Loving and hating the movies in English, German and Spanish*

Maite Taboada, Marta Carretero and Jennifer Hinnell

Simon Fraser University (Canada), Universidad Complutense de Madrid (Spain), University of Alberta (Canada)

We present a quantitative analysis of evaluative language in a genre in which it is particularly prominent, that of movie reviews. The data chosen are non-professional consumer-generated reviews written in English, German and Spanish. The reviews are analyzed in terms of the categories of Attitude and Graduation within the Appraisal framework (Martin & White, 2005). The results show a clear influence of the genre on the relative frequency of the different types of Appraisal categories. This influence accounts for a number of the similarities found in the evaluative expression across the three languages, such as the predominance of Appreciation over other Appraisal categories, and the relation between the global positive or negative polarity of the reviews and the individual polarity of the spans in the reviews. We also found cross-linguistic differences, for example, those related to word order and argumentative style.

Keywords: Appraisal, evaluation, genre, consumer reviews, English/German/Spanish

1. Introduction

After expressions of desires and needs, the linguistic manifestation of emotion and evaluation is probably one of the most basic functions of language (cf.

* We would like to thank Nicola Bergen and Patrick Larrivée-Woods for initial annotation of the English data. This work is funded by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 261104-2008; PI: Maite Taboada), and by the Spanish Ministry of Science and Innovation, under the project *Creación y validación de descripciones contrastivas (inglés-español) a través del análisis y la anotación de corpus: Aspectos lingüísticos, metodológicos y computacionales* (Creation and validation of contrastive descriptions (English-Spanish) through the analysis and annotation of corpora: Linguistic, methodological and computational aspects), with reference number FFI2008-03384 (PI: Julia Lavid, Universidad Complutense de Madrid).

Taboada, M., M. Carretero and J. Hinnell (2014) Loving and hating the movies in English, German and Spanish. *Languages in Contrast* 14 (1): 127-161. (Pre-publication version)

Halliday's (1975) personal function). One of the aspects of language that we make use of every day is the expression of our own emotions, and the evaluation of persons and objects around us. This expression of emotions and evaluations is studied under different umbrella terms in linguistics and other social sciences. Studies of affect (Batson et al., 1992), subjectivity and point of view (Banfield, 1982; Langacker, 1990; Traugott, 1995, 2010), evidentiality (Chafe & Nichols, 1986; Aikhenvald, 2004), attitudinal stance (Biber & Finegan, 1988, 1989), modality (Palmer, 1986; Bybee & Fleischman, 1995; Portner, 2009) and appraisal (Martin & White, 2005), to mention just a few in each area, all aim at explaining how we use language to convey emotions, evaluation and subjective expressions.

As this body of work shows, researchers have been interested for a long time in how we use language to express evaluation and subjectivity, but even more so recently, spurred in part by interest in the automatic extraction of opinions found online (for surveys, see Pang & Lee, 2008; Liu, 2012). At the same time, research in Systemic Functional Linguistics (Halliday, 1985; Halliday & Matthiessen, 2004) has resulted in a now mature proposal for classifying and studying evaluation, the Appraisal framework (Martin & White, 2005). With some notable exceptions, most of that research is carried out in English and, when other languages are studied, they tend to be approached in isolation, without comparison across languages.

In this paper, we present a corpus study of evaluation in a common online genre, that of movie reviews, carried out contrastively across three languages: English, German and Spanish. We collected data from similar sources and with similar characteristics, and annotated it using the categories proposed within Appraisal. Our analysis shows that, although there are differences across the languages, the common genre seems to lead to a similar breakdown of Appraisal categories. Differences across languages are centered, on the one hand, around the lower levels of language (different word order configurations) and, on the other hand, around the higher-level discourse organization.

In Section 2 we provide a summary of the Appraisal framework, and its place in Systemic Functional Linguistics. Section 3 is devoted to explaining our methodology for annotation, including a description of the corpus used. Results for each of the languages, and a discussion of these results, are provided in Section 4, followed by a general discussion and conclusions in Section 5.

2. Appraisal in English and other languages

Appraisal belongs in the systemic functional tradition 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, 2003b; Martin & White, 2005).

Martin (2000) characterizes appraisal as the set of resources used to negotiate emotions, judgements, and valuations, alongside resources for amplifying and engaging with those evaluations. He considers that appraisal resources form a system of their own within the language (following the Systemic Functional Linguistics), and divides the Appraisal system into three distinct sub-systems (see Figure 1): Attitude, Engagement and Graduation, which will be described below. Since Martin's approach is lexically rather than

grammatically based, he is primarily concerned with those words and semantic categories of words that allow a speaker to express different types of evaluations. Figure 1 summarizes the Appraisal network, with some example realizations, showing prototypical cases within the respective subsystems.

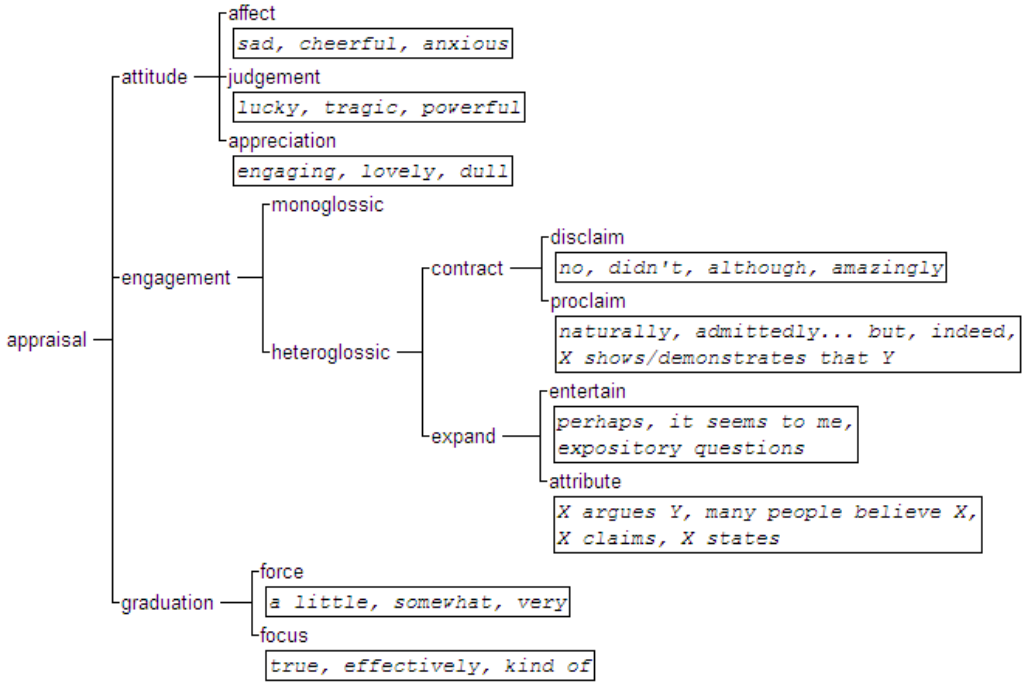

Figure 1. Appraisal system.

Let us now briefly discuss the main sub-systems of Appraisal. Attitude has three subsystems: Affect is used to construe emotional responses about the speaker or somebody else's reactions (e.g., *happiness, sadness, fear*); Judgement conveys moral evaluations of character about persons or, less commonly, non-human entities (e.g., *ethical, deceptive, brave*); and Appreciation captures aesthetic qualities, most often of objects and natural phenomena (*remarkable, desirable, harmonious, elegant, innovative*). In earlier work (Taboada & Grieve, 2004), we characterized each system as appearing in prototypical sentences: Affect may be conveyed by adjectives that appear in sentences such as *I was X* (i.e., *I was sad, I am scared*). Judgement uses *He was X* (*He was brave, He was a coward*), whereas Appreciation is seen in the *It was X* pattern (*It was interesting, It was beautiful*). Martin (2003) proposed the frames *I feel (very) X; It was X of him to do that;* and *I consider it X* for Affect, Judgement and Appreciation, respectively. Those are, naturally, simplified prototypical patterns, and all correspond to inscribed instances, those that are explicitly expressed in the text. Instances that are not inscribed are considered to be evoked, in which "an evaluative response is projected by reference to events or states which are conventionally prized" (Martin, 2000). Thus, *a bright kid* or *a vicious kid* are inscribed. On the other hand, *a kid who reads a lot* or *a kid who tears the wings off butterflies* present evoked appraisal.

The Engagement system refers to the distinction between heteroglossic and monoglossic expressions, following proposals by Bakhtin (1981). In a heteroglossic expression, inter-subjective positioning is open, because utterances

invoke, acknowledge, respond to, anticipate, revise or challenge a range of convergent and divergent alternative utterances (White, 2003b, 2003a; Martin & White, 2005). The alternative is monoglossia, where no alternative view or openness to accept one is present. Monoglossic utterances are presented as facts. Within heteroglossia, the two possibilities are contract and expand, depending on whether possibilities for different opinions are either limited or open. Contract can, in turn, take the form of disclaim (position at odds with or rejecting some contrary position) or proclaim, where a speaker or writer sets themselves against, suppresses or rules out alternative positions. Examples of disclaim are negative statements; proclaim can be expressed through adverbials such as *naturally* or *admittedly* as seen in Figure 1. When expansion is possible, the two possibilities are entertain, where the speaker represents the position as one of a number of possible positions, and attribute, where the proposition is presented as externally grounded, in the words of another speaker.

