

Annotating Opinion – Evaluation Of Blogs

Estelle Dubreil and Matthieu Vernier and Laura Monceaux and Béatrice Daille

Abstract This chapter deals with annotating opinions on a non-specific corpus of blogs. This work is motivated by a more general aim of building a generic method for detecting opinions. In accordance with this aim, we propose a linguistic model for the description of the opinion expression phenomenon.

1 Introduction

Up to now the previous approaches of sentiment analysis can be separated into two categories. The first one aimed to categorized texts according to semantic orientation [6, 11, 15, 16] and most of the time we distinguish between the positive orientation or the negative one, even neutral.

The second approach concentrated his effort on the extraction of evaluations by taking a particular attention on the conveyed information. For example, to determine who think what about which product? By this way we find a richer and more precise information [7, 8, 12, 18].

Estelle Dubreil
LINA CNRS UMR 6241, 2, rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, FRANCE ,
e-mail: Estelle.Dubreil@univ-nantes.fr

Matthieu Vernier
LINA CNRS UMR 6241, 2, rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, FRANCE,
e-mail: Matthieu.Vernier@univ-nantes.fr

Laura Monceaux
LINA CNRS UMR 6241, 2, rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, FRANCE,
e-mail: Laura.Monceaux@univ-nantes.fr

Béatrice Daille
LINA CNRS UMR 6241, 2, rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, FRANCE,
e-mail: Beatrice.Daille@univ-nantes.fr

This paper presents an annotating work based on a linguistic model to prepare such an extraction of evaluations approach. We focus on the different modality (*opinion, judgement, agreement ...*) of the evaluation expression on a corpus of weblog posts. Such corpus, also studied in [4, 10] is likely to be representative of the various language phenomenon linked to the expression of evaluation. By this study, we tend to develop a general method in order to detect evaluation expressions on every kind of evaluated subject existing on blogs.

2 Blogosphere

The blogosphere defines itself as a space bounded on Web by the completeness of blogs. Now, blogs represent a new type of media which the Internet users use to communicate around themes who they like and who arouse the expression of opinion; that is why their analysis contains a scientific stake.

2.1 *Issues of an automatic detection of opinions*

Blogs are everywhere nowadays: the media uses and refers to them, politicians resort to them and researchers use them for their work. With this popular fashion, the number of Blog platforms has increased in France since 2002, so multiplying tenfold the possible extent of information exchanges. On average 1100 Blogs are created every day on the platform 'Over-blog'¹, the French website form from which the corpus is extracted. Every visitor spends approximately 12 minutes of his or her day on its 3,5 million pages. They are read by a Blogger population that is representative of the global population, because Over-blog is not specifically marketed for teenagers, unlike the Skyblog and MSN Spaces for example. Blogs therefore constitute both a new method of information exchange and a new power of information, which can influence the opinion of readers. Therefore, Blogs represent an ideal object of study for the observation of different forms of expressions of opinion – evaluation, where evaluation includes all types of opinion. Representing a subset of the Web, Blogs are "mainly made up of published posts which are deposited and appear in a non chronological order (the most recent are found at the top of the page), and they often include external hypertexts or links" [5, p. 3].

¹ Over-blog – <http://www.over-blog.com/>

2.2 A new media

This new form of media is specific because of the abolition of the impersonal character of communication. Every reader has the possibility of answering the post published by the author of a Blog by posting his or her own comment. In addition, as the publication of the information is becoming increasingly easy, this type of media has become very popular. The interactive specificity of Blogs is based on enunciation rules and the management of the relation with the public. This has resulted in the internet community rapidly adopting Blogs as a means of expressing favourable and unfavourable private states. If the expression of a kind of evaluation is the common denominator to all Blogs, the landscape of the Blogosphere is not homogeneous, but articulated around an axis which stretches from the personal diary to pure weblogs (link archives), by way of the thematic blog.

