaluation stages, both annotators had to read the respective stylebook intensively and to ask questions, where appropriate. Requests during the annotation process were not allowed. The annotators worked on the same corpus using the identical tagset, separately. Table 3 shows an excerpt of the tag set[5].

The inter-annotator agreement is calculated for each level manually. To precise, the evaluation focused two objectives: (i) analyzing the difference in the allocation of tags on the levels typography and polarity and (ii) identifying variations in the annotation scope on the level evaluative speech act. Table 4 shows the annotation differences for the selected MCS-corpus per level and per annotator (objective (i)).

As it is evident from the results of the first evaluation (EI), the error rates for tag allocation are relatively high. Particularly, the annotation of polarities appears to be problematic (T=50.7%), especially for the annotation of diminishers (56.5%). At the level of typography, an enormous error rate (90%) for the allocation of iterations (ITER) is recognizable, particularly. However, the overall rating for tag allocation delivers a better result (T=20.7%).

[5] Due to page limitations, the entire tagset could not be presented.

Related to evaluation objective (ii), we have tested in how many cases the annotation scope (number of tokens that have been attributed to a tag) varies. Results show that at no text passage the annotation scope is identical for both annotators. Examples A1 (annotator 1) and A2 (annotator 2) show differences in the annotation of the evaluative speech act IRONIZE; the respective annotated text passages are marked in bold.

(A1) Pralerei oder **sind die Taschen zu klein**?

(A2) **Pralerei oder sind die Taschen zu klein?**

Thus, the provided annotation guidelines of the first evaluation (EI) seem to be too shallow. Comparing the annotated data of A1 and A2 for objective (ii) shows that annotation differences had to be diminished through guideline modifications. Therefore, for each tag it was re-defined

i. with which *part of speech* it can appear, e.g., diminishers (%) occur only in combination with the POS tags PIAT (attributive indefinite pronoun) and ADV (adverb);
ii. where an annotation starts and ends (*scope*), i.e., which feature terminates an annotation over multiple tokens, e.g., punctuation marks as terminating features.
iii. whether *special characters* and interaction signs are annotated within an evaluative speech act, e.g., emoticons at the end of a sentence.

For testing the modified guidelines, a second evaluation (EII) was carried out (see table 4). The results show that a significant improvement is recognizable in the tag allocation (objective (i)), which is due to modification i. Nevertheless,

the allocation rate of several tags has not improved.

An excellent result, however, provides the comparison in terms of objective (ii). Compared to the results of the first evaluation (EI), the error rate of the second evaluation (EII) is 20.6%. The improved result is probably due to the modifications ii and iii.

5.2 Pattern Recognition Study

The aim of the study is to show that the model can be used for the recognition of evaluative patterns. For this purpose, a selected corpus is created. To ensure that different formulation and evaluation styles of bloggers are considered and represented (e.g. use vs. non-use of emoticons, use of specific emoticon types), the corpus is formed with comments from different user groups (selection criteria: average number of comments in the entire corpus). From each group, 50 comments are taken.

A single annotator semi-automatically annotated the corpus. First, the automatically tokenized and POS tagged data are checked and corrected, manually. Second, blog comment data is annotated also manually on the levels polarity, evaluative speech act and graphemic using the provided tagset. The investigation focuses on the identification of evaluative speech acts that occur in combination with emoticons and their representation by regular expressions. Results of the frequency analysis are summarized in Table 5.

Blogger / #Comments	**Token**	**Emoticons**			**∑**
		+	-	**0**	
1	2,897	4	2	3	9
10	3,083	15	1	0	16
20	5,362	5	0	0	5
max	4,264	1	1	0	2
∑	**15,606**	**25**	**4**	**3**	**32**

Table 5: Distribution of emoticons over user groups.

In the analyzed corpus, a total of 32 emoticons occur, of which 25 (78.125 %) were annotated as positive and 4 (12.5 %) as negative; 3 (9.4 %) got no polarity attribution. Overall, most commonly positive connoted emoticons were used for the marking of ironic speech acts (IRONIZE), e.g., ;-), :), ;), ,) and ^^. Negatively connoted emoticons, e.g., :(, :-(and :-/, never occurred in combination with an ironic speech act. Furthermore, in the user group with ten comments per blogger, positive emoticons are used most often. The weighting is even more apparent, when the number of emoticons per user group is compared to the number of tokens per user group corpus (see table 5).

