

Discourse relations and evaluation*

Radoslava Trnavac, Debopam Das and Maite Taboada

Simon Fraser University

rtrnavac@sfu.ca, ddas@sfu.ca, mtaboada@sfu.ca

Abstract

We examine the role of discourse relations (relations between propositions) in the interpretation of evaluative or opinion words. Through a combination of Rhetorical Structure Theory or RST (Mann & Thompson, 1988) and Appraisal Theory (Martin & White, 2005), we analyze how different discourse relations modify the evaluative content of opinion words, and what impact the nucleus-satellite structure in RST has on the evaluation. We conduct a corpus study, examining and annotating over 3,000 evaluative words in 50 movie reviews in the SFU Review Corpus (Taboada, 2008) with respect to five parameters: word category (nouns, verbs, adjectives or adverbs), prior polarity (positive, negative or neutral), RST structure (both nucleus-satellite status and relation type) and change of polarity as a result of being part of a discourse relation (Intensify, Downtone, Reversal or No Change). Results show that relations such as Concession, Elaboration, Evaluation, Evidence and Restatement most frequently intensify the polarity of the opinion words, although the majority of evaluative words (about 70%) do not undergo changes in their polarity because of the relations they are a part of. We also find that most opinion words (about 70%) are positioned in the nucleus, confirming a hypothesis in the literature, that nuclei are the most important units when extracting evaluation automatically.

Keywords: discourse relations, evaluation, polarity, Rhetorical Structure Theory, Appraisal Theory

1. Introduction

In our previous work on evaluation and sentiment analysis in text (Taboada & Grieve, 2004; Voll & Taboada, 2007; Taboada et al., 2009; Taboada et al., 2011; Trnavac & Taboada, 2012, 2014), we have observed that certain discourse relations affect the interpretation of the evaluation contained therein. Consider the following examples (1) – (3) from the Simon Fraser Review Corpus (Taboada, 2008) in which the semantics of embedded evaluators within concessive, conditional and elaborative sentences is affected by the corresponding discourse relation:

- (1) A lot of people say, "This is a movie for kids, not adults, so don't be so harsh." But even my kids (10 and 14) disliked it.

* This is a pre-publication version of a paper to appear in *Corpora*. Current version: October 2015.

- (2) Sure the special effects were neat, I guess, if you are into those types of things.
- (3) This movie portrays the children's story "The Cat in the Hat" in a very colorful and original manner. The children are adorable and the cat is very amusing.

In the above examples, the concessive discourse relation marked by the concessive markers *but* and *even* in sentence (1) intensifies the negative evaluation of the word *disliked* in the subordinate clause, while the conditional relation in (2) downtones the positive semantics of the word *neat*. The positive evaluation of *original* in (3) is reinforced by the elaborations in the satellite (the second sentence).

In this study, we explore the idea that the organization of the discourse structure contributes information relevant to assessing attitude of the text and can be useful in sentiment analysis, a general method for extraction subjective content in texts. The terms evaluation, opinion and sentiment are used interchangeably in this study and can be found under different umbrella terms in literature (see Taboada et al., 2011: 268): sentiment analysis (Pang & Lee, 2008), subjectivity (Lyons, 1981; Langacker, 1985), opinion mining (Pang & Lee, 2008), analysis of stance (Biber & Finegan, 1988; Conrad & Douglas, 2000), appraisal (Martin & White, 2005) or evaluation (Hunston & Thompson, 2000; Thompson & Alba-Juez, 2014).

The paper presents a corpus study that focuses on the following two questions: (1) How different discourse relations according to Rhetorical Structure Theory (Mann & Thompson, 1988; Taboada & Mann, 2006b) modify the evaluative content of evaluative words expressed with lexical items (nouns, verbs, adjectives, adverbs); and (2) What impact the nucleus-satellite structure has on the evaluation.

Rhetorical Structure Theory describes how to split a text into spans (nuclei and satellites), each representing a meaningful part of the text. A nucleus is considered to be the span with the highest degree of importance with respect to its related spans. Satellites support the nuclei and can therefore be seen as less important spans. In this article, we use a lexicon-based approach to extract sentiment (or evaluation) from text. The Semantic Orientation CALculator software or SO-CAL (Taboada et al., 2011) uses dictionaries of words annotated with their semantic orientation (polarity and strength), and incorporates intensification and negation. SO-CAL is applied to the polarity classification task, which is the process of assigning a positive or negative label to a text that captures the text's opinion towards its main subject matter.

Most extant research efforts on sentiment classification do not use discourse information. Among the few research studies on discourse-based sentiment analysis, we single out the following. Polanyi and Zaenen (2006) emphasize that the base attitudinal valence of a lexical item can be modified by contextual shifters such as negatives, intensifiers, presuppositional items, connectors and discourse structure that includes two basic discourse relations – lists and elaborations. The basic premise is that words have a *prior polarity*, i.e., their polarity in isolation, in a sort of dictionary sense, but also *contextual polarity*, affected by the context in which they appear (Wilson et al., 2009).

Trnavac and Taboada (2012) carry out a corpus analysis on movie and book reviews to examine how nonveridical markers (i.e., negation, modals, imperatives, questions, habituais, intensional verbs, subjunctives) and discourse relations (concessive and conditional) contribute

to the expression of evaluation in discourse. The authors conclude that nonveridical elements in the majority of cases modify polarity at the local level (level of the clause), while discourse relations derive the changes from the combination of two or more clauses. Since the scope from which discourse relations derive changes is wider, the modification that nonveridical markers exercise on polarity depends much on the type of the relation where they occur. They point out that concessive relations seem to have the effect of reversing the polarity of evaluative words therein, whereas conditional relations result in either intensification or downtoning of evaluation.

