Labelling evaluative language in English and Spanish:
The case of Attitude in consumer reviews

Maite Taboada

Simon Fraser University

Vancouver, Canada

mtaboada@sfu.ca

Marta Carretero

Universidad Complutense

de Madrid

mcarrete@filol.ucm.es

6th International Contrastive Linguistics Conference
Berlin
September 30-October 2, 2010

http:www.sfu.ca/~mtaboada

Universidad
Complutense
Madrid

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

Goals and framework

- Purpose of this presentation
  - To show and illustrate the process of corpus annotation of Appraisal categories

Framework for the study
  - The CONTRANOT project (Lavid 2008; Lavid et al. 2007)
  - Related project
    - Automatic sentiment detection
    - Use of linguistic information
    - Drawing on insights provided by the Appraisal framework
    - Work being carried out at Simon Fraser University

Universidad
Complutense
Madrid

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

The CONTRANOT Project

- Current research effort of our group at UCM
- 3-year project funded by Spanish Ministry of Science and Innovation
- Research team:
  - 6 UCM members (Julia Lavid as principal investigator)
  - 1 member from Simon Fraser University (Canada)
  - International Advisory Board (4 members)
  - 1 doctoral student

CONTRANOT aims

1) Develop functional contrastive (English-Spanish) descriptions based on corpus analysis and annotation

2) Use text annotation as a tool for testing linguistic hypotheses empirically

3) Creation of online resources which will allow adequate sharing and distribution of the contrastive resources developed in the project

Corpus annotation tasks

- Main categories to be annotated contrastively (English and Spanish):
    - Modality and appraisal
    - Coherence relations (Rhetorical Structure Theory)
    - Topic and reference expressions
    - Thematic features

Specifications on evaluation

- Definition of evaluative language
  - The linguistic expressions that indicate "the subjective presence of writers/speakers in texts as they adopt stances towards both the material they present and those with whom they communicate" (Martin and White 2005: 1)
- Why evaluative language is interesting
  - We all use language to evaluate, appraise and classify objects and people on an everyday basis
  - Applied venues: computational applications
    - Opinions in the web, which are of interest to marketers, policy makers and the public in general
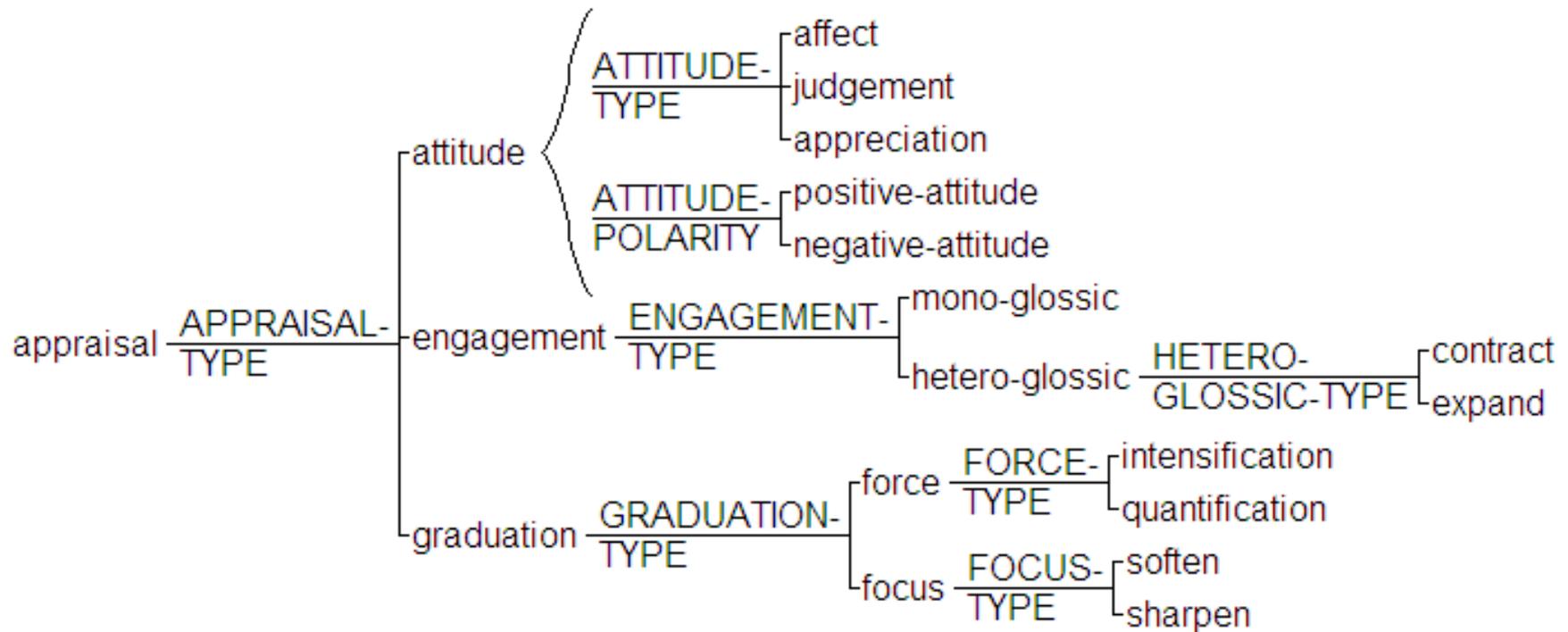
Corpus analysis of Appraisal