Regarding the morpho-syntactic function of emoticons, the result show that emoticons are in most cases used as sentence boundary signs (78.1 %), what means they are set instead of punctuation marks such as ! ?. Thereby, emoticons were set at the end of the blog comments in 44.1 % of the occurrence, i.e., behind the emoticon no more token followed. Then, on the POS level, the left neighbor of the emoticon is an internal character ($(), a noun (NN), an adverbial adjective (ADJD) or a substituting indefinite pronoun (PIS). Moreover, emoticons take over the function of internal characters such as – “ #.

Finally related to the selected user corpus, the following pattern is identified as stereotypical for the evaluative speech act IRONIZE:

$$[EMO_{+}[\text{^^ v ;) v :) }] = \$.] + [\$.\ v\ \emptyset]$$

The term implies that three types of positive emoticons (^^ ;) :)) are typically used to mark the speech act IRONIZE. These emoticons usually take over the morpho-syntactic function of a sentence boundary sign. Habitually, a further sentence boundary sign follows the emoticon or no more token occurs.

6 Summary and Future Work

In this paper, we presented a fine-grained annotation model for the analysis of evaluative expressions in blog comments, exemplarily. The results of the evaluation studies show that the model is reliable. Moreover, we have demonstrated that the model serves for the identification of evaluation patterns. Future work will focus on the further improvement of the annotation guidelines and the identification of additional evaluation patterns, such as the presented pattern for ironic speech acts.

Acknowledgments

We owe gratitude to the Excellence Initiative of the German Federal and State Government as well as Eva Reimer, Julia Ninnemann, Tariq Maqbool and Simon Rüppel for their support in data annotation and tool modification.

References

Saima Aman and Stan Szpakowicz. 2007. Identifying Expressions of Emotions in Text. In *Proceedings of TSD'07*. Berlin:196-205.

John Langshaw Austin. 1972. *Zur Theorie der Sprachakte*. Stuttgart.

Alexandra Balahur and Andrés Montoyo. 2010. Semantic Approaches to Fine and Coarse-Grained Feature-Based Opinion Mining. *Natural Language Processing and Information Systems, (5723)*:142-153.

Michael Beißwenger and Angelika Storrer. 2008. Corpora of Computer-Mediated Communication. In Anke Lüdeling and Merja Kytö, editors, *Corpus Linguistics. An International Handbook,* volume 1. de Gruyter, Berlin, New York:292-308.

Michael Beißwenger. 2009. Chattern unter die Finger geschaut: Formulieren und Revidieren bei der schriftlichen Verbalisierung in synchroner internetbasierter Kommunikation. In Vilmos Ágel and Mathilde Hennig, editors, *Nähe und Distanz im Kontext variationslinguistischer Forschung*. de Gruyter, Berlin, New York.

Michael Beißwenger. 2010. Empirische Untersuchungen zur Produktion von Chat-Beiträgen. In Tilmann Sutter and Alexander Mehler, editors, *Medienwandel als Wandel von Interaktionsformen.* Verlag für Sozialwissenschaften, Wiesbaden:47-81.

Michael Beißwenger, Maria Ermakova, Alexander Geyken, Lothar Lemnitzer, and Angelika Storrer. 2012/in press. A TEI Schema for the Representation of Computer-mediated Communication. *Journal of the Text Encoding Initiative* (TEI).

Simon Clematide and others. 2012. MLSA — A Multi-layered Reference Corpus for German Sentiment Analysis. In *Proceedings of the 14th LREC,* Istanbul, Turkey.

Stefanie Dipper. 2005. XML-based Stand-off Representation and Exploitation of Multi-Level Linguistic Annotation. In *Proceedings of BXML'05.* Berlin:39-50.

EXMARaLDA, http://www.exmaralda.org/downloads.html, cited on 03.02.2012.

Paul Ekman. 1993. Facial expression and emotion. *American Psychologist*, 48(4):384–392.

Clayton R. Fink, Danielle S. Chou, Jonathon J. Kopecky, and Ashley J. Llorens. 2011. Coarse- and Fine-Grained Sentiment Analysis of Social Media Text. *Johns Hopkins APL technocal Digest 30(1)*:22-30.

Eugenie Giesbrecht and Stefan Evert. 2009. Is Part-of-Speech Tagging a Solved Task? An Evaluation of POS Taggers for the German Web as Corpus. In *Proceedings of the 5th Web as Corpus Workshop*. San Sebastian, Spain.

Kevin Gimpel and others. 2011. Part-of-Speech Tagging for Twitter: Annotation, Features and Experiments. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. Portland. ACL.