Some studies in computational linguistics include discourse information in calculating sentiment of texts. The available methods on sentiment analysis and opinion mining are still much focused in the lower levels (up to the sentence level), with a few exceptions in the literature. Taboada et al. (2011) propose a method which predicts that opinion words found in the nuclei (more important parts) of a document are more significant for the overall sentiment. Heerschop et al. (2011) have used an RST discourse parser in order to calculate semantic orientation at the document level by weighting the nuclei more heavily. The authors hypothesize that there is a possible hierarchy in relations – the satellites of some relations may contribute more to the overall sentiment than others. Bal (2014) analyzes opinions and arguments in news editorials and op-eds by annotating a corpus with labels from the Appraisal Framework (Martin & White, 2005) and rhetorical relations (Mann & Thompson, 1988). Benamara et al. (2011; 2013), Asher et al. (2008, 2009) and Chardon et al. (2012) focus on measuring the effect of discourse structure on sentiment analysis. Using Segmented Discourse Representation Theory (SDRT) as a formal framework and shallow semantic representation, they investigate how discourse relations interact with opinions and to what extent these interactions depend on the corpus genre. They propose a new annotation schema that 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 in discourse. They select opinion verb classes and verbs which take opinion expressions within their scope and which reflect the holder's commitment to the opinion expressed. Asher et al. (2008, 2009) use five types of discourse relations: Contrast, Correction, Support, Result and Continuation. They propose that Result relations strengthen the polarity of the opinion in the second argument, while Continuation relations strengthen the polarity of the common opinion, and Contrast relations may strengthen or weaken the polarity of opinion expressions. Somasundaran et al. (2007) propose to model the discourse-level associations between related opinion topics using opinion frames. In this model, a frame is a structure composed of two opinions and their respective targets connected via two types of relation: the same and the alternative relation. However, there is no definitive mapping between opinion frames and rhetorical relations.

In this article, our objective is to classify existing types of discourse relations proposed by Rhetorical Structure Theory (Mann & Thompson, 1988) in their interaction with subjective information that is based on the Appraisal Framework (Martin & White, 2005). Our corpus analysis of the 50 positive and negative movie reviews from the Simon Fraser University Review Corpus (Taboada, 2008) extends to all RST relations and shows statistical tendencies for attraction between polarity types, types of relations, and the nucleus/satellite structures.

In the next section, we briefly introduce the Appraisal Framework (Martin & White, 2005), characterize the evaluative lexicon that we use in our analysis and describe the basic

features of the SO-CAL software. Section 3 introduces Rhetorical Structure Theory (Mann & Thompson, 1988; Taboada & Mann, 2006b), a theory of coherence in discourse that provides a list of discourse relations for this study. Section 4 presents methodology, corpus, as well as corpus annotation and the corpus study that we perform. In Section 5, we discuss the results of our study, and in Section 6, we provide some conclusions on the interaction between discourse relations and evaluation in the movie reviews.

2. Evaluation

We use a lexicon-based approach for extracting sentiment automatically and for analyzing evaluation in the Epinion corpus of movie reviews¹. Comparable to the approach of Asher et al. (2008) that categorizes sentiment based on the lexical semantic research of Wierzbicka (1987), Levin (1993) and Mathieu (2004), or the approaches that use adjectives as indicators of the semantic orientation of text (see Hatzivassiloglou and McKeown (1997); Wiebe (2000); Hu and Liu (2004); Taboada et al. (2011)), our lexical classification is based on Martin and White's (2005) Appraisal system.

Appraisal belongs to the tradition of systemic-functional analysis started by Halliday (Halliday, 1985; Halliday & Matthiessen, 2004), and has been developed mostly in Australia by Jim Martin, Peter White and colleagues (Martin, 2000; White, 2003; Martin & White, 2005). Martin (2000) characterizes Appraisal as the set of resources used to express emotions, judgements, and valuations, alongside resources for amplifying and engaging with those evaluations. Since the Appraisal Framework's approach is lexically rather than grammatically based, it is primarily focused on those words and semantic categories of words that allow a speaker to express different types of opinions. The first sub-system of Appraisal, Attitude, is concerned with our feelings, judgements of behavior and evaluations of things and is classified as Affect, Judgement and Appreciation. Affect is used to construe emotional responses about the speaker or somebody else's reactions (e.g., *happiness*, *sadness*); Judgement conveys moral evaluations of character about somebody else than the speaker (e.g., *ethical*, *deceptive*); whereas Appreciation captures aesthetic qualities of objects and natural phenomena (*remarkable*, *desirable*, *harmonious*, *elegant*, *innovative*). Attitude is complemented by the system of Graduation, which captures the upscaling and downtoning possible within the range of Attitude (*very concerned*, *kind of callous*, *interesting to a certain extent*). Finally, Engagement is concerned with the ways in which resources such as modality, polarity, modal adjuncts, conditionals and concessives position the speaker/writer with respect to the opinion being advanced. The categories of Appraisal are summarized in Figure 1:

¹ The reviews in the Simon Fraser University Review Corpus (Taboada, 2008) are extracted from the web site <http://www.epinions.com/>.

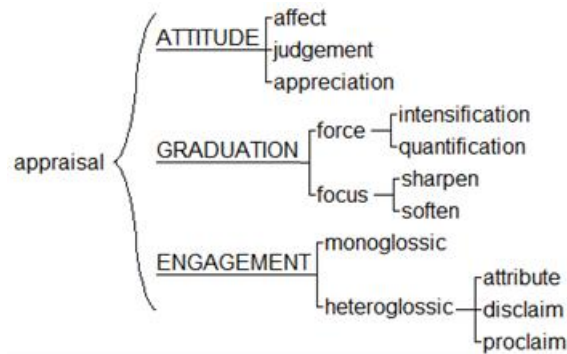


Figure 1: The Appraisal Framework (adapted from Martin & White (2005: 38))

The focus of this study is the semantic continuum that is covered by Attitude², which includes emotional, moral and aesthetic opinions, and Graduation, which is concerned with changes in intensification of opinion.