Finally, the Graduation system allows modulation of the evaluation, by using force or focus to intensify or downtone gradable words (force), or to sharpen or soften words that are usually non-gradable (focus). Examples of intensification and downtoning are *somewhat interesting* and *a little bit sad*. In *a true friend* the meaning of *friend*, usually a non-gradable word, is sharpened. On the other hand, *a kind of friend* implies a softening of the meaning.

In this paper, we are focusing on the two systems of Attitude and Graduation: The analysis will be based on the spans of Attitude and how they are emphasized or downtoned by spans of Graduation. Engagement is undoubtedly important, but, given its complexity, it falls outside the scope of our current annotation work. We are investigating, in related work, how Engagement overlaps with other areas of the linguistic system, such as modality (Carretero & Taboada, to appear) and the expression of non-veridicality (Trnavac & Taboada, 2012).

Although much ground remains to be covered, Appraisal is quite well understood in English, with a wide range of studies dealing with different genres, from political discourse/news stories (White, 1998; Coffin & O'Halloran, 2006) and different types of narratives (Macken-Horarik, 2003; Page, 2003), including those produced by school children (Martin, 1996; Coffin, 1997) and by children in the process of language acquisition (Painter, 2003) to discussions of literary texts (Love, 2006) and casual conversation (Eggins & Slade, 1997).

A few studies for other languages exist. For Spanish, Kaplan (2007) studied television news (including both sound and images) and how the news editors mark their own point of view, and Achugar (2008) tracked the construction of memory during Uruguay's military dictatorship through, in part, Appraisal analyses of historical documents. In German, Becker (2009) studied English-German political interviews, focusing on the expression of Engagement.

From a more general point of view, the three languages in this paper have been studied as pairs, although there is no work covering all three at the same time. House has extensively studied German-English differences, and has observed differences in communicative style, such as more directness and explicitness in German as opposed to English (House, 2006).

The other main work we should mention here is Johansson's study of *love* and *hate* in English and their cognates in Norwegian, which inspired the title of this paper (Johansson, 1998). Johansson found that the verbs *love* and *hate* are

used much more frequently in Norwegian translations of English than in original Norwegian texts, because of their higher frequency in English. We will see that *love* and *hate* and their equivalents in German and Spanish are actually quite infrequent in our corpus, because they express Affect, which, as we will see, is not very common in our corpus, in contrast to Appreciation.

3. Corpus and methodology

3.1 Tri-lingual comparable corpus

We collected a three-way corpus of movie reviews in English, German and Spanish. The English corpus comes from the website Epinions; the Spanish one from Ciao; and the German corpus from the German version of Ciao[1]. Most of the texts are part of the SFU Review Corpus (Taboada, 2008), a larger collection of reviews of books, movies and other consumer products, portions of it annotated with Appraisal labels and with Rhetorical Structure Theory relations (Mann & Thompson, 1988).

The reviews are all written by non-experts and posted online, with the purpose of being informative to other potential viewers. The genre differs, in formality and structural characteristics, from that of critical reviews written by professional movie critics and printed in newspapers.

For this paper, we selected 50 reviews for each language, equally divided between favourable reviews (positive reviews) and unfavourable reviews (negative reviews), as indicated by the reviewer's label of "recommended" or "not recommended". Because the texts were collected at different times, they review many different movies, typically those recently released at the time. The English corpus, collected in 2004, reviews the following movies: *Bad Santa, Calendar Girls, The Cat in the Hat, Elf, Gothika, The Haunted Mansion, The Last of the Samurai* and *Mona Lisa Smile.* The Spanish corpus, collected in 2008, contains reviews of a number of Hollywood movies (*27 Dresses, Bridge to Terabithia, Indiana Jones and the Kingdom of the Crystal Skull, Juno, Ratatouille, Sex and the City, Shrek 3, The Dark Knight* and *Wall-E* among them), as well as some Spanish ones (*El orfanato, Los crímenes de Oxford*). Finally, the German corpus was collected in 2011, and it reviews primarily 2010-2011 Hollywood hits in their German translations, among them *True Grit, Freundschaft Plus (No Strings Attached), 127 Hours, Tron Legacy, The American, Little Fockers, Gullivers Reisen (Gulliver's Travels), Unstoppable-Ausser Kontrolle (Unstoppable), The Road, The Green Hornet,* and *Avatar 3D*, as well as original German films, such as *Der letzte schöne Herbsttag*.

The English corpus has been analyzed elsewhere (Taboada, 2011) as an instance of a particular genre, with a detailed description of the stages it contains. Upon examination (but not full analysis) of the texts in the other two languages, we can state that the generic structure is quite similar to that of English. This similarity is to be expected, since the corpus is comparable, in that the genre is the same across languages; the medium, the writers and the audience are assumed to be similar; and the texts all deal with the common experience of watching a film and commenting on it. Although there are

[1] The websites are, respectively: www.epinions.com, www.ciao.es and www.ciao.de.

potential differences in the three languages and cultures, we found that the generic structure was very similar, with Description and Evaluation stages present in all texts. These stages can be sub-classified as description or evaluation of plot, characters, or other aspects of the movie. In addition, some reviews contain optional Background and Subject Matter stages, where additional details or a general summary of the movie's content are provided. For analysis of the English stages, see Taboada (2011). In terms of register (Halliday, 1989; Eggins & Martin, 1997), they all share similar properties of field, tenor and mode, with the field being discussion and evaluation of a movie; the tenor being informal (but with an unknown audience); and the mode written.

Table 1 shows a summary of the number of words and sentences for each language. The sentence count is approximate, and it is derived from UAM CorpusTool[2], the system used to annotate the data (see the next section for more detail).

Table 1. Corpus statistics.

	English	German	Spanish
Words	33,387	47,128	52,845
Words, favourable reviews	17,219	25,401	36,071
Words, unfavourable reviews	16,168	21,727	16,774
Sentences	1,766	2,608	2,593

3.2 Annotation methodology

The annotation was carried out by the three authors of the paper. To ensure that it was consistent across all annotators, we designed a procedure that outlined clearly the segments to be annotated, and provided general guidelines for the annotation.

We found early on that the most difficult task was not so much how to annotate particular segments, but to decide what to annotate, that is, what constituted a span or segment. In order to help with those decisions, and to keep annotation consistent across languages and annotators, we created an annotation manual, and have reported preliminary analyses of the English and Spanish data (Taboada & Carretero, 2012). The main aspects are summarized in the remainder of this section.

First of all, we decided that *markables* (i.e., units to be annotated) are those aspects that refer to the movie itself, not to aspects external to the movie, such as opinions of the reviewer about anything other than the movie, reasons why they went to the theatre on a particular day, or references to other movies involving the same professionals (director, actors, scriptwriters, etc.). Similarly, evaluative aspects about the characters or atmosphere of the movie are excluded. For instance, the description of a character as *this mean, controlling mother* in Example (1)[3] is not part of the artistic achievement of the movie. The description of the performances as *clichéd*, or of Mitchell's character as *a little too much* is, however, the territory of the scriptwriter and the director, and

[2] http://www.wagsoft.com/CorpusTool/

[3] Examples are reproduced verbatim, including any typos, grammatical errors and punctuation errors (such as lack of accentuation in Spanish). Because some of the spans that we reproduce are incomplete, examples do not always have final punctuation.

therefore part of the artistic merit of the movie. We identify markables with underlining.

(1) While the smaller performances of Juliet Stevenson, Marian Seldes, and Donna Mitchell as Dunst's mother were used to a good degree despite their clichés while [sic] Mitchell's character as this mean, controlling mother was a little too much. [no1]

The reason to exclude aspects not intrinsic to the movie is that the annotation work is part of a larger project that aims at extracting opinion automatically, by processing texts and providing a score that corresponds to the text's sentiment about the movie, book or product being reviewed (Taboada et al., 2011). Because opinions about aspects other than the movie are not relevant in such a task, we decided not to annotate them.

In terms of the categories to be annotated, we focused on the Attitude and Graduation systems, leaving Engagement for future work. In our initial annotations, we realized that Engagement is a much more complex system, not only because of the different types of parts of speech involved (verbs, adverbs, etc.), but also because it often entails annotating large chunks of the text, as Engagement tends to be expressed through longer phrases or even entire sentences. Despite our decision to leave Engagement out for now, in some cases Engagement was close to Attitude or Graduation. For instance, in Example (2), *definitely* expresses Engagement, but its force in conjunction with the adjective *long* is very close to that of intensifiers. Another case is presented in (3), where there are two downtoners, *I guess* and *if you are into those types of things*. The second one we have annotated as part of the expression of Attitude, but we believe that *I guess* expresses mainly Engagement, and should not be part of the markable.

(2) Definitely too long [no17]

(3) neat, I guess, if you are into those types of things [no19]

Within Graduation,, annotation was complicated in cases where there were markables containing two types of graduation. Due to limitations in CorpusTool, it is not possible to annotate both instances without separating the span. In these cases the element with wider scope took precedence. For example in (4), *einfach* ("simply") takes scope over *absolut*, and therefore it is annotated (as focus), and the intensification of *absolut* (normally force) goes unannotated.