- The larger Simon Fraser University Review Corpus: 1,600 movie, book, and consumer product reviews
    - English: 800 reviews from Epinions.com
    - Spanish: 800 reviews from Ciao.es and Dooyoo.es
- Users post reviews about products, services or art that they have used/experienced
- Current Appraisal analysis
    - Reviews of movies, books and hotels

Authors

- ## Unsophisticated writers
    - Although there is heavy use of irony and sarcasm
- ## Straightforward
    - The goal is to help other viewers, readers, consumers
    - In the Spanish site, there is also a potential monetary compensation
        - The website advertises that users may earn money if their reviews are classified as "useful" by other users

The Appraisal system

Appraisal: Attitude

- Martin & White (2005: 35) "Attitude is concerned with our feelings, including emotional reactions, judgements of behaviour and evaluation of things".

Subcategories of Attitude:

- <u>Affect</u> (emotions)
  - *sad, cheerful, anxious... like, love, hate...*
- <u>Judgement</u> (ethics; rules and regulations)
  - *honest, fair, legal... reward, punish, deserve...*
- <u>Appreciation</u> (aesthetics/value: criteria and assessment)
  - *beautiful, ugly... beauty, elegance...*

Appraisal: Engagement

- Relation between what is being stated and other actual or potential viewpoints
- Monoglossic statements
  - No alternative viewpoint, or openness to accept one
- Heteroglossic statements
  - Inter-subjective positioning is open
  - Utterances invoke, acknowledge, respond to, anticipate, revise or challenge a range of convergent and divergent alternative utterances
- (Engagement not the focus of this presentation)

Appraisal: Graduation

- **Graduation** concerns linguistic expressions used for emphasizing or downtoning other expressions.

- Expressions of Graduation differ from those of Attitude in that they **do not have intrinsic positive or negative values by themselves**, but acquire them in context.

- Examples
  - Intensifiers applied to nouns (*real, true, genuine*)
  - Intensifiers applied to adjectives (*very*, *really*)
  - Softeners (*kind of, sort of, or something*).