The Appraisal system represents a basis for creating dictionaries to calculate sentiment in texts, which are part of the Semantic Orientation CALCulator software (SO-CAL). The calculation of sentiment in SO-CAL is grounded on two assumptions (Taboada et al., 2011: 270): That individual words have what is referred to as prior polarity, that is, a semantic orientation that is independent of context; and that semantic orientation can be expressed as a numerical value. The dictionaries in SO-CAL were produced by hand-tagging all adjectives, nouns, verbs, adverbs, modifiers (amplifiers and downtoners), negation and irrealis markers which are found in a 400-text corpus of Epinion reviews extracted from eight different categories: books, cars, computers, cookware, hotels, movies, music, and phones. The opinion words were assessed on a scale ranging from -5 for extremely negative to +5 for extremely positive. “Positive” and “negative” were decided on the basis of the word’s prior polarity, that is, its meaning in most contexts. The dictionaries of SO-CAL contain 2,252 adjective entries, 1,142 nouns, 903 verbs, and 745 adverbs.

Appraisal, and in particular the Attitude system, represents the basis of our work on how evaluation is expressed in text. For the corpus extraction of opinion words, we deploy the SO-CAL software. In the next section, we describe Rhetorical Structure Theory, a system of discourse relations that we use in our analysis to capture the relationship between discourse structure and evaluation.

3. Rhetorical Structure Theory

Rhetorical Structure Theory or RST (Mann & Thompson, 1988; Taboada & Mann, 2006b) is a functional theory of text organization. It describes what parts a text is made of, what kinds of

² Comparing to this study, the emphasis in Trnavac and Taboada (2012) is mostly on the relationship between Engagement and concessive and conditional discourse relations.

relationships exist between these parts, and how these parts are organized with respect to each other to constitute a coherent piece of discourse.

Text organization in RST is primarily described in terms of relations that hold between two (or sometimes more) non-overlapping text spans. Relations can be multinuclear, reflecting a paratactic relationship, or nucleus-satellite, a hypotactic type of relation. The names nucleus and satellite refer to the relative importance of each of the relation components. Relation inventories are open, and the most common ones include names such as Cause, Concession, Condition, Elaboration, Result or Summary.

Relations in RST are defined in terms of four fields: (1) constraints on the nucleus, (2) constraints on the satellite, (3) constraints on the combination of nucleus and satellite, and (4) effect (on the reader). The *locus* of the effect, derived from the *effect* field, is identified as either the nucleus alone or the nucleus-satellite combination. An analyst builds the RST structure of a text based on the particular judgements that are specified by these four fields. An example of how a relation (here Concession) is defined in RST is provided in Table 1, reproduced from Mann and Thompson (1988: 254-255).

<i>Relation name:</i>	CONCESSION
<i>Constraints on N:</i>	W has positive regard for the situation presented in N
<i>Constraints on S:</i>	W is not claiming that the situation presented in S doesn't hold
<i>Constraints on the N + S combination:</i>	W acknowledges a potential or apparent incompatibility between the situations presented in N and S; W regards the situations presented in N and S as compatible; recognizing that the compatibility between the situations presented in N and S increases R's positive regard for the situation presented in N
<i>The effect:</i>	R's positive regard for the situation presented in N is increased
<i>Locus of the effect:</i>	N and S
Note: N = nucleus; S = satellite; W = writer; R = reader	

Table 1: Definition of Concession relation

Texts, according to RST, are built out of basic clausal units that enter into rhetorical (or discourse, or coherence) relations with each other in a recursive manner. Mann and Thompson (1988) proposed that most texts can be analyzed in their entirety as recursive applications of different types of relations. In effect, this means that an entire text can be analyzed as a tree structure, with clausal units being the branches and relations the nodes.

RST relations, based on their intended effect, are divided into two classes: *subject matter* relations and *presentational* relations. In subject matter relations, the intended effect is that the reader recognizes the relations in question. On the other hand, in presentational relations, the intended effect is to increase some inclination in the reader (positive regard, belief, or acceptance of the nucleus). The original RST taxonomy includes 16 subject matter relations and 7 presentational relations, as shown in Table 2, reproduced from Mann and Thompson (1988: 257).

Subject Matter	Presentational
Elaboration	Motivation (increases desire)
Circumstance	Antithesis (increases positive regard)
Solutionhood	Background (increases ability)
Volitional Cause	Enablement (increases ability)
Volitional Result	Evidence (increases belief)
Non-Volitional Cause	Justify (increases acceptance)
Non-Volitional Result	Concession (increases positive regard)
Purpose	
Condition	
Otherwise	
Interpretation	
Evaluation	
Restatement	
Summary	
Sequence	
Contrast	

Table 2: Taxonomy of relations in RST

For illustration purposes, we provide the RST annotation of the following short text taken from one of the movie reviews in the SFU Review Corpus (Taboada, 2008)³.

- (4) The effects and mood were done rather well. The entire theater gasped and jumped at one particular scene - which is always a good sign for this kind of movie. But the plot inconsistencies coupled with the fact that this movie beats a dead horse is enough to take away a few stars.

The graphical representation of the RST analysis of this text is provided in Figure 2.

³ Examples from the reviews are reproduced verbatim, including possible typos and unconventional grammar.