(4) …*einfach absolut unlustig* [nein1]

"…simply altogether not funny"

The main criterion to include a span as a markable was Attitude. Graduation was only annotated if it was associated to a span of Attitude. In addition, the Attitude span had to be clearly polar, that is, either positive or negative. The final system we used for annotation, excluding Engagement, but with an expanded Graduation system, is presented in Figure 2.

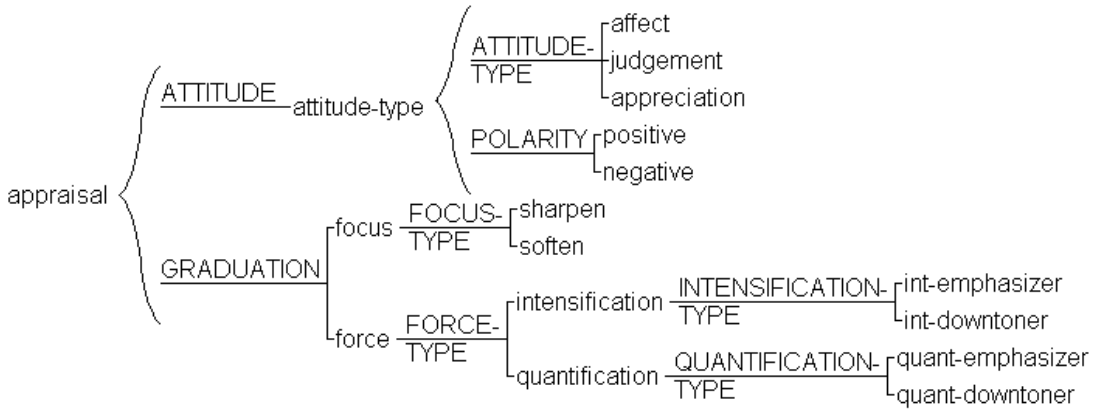

Figure 2. Appraisal system used for the annotation.

Once the topics or aspects to be annotated have been decided, a number of other decisions affect the scope of the constituents, e.g., how much of the context of a particular word or expression should be included in the annotation. In general, we include only evaluative words and those words that are part of the evaluative content. For adjective+noun combinations, only the adjective is included, unless the noun also conveys evaluation. In *a good girl,* then, *good* is the only markable. A few exceptions to this general rule are detailed below.

All of the expressions of Attitude were classified as either positive or negative. In fact, the main criterion for deciding whether a unit was a markable (an expression of Attitude) or not was whether we could attribute polarity to it. In very few cases we felt that there was Attitude but no clear polarity. The decision in those cases was to either discard them as markables, or to assign the most likely polarity. In (5), for instance, there are two such examples (*pretty standard Disney* and *par for the course*). In both cases, there is a clear evaluation, but the polarity is close to neutral. We decided to include them as markables, and classify them as positive.

(5) You begin the film with a married couple with 2 kids (pretty standard Disney stuff), and the dad is too busy at work (again, par for the course, but we have been doing this since at least "Mary Poppins," you know?) [yes18]

Below we outline the main areas where we had to make decisions as to what to annotate, or how to annotate it.

1. **Coordinated elements** with *and* are part of the same span, but elements coordinated with commas or other punctuation are separate. This is because coordinated spans with *and* tend to convey the same Attitude, whereas punctuation may indicate contrast rather than addition. Coordination with *and* tends to refer to nouns and adjectives, but some cases merit the inclusion of two separate clauses in one span. In (6), two clauses (*took the rights they had* and *ran with them*) are coordinated, with subject ellipsis in the second one. The Attitude (negative Judgement in this case) is similar. Moreover, the

example is evocative of the idiomatic expression *take the money and ran*, which would be treated as a single span (see below).

(6) In THE CAT IN THE HAT, the producers and writers took the rights they had and ran with them, [no15]

This generalization also covers cases where the two coordinated elements are both modified by the same adjective that precedes the first element, as in (7), since both adjectives are in the scope of the modifier[4].

(7) *zu zurückgezogen und introvertiert* [nein10]

"too withdrawn and introverted"

However, it can also be the case that two adjectives in a coordinated span each have different modifiers. In these cases the spans are annotated separately. One further exception to annotating coordinated elements as one span, is in strings of more than two adjectives as in "X, Y and Z", where the first pair items are separated by a comma and the second pair are separated by the conjunction *and*. In these cases the adjectives are annotated individually to yield three separate spans, since there is less likelihood of all three conveying the same Attitude.

2. **Comparatives and superlatives** are included completely, because the term of the comparison is also part of the evaluative content. In (8), the main evaluation is *sosa* ("dull"), but the comparison contributes to understanding the intensity of that evaluation, and is thus included. Similarly in (9), where the main evaluation is *dünn* ("thin"), but the comparison "as thin as Lindsay Lohan's hair extensions" identifies the intensity of the evaluation and is therefore included.

(8) *…esta vez la trama es tan sosa que permanezco impasiva toda la película.* [no1-3]

"… this time the plot is so dull that I remain impassive throughout the movie."

(9) *Die Story ist so dünn wie Lindsay Lohan's Haarverlängerungen* [nein4]

"The story is as thin as Lindsay Lohan's hair extensions"

3. **Modal auxiliaries** to an evaluative verb are included in the markable. The modal verb often affects the Graduation conveyed by the verb.

(10) … it is fantastically watchable and jam-packed with witty one-liners and should appeal to a wide ranging audience. [yes3]

In the German example in (11), the modal *möchte* in *möchte empfehlen* ("would like to recommend") downtones the recommendation.

(11) *Ich möchte den Film empfehlen (er ist aber beileibe nichts für einen gemütlichen Familien Kinoabend)* [ja12]

"I would like to recommend the film (though it is by no means something for a cozy family movie night)"

[4] Translations into English of the German and Spanish examples are our own.

4. An entire **idiomatic expression** is a markable, since it cannot usually be decomposed into constituent parts. In (12), we see an idiomatic expression in Spanish, *tirar por la borda* (lit. "to throw something overboard", i.e., "to not make good use of something").

   (12) *una historia que puede dar tanto de sí y que se tira por la borda* [no2-19]

   "a story with so much potential and that is thrown overboard"

   The entire idiomatic expression is considered as a markable, even if its constituents are separated by other constituents. That is to say, if the word order had been *se tira una historia por la borda* (lit, "is thrown a story overboard"), the span would have included *la historia*. An alternative would have been to consider *se tira* and *por la borda* as separate constituents. This procedure, however, would lead to an artificially high number of markables, since those two items are properly part of the same expression.

5. **Repetitions** are sometimes used for emphasis. They are considered as a single span, and the repetition is annotated as Graduation:force: intensification.

   (13) *una película sosa sosa* [no1-1]

   "a dull dull movie"

6. **Conditionals** and similar constructions recommending a movie are included in their entirety, since it is the combination of protasis and apodosis that conveys the evaluation. In (14), it is the combination of the two clauses that results in a negative evaluation of the movie under review, *The Cat in the Hat.*

   (14) A dog even pees on someones food. If this is what Hollywood thinks is quality childrens entertainment… they are mistaken [no16]

   This includes imperative-like conditionals, which occur frequently in the German corpus as in (15).

   (15) *Spart euch das Geld und die Zeit und investiert es sinnvoller* [nein9]

   "Save yourselves the money and time and invest it more wisely"

   Also included as full spans are recommendations (or non-recommendations as in the case below), where the sentiment results from the main clause containing a setup, followed by a subordinate (here comparative) clause, as in (16)

   (16) *Diese Aneinanderreihung von Banalitäten interessieren genauso viel wie die Tatsache, dass nach dem Kinobesuch mein Schuhband aufgegangen war.* [nein1]

"This sequence of banalities is about as interesting as the fact that my shoelace came undone after the movie."

7. **Causality**. Causation contains two parts, one for the causer and another for the effect. The word expressing cause, if there is one, is excluded. But the division should be such that none of the spans is discontinuous. If the expression of Affect is modified by a comparison, it is an expression of Graduation under its span. In (17), *this confusion* is the causer, and therefore not included. The span here is *bigger laughs*, the effect (Affect, with force:intensification).

   (17) However this confusion was the cause of some of the film's bigger laughs [yes1]

8. **Word order.** In German it is frequently the case that spans are quite long, due to aspects of morphosyntax such as the presence of separable prefix verbs, or the verb second constraint. In the case of separable prefix verbs, the span has to include the verb stem since the prefix is not really intelligible without the verb, as in (18), where the infinitive *aussehen* is separated to *sieht … aus* in its finite form. The example includes the separable prefix *aus* in the middle of the span, but also the verb *sieht* at the beginning, as the prefix without the verb would make no sense. Compare this with the translation of this span into English, where the verb "looks" would not be included in the markable if this were an original English sentence.

   (18) *Steven Seagal ist ganz schön fett geworden und sieht so schmierig aus wie ein Liter Motoröl* [nein4]

   "Steven Seagal has gotten really fat and looks as slimy as a litre of motor oil."