Kristin Luckhardt. 2009. *Stilanalysen zur Chat-Kommunikation. Eine korpusgestützte Untersuchung am Beispiel eines medialen Chats*. Diss., TU Dortmund. Digitale Ressource: http://hdl.handle.net/2003/26055

Anke Lüdeling, Maik Walter, Emil Kroymann, and Peter Adolphs. 2005. Multi-level error annotation in learner corpora. In *Proceedings of Corpus Linguistics,* Birmingham:105-115.

Anke Lüdeling. 2011. Corpora in Linguistics: Sampling and Annotation. In *Going Digital. Evolutionary and Revolutionary Aspects of Digitization.* Stockholm: 220-243.

Malik M.S. Missen, Mohand Boughanem, and Guillaume Cabanac. 2009. Challenges for Sentence Level Opinion Detection in Blogs. In *Proceedings of ICIS'09*. Shanghai:347-351.

Gilad Mishne. 2005. Experiments with Mood Classification in Blog Posts. In *Style2005: The 1st Workshop on Stylistic Analysis of Text for Information Access. SIGIR'05*. Salvador, Brasil.

Melanie Neunerdt, Bianka Trevisan, Rudolf Mathar, and Eva-Maria Jakobs. 2012/in press. Detecting Irregularities in Blog Comment Language Affecting POS Tagging Accuracy. In *Proceedings of 13th CICLing*. Springer, New Delhi.

Robert Remus, Uwe Quasthoff, and Gerhard Heyer. 2010. SentiWS – a Publicly Available Germanlanguage Resource for Sentiment Analysis. In *Proceedings of the 7th LREC*. Malta:1168–1171.

Robert Remus and Christian Hänig. 2011. Towards Well-Grounded Phrase-Level Polarity Analysis. In *Proceedings of 11th CICLing.* Tokyo:380-392.

Barbara Sandig. 2003. Formen des Bewertens. *Anabasis.* In *Festschrift für Krystyna Pisarkowa,* Lexis, Kraków: 279-287.

Sandig, Barbara. 2006. *Textstilistik des Deutschen*, volume 2. de Gruyter, Berlin.

Anne Schiller, Simone Teufel, and Christine Stöckert. 1999. *Guidelines für das Tagging deutscher Textkorpora mit STTS. Technischer Bericht.* Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart und Seminar für Sprachwissenschaft, Universität Tübingen.

Peter Schlobinski and Torsten Siever. 2005. Sprachliche und textuelle Aspekte in Weblogs. Ein internationales Projekt. *Networx* (46).

Helmut Schmid. 1995. Improvements in Part-of-Speech Tagging with an Application to German. In *Proceedings of SIGDAT'95.* Dublin. ACL.

Angelika Storrer. 2009. Rhetorisch-stilistische Eigenschaften der Sprache des Internets. In Ulla Fix,

Andreas Gardt, and Joachim Knape, editors, *Rhetorik und Stilistik – Rhetorics and Stilistics*. de Gruyter, Berlin, New York.

Veselin Stoyanov and Claire Cardie. 2008. Annotating Topics of Opinions. *Proceedings of the 6th LREC*, Marrakech, Morocco.

Carlo Strapparava and Rada Mihalcea. 2008. Learning to Identify Emotions in Text. In *Proceedings of the SAC'08*. Fortaleza, Brasil:1556-1560. ACM.

Dirk Thorleuchter, Dirk van den Poel, and Anita Prinzie. 2010. Extracting Consumers Needs for New Products - A Web Mining Approach. In *Proceedings of WKDD '10*. Phuket:440-443.

TreeTagger http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/, cited on 01.02.2012.

Bianka Trevisan and Eva-Maria Jakobs. 2010. Talking about Mobile Communication Systems. Verbal Comments in the Web as a Source for Acceptance Research in Large-scale Technologies. In *Proceedings of the IPCC 2010*. University of Twente (NL):93-100.

Bianka Trevisan and Eva-Maria Jakobs. 2012. Probleme und Herausforderungen bei der Identifikation von Bewertungen in großen Textkorpora. Am Beispiel Mobilfunk. In Sabrina Braukmeier, Julia Burkhardt, and Fleur Pfeifer, editors, *Wege in den Sprachraum*. Lang, Frankfurt/Main:189-209.

Janyce Wiebe, Teresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165-210.

Theresa Ann Wilson. 2008. *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*. University of Pittsburgh.