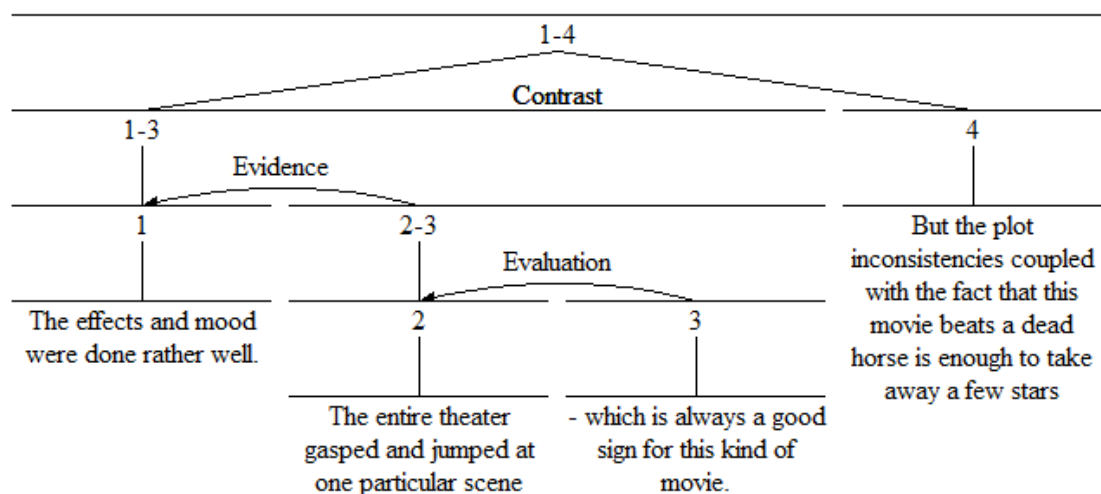


Figure 2: Graphical representation of an RST analysis

The RST analysis shows that the text comprises four spans which are represented in the diagram (in Figure 2) by the cardinal numbers, 1, 2, 3, and 4, respectively. In the diagram, the arrowhead points to a span called the nucleus, and the arrow points away from another span called the satellite. Straight lines above a span mean that it is a nucleus. Span 3 (as a satellite) is connected to Span 2 (here nucleus) by an *Evaluation* relation, and together they make a combined Span 2-3. Then, Span 2-3 (as a satellite) is linked to Span 1 (here nucleus) by an *Evidence* relation, and together they make a combined Span 1-3. Finally, Span 1-3 (as a nucleus) connects to Span 4 (another nucleus) by a multinuclear *Contrast* relation.

The account of RST presented here is very simplified, and omits many details. For a more extensive introduction to RST, see, along with the original description of RST (Mann & Thompson, 1988), the RST website (Mann & Taboada, 2015) and reviews of research on RST (Taboada & Mann, 2006b) and applications of RST (Taboada & Mann, 2006a).

4. Methodology

In this study, we used the SFU Review Corpus (Taboada, 2008) as our source of data. The SFU Review Corpus includes a collection of 400 reviews of movies, books, music, hotels, and consumer products (cars, telephones, cookware, computers, etc.). The reviews were originally posted on the web site Epinions. We analyzed all the 50 movie reviews from the corpus which are equally distributed into 25 positive reviews and 25 negative reviews. The 50 texts contain 33,425 words.

We extracted 3,244 opinion words or phrases expressing evaluation from the 50 movie reviews using SO-CAL, a calculator of semantic orientation for words, sentences and texts (Taboada et al., 2011). SO-CAL includes a dictionary of opinion words divided into four categories: nouns, adjectives, verbs and adverbs, which are manually ranked (on a scale from □5

to +5). The extracted 3,244 words are distributed into 798 nouns, 1,556 adjectives, 581 verbs and 309 adverbs.

4.1. Types of Annotation

We annotated the extracted opinion expressions with respect to five parameters. First, we annotated them by their *word class* such as noun, adjective and verb, and also by their *polarity*: positive or negative. These values were provided by SO-CAL, which uses the Brill tagger for part-of-speech tagging (Brill, 1995). Then, the expressions were annotated with respect to the *RST structure*. Here, we examined, for each target evaluative word or phrase, the clause or sentence (or sometimes a group of sentences), and the rhetorical relations into which the clause entered, at the most local level of discourse. This means that we only annotated the clause/sentence (span) containing the evaluative word, and its connection to neighbouring spans, but we did not continue annotating until all units of discourse were considered. The latter is the usual procedure in RST, and results in a tree for the entire text. Since we were interested in changes in polarity at the very local level, i.e., at the level of evaluative words and phrases, we only carried out local annotation. The advantage of doing that is that RST annotations tend to be more reliable at the local level, and less when the entire text needs to be annotated (Carlson et al., 2001; Soricut & Marcu, 2003). In the RST analysis, we additionally annotated the evaluative expression with two parameters: its *position within nucleus-satellite structure* (i.e., whether the target word or phrase occurs in the nucleus or satellite span) and the *type of coherence relation* (Contrast, Elaboration, Purpose, etc.) by which the text spans which host the evaluative expression connect together.

Finally, we examined the opinion expressions as they are affected by the relational context. In particular, we annotated the evaluative words or phrases with respect to the *change* that their prior polarity undergoes by virtue of being in a coherence relation. In this stage of annotation, we assigned a target word or phrase one of the following four possible values:

- Reversal: The evaluative load of an opinion expression is reversed (e.g., from negative to positive).
- Intensification: The evaluative load of an opinion expression is modified towards a higher value (i.e, more positive if the word is originally positive; more negative if the word is originally negative).
- Downtoning: The evaluative load of an opinion expression is modified towards a lower value, at either end of the scale.
- No change: The evaluative load of an opinion expression remains unchanged.

In some cases, the annotation of polarity is somewhat subjective, and we went through a process of independent annotation by each of the authors and comparison of the annotation, until we were satisfied that we agreed on how to label examples.

4.2. Annotation scheme

In our annotation, there are five main parameters: word class, polarity, position within nucleus-satellite structure, RST relation type and change, each with a set of possible values of its own. All these parameters and their values are organized systematically in a hierarchical structure in our annotation scheme. The hierarchical organization of the annotation scheme is provided in Figure 3. Note that subcategories under (RST) RELATION are only illustrative, not exhaustive⁴.

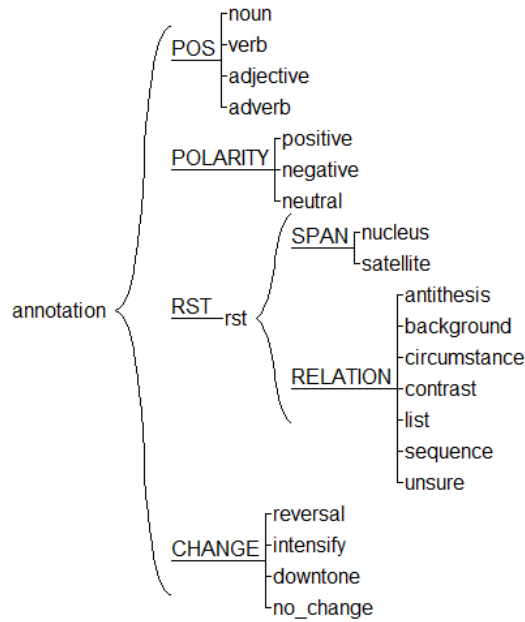


Figure 3: Hierarchical taxonomy of parameters in appraisal annotation

As Figure 3 illustrates, the parameter *word class* (POS) has four possible values: noun, verb, adjective and adverb. The parameter *polarity* (POLARITY) includes three possible values: positive, negative and neutral. *Position within nucleus-satellite structure* (SPAN) represents two possible values: nucleus or satellite. *RST relation type* (RELATION) includes a set of 25 possible relations: Antithesis, Background, Circumstance, Concession, Condition, Elaboration, Enablement, Evaluation, Evidence, Interpretation, Justify, Motivation, Non-volitional Cause, Non-volitional Result, Otherwise, Purpose, Restatement, Solutionhood, Summary, Volitional Cause, Volitional Result, Contrast, List, Sequence and Unsure. Finally, the parameter *change* (CHANGE) represents four possible values: reversal, intensify, downtone and no change.

⁴ In our RST annotation, we used a set of 25 RST relations, including the 23 relations in the original RST taxonomy (Mann & Thompson, 1988) presented in Table 2, plus two additional relations: List (e.g., to annotate a sequence of items) and Unsure (to annotate those situations in which no relevant relations were found).

4.3. Annotation tool

We mainly used UAM CorpusTool (O'Donnell, 2008) to perform our appraisal annotation task. UAM CorpusTool is a text annotation software which provides annotation at multiple levels defined by the user (document layer, semantic-pragmatic level, syntactic level, etc.). In our task, there is only one level of annotation; however, our annotation scheme includes multiple layers of parameters and their values organized into a hierarchical structure. We chose to use UAM CorpusTool for our purposes, as it conveniently supports hierarchically-organized tagging schemes such as ours.

Additionally, we used RSTTool (O'Donnell, 1997) to annotate the RST structure of texts of the movie reviews because it provides the most convenient representation of the relational annotation of texts. Using RSTTool, we segmented a text into its elementary discourse units, usually clauses (Tofiloski et al., 2009), identified them as being either nuclei or satellites, and then identified the relevant coherence relations connecting those spans. Next, we used the information about the RST structures produced by RSTTool to annotate the target opinion expressions in UAM CorpusTool with respect to their membership to nucleus or satellite spans and also to the relation types hosting those expressions.

4.4. An example of appraisal annotation

We provide annotation of a few opinion expressions in a text excerpt taken from one of the movie reviews (file number: no22) in the SFU Review Corpus (Taboada, 2008). The text is provided below, with the annotated opinion expressions (words or phrases) underlined.

- (5) Overall, this movie is probably worth seeing. I am left with an empty feeling where the creepiness should be after a good ghost story, so I'm a bit disappointed.

The SO-CAL output for the opinion expressions in the text is provided in Table 3.

#	Opinion Expression	Word class	Polarity
1.	worth	Adjective	1.0
2.	empty	Adjective	-3.0
3.	creepiness	Noun	-1.5
4.	good	Adjective	0.6
5.	a bit disappointed	Verb	-2.1

Table 3: SO-CAL output of text in file no22

Next, the RST structure of the text (produced by RSTTool) is provided in Figure 4.

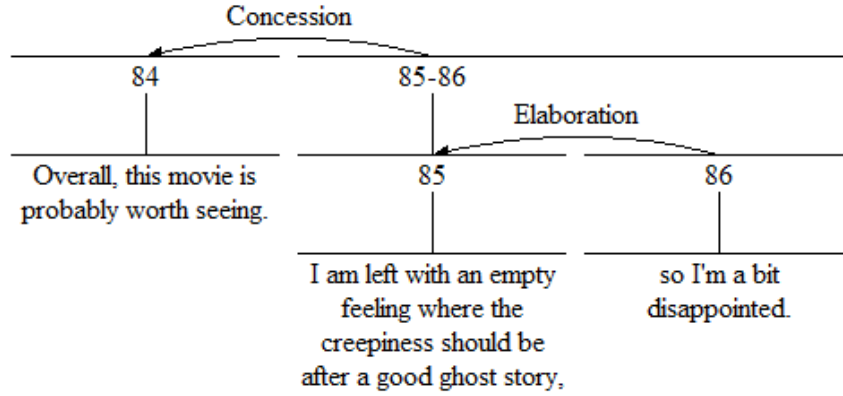


Figure 4: RST analysis of example text

With respect to the parameter *change*, we considered how the opinion expression changes because of the fact that it is embedded in a particular rhetorical relation. For instance, the word *worth* in the first span has a positive prior polarity. The prior polarity is its polarity in the dictionary sense, confirmed by native speakers, and captured in the value 1.0 in the SO-CAL dictionary. Its prior polarity is, however, downtoned, by virtue of being the nucleus of a Concession relation in this particular example. The author recommends the movie as worth seeing, but downtones that opinion in the satellite of the Concession (felt empty, absence of creepiness, disappointed). The next word, the adjective *empty* has a negative prior polarity of $\square 3.0$. This negativity is intensified because *empty* is in the nucleus of an Elaboration relation, and the satellite of that relation contributes to making the negative stronger, by adding *disappointed* to it. The words *creepiness* and *good*, in our opinion, do not undergo a change from their prior polarity. They are merely descriptive of what should happen in a movie, and the overall evaluation conveyed by the Concession and Elaboration do not seem to affect them. Finally, *a bit disappointed* is intensified in the same way as *empty*, because the two words together, in the Elaboration, intensify each other.

Compiling all these information from Table 3, Figure 4 and the immediately preceding paragraph above, the complete annotation of these five opinion expressions with respect to all five parameters is given in Table 4.

#	Expression	Word class	Polarity	Nucleus-satellite	Relation	Change
1.	worth	Adjective	positive	nucleus	Concession	downtone
2.	empty	Adjective	negative	nucleus	Elaboration	intensify
3.	creepiness	Noun	negative	nucleus	Elaboration	no change
4.	good	Adjective	positive	nucleus	Elaboration	no change
5.	a bit disappointed	Verb	negative	satellite	Elaboration	intensify

Table 4: Complete annotation of opinion expressions in text in Example (5)

A snapshot of the annotation window in UAM CorpusTool is provided in Figure 5.

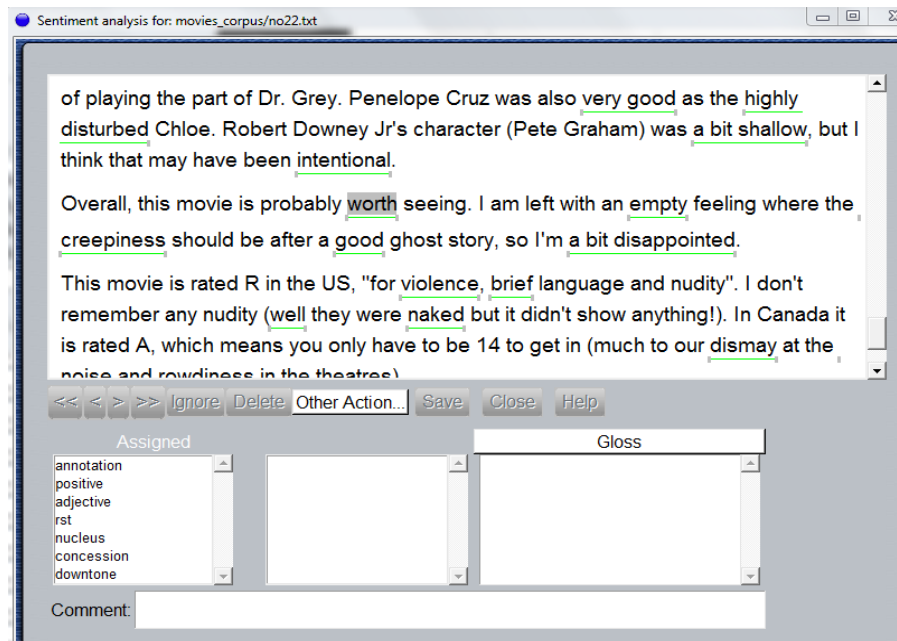


Figure 5: Appraisal annotation in UAM CorpusTool

5. Results

The results of this study should answer two questions: (1) How do different discourse relations modify the evaluative content of opinion words?; and (2) What impact does the nucleus-satellite structure have on the polarity of opinion words? The distribution of opinion words within the discourse relations represented in Table 5 demonstrates that most frequently they occur in our corpus in relations such as Elaboration, followed by List, Concession and Evaluation.

Relation	Number of sentiment words	Relation	Number of sentiment words
Antithesis	70	Justify	88
Background	82	List	726
Circumstance	202	NV-cause	6
Concession	253	Otherwise	1
Condition	170	Purpose	75
Contrast	123	Restatement	8
Elaboration	902	Sequence	148
Enablement	2	Solutionhood	9
Evaluation	240	V-cause	7
Evidence	32	V-result	11
Interpretation	28	Unsure	61

Table 5: Distribution of sentiment words in discourse relations

Table 6 presents the results that are related to the nucleus-satellite structure and the type of polarity change. The deletion test predicts that when removing the nuclear unit, the overall message of the discourse relation typically becomes quite difficult to infer. In accordance with that idea, our results show that the majority of opinion words are present in the nuclei of relations (70.78%). However, both structures with opinion words reflect the same tendencies: While ‘No change’ prevails, the second largest group of relations belongs to ‘Intensify’, followed by the ‘Downtone’ and ‘Reversal’ types of polarity change.

Nucleus / satellite	Downtone (n)	Intensify (n)	Reversal (n)	No Change (n)
Nucleus	116	550	35	1595
Satellite	62	195	9	682

Table 6: Nucleus-satellite structure and type of change

Based on the data in Table 6, we created a table to test whether there is a statistically significant difference between nucleus and satellite and presence or absence of polarity change. That is, we want to find out whether words in the nucleus are more likely (or not) to see their polarity changed; and likewise for the satellite. Table 7 shows the overall frequencies used to test that hypothesis, using a chi-square test.

Nucleus / satellite	Presence of polarity change	Absence of polarity change
Nucleus	701	1595
Satellite	266	682

Table 7: Chi-square test (nucleus-satellite distinction and presence/absence of polarity change)

The chi-square test does not show statistical significance ($\chi^2=1.96$, $df=1$, $p < 0.05$). This means that the null hypothesis cannot be rejected, that the nucleus-satellite distinction has a bearing on whether a word will change polarity or not. In other words, there is no correlation between the nucleus-satellite distinction and presence of polarity change/absence of polarity change. The saliency of the textual context⁵ does not have a direct influence on the occurrence of polarity. In addition, all four analyzed word classes which mark sentiment occur in both the nucleus and the satellite positions. Consider the examples in (6) – (8):

- (6) [This movie is allegedly for kids]_S, [but I do not think it's suitable for most kids under 10.]_N
- (7) [Wounded and in the hands of the enemy]_S, [Algren is surprised by the relative kindness he is shown.]_N
- (8) [Topics like the history of broccoli and carpets tend to predominate the presentations, and right from the beginning, we find ourselves laughing along with our two main protagonists, Annie (Julie Walters) and Chris (Helen Mirren)]_N, [as they suffer through the deadly dull pontifications that make up the WI meetings.]_S

In example (6), an opinion word, an adverb *allegedly*, is positioned within the satellite of the Antithesis relation. The same is observed with the verb *suffer* within the Circumstance relation in (8), while the noun *kindness* in (7) is found in the nucleus span of the Background relation.

We then investigated whether it's not the nucleus-satellite distinction, but the type of relation that shows a correlation with type of change (or no change). As illustrated in Table 8, the choice of a polarity type does not seem to be modified by the type of a relation in the majority of cases. Out of 3,244 contexts with opinion words in our corpus of movie reviews, 2,277 (70.19%) did not change their polarity under the influence of a discourse relation, while 967 (29.8%) underwent some type of change.

⁵ Mann and Thompson (1988: 266) characterized the nucleus of a relation as being “more essential to the writer’s purpose than others”.