   The verb-second constraint in German also created some overly long spans. For example, in (19), the conjugated modal (*hätte*) occurs in second position in the sentence, and the rest of the verb phrase, including an additional modal, is in sentence final position. This would normally mean that the full sentence is annotated; however, in order to avoid including unnecessary information in the markable, the modal carrying tense was omitted from the span, as the modality was present in the span through the second modal. Note that in the English translation of the German double-modal construction contains only the modal 'should'.

   (19) *Hier hätte man die Figuren durchaus stärker beleuchten dürfen* [nein10]

   "Here one should really have lit the characters powerfully."

9. **Information structure.** In German it became important to mark spans within spans. German tends to present more information in one clause; English by contrast favours subordinate clauses for this purpose. In English, the sentence in (20) would have been more authentically written as (21). In general, German tends to be accommodating of long passages, so double-layered annotations such as the one in (20) were needed more frequently.

(20) *Barbara Hershey als Ninas Mutter […] ist ebenso überzeugend wie der großartige Vincent Cassel als Mischung aus gnadenlosem Intendanten und erotischem Lehrer*. [ja15]

"Barbara Hershey as Nina's mother is just as convincing as the amazing Vincent Cassel as a mix of merciless director and erotic teacher"

(21) Barbara Hershey as Nina's mother is just as convincing as Vincent Cassel, who was amazing as a mix of merciless director and erotic teacher.

We completed several iterations of the annotation, comparing first the markables, to check that the spans selected were comparable, and then the annotations themselves, to make sure we were all consistent in our annotation procedure. Once we were satisfied with the level of comparability across annotators, we proceeded with the annotation of the full corpus, the results of which are described in the next section.

4. Results

Table 2 summarizes the number of spans annotated for each language, and the breakdown in terms of the three main categories annotated (Affect, Judgement and Appreciation within the Attitude category). The analysis was also broken down into reviews that provided a favourable assessment of the movie ("positive") and those that provided an unfavourable one ("negative").

The reviews, overall, are dominated by spans of Appreciation. This is to be expected, as the reviews are about objects of art, the prime example of Appreciation. Nevertheless, Affect and Judgement contribute to the total, with a somewhat surprising higher number of spans of Judgement. One may expect for Affect to be second after Appreciation, since the reviews are likely to convey the emotions that the movies triggered in the authors. However, Judgement about the cast and director, and to a certain extent about the characters, takes precedence, and for German the Judgement proportion is significantly higher. In other words, the reviewers focus on the object to be reviewed, including cast and director (and sometimes criticism of audience members), over their own emotions when watching the movies. We could say that the reviews are not about "loving and hating the movies", but about the artistic merit of the movies as works of art. The rest of this section summarizes results for each language.

Table 2. Summary of Appraisal realization in the three languages.

	English		**German**		**Spanish**	
	n	**%**	***n***	**%**	***n***	**%**
Spans (Attitude)	1,312		1,882		1,950	
Affect	203	15.47	121	6.43	389	19.95
Judgement	400	30.49	447	23.75	481	24.67
Appreciation	709	54.04	1,314	69.82	1,080	55.38
Graduation	553	42.14	1,030	54.73	965	49.49
Total number of words	33,387		47,128		52,845	

4.1 English

The reviews in English are, as in the other languages, dominated by spans of Appreciation. Table 3 shows that, in most cases, the distribution of the three subcategories is very similar in positive and negative reviews. In both, Appreciation still accounts for the majority of Attitude spans, followed by Judgement, with Affect a distant third.

Unlike in Spanish (see Section 4.5 below), the number of spans in positive and negative reviews are quite similar, as are the number of words in each (17,034 in positive reviews; 15,971 for negative).

Table 3. Appraisal in positive and negative reviews (English).

	Positive		**Negative**	
	n	**%**	***n***	**%**
Spans (Attitude)	649		663	
Affect	109	16.80	94	14.18
Judgement	195	30.05	205	30.92
Appreciation	345	53.16	364	54.90
Polarity	649		663	
Positive	509	78.43	167	25.19
Negative	140	21.57	496	74.81
Spans w/o Graduation	362	55.78	397	59.88
Spans w. Graduation	287	44.22	266	40.12
Focus	38	13.24	48	18.05
Force	249	86.76	218	81.95
Focus type				
Sharpen	32	84.21	32	66.67
Soften	6	15.79	16	33.33
Force type				
Intensification	173	69.48	136	62.39
Quantification	76	30.52	82	37.61
Intensification type				
Int. emphasizer	113	65.32	106	77.94
Int. downtoner	60	34.68	30	22.06
Quantification type				
Qu. emphasizer	50	65.79	72	87.80
Qu. downtoner	26	34.21	10	12.20

Naturally, the proportions of positive and negative spans are reversed in each type of review. However, positive reviews still contain a fair amount of negative evaluation (over 21%), and, conversely, negative reviews contain about 25% of positive spans. The reasons for this are twofold: First of all, even though the overall evaluation may be, for instance, negative, there are often positive aspects in the movie that the authors point out. The second related reason is that the reviews typically follow a rhetorical pattern where positive aspects are mentioned first, and then the writer moves on to negative ones. This is a common pattern in (North American) English discourse, in both written and spoken informal discourse. A negative evaluation by itself is perceived as too blunt, and is often softened by a few initial words of praise. A short example of this pattern is presented in (22), part of a negative review. This is the second paragraph in the review, one that the author uses to mention “the good stuff”. The rest of the review contains a negative appraisal of the plot, the length and the movie overall. Even within this paragraph, the pattern is repeated, with a positive appraisal first and a criticism second.

(22) The good stuff….the visual production itself with its ultra-stylized appearance. It looks nice, but did the budget of a third world nation need to be spent to create this film? [no17]

This positive-first, negative-mostly pattern does not seem to be as widespread in its reverse form when the overall appraisal is meant to be positive, but we still find some examples with a similar structure. In (23) we can read the first few sentences of a very positive review. It starts with the author's misgivings about the film before he or she saw it, but soon the tone changes to one of positive evaluation.

(23) I've got to admit that the first time I saw the trailer for Calendar Girls I wasn't exactly charged with anticipation. The trailer did manage to make the film look fun, but the concept strongly reminded me of The Full Monty, which is a film I personally found a little overrated. If I wanted to add further concerns I could tell you about how I decided to search Epinions for some reviews on the film, and my search revealed nothing but a small selection of more, adult titles. Nevertheless I recognized that not only did the trailer look fun, but I had been pleasantly surprised by another British comedy that didn't appeal to me, Bend It Like Beckham. With those thoughts running through my mind I decided to throw caution to the wind and go down to the local cinema to see Calendar Girls.

I'm glad I went to see it now, because the film has remained just that. Fun! … [yes1]

Expressions of Attitude span different parts of speech, with adjectives being most frequent (*convincing, worth seeing, pathetic, irritating*), followed by mental processes (*really enjoy*) and behavioural processes (*hadn't laughed that hard or that consistently at just a preview in a very very long time*) and other processes (*groaned, hailed, can never stop laughing*), nouns (*masterpiece, clichés, scene-stealer*) and adverbials (*feebly, willy-nilly*). In some cases, an entire sentence conveys the opinion and is therefore a span (*Go see this movie*). There are frequent ready-made phrases (*two thumbs way up*), but also interesting new expressions (*[Jon must] think that if he chops up a story up into a bunch of little skits, inserting a ton of physical comedy, then a movie will magically rise from it all*).

4.2 Graduation in English

About 42% of the expressions of Attitude in the English reviews contained some type of Graduation. We devote this section to the complexities of Graduation.

The overwhelming majority of expressions of Graduation convey force, rather than focus, in both the positive and negative reviews. Force has two different aspects, intensification and quantification, with the former being much more frequent, in similar proportions for the negative and positive reviews. It seems to us that this may be a general phenomenon in English. Furthermore, emphasizing intensifiers are more common than downtoning intensifiers.

Intensifiers are typically adverbs of graduation, such as *very, incredibly, extremely* or *somewhat, kind of/kinda*. In some cases, more creativity is deployed, as in (24), where an expression of attitude (*hilariously)* becomes an expression of Graduation and intensifies the adjective *funny*. In this example, and all that follow, underlining indicates the entire markable, and bold the part that conveys Graduation. In (25), the concession expressed through *still* serves as a downtoner.

(24) That's what makes Bad Santa such a **hilariously** funny movie… [yes21]

(25) It doesn't really deserve many nominations, much less the wins it's somewhat likely to receive. Nevertheless, it's **still** above-average in my opinion, and though it has a lot going against it, it has a lot going for it as well. [yes6]

Most expressions specialize and act as either intensifiers or downtoners, but *few* and *a few* appear in both cases. For instance, *a few mistakes* is a downtoned expression, whereas *a few brilliantly crafted scenes* seems to use *a few* as an intensifier.

As for quantification, the most common realization is in superlatives (*one of the best movies out*), expressions with *too* (*too many, too much*), and expressions of quantity (*exactly one funny moment).* Example (26) quantifies by referring to audience age.