Demands on the analysis and annotation

- **Rigour**, for future reliability and consistency (intra-rater and inter-rater)

- **Simplicity**: this analysis will be the basis of an annotation system to be applied to large amounts of discourse / texts

Inter-annotator reliability

- Necessary that annotators perform annotation reliably
  - Clarity of annotation categories
  - **First pass: determining markables**
  - Second pass: Appraisal annotation

Problems in the application of Appraisal

1. Key features for labelling Attitude
   - ethics vs. aesthetics
   - entity evaluated
2. Context dependence of evaluative meanings
3. Implicit evaluation: comparison and metaphor
4. Positive or negative polarity of the evaluation, and its dependence on the main entity evaluated
5. Size of evaluative spans: accuracy versus simplicity
   - Annotator agreement study

1. Key features for labelling Attitude

- Judgement versus Appreciation:
  - ethics vs. aesthetics
  - entity evaluated

- Doubtful cases:
  - … sin embargo, cuando un actor demuestra su valía es realmente en solitario. Y, en este sentido, Will Smith **aprueba y con nota.** (Películas yes4.2).

    '…however, when an actor displays his worth, it is really on his own. And, in this sense, Will Smith **passes with a high mark'**
  - Before I finish I will give mention to the film's nudity. It's never strong or prolonged but there are a few moments of nudity in the film that will likely earn the same sort of **childish** comments as Cathy Bates infamous scene in About Schmidt. (Movies, yes1)

2. Context-dependence of evaluative meanings

- Some expressions acquire evaluative meaning due to the context in which they occur

  - A la protagonista nos la presentan como a la **típica** mujer que sabe que consigue más luciendo carne, que utilizando el cerebro (Libros no1.14).

    'The protagonist is presented as the **typical** woman who knows that she achieves more by showing flesh than by using her brains'

Universidad Complutense Madrid

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

3. Implicit evaluation (I)

- White (2003): Choice between
  - 'Token' Judgement, an implicit evaluation (1), in which the facts are to be evaluated by the reader
  - Explicit evaluation by the author of those same facts (2)

  (1) They shot the man in the head at point-blank range.
  (2) They murdered him, **heinously**, **callously** and **in cold blood**.

- Our annotation system will include (2), but not (1)

3. Implicit evaluation (II): Metaphor

Metaphor is included in evaluation:

- Helen Mirren **sparkles** as Chris, portraying a small-town woman with *an abundance of energy, enthusiasm, and grit*. (Films yes2)

- **… a totally generic 'shoebox'** of a room that even a discount chain hotel would be *embarrassed* to lay claim to. (Hotels no4)

3. Implicit evaluation (III)

Other cases included in evaluation:

- ## Comparison

  The shower drain was *rusty* and *unclean*. I know this is pretty standard, but the water pressure was **more like a spritzer bottle than a shower**! (Hotels no4)

  The sounds of a *grand* piano drew me toward a *beautiful* high-ceiling room that looked **almost as much like a cathedral as it did a hotel.** (Hotels yes5)

- ## Rhetorical questions

  There was a "security guard" there that also had no idea where our rooms were! That certainly didn't make me feel secure in any way-- **what kind of "security guard" doesn't know the lay of the land in the hotel he works?** (Hotels no4)

4. Polarity of the Appraisal (I)

- ## Positive polarity inferred from context

  but the **very ordinariness** of many of the other women, and the *art* with which they are presented in the calendar is a *refreshing* reminder that beauty can be more than we see in Cosmopolitan and Access Hollywood. (Films yes2)

- ## Polarity opposite to the lexical meaning of the evaluative item

  A few conflicts were developed that never found any resolution in the film. Of course, there is no absolute necessity to wrap everything up neatly by the end of an hour and a half, but it is **nice** to see at least the beginnings of a resolution on most of the issues. (Films yes2)

4. Polarity of the Appraisal (II)

- 'Official' classification of entities against real quality (irony)

A few more minutes before 4:00 another person is banging on the door. This is really unbelievable. I specifically said not to disturb us before 4:00 nap time was over. So much for **first-class** service. (Hotels no4)

We had a *terrific* time in Key West as long as we didn't spend a minute more than we had to in our so-called '**luxury**' rooms. (Hotels no4)

5. Size of the evaluative spans

- Exclusion of the elements within the scope of the evaluative expression (e.g., entities)

    I would definitely recommend the Golden Nugget (oh, and did I mention **free parking for hotel guests**?) (Hotels yes22)