2.3 The personal thematic blogs

We worked on the personal thematic blogs written in French by adults, defined as “sites where cultural productions are critically evaluated [...], because the enunciator declares him/herself initially as an individual capable of judgement and analysis, and who can evaluate and exchange on the subject of public objects” [1, p. 40].

Blogs represent a new kind of type of text which concentrates a large diversity of subjects, lexicon, morpho-syntactic patterns, which made that we are not limited to a specific domain. The subjects on which we are working on are as different as Harry Potter, Wii, Vladimir Poutine (Named entities), strike, ecology (concept).

This work presents the annotation of a corpus of posts and their comments with the aim of developing a blog monitoring tool which is able to automatically detect the evaluation of bloggers with regard to a given entity. This manual labelling was made by tagging XML via the creation of a DTD which contains the notions of “subject” and “evaluation”. As detailed below.

3 Description of evaluation

The goal of the annotation scheme is to distinguish the different forms of opinions that a speaker can express about a subject. So, the annotation of the Blogoscopy corpus concerns these two categories of elements: the subjects contained in posts and comments, and evaluations emitted by the author of the post or the comment about the subjects.

3.1 Subject

The notion of “subject” contains two categories: the *concepts* and the *named entity*.

A concept is generally a noun or a nominal group, occurring in the text, and which is representative of the theme of the post or the comment which has been tagged. We defined three types of concepts: the *Concerned Concepts* – CC corresponding to the nature of the referent and answering the question “What is the post or the comment about?”, the *Associated Concepts* – AC associated in field of the CC, and the *Non associated Concepts* – NC which contain an evaluation independent to the general theme of the post or the comment (Ex: post “Sin City”: CC: info, film, upload ; AC: date de sortie/release date, réalisateur/producer ; NC: ville/town, quartier/area, flics/police).

A named entity is generally a proper noun (Ex: Sarkozy, Weeds), occurring in the text, which inevitably contains an evaluation. We defined two types of named entity, the *associated named entity* – IA, associated with the semantic field of the post or the comment, and the *non associated named entity* – IN, which is not associated with the semantic field of the post or the comment.

3.2 Linguistic model

The notion of *evaluation* brings together linguistic data which can be observable through the phenomena of modalisation described by Charaudeau [3], among which we extract the descriptions specified for the elocutive modality of enunciation (dedicated to the speaker), which he put to use in his study on the film criticism [2]. The choice of this linguistic model is made legitimate by the descriptions of the evaluation proposed, firstly because it enables researchers to find answers to traditional questions of this type [14]: “What can be qualified as evaluation?”, “How can we rank evaluations according to their positive or negative polarity?”, “How does the context change the polarity and the strength of evaluation?” Secondly, these descriptions are significant, because they come into being through a list of occurrences, and are important benchmarks, as was shown by the procedure of inter-annotations by Wiebe et al. [17].

Concretely, the elocutive modality of enunciation is divided into twelve modalities [3]. Five of these modalities (opinion, appreciation, acceptance-refusal, agreement-discord and judgment) refer to a type of evaluation: Each of these modalities contain many subcategories, also defined and clarified. For example, ‘opinion’ – OP is divided into five subcategories:

- *conviction* (Ex : je suis persuadé) ; I am absolutely sure
- *supposition certitude high* (Ex : je me doute) ; I am almost sure
- *supposition certitude medium* (Ex : je crois) ; I think
- *supposition certitude low* (Ex : je doute) ; I am not sure
- *supposition premonition* (Ex : je sens). I feel

In the same way, ‘appreciation’ is divided into six subcategories:

- *explicit appreciation favourable* – EAF (Ex : je suis satisfait) ; I am satisfied
- *explicit appreciation unfavourable* – EAU (Ex : je suis triste que) ; I am sad that...
- *explicit appreciation exclamative form favourable* – EAEF (Ex : Youpi!) ; I’m very happy !
- *explicit appreciation exclamative form unfavourable* – EAEU (Ex : Merde!) ; Dam !
- *implicit appreciation favourable* – IAF (Ex : c’est vraiment intéressant) ; This is very interesting
- *implicit appreciation unfavourable* – IAU (Ex : c’est vraiment mauvais). This is really bad