(26) I know that, **even at 11**, I would have been disturbed by the sexual jokes in it, had I understood them. [no15]

In both intensification and quantification, emphasizers are more numerous than downtoners. This seems to be a result of the straightforward nature of the reviews, which tend to lack subtlety. When downtoning is used, it tends to be an instance of litotes, the negation of the opposite opinion, such as *not really good*, where *really*, usually an intensifier, becomes a downtoner because it is part of the negation.

Example (27) shows quantification realized as a conditional clause. The positive assessment of *good* is quantified, and downtoned, by the conditional clause *only if you know when to stop*. Similarly, in (28), the endorsement in *suggest* is downtoned by the conditional.

(27) OK, I'll give credit here, although I really didn't like the path they took to get there, the story line was basically the same. Fun is good, **but only if you know when to stop**. [no14]

(28) As for home viewing, I would suggest it **if you have a good bass backing**. [yes20]

We have found that conditional and concessive relations have an important role to play in the interpretation of evaluation, but that clearly determining how the relation affects the evaluation is quite complex. See Trnavac and Taboada (2012) for a study of such relations in this corpus.

Within focus, most of the instances correspond to sharpen. Interestingly, negative reviews contain a few more expressions of soften than positive ones. We believe that this is because writers are more likely to soften a negative opinion than a positive one. In the following examples, we see softening expressions. An interesting case is (29), where a positive evaluation is softened by *depending on what film she does*, but preceded by a strong endorsement, *without a doubt* (which is not part of the markable because it conveys Engagement).

(29) Julia Roberts without a doubt, is a good actress **depending on what film she does**. [no1]

(30) The **so-called** 'humor,' as mentioned above, is so pandering, so foul, and so UN-SEUSS, [no10]

(31) The setting wasn't **exactly** what I had envisioned either. [no13]

Most of the cases of Graduation involved one of the two systems, either force or focus. In some cases, however, both types were present. As the CorpusTool annotation program does not allow for double annotation of items within the same system (see Section 3.2), we had to choose the main label under which to classify these cases. We made a note of them, and found in total about 20 examples in English. In (32), we find a graduation of *original* both in terms of force:quantification (*wholly*) and force:intensification (*at all*). We gave priority to force:quantification for *wholly* because it has scope over the entire expression.

(32) It may not be **wholly** original **at all** [yes8]

4.3 German

The German reviews yielded almost 1,900 spans. Like English (and Spanish), the German reviews are dominated by Appreciation spans. Notably, however, Appreciation accounts for almost 70%, where for the other two languages Appreciation spans account for around 55%. The higher percentage of Appreciation spans in contrast to the other languages is balanced in the lower percentage of Affect spans in German, at 6.43%.

Table 4 shows the distribution of spans across positive and negative reviews. With 214 additional spans in the negative reviews, there is a greater difference in the number of spans in the positive and negative contexts than in English (almost equal) and Spanish (88 more spans in the positive reviews). Word count differences are more pronounced than they are in the English corpus, but still closer than in the Spanish reviews, with 25,401 words in the positive reviews, and 21,727 in the negative.

Table 4. Appraisal in positive and negative reviews (German).

	Positive		**Negative**	
	n	**%**	***n***	**%**
Spans (Attitude)	834		1048	
Affect	47	5.60	74	7.06
Judgement	170	20.40	277	26.43
Appreciation	617	74.00	697	66.51

Polarity	834		1048	
Positive	669	80.20	224	21.37
Negative	165	19.80	824	78.63
Spans w/o Graduation	374	44.84	477	45.52
Spans w. Graduation	460	55.16	571	54.48
Focus	97	21.10	93	16.29
Force	363	78.90	478	83.71
Focus type	97		93	
Sharpen	79	81.40	72	77.42
Soften	18	18.60	21	22.58
Force type	363		478	
Intensification	220	60.60	281	58.79
Quantification	143	39.40	197	41.21
Intensification type	220		281	
Int. emphasizer	205	93.20	245	87.19
Int. downtoner	15	6.80	36	12.81
Quantification type	143		197	
Qu. emphasizer	104	72.70	118	59.90
Qu. downtoner	39	27.30	79	40.10

In German, there is only a small difference in the occurrence of negative comments in positive reviews and vice versa, with marginally more positive comments in the negative reviews (21.37% vs. 19.80%). Thus the positive-first pattern for negative comments described above for English does not hold to the same degree. Rather it seems that Germans tend to balance their negative comments in positive reviews just as much as their positive comments in negative reviews.

Whereas positive reviews contain more Appreciation, negative reviews contain a higher percentage of both Judgement and Affect. In fact, there are two negative reviews with more Judgement than Appreciation. This means that rather than appraising the film, the idea, the acting, etc., the writers of negative reviews direct more comments towards the creators of these elements, i.e., the director, the script-writer, the actors. The following example demonstrates the positive/negative alternation mentioned above. It also shows verbs frequently used to express Judgement in the German reviews, namely *gelingen* (“succeed”) (or *gelingt nicht*, “not succeed”), and *scheitern* (“to fail”).

(33) *Grundsätzlich ist die Idee der Story ja wirklich **sehr** interessant und könnte auch wahrscheinlich richtig gut auf die Leinwand gebracht werden. Aber Regisseur und Drehbuchautor George Nolfi gelingt dies leider nicht.* [nein14]

“The idea for the story is fundamentally really **very** interesting and could probably come across really well on screen. But director and script-writer George Nolfi unfortunately doesn’t manage this.”

Adjectives can also be used to express Judgement, as in (34) below, where the comment is about the director’s achievement:

(34) *Der gelungene Umgang mit der Hauptfigur…* [nein10]

“The successful manipulation of the main character…”

Schauspielern (“to act”) is similarly used to express Judgement, as in (35), and here the polarity is added by the modal *kann* (“can”). The positive comment preceding the negative evaluation highlights the contrast between the author’s expectation and her evaluation of the actor’s performance.

(35) *Guillaume Canet kann eigentlich schauspielern. Warum er in diesem Fall **so** zurückhaltend ist, **so** unglaublich falsch und deplatziert wirkt und **vor allem** so verklemmt, dass weiß kein Mensch.* [nein21]

"Guillaume Canet can act. Why in this case he is **so** withdrawn, comes across as **so** unbelievably badly cast and out of place, and **above all** so uptight, no one knows."

Modals and full sentences are also used to express Judgment, as in (36) and (37) respectively.

(36) *Es bleibt **einfach** langweiliger Schrott, den man nicht filmen sollte.* [nein21]

"It is **simply** boring junk, that no one should film."

(37) *Das hätten auch noch Erstsemester an einer Filmakademie mit Laiendarsteller geschafft* [nein21]

"This could also have been achieved by first-semester students at a film school with amateur actors."

In German, Affect was frequently marked by the very colloquial expression *schade*, ("too bad", "what a shame", "bummer"), often standing on its own, or incorporated at the beginning of a sentence as in (38), where the annotation is for negative Affect, and focus-sharpen:

(38) *Schade **eigentlich**, ich hatte mir **echt mehr** versprochen.* [nein10]

"Too bad, **really**. I had expected a lot **more**."

4.4 Graduation in German

Annotating for graduation in German proved to be very complex. First, we examine the distribution presented above in Table 4. Then we will elucidate some of the complexities particular to German.

As Table 4 shows, there were 460 instances of Graduation annotated in the positive reviews, and 571 in the negative reviews. Of these, the overwhelming majority were force (over three-fourths in both cases). Focus accounted for the remainder, 21.10% in the positive and 16.29% in the negative reviews. As for force type, German, too, showed that intensification was more prevalent in both the positive and negative reviews, though the difference between the frequency of intensification and quantification frequency was not as great as in English. With regard to emphasizing and downtoning force, emphasizing is, as in English, more frequent than downtoning. This pattern is more marked for intensification than for quantification, where intensifiers were downtoned only 6.8% and 12.81% in the positive and negative reviews respectively, and quantification 27.3% and 40.1% respectively. Finally, focus was more often sharpened than softened, and in comparable frequencies in both sets of reviews.

In the German corpus, intensification emphasizers were typically adverbs such as *deutlich* ("clearly"), *wirklich* ("really"), comparatives and superlatives, temporal adverbial phrases such as *zu keiner Zeit* ("at no time") and either/or expressions (*weder/noch*) and other phrases such as *alles andere als* ("anything but"). *Nie* ("never") was a frequent emphasizer, as it is more emphatic than the

regular negation with *nicht* ("not"), as is the commonly used *überhaupt nicht* ("not at all"). There were also more creative emphasizers, as in the intensification *mit Müh und Not* in (39), which emphasizes the negative appreciation of the review.

(39) ***Mit Müh und Not*** *einen von fünf Felsblöcken, die alles unter sich begraben, was diesen Film gut hätte werden lassen können.* [nein2]

"**Barely** [lit., "with effort and hardship"] one out of five stones [stars], which bury everything that could have made this film a good one."

