Distilling Opinion in Discourse: A Preliminary Study

Nicholas Asher and Farah Benamara
IRIT-CNRS Toulouse,
France
{asher, benamara}@irit.fr

Yvette Yannick Mathieu
LLF-CNRS Paris,
France
yannick.mathieu@linguist.jussieu.fr

Abstract

In this paper, we describe a preliminary study for a discourse based opinion categorization and propose a new annotation schema for a deep contextual opinion analysis using discourse relations.

1 Introduction

Computational approaches to sentiment analysis eschew a general theory of emotions and focus on extracting the affective content of a text from the detection of expressions of sentiment. These expressions are assigned scalar values, representing a positive, a negative or neutral sentiment towards some topic. Using information retrieval, text mining and computational linguistic techniques together with a set of dedicated linguistic resources, one can calculate opinions exploiting the detected "bag of sentiment words". Recently, new methods aim to assign fine-grained affect labels based on various psychological theories–e.g., the MPQA project (Wiebe et al., 2005) based on literary theory and linguistics and work by (Read et al., 2007) based on the Appraisal framework (Martin and White, 2005).

We think there is still room for improvement in this field. To get an accurate appraisal of opinion in texts, NLP systems have to go beyond positive/negative classification and to identify a wide range of opinion expressions, as well as how they are discursively related in the text. In this paper, we describe a preliminary study for a discourse based opinion categorization. We propose a new annotation schema for a fine-grained contextual

opinion analysis using discourse relations. This analysis is based on a lexical semantic analysis of a wide class of expressions coupled together with an analysis of how clauses involving these expressions are related to each other within a discourse. The aim of this paper is to establish the feasibility and stability of our annotation scheme at the subsentential level and propose a way to use this scheme to calculate the overall opinion expressed in a text on a given topic.

2 A lexical semantic analysis of opinion expressions

We categorize opinion expressions using a typology of four top-level categories (see table 1): RE-PORTING expressions, which provide an evaluation of the degree of commitment of both the holder and the subject of the reporting verb, JUDG-MENT expressions, which express normative evaluations of objects and actions, ADVISE expressions, which express an opinion on a course of action for the reader, and SENTIMENT expressions, which express feelings (for a more detailed description of our categories see (Asher et al, 2008)).

Our approach to categorize opinions uses the lexical semantic research of (Wierzbicka, 1987), (Levin, 1993) and (Mathieu, 2004). From these classifications, we selected opinion verb classes and verbs which take opinion expressions within their scope and which reflect the holder's commitment on the opinion expressed. We removed some verb classes, modified others and merged related classes into new ones. Subjective verbs were split into these new categories which were then extended by adding nouns and adjectives.

Our classification is the same for French and English. It differs from psychologically based classifications like Martin's Appraisal system : in ours

Groups	SubGroups	Examples
Reporting	a) Inform	*inform, notify, explain*
	b) Assert	*assert, claim, insist*
	c) Tell	*say, announce, report*
	d) Remark	*comment, observe, remark*
	e) Think	*think, reckon, consider*
	f) Guess	*presume, suspect, wonder*
Judgment	g) Blame	*blame, criticize, condemn*
	h) Praise	*praise, agree, approve*
	i) Appreciation	*good, shameful, brilliant*
Advise	j) Recommend	*advise, argue for*
	k) Suggest	*suggest, propose*
	l) Hope	*wish, hope*
Sentiment	m) Anger/CalmDown	*irritation, anger*
	n) Astonishment	*astound, daze, impress*
	o) Love, fascinate	*fascinate, captivate*
	p) Hate, disappoint	*demoralize, disgust*
	q) Fear	*fear, frighten, alarm*
	r) Offense	*hurt, chock*
	s) Sadness/Joy	*happy, sad*
	t) Bore/entertain	*bore, distraction*
	u) Touch	*disarm, move, touch*

Table 1: Top-Level opinion categories.

the contents of the JUDGMENT and SENTIMENT categories are quite different, and more detailed for SENTIMENT descriptions with 14 sub-classes. Ours is also broader: the REPORTING and the AD-VISE categories do not appear as such in the Appraisal system. In addition, we choose not to build our discourse based opinion categorization on the top of MPQA (Wiebe et al, 2005) for two reasons. First, we suggest a more detailed analysis of private states by defining additional sets of opinion classes such as HOPES and RECOMMENDATIONS. We think that refined categories are needed to build a more nuanced appraisal of opinion expressions in discourse. Second, text anchors which correspond to opinion in MPQA are not well defined since each annotator is free to identify expression boundaries. This is problematic if we want to integrate rhetorical structure into opinion identification task. MPQA often groups discourse indicators (but, because, etc.) with opinion expressions leading to no guarantee that the text anchors will correspond to a well formed discourse unit.

3 Towards a Discursive Representation of Opinion Expressions

Rhetorical structure is an important element in understanding opinions conveyed by a text. The following simple examples drawn from our French corpus show that discourse relations affect the strength of a given sentiment. S1 *: [I agree with you]$_a$ even if I was shocked* and S2 *: Buy the DVD, [you will not regret it]$_b$.* Opinions in S1 and S2 are positive but the contrast introduced by *even* in S$_1$ decreases the strength of the opinion expressed in (a) whereas the explanation provided by (b) in S2 increases the strength of the recommendation.

Using the discourse theory SDRT (Asher and Lascarides, 2003) as our formal framework, our four opinion categories are used to label opinion expressions within a discourse segment. For example, there are three opinion segments in the sentence S3: *[[It's poignant]$_d$, [sad]$_e$]$_g$ and at the same time [horrible]$_f$*

We use five types of rhetorical relations: CON-TRAST, CORRECTION, SUPPORT, RESULT and CONTINUATION (For a more detailed description see (Asher et al, 2008)). Within a discourse segment, negations were treated as reversing the polarities of the opinion expressions within their scope. Conditionals are hard to interpret because they affect the opinion expressed within the consequent of a conditional in different ways. For example, conditionals,expressions of ADVISE can block the advice or reverse it. Thus *if you want to waste you money, buy this movie* will be annotated as a recommendation not to buy it. On the other hand, conditionals can also strengthen the recommendation as in *if you want to have good time, go and see this movie*. We have left the treatment of conditionals as well as disjunctions for future work.

3.1 Shallow Semantic Representation

In order to represent and evaluate the overall opinion of a document, we characterize discourse segments using a shallow semantic representation using a feature structure (FS) as described in (Asher et al, 2008). Figure 1 shows the discursive representation of the review movie S4: *[This film is amazing.]$_a$. [[One leaves not completely convinced] $_{b.1}$, but [one is overcome]$_{b.2}$]. [[It's poignant]$_{c.1}$, [sad]$_{c.2}$] and at the same time [horrible]$_{c.3}$].[Buy it]$_d$. [You won't regret it]$_e$.*
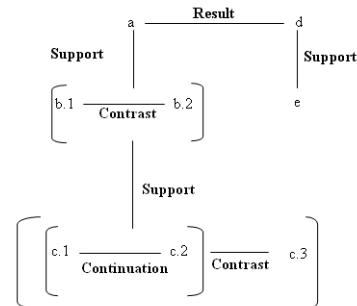

Figure 1: Discursive representation of S4.

Once we have constructed the discursive representation of a text, we have to combine the different FS in order to get a general representation

that goes beyond standard positive/negative representation of opinion texts. In this section, we first explain the combination process of FS. We then show how an opinion text can be summarized using a graphical representation.

The combination of low-level FS is performed in two steps: (1) combine the structures related by coordinating relations (such as CONTRAST and CONTINUATION). In figure 1, this allows to build from the segments b.1 and b.2 a new FS ; (2) combine the strutures related via subordinating relations (such as SUPPORT and RESULT) in a bottom up way. In figure 1, the FS of the segment a is combined with the structure deduced from step 1. During this process, a set of dedicated rules is used. The procedure is formalized as follows. Let a, b be two segments related by the rhetorical relation R such as: $R(a, b)$. Let S_a, S_b be the FS associated respectively to a and b i.e S_a : $[category : [group_a : subgroup_a], modality : [polarity : p_a, strength : s_a] \cdots]$ and S_b : $[category : [group_b : subgroup_b], modality : [polarity : p_b, strength : s_b] \cdots]$ and let S : $[category : [group], modality : [polarity : p, strength : s] \cdots]$ be the FS deduced from the combination of S_a and S_b. Some of our rules are:

CONTINUATIONS strengthen the polarity of the common opinion. One of the rule used is: if $(group_a = group_b)$ and $(subgroup_a \neq subgroup_b))$ then if $((p_a = neutral)$ and $(p_b \neq neutral))$ then $group = group_a$ and $p = p_b$ and $s = max(s_a, s_b)$, as in *moving and sad news*.