Relations	Downtone (n)	Intensify (n)	Reversal (n)	No Change (n)
Antithesis	6	11	5	48
Background	4	12	1	65
Circumstance	8	25	4	165
Concession	43	59	0	151
Condition	17	23	1	129
Contrast	19	21	3	80
Elaboration	31	236	11	624
Enablement	-	-	-	2
Evaluation	15	73	5	147
Evidence	0	14	0	18
Interpretation	0	3	0	26
Justify	6	17	1	64
List	22	203	9	492
NV-cause	0	2	0	3
Otherwise	0	0	0	1
Purpose	1	6	2	66
Restatement	1	4	0	3
Sequence	3	18	2	125
Solutionhood	1	0	0	8
V-cause	0	2	0	5
V-result	0	4	0	7
Unsure	1	12	0	48
Total	178	745	44	2277

Table 8: Discourse relations and type of change

Table 9 shows the relations which most frequently modify the semantics of the evaluative content contained in them in our corpus. The column ‘total change’ refers to the total percentage of the use of a relation in this corpus that changes the polarity of the opinion words. The table breaks down all the individual types of polarity change associated with these discourse relations.

Discourse relation	Downtone	Intensify	Reversal	Total change
Concession	17%	23%	0%	40%
Contrast	15%	17%	3%	35%
Elaboration	3%	26%	2%	31%
Evaluation	6.5%	30.5%	2%	39%
Evidence	0%	44%	0%	44%
Restatement	12.5%	50%	0%	62.5%
V-result	0%	36%	0%	36%

Table 9: Discourse relations that typically modify the semantics of the evaluative content

The discourse relation which most frequently modifies the meanings of opinion words in our corpus is Restatement, followed by Evidence, Evaluation and Concession. Their preferred type of change is ‘Intensify’. Consider the examples (9 – 12) below, and how the underlined words are intensified because of the type of relation they are found in.

- (9) [The Knapely chapter of the Women's Institute (W.I.) isn't very interesting.]_N [In fact, it's downright boring.]_S (Restatement)
- (10) [Although a comedy, this movie does have some touching moments.]_N [Heck, it made me cry...]_S (Evidence)
- (11) [But I was quite shock to see her play such a role that was different that her usual roles.]_N [It was great!]_S (Evaluation)
- (12) [Betty isn't fond of Watson's subversive tactics]_N [although the rest of her classmates including valedictorian Joan Brandwyn (Julia Stiles) find them refreshing.]_S (Concession)

‘Intensify’ is the prototypical type of polarity change for all of the discourse relations. The discourse relation with this polarity type usually contains several units and several sentiment words. Only Contrast, Condition and Concession relations have a relatively similar number of ‘Intensify’ and ‘Downtone’ types, as was shown in Table 8. Consider the examples of Condition relation with ‘Intensify’ and ‘Downtone’ polarity types in (13) and (14).

- (13) [If this is what Hollywood thinks is quality childrens entertainment...]_S [they are mistaken.]_N (Condition, Intensify)
- (14) [The risk factor is minimal]_N [unless you overplay and come off as a caricature.]_S (Condition, Downtone)

All the discourse relations in the movie corpus have reversal as the least prototypical type of polarity change. In other words, a relation rarely changes the polarity of evaluative words to the opposite pole. Changes in polarity tend to be more subtle than full reversal.

6. Conclusion

In this corpus study, we investigated the interaction between evaluation (subjective content) and discourse relations in the movie reviews of the SFU Review Corpus (Taboada, 2008). The results show that relations such as Concession, Contrast, Evaluation and Result most frequently modify the polarity of the opinion words in terms of intensification of the evaluation. This modification is reached through creating the discourse relation out of several discourse units (clauses or sentences) or by using several opinion words within the discourse relation, or, as in the case of

Concession, through counterexpectational semantics. The study as well demonstrates that the majority of discourse relations (70.19%) do not change the polarity of opinion words that they contain. In cases where there is modification of sentiment, ‘Intensify’ is the preferred type of polarity change. In terms of the interaction between the nucleus-satellite structures and the type of polarity, although most opinion words are positioned in the nucleus, both nucleus and satellite show a similar distribution of polarity types with ‘No change’ being the leading polarity type, followed by ‘Intensify’ and then ‘Reversal’ being the least employed polarity. Chi-square tests showed that there is no correlation between the nucleus-satellite distinction and presence of polarity change/absence of polarity change, but that some relations show statistically significant tendencies to change the polarity of the words in their scope. Polarity change can be predicted for seven relations: Circumstance, Evaluation, Interpretation, Purpose, Restatement and Sequence (which are more likely to intensify evaluative words in their scope), and Concession (more likely to downtone evaluative words).

The results of this work are useful in practical applications, such as sentiment analysis. Determining that ‘no change of polarity’ is the most frequent type of polarity in our corpus and that opinion words are usually situated in the nucleus of a discourse relation suggests that nuclei can be the focus of the automatic extraction of evaluation.

Acknowledgements

This research was funded by the Social Sciences and Humanities Research Council of Canada. We thank Mara Katz for her help with the annotation.

References

- Asher, N., Benamara, F., & Mathieu, Y. (2008). *Distilling opinion in discourse: A preliminary study*. Paper presented at the Computational Linguistics (CoLing), Manchester, UK.
- Asher, N., Benamara, F., & Mathieu, Y. (2009). Appraisal of opinion expressions in discourse. *Linguisticae Investigationes*, 32(2), 279-292.
- Bal, K. B. (2014). Analyzing Opinions and Argumentation in News Editorials and Op-Eds. *International Journal of Advanced Computer Science and Applications, Special Issue on Natural Language Processing*, 22-29.
- Benamara, F., Chardon, B., Mathieu, Y., & Popescu, V. (2011). *Towards Context-Based Subjectivity Analysis*. Paper presented at the the International Joint Conference on Natural Language Processing (IJCNLP).
- Benamara, F., Popescu, V., Chardon, B., Asher, N., & Mathieu, Y. (2013). Assessing Opinions in Texts: Does Discourse Really Matter? In M. Taboada & R. Trnavac (Eds.), *Nonveridicality and Evaluation: Theoretical, Computational and Corpus Approaches*. Leiden: Brill.
- Biber, D., & Finegan, E. (1988). Adverbial stance types in English. *Discourse Processes*, 11(1), 1-34.