Common downtoners included modals, as in Example (40):

(40) *Ich* ***möchte*** *den Film empfehlen (er ist aber* ***beileibe nichts*** *für einen gemütlichen Familien Kinoabend)* [ja12]

"I **would like to** recommend the film (though it is **by no means** something for a cozy family movie night)"

The German particle *zwar* sometimes functions as a downtoner, since it adds a sense of hedging to the statement, as in (41):

(41) *Der Handlungsverlauf ist* ***zwar*** *vorhersehbar* [ja10]

"**While** the plot development is predictable"

Quantification included adjuncts relating to an amount or quantity, broadly construed, such as *mehr oder weniger* ("more or less"), *mehr und mehr* ("more and more"), *zeitweise* ("occasionally") or a whole, such as *völlig, komplett, ganz,* and *gesamt,* all of which can be loosely translated as "entirely", "completely" or "absolutely". Temporal adverbs such as *ab und an/ab und zu* ("now and then") also downtoned quantification. In negative contexts quantification also included *gar nicht* ("not at all") and *an keiner Stelle* ("at no point"), which is similar to the absolute negative *nie* ("never"), and *kaum* ("hardly"). Phrases such as *auf ganzer Linie* ("in its entirety") that refer to a quantity were also common.[5]

The German negation particle *kein* proved complex to annotate. Depending on context it can operate as a simple negator with no graduation, where it negates a noun (*kein gutes Schauspielen* is simply "not good acting", where the evaluation is a straightforward negative polarity). At other times *kein* evokes a quantification as in (42) below. Here, a simple negation would have been negating the sentence using *nicht*, as in *er hat nicht das Gefühl* ("he didn't have the feeling"), rather than "at no point did he have the feeling".

(42) *Man hat* ***an keiner Stelle*** *das Gefühl, auf die beiden würde ein lohnenwertes Ziel warten* [nein12]

"One has **at no point** the feeling that a worthwhile goal awaited the two [characters]"

[5] Note that one of the complexities in annotating graduation in German was whether to consider emphasized negation using *nichts* and the phrases *gar nicht* and *überhaupt nicht* as quantification or intensification. The judgements in these cases depended on whether the noun or verb that was being negated could be construed to have quantity in any sense.

Kein was also involved in many cases of intensification as a downtoner:

(43) …***keine wirklich grosse** Überraschung* [nein15]

"…**not** a really big surprise"

Litotes, the practice of expressing an affirmative by negating its opposite, was a tricky phenomenon to annotate in German. We decided, for all languages, not to annotate litotes on its own, as it is too difficult to determine what the alternative to the negative statement is. However, very often, litotes expressions included quantification by using words like *viele* ("many"). When these quantity expressions were negated, they served as downtoners, and required annotation, as in (44).

(44) *Ich erwarte eigentich nicht **viel*** [nein1]

"I actually don't expect **much**"

This could also have been expressed in the positive, as either *I expect little* or *I expect something*. As mentioned, the affirmative is hard to deduce precisely from the negative. However, here it is annotated as negative/downtoned due to the quantity word *nicht viel* ("not much").

It was difficult to assess litotes involving *kein* in German. As mentioned previously, *kein* can indicate quantification emphasis (as in English "none"), or a simple negation ("no"). It does serve in litotes, as shown in (45). However, where the sentence negation occurs before an intensifier or focus-sharpening adverb, as in (46) below, *kein* serves as a simple negator.

(45) ***kein so** schlecht gewählter Titel* [nein16]

"not such a badly chosen title"

(46) *Nolfi (direktor) verfolgt in meinen Augen kein **richtiges** Ziel fuer seine Geschichte* [nein14]

"In my eyes, Nolfi (the director) doesn't follow a real goal for his story"

For the first of the pair, the alternative would have been to say "this is a well-chosen title", which results in the example being annotated as positive appreciation with no graduation. The second, however, is not litotes and is simply negative judgement with a sharpened focus resulting from *richtiges* ("real").

Concerning the two categories of focus, focus:soften was marked in German primarily by particles *eher* ("rather"), *etwas* ("somewhat"), *eine Art* ("a kind of"). Focus:sharpen was characterized by the adverbs *einfach* ("simply"), often in combination with *nur*, as in *einfach nur* ("simply/just"), *eigentlich* ("actually"), *ziemlich* ("quite"), and *ganz* ("quite" as in *ganz gut*, "quite good". Note that this use of *ganz* differs from the meaning listed above under quantification in which it refers to a whole.

Finally, there were frequently two types of graduation in one phrase. It was agreed to always annotate the graduation with the widest scope. In Example (47) below, *in manchen Stellen* ("in some places") has wider scope than the comparative *zu zurückgezogen und introvertiert* ("too withdrawn and introverted"). Thus in (47) the span is marked for

force:quantification:downtoner of *in manchen Stellen*, and not force:intensification:emphasizer as it would be for the comparison.

(47) …***in manchen Szenen*** *in meinen Augen zu zurückgezogen und introvertiert* [nein10]

"…**in some scenes** in my opinion too withdrawn and introverted"

In sum, the German reviews were characterized by a higher proportion of Appreciation spans than English and Spanish; however, negative reviews favour Judgement and Affect more than positive reviews do. Furthermore, graduation in German proved to be complex and very dependent on context, rather than on particular phrases. Unique to German in particular were the nuances described above in annotating negation with the particles *nicht* and *kein.*

4.5 Spanish

The results of the annotation of the Spanish movies are displayed in Table 5. The number of spans for positive reviews (1,019) is higher than for negative reviews (931); however, we must remember that the total number of words of the positive reviews is much higher: The ratio of words per span is 35.40 in the positive reviews and 18.02 in the negative reviews. This difference in density is largely due to the greater length of the description stages in the positive reviews, in which reviewers, often driven by their enthusiasm, write lengthy paragraphs that contain few or no evaluative spans about the plot or technical details.

Table 5. Appraisal in positive and negative reviews (Spanish).

	Positive		**Negative**	
	n	**%**	***n***	**%**
Spans (Attitude)	1019	52.26	931	47.74
Affect	235	23.06	154	16.54
Judgement	202	19.82	279	29.27
Appreciation	582	57.11	498	53.49
Polarity	1019		931	
Positive	843	82.73	204	21.91
Negative	176	17.17	727	78.09
Spans w/o Graduation	494	48.48	479	51.45
Spans w. Graduation	513	50.34	452	48.55
Focus	50	9.75	47	10.40
Force	463	90.25	405	89.60
Focus type	50		47	
Sharpen	26	52.00	24	51.06
Soften	24	48.00	23	48.94
Force type				
Intensification	284	61.34	227	56.05
Quantification	179	38.66	178	43.95
Intensification type				
Int. emphasizer	208	73.24	165	72.69
Int. downtoner	76	26.76	62	27.31
Quantification type				
Qu. emphasizer	139	77.65	150	84.27
Qu. downtoner	40	22.35	28	15.73

With regard to the kind of spans, Appreciation amounts to more than half (57.11% in the positive reviews and 53.49% in the negative), which is not

surprising, since the main object of the reviews are non-human entities (the movies). The next most frequent spans are those of Judgement, which, summing up the two types of reviews, total 481, while those of Affect add up to 389. The percentage of Appreciation is 3.62% higher in the positive reviews. The difference is larger in the case of Affect, whose percentage is 6.52% higher in the positive reviews: writers of positive reviews tend to express their positive feelings (and, less commonly, those of other viewers) by several kinds of devices: verbs of mental processes of emotion such as *gustar* ("like"), *encantar* ("love"), *disfrutar* ("enjoy"), *alegrar(se)* ("make glad / be glad"), *emocionarse* ("get excited") or *sorprender* ("surprise") or denials of negative verbs such as *no defraudar* ("not dissapoint"); adjectives such as *contento* ("glad"), *identificado* ("identified"), *satisfecho* ("satisfied"), *fascinado* ("fascinated") or *atento* ("attentive"); nouns such as *sonrisa* ("smile"), *risa* ("laugh") or *encanto* ("charm"), or idiomatic expressions such as *partirse de risa* ("burst out laughing") or *saltarse las lágrimas* ("shed tears").

In contrast to Affect and Appreciation, Judgement is over 10% more frequent in the negative reviews. Reviewers tend to evaluate the unsatisfactory performance of the director or the cast with a number of common devices: adjectives such as *pobre* ("poor"), *pésimo* ("very bad") or *absurdo* ("absurd"); nouns or nominal expressions such as *mediocridad* ("mediocrity"), *poca credibilidad* ("little credibility") or verbs referring to actions concerning a better original book or script, such as *destrozar* ("spoil") or *destruir* ("destroy"). Some expressions refer to moral issues, such as *sabe embaucar* ("knows how to fool") or *auténtico timo* ("real swindle"). In certain cases, the spans occupy whole sentences, as in *Yo con mi cámara super 8 y un muñeco de plastilina lo hubiera hecho mejor* ("I would have done a better job with my super 8 camera and a play-doh puppet").

Concerning polarity, positive spans outnumber negative ones. In the positive reviews, positive spans sum up more than 80% of the cases for the three subtypes of Attitude, while in the negative reviews negative spans amount to 78% of the cases. The percentage of spans with contrary polarity to that of the overall review is slightly smaller than for the English reviews, both in the positive and the negative reviews. The rhetorical pattern of initial spans with contrary polarity for the English reviews, mentioned in Section 4.1, was not significant for the Spanish reviews. Rather, these spans occur in no fixed position, and mostly refer to: individual performances of actors or actresses, as in (48), which was found in a negative review; concrete aspects or parts of the movie that do not match its overall quality, such as (49), which again was in a negative review; or reported opinions by other people than the reviewer, as in Example (50).

(48) *Beth, interpretada por Julie Cox, me sorprendió por su excelente interpretación* [no2-12]

"Beth, interpreted by Julie Cox, surprised me with her excellent performance."

(49) *muy bien* ***a nivel de efectos especiales y música*** [no2-20]

"very well **with regard to special effects and music**"

(50) *La película dura 2 horas y 20 minutos. Hay gente que dice que se le hace pesado porque están acostumbrados/-as a ver un solo capítulo.* [yes5-5]

"The movie lasts 2 hours and 20 minutes. There are people who say that it feels tedious because they are used to seeing one chapter only."

4.6 Graduation in Spanish

With regard to Graduation inscribed in Attitude, about half the spans are graduated (50.34% for positive reviews and 48.48% for negative reviews). Within Graduation, roughly 10 percent of the spans display focus and 90 per cent display force for both kinds of reviews. Further quantitative research on the cases of focus showed that, out of the total 97, 65 of them (67.01%) correspond to spans of Appreciation. That is, focus occurs mostly in spans that evaluate the movie itself or other non-human entities, or else aesthetic features of the movies. Examples of these spans of Appreciation graduated with focus are *auténtico disparate* ("real nonsense"), *patética por un lado* ("pathetic on the one hand"), *sencillamente magistral* ("simply masterful"), *perfectas en todos los aspectos* ("perfect in all aspects"). With regard to the subtypes of focus, cases of sharpen occur with roughly the same frequency as those of soften. Although the quantity of focus spans in Judgement is small, totalling 16 occurrences, it is worth pointing out that 12 of them (75%) are of sharpen. That is to say, the spans concerning the director of the cast of the movie, as well as ethical issues, are more often emphasized than downtoned in terms of prototypicality (including completeness). Some examples of Judgement spans graduated with focus: sharpen are *completamente prohibidas* ("completely forbidden"), *totalmente plana* ("totally flat") or *claro ejemplo de amistad, de superación y de optimismo* ("clear example of friendship, of will and of optimism").

Concerning force, intensification predominates over quantification in both categories, especially in the positive reviews. This predominance is particularly strong for spans of Judgement in the positive reviews (71.76%). These often display favourable evaluative spans about the cast consisting of adjectives or adverbs modified by intensifying adverbs, mostly *muy* ("very") or *tan* ("so, such"). Other common devices are the span with the superlative adjective *el mejor* ("the best") and nouns modified by the adjective *gran*. Examples of these spans are *muy creíble* ("very believable"), *muy, muy simpático* ("very, very friendly"), *tan bonito* ("so pretty"), *tan bien* ("so well"), *gran magnetismo* ("great magnetism") or *gran estrella* ("great star"). Affect differs from the other subcategories in that it displays a predominance of quantification over intensification (52.00% in the positive reviews and 60.94% in the negative ones). The reason is that Affect is commonly realized by verbs of emotion modified by expressions of quantity. Instances of spans of this kind are *gustó mucho* ("pleased a lot"), *disfruté mucho muchísmo* ("I enjoyed a lot really a lot"), *esperaba bastante* ("I expected quite a lot"), *tanto me defraudó* ("disappointed me so much").

With regard to the subtypes of intensification, emphasizers are more common than downtoners. That is to say, reviewers tend to reinforce their evaluative judgements rather than to downtone them. The global percentages are

similar in the positive and negative reviews. Research on subcategories of Attitude shows that this predominance is lowest in Affect, especially in the positive reviews (the number of downtoners is 18, totalling 37.5%), due to the frequent cases of litotes, i.e., downtoning by means of negation, such as *la sensación final no fue la que deseaba* ("the final impression was not what I wanted"), *no me ha desagradado* ("I haven't disliked it"), or *no es un film que te asombra por su excelencia* ("it is not a movie that astonishes because of its excellence"). The proportion of emphasizers is highest in Judgement, especially in positive reviews (51 cases, 83.61%), due to the frequency of intensified spans such as those mentioned in the previous paragraph. From a qualitative point of view, it must be noted that the partitive turns an evaluative span with a superlative expression as a downtoner, as in (51).

(51) ***De lo que más*** *me ha gustado son las conversaciones que tiene con Kevin* [no1-17]

"**Among the things** that I liked **best** are the conversations that she has with Kevin"

Concerning quantification, emphasizers predominate over downtoners, which is consistent with the reviewers' tendency to reinforce their evaluations. This predominance is almost the same as that found in intensification in the positive reviews, but it is markedly higher in the negative reviews. The reason is that quantifying emphasizers are often used in negative spans so as to create a humorous effect, as in *todo decaía* ("everything was decaying"), *se pasan la mayor parte del tiempo a grito limpio* ("they spend most of the time shouting like mad") or *nada ni nadie en la película pasa de ser plano como el encefalograma de un muerto* ("nobody or nothing in the movie is anything but flat like a corpse's encephalogram").

As a global comment, we may state that the most striking differences were found between spans of Affect and those of Judgement; Appreciation being in a middle position in many respects.

5. Discussion and Conclusions

In all three languages, the distribution of Appraisal categories was found to be similar, with Appreciation being the most frequent type of Attitude, followed by Judgement and then Affect. In other words, the reviews are not actually about *loving and hating* movies, but rather about movies as works of art. Expressions of love and hate, which would be categorized as Affect, are actually quite infrequent.

In Spanish, positive reviews may be considered as an exception to this tendency. As we discussed in 4.5 and 4.6, expressions of positive feelings did occur with some frequency. However, there is a caveat: The two most common positive verbs of Affect, *gustar* ("like") and *encantar* ("love") have a different syntactic pattern from that of their English equivalents. Their pattern resembles that of the English verb *please,* in that the subject is the entity that provokes the feeling and the object is the entity that feels. No occurrences were found of the verb *amar*, whose syntax is the same as that of its English equivalent *love.* In this way, the role of the movie (or an entity related to it) as initiator of the

emotion, is emphasized, while viewers play a receptive role as undergoers of these emotions. Thus, in spite of the relatively frequent expressions of Affect, Spanish positive reviews are not mainly about loving or hating movies, but rather about being pleased, impressed or moved by them. This perspective is also visible in the Spanish negative reviews, in which the overall percentage of Affect was lower. No occurrences were found of *odiar*, the most direct correlate of the English verb *hate*, nor of any close synonym such as *detestar* ("detest") or *aborrecer* ("loathe"). Instead, negative affect was very commonly expressed by negating *gustar*; to a lesser extent, other expressions were also used, such as *defraudar* ("disappoint"), *aburrirse* ("get bored") or *enfadado* ("angry").

The genre that all the texts belong to seems to drive the types of Attitude being expressed. Similarly, the high level of Graduation (over 40% for the three languages), is a reflection of the evaluative nature of this genre, in which reviewers feel the need to graduate the strength or the applicability of the evaluative expressions that they use. In the three languages, intensification is more common than quantification, and this predominance is stronger in the positive than in the negative reviews. This suggests that many negative spans express a semantic element of "not anything" (quantifying emphasizer) or "not much" (quantifying downtoner).

As well as these cross-linguistic similarities, we also found some interesting differences across the three languages. In particular, there are a number of characteristics of the German reviews that vary from their English and Spanish counterparts. Differences in span length resulted from the particulars of German word order, in addition to the fact that, generally, German sentences are more accommodating of long passages. In some instances, double layers of annotations were required. In comparison to the other languages, the distribution of Evaluation in German was also different. Appreciation spans account for almost 70%, where for the other two languages Appreciation spans account for around 55%. Affect spans are markedly less frequent than in the other two languages. Positive reviews contained more Appreciation and less Judgement than the negative reviews. The significantly high percentage of Judgement in negative reviews, as compared to positive reviews, is shared by the Spanish part of the corpus. This suggests that, in these two languages, criticism is expressed with regard to the (lack of) achievement or performance of individuals such as the director or the cast members, while expressions of praise focus on the object of success, such as the camera work, script, or scenery (rather than the person). Interestingly, the Spanish and German negative reviews also share almost the same distribution of positive comments in negative reviews (though the Spanish positive reviews have fewer negative comments than the German ones do). English showed a higher frequency of positives in negative reviews than the other way around. Thus the 'positive-first' feature described for English is less striking in German and Spanish.

As discussed earlier, Graduation in German proved complex. In fact, over 54% of the spans in the German reviews contained Graduation as compared to about 42% for English and approximately 49% for Spanish. Though in all languages the majority of cases were force (in which intensification was more frequent than quantification), there was more focus (18.7%) in the German reviews than in English (15.6%), and the difference is even larger with Spanish (under 10%). German contains many particles that serve to sharpen and soften focus, such as *eher* ("rather"), *etwas* ("somewhat"), *einfach* ("simply"), *einfach*

nur (“simply/just”), *eigentlich* (“actually”), *ziemlich* (“quite”). German also differs from English and Spanish in the distribution of its downtoners: the intensifying downtoners are remarkably less frequent (roughly 13%, as compared to 22 % for English and 27% for Spanish), while the quantifying downtoners are considerably more common (about 40%, as compared to 12% for English and 16% for Spanish).

In other words, at the lexicogrammatical level, German is different from English and Spanish, in word order and frequency and type of Graduation. On the other hand, the overall organization and distribution of positive and negative evaluations showed that English is the outlier, with a recurring structure in reviews where positive aspects are presented first, before a negative evaluation is presented. We have, elsewhere, referred to this pattern as a form of vernacular argumentation (Taboada & Gómez-González, 2012).

Concerning Spanish, the higher frequency of expression of Affect in positive reviews in comparison to the other two languages was already mentioned above in this section. Another remarkable feature lies in the category of focus, in which the number and percentage of sharpeners and softeners is almost the same, while English and German displayed a considerably higher number of sharpeners. Examples of Spanish softeners are *en parte* (“in part”), *más o menos* (“more or less”) or *por así decirlo* (“so to say”).

The findings reported above leads to suggest a number of possible extensions for the research presented here. For example, a contrastive analysis of English, German and Spanish movie reviews of this kind might also be applied to the category of Engagement; the analysis of Attitude could include more languages, so that a wider crosslinguistic perspective of this category in movie reviews would be obtained; and finally, the analysis could be extended to reviews of different products, which would reveal the ways in which the expression of Attitude is influenced by the kind of item reviewed.

References

Achugar, Mariana. (2008). What We Remember: The Construction of Memory in Military Discourse. Amsterdam: John Benjamins.

Aikhenvald, Alexandra. (2004). Evidentiality. Oxford: Oxford University Press.

Bakhtin, Mikhail. (1981). Discourse in the novel (C. Emerson & M. Holquist, Trans.). In M. Holquist (Ed.), The Dialogic Imagination: Four Essays by M. M. Bakhtin (pp. 259-422). Austin: University of Texas Press.

Banfield, Ann. (1982). Unspeakable Sentences: Narration and Representation in the Language of Fiction. Boston: Routledge and Kegan Paul.

Batson, C. Daniel, Shaw, Laura L., & Oleson, Kathryn C. (1992). Differentiating affect, mood, and emotion: Toward functionally based conceptual distinctions. In M. S. Clark (Ed.), Emotion. Review of Personality and Social Psychology (pp. 294-326). Newbury Park, CA: Sage.

Becker, Annette. (2009). Modality and Engagement in British and German political interviews. Languages in Contrast, 9(1), 5-22.

Biber, Douglas, & Finegan, Edward. (1988). Adverbial stance types in English. Discourse Processes, 11(1), 1-34.

Biber, Douglas, & Finegan, Edward. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. Text, 9(1), 93-124.

Bybee, Joan, & Fleischman, Suzanne (Eds.). (1995). Modality in Grammar and Discourse. Amsterdam and Philadelphia: John Benjamins.

Carretero, Marta, & Taboada, Maite. (to appear). The annotation of Appraisal: How attitude and epistemic modality overlap in English and Spanish consumer reviews. In J. R. Zamorano (Ed.), Thinking Modally: English and Contrastive Studies on Modality. Berne: Peter Lang.

Chafe, Wallace, & Nichols, Johanna. (1986). Evidentiality: The Linguistic Coding of Epistemology. Norwood, NJ: Ablex.

Coffin, Caroline. (1997). Constructing and giving value to the past: An investigation into secondary school history. In F. Christie & J. R. Martin (Eds.), Genre and Institutions: Social Processes in the Workplace and School (pp. 196-230). London: Cassell.

Coffin, Caroline, & O'Halloran, Kieran. (2006). The role of appraisal and corpora in detecting covert evaluation. Functions of Language, 13(1), 77-110.

Eggins, Suzanne, & Martin, James R. (1997). Genres and registers of discourse. In T. A. v. Dijk (Ed.), Discourse as Structure and Process. Discourse Studies: A Multidisciplinary Introduction (pp. 230-256). London: Sage.

Eggins, Suzanne, & Slade, Diana. (1997). Analysing Casual Conversation. London: Cassell.

Halliday, Michael A. K. (1975). Learning How to Mean. London: Edward Arnold.

Halliday, Michael A. K. (1985). An Introduction to Functional Grammar (1st ed.). London: Arnold.

Halliday, Michael A. K. (1989). Context of situation. In M. A. K. Halliday & R. Hasan (Eds.), Language, Context and Text (pp. 3-14). Oxford: Oxford University Press.

Halliday, Michael A. K., & Matthiessen, Christian M.I.M. (2004). An Introduction to Functional Grammar (3rd ed.). London: Arnold.

House, Juliane. (2006). Communicative styles in English and German. European Journal of English Studies, 10(3), 249-267.

Johansson, Stig. (1998). Loving and hating in English and Norwegian: A corpus-based contrastive study. In D. Albrechtsen, B. Henriksen, I. M. Mees & E. Poulsen (Eds.), Perspectives on Foreign and Second Language Pedagogy. Essays Presented to Kirsten Haastrup on the Occasion of her Sixtieth Birthday (pp. 93-103). Odense: Odense University Press.

Kaplan, Nora. (2007). La construcción discursiva del evento conflictivo en las noticias por televisión. PhD dissertation, Universidad Central de Venezuela, Caracas, Venezuela.

Langacker, Ronald W. (1990). Subjectification. Cognitive Linguistics, 1(1), 5-38.

Liu, Bing. (2012). Sentiment Analysis and Opinion Mining: Morgan & Claypool.

Love, Kristina. (2006). APPRAISAL in online discussions of literary texts. Text and Talk, 26(2), 217-244.

Macken-Horarik, Mary. (2003). APPRAISAL and the special instructiveness of narrative. Text, 23(2), 285-312.

Mann, William C., & Thompson, Sandra A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. Text, 8(3), 243-281.

Martin, James R. (1996). Evaluating disruption: symbolising Theme in junior secondary narrative. In R. Hasan & G. Williams (Eds.), Literacy in Society (pp. 124-171). London: Longman.

Martin, James 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, James R. (2003). Introduction: Special issue on Appraisal. Text, 23(2), 171-181.

Martin, James R., & White, Peter R. R. (2005). The Language of Evaluation. New York: Palgrave.

Page, Ruth E. (2003). An analysis of APPRAISAL in childbirth narratives with special consideration of gender and storytelling style. Text, 23(2), 211-237.

Painter, Claire. (2003). Developing attitude: An ontogenetic perspective on APPRAISAL. Text, 23(2), 183-209.

Palmer, Frank. (1986). Mood and Modality. Cambridge: Cambridge University Press.

Pang, Bo, & Lee, Lillian. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.

Portner, Paul. (2009). Modality. Oxford: Oxford University Press.

Taboada, Maite. (2008). SFU Review Corpus [Corpus]. Vancouver: Simon Fraser University, http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html.

Taboada, Maite. (2011). Stages in an online review genre. Text and Talk, 31(2), 247-269.

Taboada, Maite, Brooke, Julian, Tofiloski, Milan, Voll, Kimberly, & Stede, Manfred. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2), 267-307.

Taboada, Maite, & Carretero, Marta. (2012). Contrastive analyses of evaluation in text: Key issues in the design of an annotation system for Attitude applicable to consumer reviews in English and Spanish. Linguistics and the Human Sciences, 6(1-3), 275-295.

Taboada, Maite, & Gómez-González, María de los Ángeles. (2012). Discourse markers and coherence relations: Comparison across markers, languages and modalities. Linguistics and the Human Sciences, 6(1-3), 17-41.

Taboada, Maite, & Grieve, Jack. (2004). Analyzing appraisal automatically. In Y. Qu, J. G. Shanahan & J. Wiebe (Eds.), Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07) (pp. 158-161). Stanford University, CA: AAAI Press.

Traugott, Elizabeth Closs. (1995). Subjectification in grammaticalisation. In D. Stein & S. Wright (Eds.), Subjectivity and Subjectivisation: Linguistic Perspectives (pp. 31-54). Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs. (2010). (Inter)subjectivity and (inter)subjectification: A reassessment. In K. Davidse, L. Vandelanotte & H. Cuyckens (Eds.), Subjectification, Intersubjectification and Grammaticalization (pp. 29-74). Berlin: De Gruyter Mouton.

Trnavac, Radoslava, & Taboada, Maite. (2012). The contribution of nonveridical rhetorical relations to evaluation in discourse. Language Sciences, 34(3), 301-318.

White, Peter R. R. (1998). Telling Media Tales: The News Story as Rhetoric. Ph.D. dissertation, University of Sydney, Sydney.

White, Peter R. R. (2003a). Beyond modality and hedging: A dialogic view of the language of intersubjective stance. Text, 23(2), 259–284.

White, Peter R. R. (2003b). An Introductory Course in Appraisal Analysis Retrieved March 16, 2009, from http://www.grammatics.com/appraisal

Authors' addresses

Maite Taboada
Department of Linguistics
Simon Fraser University
8888 University Dr.
Burnaby, B.C., V5A 1S6
Canada
mtaboada@sfu.ca

Marta Carretero
Departamento de Filología Inglesa I
Facultad de Filología
Universidad Complutense de Madrid
Ciudad Universitaria s/n
28040 Madrid
Spain
mcarrete@filol.ucm.es

Jennifer Hinnell
Department of Linguistics
University of Alberta
2-40 Assiniboia Hall
Edmonton, AB, T6G 2E7
Canada
hinnell@ualberta.ca