4 Data : Blogoscopie Corpus

The Blogoscopie Corpus was annotated according to the annotation scheme described above. At present it contains 200 posts and their comments, which represents 83500 words, extracted in June 2007 among 33 of 43 themes proposed on the website, and selected in accordance with personal thematic blogs (Ex: current events, blogzines, business, cinema, consumption/buying, beliefs, music, politics, etc.). With an aim to representing of the interests of bloggers, this extraction focused on the 10 most visited blogs according to each theme, and more particularly the first 10 posts published and their comments (maximum 10). The work of annotation is finished and the corpus will be made public at the end of 2008.

5 Annotating methodology

The annotation of 200 posts was carried out in four phases: the application of the linguistic model phase, the confrontation of the data phase, the consolidation of the annotation scheme phase, and the increase of the volume of data phase. The annotation instructions do not specify either the formal criteria for the identification of the various forms of evaluation, or the type of words to be annotated (Ex: verbs or adjectives, parts of speech, word classes). However the annotation does take into account the syntagmatic level.

5.1 Application of the linguistic model

The application phase for the linguistic model was simultaneously carried out by 4 annotators on 12 posts, chosen among very different themes for their linguistic

difficulties. The aim was to evaluate the relevance of the annotation scheme so that common rules of annotation could emerge.

Regarding the entities, among the agreed inter-annotators principles, the most important entities are the annotations of all the AC even if they do not contain evaluation and, secondly, the annotations of the IA, only if they are the object of an evaluation. Next, if an evaluation has an impact on several entities, they have to all appear in the attribute “form”; the separator being the comma.

Regarding the evaluations, it seemed necessary to annotate the polarity according to the context of enunciation, to annotate the phrase logic forms (Ex: *bailler aux corneilles/dead tired, regretter amèrement/bitterly regret*), and to deconstruct the evaluations in case of a conjunction of modalities (Ex: *cette femme [CC] **brillante** [implicit appreciation favourable] et **ravissante** [implicit appreciation favourable]/this brillante and charming woman*).

5.2 Confrontation of the data

The confrontation phase of the data aimed to stabilise the annotation rules updated for the concepts. A team of 3 annotators were in charge of annotating concepts only on 64 posts. The annotation was carried out in three stages. Every post was annotated by two linguists, working individually. With a vantage to objectivity, a harmonisation stage was carried out by the third linguist, who had not participated in the annotation. This refined the annotation principles. From now on, when a text contains a spelling variant or a synonymic occurrence of a concept, it has to be annotated with the same identifier (Ex: *série, séries, sitcom*). When necessary, the AC are annotated as hyponymous of CC (Ex: *fruits secs/ dried fruit [CC], raisins/grapes [AC]*). Finally, the annotation of the concepts is made on the longest nominal group in the presence of a preposition (Ex: *feuille de laurier/ laurel leaf, or in French ' leaf of the laurel*). The first annotation showed great differences between the annotators. The first one considered in average 1.5 CC (up to 5) and 9 CA (up to 42) by post. The second one indicated 1.5 CC (up to 7) and 6.7 (up to 25) CA by post. The agreement between annotators was low: only 50% for the CC and 44% for the CA.

5.3 Consolidation of the annotation scheme and increase of the volume of data

The consolidation phase of the annotation scheme aimed to stabilize the annotation rules updated for the named entity and the evaluations. The annotation was made by a single annotator, but every question or problem was the object of a discussion with 4 other linguists and computer specialists. This work also allowed us to refine the annotation principles. For example, it seemed necessary to create an attribute “irony” (Ex: *Ah bon ? parce que ça marche moyen/ Really? Because it doesn't really*

work [*implicit appreciation unfavourable*] entre Nico [IA] et Cécilia [IA] ? Quelle blague! / what a joke!), to integrate personal pronoun subjects into the prospect, so that they could be used later as markers of intensity (Ex: **Je suis persuadé que / I am sure that**[opinion conviction]), so as to annotate the exclamatory forms of the appreciation. Even when the exclamation mark is distant or even missing (Ex: **Contente/Happy** [*explicit appreciation exclamative form favourable*] de t'avoir fait rire / to have made you laugh! [NC]!!!). The new rules allow an improvement in these figures with an agreement of 63% for the CC and 52% for the CA.

The increase of the volume of data phase aimed to confirm the general annotation principle. The annotation was made by a single annotator on 124 blogs.

6 Statistics and observations

A part of our corpus is constituted by 100 blog pages talking about different kind of evaluated subject : a movie ("Le coeur des hommes 2"), a wine ("Beaujolais"), a person ("Raymond Domenech"), a social event ("The strike"), a book ("Harry Potter"), a polemical french law ("LRU"), an object ("Wii") and two conceptual subject ("Sustainable development" and "Nuclear"). Tab bellow presents the repartition of the evaluation modality observed after the annotating process in these weblogs posts and in theirs comments.

	OP	EAF	EAU	IAF	IAU
Eval. markers in posts	68 5.84%	43 3,7%	19 1,6%	434 37,25%	472 40,52%
Eval. markers in comments	22 6.5%	31 9,1%	8 2,4%	103 30,3%	131 38,5%
Movie	14	15	4	51	32
Wine	4	4	2	98	31
Person	12	9	6	80	115
Social event	8	4	1	32	57
Book	7	25	5	72	33
Law	24	8	5	84	155
Object	5	6	2	32	31
Sustainable development	5	3	2	57	49
Nuclear	11	0	0	31	100

Table 1 Repartition of the evaluation modality.

Through these different kinds of subjects, we observe regularities used to formulate an evaluation. Proportion of modalities seems to be separate from the analysed subject. In this way, we notice that "Implicit appreciation" modality is the most frequently used by bloggers. Next, we find more explicit modality of evaluation by the using of phrase indicating a "judgement" or an "opinion". On an other hand, ac-

ording to analysed subject, we can already note which subjects have a favourable evaluation and those who don't.

All regularities in the observed modalities and the fact that these regularities are separated from the kind of subject treated will be used to automatically detect evaluations in new weblogs and subjects of any kinds.

At the same time, first all kind of the annotated evaluations were got back to establish lexical resources, which specified the positive or negative value of the data. Secondly, we study the relation between an evaluation and the evaluated subject.

For example:

Vladimir Poutine, qui jouit d'une forte popularité en Russie/Vladimir Putin, who enjoys a strong popularity in Russia.

<IA cc="Président, Russie">**Vladimir Poutine**</IA>, **qui** <Appreciation type="IAF" subject="Poutine">**jouit d'une forte popularité**</Appreciation> en Russie

Harry Potter 3 est une réussite totale/ Harry Potter 3 is a total success.

<IA cc="film, saga">**Harry Potter 3**</IA> <Appreciation type="IAF" forme="Harry Potter 3">**est une réussite totale**</Appreciation>

We also establish automatically a list of regular morpho-syntactic patterns as:

- [Subject] + qui + [evaluation] (cf. beside) ;
- [Subject] + être + [evaluation]

All these patterns confine on the intra-phrastic level. Moreover by using intra-phrastic patterns, [8] showed that it was possible to find a relation subject-evaluation in new posts with a precision of 0,56 and a recall of 0,53. These first observations tend towards a sketch of a grammar of the expression of the evaluation in blogs, just like the specifications obtained by Legallois and Ferrari [9] on of cultural objects evaluations. These regularities will be automatically found in new posts by a bootstrapping method [13].