For CONTRAST, let OW_i be the set of opinion words that belongs to a segment S_i. We have for $OW_a = \emptyset$ and $OW_b \neq \emptyset$: $group = group_b$, $p = p_b$ and $s = s_b + 1$, as in *I don't know a lot on Edith Piaf's life but I was enthraled by this movie*.

Finally, an opinion text is represented by a graph $G = (\aleph, \Re)$ such as:
– $\aleph = H \cup T$ is the set of nodes where : $H = \{ho_i/ho_i$ is an opinion holder$\}$ and $T = \{to_i : value/to_i$ is a topic and value is a $FS\}$, such as : $value = [Polarity : p, Strength : s, Advice : a]$, where: $p = \{positive, negative, neutral\}$ and $s, a = \{0, 1, 2\}$.
– $\Re = \Re_H \cup \Re_T \cup \Re_{H-T}$ where: $\Re_H = \{(h_i, h_j)/h_i, h_j \in H\}$ means that two topics are related via an ELABORATION relation. This holds generally between a topic and a subtopic, such as a movie and a scenario ; $\Re_T = \{(t_i, t_j, type)/t_i, t_j \in T$ and $type = support/contrast\}$ means that two holders are related via a CONTRAST (holders h_i and h_j have a contrasted opin-

ion on the same topic) or a SUPPORT relation (holders share the same point of view) ; and $\Re_{H-T} = \{(h_i, t_j, type)/h_i \in H$ and $t_j \in T$ and $type = attribution/commitment\}$ means that an opinion towards a topic t_j is attributed or committed to a holder h_i. For example, in *John said that the film was horrible*, the opinion is only attributed to John because verbs from the TELL group do not convey anything about the author view. However, in *John infomed the commitee that the situation was horrible*, the writer takes the information to be established. The figure 2 below shows the general representation of the movie review S4.
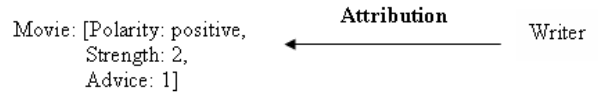

Movie: [Polarity: positive, Strength: 2, Advice: 1] ← **Attribution** — Writer

Figure 2: General representation of S4.

4 Annotating Opinion Segments: Experiments and Preliminary Results

We have analyzed the distribution of our categories in three different types of digital corpora, each with a distinctive style and audience : movie reviews, Letters to the Editor and news reports in English and in French. We randomly selected 150 articles for French corpora (around 50 articles for each genre). Two native French speakers annotated respectively around 546 and 589 segments. To check the cross linguistic feasability of generalisations made about the French data, we also annotated opinion categories for English. We have annotated around 30 articles from movie reviews and letters. For news reports, the annotation in English was considerably helped by using texts from the MUC 6 corpus (186 articles), which were annotated independently with discourse structure by three annotators in the University of Texas's DISCOR project (NSF grant, IIS-0535154); the annotation for our opinion expressions involved a collapsing of structures proposed in DISCOR.

The annotation methodology is described in (Asher et al, 2008). For each corpus, annotators first begin to annotate elementary discourse segments, define its shallow representation and finally, connect the identified segments using the set of rhetorical relations we have identified. A segment is annotated only if *it explicitly* contains an opinion word that belong to our lexicon or if it bears a rhetorical relation to an opinion segment.

The average distribution of opinion expressions in our corpus across our categories for each language is shown in table 2. The annotation of movie reviews was very easy. The opinion expressions are mainly adjectives and nouns. We found an average of 5 segments per review. Opinion words in Letters to the Editor are adjectives and nouns but also verbs. We found an average of 4 segments per letter. Finally, opinions in news documents involve principally reported speech. As we only annotated segments that clearly expressed opinions or were related via one of our rhetorical relations to a segment expressing an opinion, our annotations typically only covered a fraction of the whole document. This corpus was the hardest to annotate and generally contained lots of embedded structure introduced by REPORTING type verbs.

To compute the inter-annotator agreements (IAA) we did not take into account the opinion holder and the topic as well as the polarity and the strength because we chose to focus, at a first step, only on agreements on opinion categorization, segment idendification and rhetorical structure detection. We computed the agreements only on the French corpus. The French annotators performed a two step annotation where an intermediate analysis of agreement and disagreement between the two annotators was carried out. This analysis allowed each annotator to understand the reason of some annotation choices. Using the Kappa measure, the IAA on opinion categorization is 95% for movie reviews, 86% for Letters to the Editors and 73% for news documents.

Annotators had good agreement concerning what the basic segments were (82%), which shows that the discourse approach in sentiment analysis is easier compared to the lexical task where annotators have low agreements on the identification of opinion tokens. The principal sources of disagreement in the annotation process came from annotators putting opinion expressions in different categories (mainly between PRAISE/BLAME group and APPRECIATION group, such as *shame*) and the choice of rhetorical relations. Nevertheless, by using explicit discourse connectors, we were able to get relatively high agreement on the choice of rhetorical relations. We also remained quite unsure how to distinguish between the reporting of neutral opinions and the reporting of facts. The main extension of this work are to (1) deepen our opinion typology, specifically to include modals

Groups	Movie (%)		Letters (%)		News (%)	
	French	English	French	English	French	English
Reporting	2.67	2.12	14.80	13.34	43.91	42.85
a	0	0	0.71	1.33	4.02	4.76
b	0.53	0	0	4	5.83	0
c	0	0	1.79	0	4.51	35.71
d	0.88	0	2.17	0	11.82	0
e	1.33	0	10.12	6.67	5.89	1.34
f	0	2.12	0	1.34	11.77	0
Judgment	60.53	40.52	52.50	73.34	39.23	33.34
g	0.54	0	6.32	26.66	13.69	16.67
h	2.45	2.12	7.54	20	1.81	4.76
i	54.49	38.29	33.48	26.87	23.72	11.90
Advise	6.92	10.63	10.05	13.34	7.27	9.52
j	6.26	8.51	0.70	5.33	1.37	0
k	0.66	2.12	3.94	1.33	3.61	0
l	0	0	5.38	6.67	2.28	9.52
Sentiment	27.30	34.04	33.08	2.67	11.35	16.67
m	0.54	0	3.23	0	0,90	0
n	2.23	6.38	3.96	2.66	0,90	7.14
o	7.38	4.25	3.74	0	1,87	9.52
p	4.97	2.12	5.03	0	2,72	0
q	2.23	0	5.03	0	1,86	0
r	0.89	0	7.17	0	2,28	0
s	3.79	4.25	2.87	0	0.88	0
t	1.33	14.9	0	0	0	0
u	4.46	2.12	2.15	2.12	0	0

Table 2: Average distribution of our categories.

and moods like the subjunctive, and to (2) provide a deep semantic representation that associates for each category of opinion a lambda term involving the proferred content and a lambda term for the presuppositional content of the expression, if it has one. In terms of automatization, we plan to exploit a syntactic parser to get the argument structure of verbs and then a discourse segmenter like that developed in the DISCOR project, followed by the detection of discourse relations using cue words.

References

Asher N. and Benamara F. and Mathieu Y.Y. 2008. *Categorizing Opinions in Discourse*. ECAI08.

Asher N. and Lascarides A. 2003. *Logics of Conversation*. Cambridge University Press.

Levin B. 1993. *English Verb Classes and Alterna-tions: A Preliminary Investigation*. University of Chicago Press

Martin J.R and White P.R.R. 2005. *Language of Evaluation: Appraisal in English*. Palgrave Macmillan.

Mathieu Y. Y. 2004. *A Computational Semantic Lexicon of French Verbs of Emotion*. In Shanahan, G., Qu, Y., Wiebe, J. (eds.): Computing Attitude and Affect in Text. Dordrecht.

Read J., Hope D. and Carroll J. 2007. *Annotating Expressions of Appraisal in English*. The Linguistic Annotation Workshop, ACL 2007.

Wiebe J., Wilson T. and Cardie C. 2005. *Annotating Expressions of Opinions and Emotions in Language*. Language Resources and Evaluation 1(2).

Wierzbicka A. 1987. *Speech Act Verbs*. Sydney: Academic Press.