    … la **decrepitud** *de Ender* (Libros no1.1)
    'Ender's **decrepitude**'

- Inclusion of expressions of Graduation which emphasize or downtone Attitude:

    <u>**such a** **disappointing**</u> film <u>**in many levels**</u> (Films no1)
    Julia Roberts without a doubt, is a **good** actress <u>**depending on what film she does**</u> (Films no1)
    <u>**for the most part**</u> the staff was **friendly and helpful** (Hotels no4)
    the whole experience is <u>**rife**</u> **with humor** (Films yes2)

Agreement study: size of evaluative spans

- Two annotators
- Annotated small set of 6 texts
  - About 7,700 words
- Performed separate labelling of markables
  - Highest number of markables: 348
  - Precision: 83.05
  - Recall (partial): 93.68
  - Recall (total): 84.77

Agreement with respect to markables (I)

- Include nouns?
  - As to plot and character development - given the near absence of any meaningful interior monologues - characters blather at each other for page after page in **stilted and unnatural prose.** (Books, no1)
  - Annotator 1: stilted and unnatural
  - Annotator 2: stilted and unnatural prose
    - The noun does not have evaluative meaning, unless it is used ironically
    - Cf. *stilted and unnatural gibberish*

Agreement with respect to markables (II)

- Graduation
  - Bien, lo cierto es que el final de todas estas situaciones que os he comentado es **bastante previsible en todos los casos** (Películas, yes4.2)

    'Okay, the fact is that the end of all these predictable situations that I have told you about is quite predictable in all cases.'
  - A1: bastante previsible en todos los casos
  - A2: previsible
- Non-veridical or irrealis elements
  - Quiza a los niños les **gustara** pero la pelicula en cuestion es de lo mas sosa y aburrida (Películas no1.1)

    'Maybe children liked it, but the movie is absolutely bland and boring'

Agreement: language contrasts

- ## Signals of importance in Spanish (usually positive)
  - En la página web no podían faltar los juegos. **Destaco** uno, el de la rueda de la muerte, … (Películas, yes4.2)

    'On the web page there are, of course, games. I **highlight** one, the one with the death wheel…'
- ## If-clauses conditioning an evaluation (mostly in English)
  - If you're up for a good read that will have you constantly wanting more and getting it sooner or later, then here's your author and here's your book. (Books yes, 1)

Universidad Complutense Madrid

SFU SIMON FRASER UNIVERSITY THINKING OF THE WORLD

Conclusions

- Review and careful consideration of a number of difficult issues in applying Appraisal to a corpus

- Initial steps in corpus annotation and cross-annotator agreement indicate that the task is feasible

- Some aspects of the evaluative strategies are language-dependent

Universidad Complutense Madrid

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

References (I)

- Bednarek, M. (2009) Language patterns and Attitude. *Functions of Language* 16.2, 165-192.
- Biber, D. (1988) *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. (1995) *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge: Cambridge University Press.
- Biber, D. (2006) Stance in spoken and written university registers. *Journal of English for Academic Purposes* 5, 97-116.
- Biber, D. & E. Finegan (1989) Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text,* 9 (1), 93-124.
- Bloom, K., G. Navendu & S. Argamon (2007) Extracting appraisal expressions. In *Proceedings of HLT/NAACL* (pp. 308-315). Rochester, NY.
- Carretero, M. (2007). Subjectivity in English epistemic modality: a two-resource based approach. BELL (Belgian Journal of English Language and Literatures) New Series 5, 97-111.
- Carretero, M. & Taboada, M. (2009). Contrastive Analyses of Evaluation in Text: Key Issues in the Application of Appraisal Theory to Consumer Reviews in English and Spanish. Paper presented at the I Congreso Internacional de Lingüística de Corpus (CILC I), University of Murcia (Spain), May 7-9 2009.
- Carretero, M., Zamorano, J. R.; Nieto, F.; Alonso, C.; Arús, J. and A. Villamil. 2007. "An Approach to Modality for Higher Education Studies of English". In M. Genís, E. Orduna and D. García-Ramos (eds.), *Panorama de las lenguas en la enseñanza superior. ACLES 2005, Universidad Antonio de Nebrija.* CD-ROM. Madrid: Universidad Antonio de Nebrija. 91-101.
- Esuli, A. & F. Sebastiani (2006) Determining term subjectivity and term orientation for opinion mining. In *Proceedings of EACL-06, the 11th Conference of the European Chapter of the Association for Computational Linguistics*. Trento, Italy.
- Goldberg, A.B. & X. Zhu (2006) Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. In *Proceedings of HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing*. New York.