7 Conclusion

Starting from a linguistic model for the description of evaluation, we create an annotation system enable to list the specific linguistic marks resulting from the expression of evaluation in blogs. This annotation scheme distinguishes between the subject and the evaluations concerning these subjects. By this way, we applied this scheme to posts and comments. Also, we have henceforth a corpus under format xml constituted by 200 annotated blogs and of their associated comments, what represents 83500 words. Besides a stabilized annotation scheme, we also have rules of annotation the relevance of which was validated, by an inter-annotators agreement. The first observations led on the corpus show that the proportion of modalities seems to be separate from the analysed subject, because we notice the most important frequency of the implicit appreciation compared to the other expression of evaluation.

At the moment, we build first a lexical resource, which specified the positive or negative value of the data, and a list of regular morpho-syntactic patterns, for tending towards a sketch of a grammar of the expression of the evaluation, which will be automatically found in new posts by a bootstrapping method.

References

1. Cardon, D. and Delaunay-Téterel, H. (2006). La production de soi comme technique relationnelle : un essai de typologie des blogs par leurs publics. *Les blogs, Réseaux*, Vol. 24, N. 138/2006, pp. 15–71.
2. Charaudeau, P. (1988). La critique cinématographique : faire voir faire parler. *Dans La Presse, Produit, Production, Reception*, Paris, Didier Erudition, pp. 47-70.
3. Charaudeau, P. (1992). *Grammaire du sens et de l'expression*. Paris : Hachette éducation.
4. Chesley, P. and Bruce, V. and Li, X. and Rohini, S. (2006). Using Verbs and Adjectives to Automatically Classify Blog Sentiment. To appear in *AAAI Spring Symposium Technical Report SS-06-03*.
5. Fievet, C. and Turrettini, E. (2004). *Blog Story*. Eds. Eyrolles (305 p.).
6. Gamon, M. and Aue, A. (2005). Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms. In: *Proceedings of the ACL-05 Workshop on Feature Engineering*.
7. Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In: *Proceedings of the Tenth International Conference on Knowledge Discovery and Data Mining (KDD)*.
8. Kobayashi, N. and Kentaro, I. and Matsumoto, Y. (2007). Extracting aspect-evaluation and aspect-of relations in opinion mining. In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.
9. Legallois, D. and Ferrari, S. (2006). Vers une grammaire de l'évaluation des objets culturels. *Schedae*, Prépublication n. 8 (fascicule n.1, pp. 57–68).
10. Mishne, G. and Glance, N. (2006). *Predicting movie sales from blogger sentiment*. AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006).
11. Pang, B. and Lee, L. and Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques (pp. 79–86). In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
12. Popescu, A. and Etzioni, O. (2005). Extracting product features and opinions from reviews. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
13. Riloff, E. and Patwardhan, S. and Wiebe, J. (2006). Feature Subsumption for Opinion Analysis. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, (EMNLP-06).
14. Schröder, M. and Pirker, H. and Lamolle, M. (2006). First suggestions for an emotion annotation and representation language (pp. 88–92). In: Deviller, L. et al. (Ed.), *Proceedings of LREC'06 Workshop on Corpora for Research on Emotion and Affect*, Genoa, Italy.
15. Torres-Moreno, J.-M. and El-Bèze, M. and Béchet, F. and Camelin, N. (2007). Comment faire pour que l'opinion forgée à la sortie des urnes soit la bonne? *Application au défi fouille de textes 2007. DEFT07*, pp. 119–133, AFIA 2007, Grenoble, France.
16. Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to supervised classification of reviews. In: *40th Annual Meeting of the Association for Computational Linguistics, Philadelphia*.
17. Wiebe, J. and Wilson, T. and Cardie, C. (2005). Annotating Expressions of opinions and Emotions in Language. *Language Ressources and Evaluation*, Vol. 39 (2-3), pp. 165–210.

18. Yi, J. and Nasukawa, T. and Bunescu, R. and Niblack, W. (2003). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In: *Proceedings of the third IEEE International Conference on Data Mining (ICDM)*.