- Brill, E. (1995). Transformation-based error-driven learning and Natural Language Processing. *Computational Linguistics*, 21(4), 543-565.
- Carlson, L., Marcu, D., & Okurowski, M. E. (2001). *Building a discourse tagged corpus in the framework of Rhetorical Structure Theory*. Paper presented at the Second SIG dial Workshop on Discourse and Dialogue (SIGdial-2001), Aalborg, Denmark.
- Chardon, B., Benamara, F., Mathieu, Y., Popescu, V., & Nicholas, A. (2012). *Measuring the effect of discourse structure on sentiment analysis*. Paper presented at the International Conference on Intelligent Text Processing and Computational Linguistics (CICLING).
- Conrad, S., & Douglas, B. (2000). Adverbial marking of stance in speech and writing. In S. Hunston & G. Thompson (Eds.), *Evaluation in Text: Authorial Distance and the Construction of Discourse* (pp. 56–73). Oxford: Oxford University Press.
- Halliday, M., & Matthiessen, C. (2004). *An Introduction to Functional Grammar* (Third ed.). London: Arnold.
- Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Arnold.
- Hatzivassiloglou, V., & McKeown, K. (1997). *Predicting the semantic orientation of adjectives*. Paper presented at the the eighth conference on European chapter of the Association for Computational Linguistics.
- Heerschop, B., Goosen, F., Hogenboom, A., Frasincar, F., Kaymak, U., & de Jong, F. (2011). *Polarity analysis of text using discourse structure*. Paper presented at the the 20th ACM.
- Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews*. Paper presented at the the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).
- Hunston, S., & Thompson, G. (Eds.). (2000). *Evaluation in Text: Authorial Distance and the Construction of Discourse*. Oxford: Oxford University Press.
- Langacker, R., W. . (1985). Observations and Speculations on Subjectivity. In J. Haiman (Ed.), *Iconicity in Syntax* (pp. 109-150). Amsterdam & Philadelphia: John Benjamin.
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press,.
- Lyons, J. (1981). *Language, Meaning and Context*. London: Fontana.
- Mann, W. C., & Taboada, M. (2015). RST Web Site, from <http://www.sfu.ca/rst>
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3), 243-281.
- Martin, J. R. (2000). Beyond exchange: appraisal systems in English. In S. Hunston & G. Thompson (Eds.), *Evaluation in Text: Authorial Distance and the Construction of Discourse* (pp. 142–175). Oxford: Oxford University Press.
- Martin, J. R., & White, P. R. R. (2005). *The Language of Evaluation*. New York: Palgrave.
- Mathieu, Y. (2004). A Computational Semantic Lexicon of French Verbs of Emotion. In G. Shanahan, Y. Qu & J. Wiebe (Eds.), *Computing Attitude and Affect in Text* (pp. 109-124). Dordrecht: Springer.
- O'Donnell, M. (1997). RSTTool, from <http://www.wagsoft.com/RSTTool/>
- O'Donnell, M. (2008). *The UAM CorpusTool: Software for corpus annotation and exploration*. Paper presented at the XXVI Congreso de AESLA, Almeria, Spain.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1–135.

- Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. In J. G. Shanahan, Y. Qu & J. Wiebe (Eds.), *Computing Attitude and Affect in Text: Theory and Applications* (pp. 1-10). Dordrecht: Springer.
- Somasundaran, S., Ruppenhofer, J., & Wiebe, J. (2007). *Detecting arguing and sentiment in meetings*. Paper presented at the the SIGdial Workshop on Discourse and Dialogue.
- Soricut, R., & Marcu, D. (2003). *Sentence level discourse parsing using syntactic and lexical information*. Paper presented at the Human Language Technology and North American Association for Computational Linguistics Conference (HLT-NAACL'03) Edmonton, Canada.
- Taboada, M. (2008). SFU Review Corpus [Corpus]. from Simon Fraser University http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html
- Taboada, M., Brooke, J., & Stede, M. (2009). *Genre-based paragraph classification for sentiment analysis*. Paper presented at the the 10th Annual SIGDIAL Meeting on Discourse and Dialogue, London, UK.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics* 37(2), 267-307.
- Taboada, M., & Grieve, J. (2004). *Analyzing appraisal automatically*. Paper presented at the AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07), Stanford University, CA.
- Taboada, M., & Mann, W. C. (2006a). Applications of rhetorical structure theory. *Discourse Studies*, 8(4), 567–588.
- Taboada, M., & Mann, W. C. (2006b). Rhetorical Structure Theory: Looking Back and Moving Ahead. *Discourse Studies*, 8(3), 423-459.
- Thompson, G., & Alba-Juez, L. (Eds.). (2014). *Evaluation in Context*. Amsterdam: John Benjamins.
- Tofiloski, M., Julian, B., & Taboada, M. (2009). *A Syntactic and Lexical-Based Discourse Segmenter*. Paper presented at the 47th Annual Meeting of the Association for Computational Linguistics, Singapore.
- Trnavac, R., & Taboada, M. (2012). The contribution of nonveridical rhetorical relations to evaluation in discourse. *Language Sciences* 34(3), 301-318.
- Trnavac, R., & Taboada, M. (2014). *Discourse relations and affective content in the expression of opinion in texts*. In G. Kotzoglou et al. (eds), *Selected Papers of the 11th International Conference on Greek Linguistics*. Rhodes, Greece. 1705-1715.
- Voll, K., & Taboada, M. (2007). *Not all words are created equal: extracting semantic orientation as a function of adjective relevance*. Paper presented at the the 20th Australian Joint Conference on Artificial Intelligence, Gold Coast, Australia.
- White, P. R. R. (2003). Beyond modality and hedging: a dialogic view of the language of intersubjective stance. *Text*, 23(2), 259–284.
- Wiebe, J. (2000). *Learning subjective adjectives from corpora*. Paper presented at the 17th National Conference on Artificial Intelligence (AAAI), Austin, TX.
- Wierzbicka, A. (1987). *Speech Act Verbs*. Sydney: Academic Press.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3), 399-433.