- Kennedy, A. & D. Inkpen (2006) Sentiment classification of movie and product reviews using contextual valence shifters. *Computational Intelligence, 22* (2), 110-125.
- Lavid, J. (2008) CONTRASTES: An online English-Spanish textual database for contrastive and translation learning. In B. Lewandowska-Tomaszczyk (ed), *Corpus Linguistics, Computer Tools, and Applications –State of the Art*. Frankfurt: Peter Lang Verlag.
- Lavid, J., Arús, J. and J.R. Zamorano (2007) Working with a bilingual English-Spanish textual database using SFL. *Paper presented at the 34th International Systemic-Functional Congress.* Odense (Denmark), July 2007.
- Martin, J.R. (2000) Beyond exchange: Appraisal systems in English. In S. Hunston and G. Thompson (Eds.), *Evaluation in Text: Authorial Distance and the Construction of Discourse* (pp. 142-175). Oxford: Oxford University Press.
- Martin, J.R. & P.R.R. White (2005) *The Language of Evaluation*. New York: Palgrave.
- Pang, B., L. Lee & S. Vaithyanathan (2002) Thumbs up? Sentiment classification using Machine Learning techniques*. In *Proceedings of Conference on Empirical Methods in NLP* (pp. 79-86).
- Stubbs, M. (1986) 'A matter of prolonged field work': Notes towards a modal grammar of English. *Applied Linguistics, 7* (1), 1-25.
- Taboada, M. (2008) SFU Review Corpus [Corpus]. Vancouver: Simon Fraser University, http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html.
- Taboada, M. & J. Grieve (2004) Analyzing appraisal automatically*. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07)* (pp. 158-161). Stanford University, CA.
- Taboada, M., C. Anthony & K. Voll (2006a) Creating semantic orientation dictionaries*. In *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC)* (pp. 427-432). Genoa, Italy.

References (III)

- Taboada, M., M.A. Gillies & P. McFetridge (2006b) Sentiment classification techniques for tracking literary reputation. In *Proceedings of LREC Workshop, "Towards Computational Models of Literary Analysis"* (pp. 36-43). Genoa, Italy.
- Taboada, M., K. Voll & J. Brooke (2008a) *Extracting Sentiment as a Function of Discourse Structure and Topicality* (Technical Report No. 2008-20): Simon Fraser University.
- Taboada, M., C. Anthony, J. Brooke, J. Grieve & K. Voll (2008b) SO-CAL: Semantic Orientation CALculator. Vancouver: Simon Fraser University.
- Turney, P. (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of 40th Meeting of the Association for Computational Linguistics* (pp. 417-424).
- Voll, K. & M. Taboada (2007) Not all words are created equal: Extracting semantic orientation as a function of adjective relevance. In *Proceedings of the 20th Australian Joint Conference on Artificial Intelligence* (pp. 337-346). Gold Coast, Australia.
- White, Peter R.R. (2002) Appraisal. In J. Verschueren et al. (eds.). *Handbook of pragmatics*, 1–27. Amsterdam: Benjamins.
- White, P.R.R. (2003) *An Introductory Course in Appraisal Analysis*. http://www.grammatics.com/appraisal

Labelling evaluative language in English and Spanish:
The case of Attitude in consumer reviews

Maite Taboada

mtaboada@sfu.ca

Marta Carretero

mcarrete@filol.ucm.es

http://www.sfu.ca/~mtaboada/

Universidad Complutense Madrid

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD