Form and function in evaluative language

The use of corpora to identify contextual valence shifters in a linguistically-motivated sentiment analysis system

Antonio Moreno Ortiz and Chantal Pérez Hernández
University of Málaga, Spain

In this paper we summarise current research in the analysis of evaluative language, both from a linguistic and a NLP perspective, and offer a description of the approaches that have been proposed, to focus on the use of text corpora as a common resource employed by practitioners of both disciplines. We then present some of the work carried out by our research team, Tecnolengua, which has concentrated on the construction of a domain-neutral, linguistically-motivated sentiment analysis tool that makes extensive use of lexical resources of various types. Among such resources, the inclusion of multiword expressions in our database is a key factor in the improved performance of the analyser, as they play an integral role in the creation of the context rules that we use to account for those cases in which the assigned valence of an individual word is modified by its linguistic environment.

1. Introduction

Emotions and opinions condition how humans communicate with each other and how they motivate their actions, so it is only natural that in the last twenty years the study of evaluative language has attracted the attention of a wide range of disciplines, from behavioural psychology to cognitive anthropology, with, of course, very different applications in mind (Janney 1996). From a purely linguistic perspective, different models have attempted to account for the relationship between the referential content we communicate and the "more affective contour" we may add to what we say (Carter 2004, 11). This relationship is not, however, easily identifiable, to the extent that some authors have gone as far as recognizing that "every utterance is characterised by the speaker's subjective emotional evaluation of the referentially semantic content" (Bakhtin (1953[1986], 84). Difficult as it may be, most linguistic descriptions include some reference to this functional dichotomy: representative/expressive (Buhler 1934), descriptive/expressive (Lyons 1977), ideational/interpersonal (Halliday 1994), expressive/relational (Fairclough 2001), to name but a few. Although most authors have focused on different text types and features of evaluative language, some theories have offered more comprehensive accounts of the nature of attitudinal language, such as "Modality and modulation" (Halliday 1994); "Evaluative orientations" (Lemke 1998); "Stance" (Biber et al. 1999; Conrad and Biber 2000); "Evaluation" (Thompson and Hunston 2000; Bednarek 2006; 2008) or "Appraisal Theory" (Martin and White 2005).

From a Natural Language Processing (NLP) point of view, the field of *sentiment analysis* (also known as *opinion mining*), which deals with the computational treatment of opinion and subjectivity in texts, has attracted increasing attention during the last few years (Pang and Lee 2008). With the advent of the Web 2.0 and the widespread adoption of social networking sites, it is easier than ever before to get access to vast amounts of emotion-loaded texts. Among these, product reviews are particularly interesting for companies to monitor, whilst different opinions and trends in political or social issues can be identified. Companies and organisations have traditionally employed public relations teams to do this job manually, but now these teams are increasingly relying on software tools that allow the automatic analysis of a massively growing amount of data. It is therefore hardly surprising that many companies have decided to add sentiment analysis tools to their social media measurement and monitoring resources, with a view to improving their business.

Most sentiment analysis systems, however, have been specifically developed for a particular subject domain and are based on supervised, statistical machine learning techniques (Pang and Lee 2004; 2005; Aue and Gamon 2005). Machine learning algorithms have indeed proved to be extremely useful, not only in the field of sentiment analysis, but in most text mining and information retrieval applications, although their obvious disadvantage in terms of functionality is their limited applicability to subject domains other than the one they were designed for. On the other hand, a growing number of initiatives in the area have explored the possibilities of employing unsupervised, knowledge-based approaches. These rely on a dictionary where lexical items have been assigned a valence (positive or negative), either extracted automatically from other dictionaries, or, more uncommonly, manually acquired. The degree of success of such approaches varies depending on a number of variables, of which the most salient is no doubt the quality and coverage of the lexical resources employed. In general, statistics-based approaches tend to be of limited application and achieve good recall, but low precision, whereas knowledge-based approaches usually display the opposite results: they are good at precision but may miss many sentiment-laden text segments (Andreevskaia and Bergler 2007).

The work carried out by our research team, Tecnolengua, has concentrated on the construction of a domain-neutral, linguistically-motivated sentiment analysis tool, which makes extensive use of lexical resources of various types. Of these, *context rules*, or *contextual valence shifters,* which determine the high success rate of our system, have been acquired both semi-automatically and manually using corpus techniques. In this paper we describe the current state of the art in the analysis of evaluative language, both from a linguistic and a NLP perspective. After offering a description of the approaches that have been proposed and the results obtained, both theoretically and in practical applications, we focus on the use of text corpora as a common resource employed by practitioners of both disciplines, albeit with different exploitation strategies. Finally, we show how the study of form-function interaction for the study of evaluative language and its applications can only be tackled successfully by studying language in context.

2. Linguistic approaches to the study of evaluation

Bednarek (2008, 9) offers a very complete summary of the different perspectives that can be found in the various approaches to the study of affect or emotion within the field of Linguistics, which, in turn, can be related to different sub-disciplines of linguistic research. She lists, among others, the following:

- The *cognitive approach*, which focuses on the study of words that refer to emotions, how emotions are conceptualised and the relation between emotions and their linguistic labels (i.e., Kövecses 2000).
- The *cross-linguistic approach*, concerned with the study of emotion terms across languages and the cultural values that determine the expression of emotions (Harkins and Wierzbicka 2001).
- The *functional approach*, which corresponds with the well-established tradition of research on the functions of language dating back to Bühler (1934) and more recent studies of expressive language (Leech 1994).
- The *syntactic approach*, exemplified by the work of Dirven (1997), concerned with the syntax of emotion terms and, to some extent, by the grammatical descriptions of attitude that can be found in Quirk et al. (1985), Biber et al. (1999) or Huddleston and Pullum (2002).
- The *conversation analytic approach*: these studies focus on the display of emotions in discourse and its structural organisation (Goodwin and Goodwin 2000).

- The *psycholinguistic approach*, concerned, among other things, with the study of the development of emotion-related language in childhood (Painter 2003).
- The *pragmatic/textlinguistic approach* is, perhaps, the most varied, active and heterogeneous of all the approaches proposed by Bednarek (2008, 9). Studies in this area are interested in many aspects of language and emotion, from the conventional displaying of affect through linguistic means, to the analysis of the influence of attitude on communicative decisions (Caffi and Janney 1994), the emotive prosody of texts (Bublitz 2003) used by the speakers to convey attitudes or the connection between emotion and speech acts (Weigand 2004).
- The *systemic-functional approach*, which could fit perfectly under the general umbrella of functional approaches mentioned above, although it is listed separately, to refer specifically to *appraisal theory*, the systemic-functional theory proposed by Martin and White (2005, *inter alia*) to describe the interpersonal metafunction of language, modelled in terms of systems of choices that impress attitude, emotion and evaluation in discourse.

The work of many scholars can, of course, be related to more than one of these approaches. For instance, Van Dijk's model, which focuses on the structure of journalistic discourse (Van Dijk 1998), can be regarded as both functional and textlinguistic). Conversely, many other areas of research that deal with language and emotions are excluded from this brief outline, such as those on intonation and prosody, or studies on the interaction between verbal and non-verbal characteristics of emotions and their expression (Selting 1994).

From the point of view of language-related disciplines, the sheer number of different terms used to refer to the phenomenon under study is remarkable. Most linguistic studies on evaluation include a preliminary note on the relevant terminology they use, together with a definition of their understanding and delimitation of the object of study, which is an indication of how little agreement there is on the topic. We summarise the most widely used terms and their implications in the following section.

2.1 Evaluating evaluation terms: Attitude, affect, stance, appraisal and evaluation

Thompson and Hunston (2000, 5) define *evaluation* as "the broad cover term for the expression of the speaker's or writer's attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about". The authors relate this attitude to a series of values such as certainty, obligation or desirability, which are subjective and culturally determined (Hunston 1994, 210). However, as the authors point out, some of the terms used in their definition (*stance, attitude, viewpoint*) have also been used by other scholars, sometimes as synonyms, sometimes to reflect different perspectives on the study of this phenomenon. Julian (2009, 52) uses the term *attitude* as a hypernym, to cater for a complex network of mental and emotional states which include "affects, beliefs, certainty, commitment, dispositions, emotions, ideology, standpoint, state of mind, or any other inner condition passing or permanent- of the kind". These aspects of our psychological, intellectual and emotional states must play a part in the way we view and judge the world and, accordingly, they get imprinted in our verbal interaction.

On the other hand, the term *stance*, associated with the work of Biber and his colleagues (Biber and Finnegan 1998; Conrad and Biber 2000; Hyland 2009) refers to the lexical and grammatical expression of an author's or speaker's attitudes, feelings, judgments and commitments concerning the propositional content of a message (Biber and Finnegan 1998, 93). Biber (2006) focuses on the differences between spoken and written registers, comparing frequencies of certain linguistic features in two or more corpora, while Conrad and Biber (2000) restrict their study to the grammatical devices used to frame a proposition using an adverbial expression, classifying those adverbials in three categories: *epistemic stance* (exemplified by the use of adverbials such as *probably*, *perhaps* or *undoubtedly*); *attitudinal stance* (with adverbials such as *fortunately, surprisingly* or *amazingly*); and *style stance* (the use of adverb phrases such as *quite frankly* or the noun phrase *no doubt*).

In this respect, although both the terms *stance* and *evaluation* are related to "evaluative language", *evaluation* refers more generally to the ascription of a value to an entity, whether inside or outside the text, while *stance* focuses on the study of possible indications in the text that a speaker/writer is communicating interactively with their prospective listener/reader. As Hunston (2011, 51) points out, *evaluation* is more likely to be expressed implicitly in a text, so a detailed, fine-grained analysis is required to study this phenomenon, whereas markers of interaction are more easily identified in formal, tractable terms, and thus, more related to quantitative approaches.[1]

Affect is another term usually found in the literature, and its relationship with evaluation is quite complex: whereas evaluation deals with the expression of opinion, the term *affect* is mostly used to talk about the expression of emotions or feelings (Bednarek 2006, 19) and although both *affect* and *evaluation* are related

1. Also under the "stance" research realm, the work of other scholars, such as Englebretson (2007) can be considered more quantitative in nature. Some of the contributions in that volume concentrate on conversation analysis of "stance-taking", which highlights the importance of stance as an activity, rather than as a set of traceable markers in a text.

to the expression of speaker approval or disapproval, evaluation is not necessarily concerned with how far speakers are emotionally engaged in discourse, nor with the kinds of expressions that may be used to arouse the hearer's emotions.[2]

One last term deserves special attention when it comes to the linguistic study of evaluative language: that of *appraisal.* Appraisal theory is set within the systemic-functional tradition (e.g., Halliday 1994). It was initially developed by Martin and White and the initiative is now being continued by a large number of scholars (Martin 2000; Martin and White 2005; Rothery and Stenglin 2000). In systemic-functional terms, a language performs three major functions: ideational (it constructs a world of experience), interpersonal (it creates relations between people), and textual (it organises instances of discourse). Language is also seen as a system of choices, in which meaning is created by making one choice out of a set of possibilities.

Appraisal theory places the interpersonal function of language at the centre of communicative interaction, regarded as a system of choices to impress attitude, emotion, and evaluation in discourse. Martin and White (2005, 34) define *appraisal* as "one of the three major discourse semantic resources construing interpersonal meaning (alongside involvement and negotiation)". Appraisal is divided into three interacting domains (or systems): ATTITUDE, concerned with our feelings, emotional reactions, judgments of behaviour, and evaluation of things; ENGAGEMENT, dealing with sourcing attitudes and the play of voices around opinion in discourse; and GRADUATION, concerned with values by which a speaker increases or diminishes the intensity of an utterance or the focus of their semantic categorisations.

ATTITUDE is further subdivided into three sub-systems: AFFECT, JUDGMENT and APPRECIATION. The first one, AFFECT, characterises phenomena by reference to emotions (*I'm happy, she's frightened of sharks*), whereas JUDGMENT evaluates morally human behaviour, by reference to a set of norms (*a clever/stupid person, a moral/immoral action*) and APPRECIATION includes resources used to evaluate the quality of processes, things, products and people (*a wonderful book*). These three sub-systems can be positive or negative (i.e. admiration vs. criticism in JUDGMENT) and are, of course, interrelated to the extent that borders between them are far from clear (Martin and White 2005, 57). In fact, Martin (2000, 147) regards AFFECT as the most basic system, and both JUDGMENT and APPRECIATION are recontextualisations or institutionalisations of affect (in some sort of evaluation matrix), with a view to controlling what people do or achieve.

2. In fact, Thompson and Huston (2000) regard *affect* as a type of evaluation, whereas Bednarek (2006, 20) claims that affect should be considered "a cover term for various approaches analysing the relationship between language and emotion."

Appraisal theory has been applied so far mostly to individual texts or relatively small corpora[3] and the assignment of a word or phrase to each category in the exhaustive taxonomies they propose may be rather controversial.[4] It does represent, however, a complete theoretical basis to account for the very wide scope covered by the notion of evaluative language. Even more, one of the greatest strengths of this theory is its treatment of the problem of implied evaluation, as it distinguishes between "inscribed" (explicit) and "evoked" (implicit) appraisal (the use of the explicit textual markers of assessment such as lexical choices, as opposed to contextual resources that "trigger" evaluative responses in the receiver).

As we see, linguistic approaches to evaluation differ not only in terminology, but also in scope, although they all seem to share a number of assumptions, or common ground, that permeate their theoretical underpinnings. Hunston (2011, 12–19) refers to these assumptions as "points of agreement", which can be summarised as follows:

– Evaluation is both subjective and intersubjective, and takes place within a social and ideological framework shared by writer and reader. Evaluative utterances express a personal opinion and display a broad range of lexical and other indicators of evaluative meaning, although some of them are highly context dependent. On the other hand, evaluation may be implied rather than stated explicitly.
– Evaluation is both contextual (in purely Firthian terms: it must be attested in language usage, because the immediate context of a word may change its polarity from positive to negative or *vice versa*) and cumulative: evaluative meanings tend to cluster together; in a customer review, for instance, the assessment of the product is the accumulation of all the different things said about it.
– Evaluation involves a target, or object, and a source: a person evaluates an object. The status of something constrains the criteria or grounds on which it can be given value. The source of evaluation is apparently simple, but in practice complicated by the process of attribution (an evaluation may be attributed to speakers other than the author of the text).
– The last point of agreement was already mentioned in our introduction: once we start the task of identifying evaluation "it becomes difficult reliably to identify anything that is not evaluative. Indeed it may be said that subjectivity and ideological value permeate the most objective discourse" (Hunston 2011, 19).

3. See Bednarek (2006; 2008) for a very interesting development of *Appraisal* theory and its application to the (manual) analysis of a large-scale corpus of newspaper articles (Bednarek 2006) and to a large section of the British Register Corpus (Bednarek 2008).

4. Hunston (2011, 19–22) or Bednarek (2006, 32–39) offer critical reviews of *appraisal* theory, highlighting its potential problems and the areas that deserve further attention.

2.2 Corpus approaches to evaluation: Grammar patterns and local grammars

The contribution of corpus techniques to the analysis of evaluative language has a long tradition, for instance, in the study of *stance*, which has concentrated mainly on register variation in the research carried out by Biber and his colleagues (see references in the previous section). Another fruitful area of application has been the study of lexico-grammatical patterns that are associated with evaluative expressions, in order to devise grammar patterns and local grammars of evaluation. These studies are based on statistical significance of co-occurrence, i.e., a combination of words that is found with a frequency higher than expected, compared to the relative frequencies of the component words.

Since the pioneering work of the British linguist J. R. Firth, frequency of co-occurrence has been the focus of many lexico-grammatical corpus-based studies carried out by leading scholars in the field of Corpus Linguistics. Some twenty years ago, Sinclair discussed the existence of two opposed models to describe the construction of meaning in language: the *open choice principle* and the *idiom principle* (Sinclair 1991, 109–121). Most grammatical descriptions seem to operate under the former, inasmuch as they separate the lexical and semantic aspects of words from the organisation of syntax in a slot-and-filler fashion. However, lexico-grammatical studies have shown that words do not occur at random in a text, and that we tend to co-select lexical items, so that they appear together in discourse with a statistically significant frequency. On the other hand, the idiom principle states that language users have available to them a "large number of semi-preconstructed phrases that constitute single choices" (Sinclair 1991, 110) even though they could be analysed in smaller lexico-grammatically meaningful segments. These preconstructed phrases can be totally fixed, as in *of course, to and fro*; however, this tendency to co-select words in larger chunks is far more pervasive than it might appear and, although they constitute units of meaning, they display different degrees of variation: some phrases have an indeterminate extent (they may be associated with particular types of subjects or objects, for instance), some of them allow internal lexical variation (*set sth. on fire* or *set fire to sth.*), others may allow internal syntactic variation or different word order, and still constitute what Sinclair (1996) or Stubbs (2001) call "extended units of meaning" (the phrase *it is not in his nature to…* might appear in different tenses and with different subjects, the negative particle can be replaced by a negative adverb, *it is hardly in his nature to…*, but certainly the elements in this construction have a probabilistic relation.

Stubbs (2001, 81) describes these "extended units of meaning" in terms of different "strengths of attraction", and expands Sinclair's original proposal to put forward a model that has proved to be very relevant for the study of evaluation in language. He distinguished the following:

– Collocations: this is a pure node-collocate relation, it refers to individual word forms or lemmas.
– Colligation: this is the term originally used by Firth (1957) to refer to the strong statistical association of a word and a particular grammatical category (a noun predetermined by demonstrative rather than by possessive deictics, for instance).
– Semantic preference: when a word collocates with a lexical set or a class of semantically related word-forms or lemmas.
– Discourse (or semantic) prosody: which express the speaker's attitude. Sinclair (1991, 74) gives the example of the verb "set in" used to refer to unpleasant states of affairs. Stubbs (2001) and Louw (1993) provide further examples of expressions being associated with positive or negative connotations (the verbs "cause" or "happen", for example, being associated with negative events). Louw states that semantic prosody is the "consistent aura of meaning with which a form is imbued by its collocates" (Louw 1993, 157).

It is interesting to note that Stubbs suggested that as well as collocating with purely positive or negative semantic groupings of words, words can also collocate with semantic sets that share an evaluative component. Since they are evaluative, prosodies express the speaker's reason for making the utterance, therefore identifying functional discourse units (Steward 2010). As we will see in the following sections, this is particularly relevant in the field of Sentiment Analysis in general, and in Sentitext's use of its lexical resources (see Sections 4.1 and 5) in particular, because in some cases, a number of multiword expressions were included in the database due to changes in the polarity of a word when it appeared as a component of a larger lexical unit.

Corpus-based identification of grammar patterns has traditionally been a productive area to explore the relationship between form and meaning in language, and only recently applied to the study of evaluation. Like collocations, the emergence of lexico-grammatical patterns is just a direct consequence of observing language from the *idiom principle* perspective (Hunston 2011, 123). Originally, the grammar pattern project was created as a coding system for the Collins Cobuild English Dictionary (Sinclair 1991). Since then, much work has been carried out to expand the project, both in the series of grammar pattern books published by Collins Cobuild for nouns, verbs and adjectives (Francis et al. 1996; 1998), in academic publications (Hunston and Francis 1999) and in the conception of local grammars proposed by Barnbrook and Sinclair (1995) or Hunston and Sinclair (2000).

Grammar patterns aimed at capturing in simple descriptions the recurrent behaviour of a word or a group of semantically related words. For example, Francis et al. (1996) collected together all the verbs in CCED that have the coding 'V *about* N' and grouped them to highlight semantic congruence (Hunston 2011, 123). Verbs were then grouped under six categories, for instance those that indicate mental processes such as *thinking* or *feeling*, including *forget*, and other verbs such as *agonise*, *agree*, *bother*, *brood*, etc. The approach has been criticised because it does not make any claims about the kind of relationship that holds between pattern and semantic class. Semantic groupings were created in an *ad hoc* manner, and the authors recognise that they did not make any claims in relation to the mental processing of grammar or the existence of semantic classes in the mind of the speaker (Hunston 2011, 123). Another problem area involved the status of the elements identified in each pattern through corpus analysis: only those items in the co-text that constitute defining characteristics of a particular verb or noun were mentioned (for the verb *recover*, for instance, it includes the pattern V *from* N (as in "recover from an illness"), but not the pattern V *in* N, (as in "recover in hospital"); both are prepositional phrases, and both appear with a high frequency in the corpus, but only the first one is considered a defining feature of the verb (the prepositional phrase with *from* was used to identify a class of verbs and plays a core semantic role in the pattern, whereas prepositional phrases indicating place with the preposition *in* appear with many verbs).

Several attempts have been made to complement the grammar pattern approach with other theoretical frameworks, such as Appraisal Theory or Fillmore's FrameNet Project:[5] Bednarek (2008) offers a complete account of the patterns of emotion terms, combining Hunston's patterns with FrameNet's twelve "emotion frames" to create what she terms "emotion profiles" and Hunston (2011, 130–138) complements Appraisal Theory with a corpus-based analysis of English adjectives and their lexico-grammatical patterns, and also compares Hunston and Sinclair's (2000) proposal of a local grammar of evaluation with FrameNet's semantic frames. The conclusions she draws from both experiments are quite realistic with regard to the possibilities of implementing large-scale systems to identify functional roles in unannotated texts (Hunston 2011, 150). What she rightly stresses as a direct conclusion of her work is the importance that phraseological units have in the expression of evaluation, particularly in the identification of intensifying phrases. She analyses a series of expressions that are associated with positive or negative nouns, whose function is just to add evaluative strength, as in "the depths of my ignorance" or "on the verge of a heart attack".

5. http://framenet.icsi.berkeley.edu

As we will see in the following sections, connections clearly exist between the research carried out by Hunston and the treatment given by our system, Sentitext, to multiword units. One concept in particular, that of *semantic reversal*, originally proposed by Sinclair (2004), which she does not fully explore, is directly connected to our understanding of Contextual Valence Shifters (see Section 5). Before we are able to put forward this concept, it is necessary to describe in more depth our approach to the computational implementation of a sentiment analysis system.

3. Computational approaches to sentiment analysis

Sentiment Analysis is tackled within Natural Language Processing (NLP) from the broader field of text mining (or text analytics), whose ultimate aim is to distil quantifiable data from raw text input. The emergence of text mining is motivated by the ever-increasing amount of text that Internet users generate, and the obvious benefits that businesses and organisations could obtain from tools capable of making sense of that text. The evaluative component of user-generated text is both high and relevant, thus calling for the emergence of a specific subfield.

The NLP perspective is obviously very different from the linguistic one. Generally speaking, there is no interest in discovering and analysing the cognitive and linguistic mechanisms that intervene in the thought-to-speech process, or in providing plausible explanations and grand schemes behind those processes. The aim is simply to turn text into computationally tractable data that tell something about the meaning of that text, employing whatever means yield faster and more reliable results. Not surprisingly, therefore, the field of Artificial Intelligence (AI) in general, and NLP in particular, has experienced an obvious shift from traditional, cognitive AI, concerned with thought processes, reasoning and cognitive modelling, to a more data-driven perspective, focused on performance rather than the underlying cognitive engine. In short, using the terms in Russell and Norvig (2010), there is now more interest in making tools that *act* humanly or rationally than those that *think* humanly or rationally.

This shift has meant, in effect, that the vast majority of present-day research in NLP relies heavily on statistics and, specifically, on Machine Learning algorithms, which provide the means to obtain objectively good results with little or no knowledge (linguistic or otherwise), using a set of well-established, off-the-shelf learning algorithms that have proved to offer very good results for an incredibly wide range of applications. Sentiment Analysis, as a subfield of text mining and NLP, is no exception to this trend, where the usual methodological approach involves supervised, statistical machine learning techniques.

Such approaches have indeed yielded very good results in the past (Pang and Lee 2004; Pang and Lee 2005). In fact, machine learning techniques, in any of their flavours, have proved extremely useful, not only in the field of Sentiment Analysis, but in most text mining and information retrieval applications, as well as a wide range of data-intensive computational tasks. However, their obvious disadvantage in terms of functionality is their limited applicability to subject domains other than the one they were designed for. In fact, it has become a commonplace assertion that successful results depend to a large extent on developing systems that have been specifically developed for a particular subject domain. Although interesting research has been done aimed at extending domain applicability (Aue and Gamon 2005), such efforts have shown limited success. An important variable for these approaches is the amount of labelled text available for training the classifier, although they perform well in terms of recall even with relatively small training sets (Andreevskaia and Bergler 2007).

On the other hand, a growing number of initiatives in the area have explored the possibilities of employing unsupervised lexicon-based approaches. These rely on dictionaries where lexical items have been assigned either a *polarity* or a *valence* tag,[6] extracted either automatically from other dictionaries, or, more uncommonly, manually. The works by Hatzivassiloglou and McKeown (1997) and Turney (2002) are perhaps classical examples of such an approach. The most salient work in this category is Taboada et al. (2011), whose dictionaries were created manually, and use an adaptation of Polanyi and Zaenen's (2006) concept of Contextual Valence Shifters to produce a system for measuring the semantic orientation of texts, which they call SO-CAL(culator). This is exactly the approach we used in our Sentitext system for Spanish (Moreno-Ortiz et al. 2010; 2011).

Combining both methods (machine learning and lexicon-based techniques) has been explored by Kennedy and Inkpen (2006), who also employed contextual valence shifters, although they limited their study to one particular subject domain (the traditional movie reviews), using a "traditional" sentiment lexicon (the General Inquirer), which resulted in the "term-counting" (in their own words) approach.

The degree of success of knowledge-based approaches varies depending on a number of variables, of which the most relevant is no doubt the quality and coverage of the lexical resources employed, since the actual algorithms employed to weigh positive against negative segments are in fact rather simple.

6. Although the terms *polarity* and *valence* are sometimes used interchangeably in the literature, especially by those authors developing binary text classifiers, we restrict the usage of the former to non-graded, binary assignment, i.e., positive or negative, whereas the latter is used to refer to an *n*-point semantic orientation scale.

4. Sentitext: A sentiment analysis system for Spanish[7]

Work within the field of Sentiment Analysis for Spanish is, by far, scarcer than that for English. Cruz et al. (2008) developed a document classification system for Spanish similar to Turney (2002), i.e. unsupervised, though they also tested a supervised classifier that produced better results. In both cases, they used a corpus of movie reviews taken from the Spanish *Muchocine* website. Boldrini et al. (2009) carried out a preliminary study in which they used machine learning techniques to mine opinions in blogs. They created a corpus for Spanish using their Emotiblog system, and discussed the difficulties they encountered while annotating it. Balahur et al. (2009) also presented a method of emotion classification for Spanish, this time using a database of culturally dependent emotion triggers.

Finally, Brooke et al. (2009) adapted a lexicon-based sentiment analysis system for English (Taboada et al. 2011) to Spanish by automatically translating the core lexicons and adapting other resources in various ways. They also provide an interesting evaluation that compares the performance of both the original (English) and translated (Spanish) systems using both machine learning methods (specifically, SVM) and their own lexicon-based semantic orientation calculation algorithm, the above mentioned SO-CAL. They found that their own weighting algorithm, which is based on the same premises as our system (see below), achieved better accuracy for both languages, but the accuracy for Spanish was well below that for English.

Our system, Sentitext (Moreno-Ortiz et al. 2010; 2011), is very similar to Brooke et al.'s in design: it is also lexicon-based and it makes use of a similar calculation method for semantic orientation. It differs in that the lexical knowledge has been acquired semi-automatically and then fully manually revised from the ground up over a long period of time and with a strong commitment to both coverage and quality. It makes no use of user-provided, explicit ratings that supervised systems typically rely on for the training process, and it produces an index of semantic orientation based on weighing positive against negative text segments, which is then transformed into a ten-point scale and a five-star rating system.

From an implementation perspective, Sentitext is a web-based, client-server application written in C++ (main code) and Python (server). The only third-party component in the system is Freeling (Atserias et al. 2006; Padró 2011), a powerful, accurate, multi-language NLP suite of tools, which we use for basic morphosyntactic analysis. Currently, only one client application is available, developed in Adobe Flex, which takes an input text and returns the results of the analysis in

7. The application can be accessed and tested online at http://tecnolengua.uma.es/sentitext.

several numerical and graphical ways, including visual representations of the text segments that were identified as sentiment-laden. Lexical information is stored in a relational database (MySQL).

Being a linguistically-motivated sentiment analysis system, special attention is paid to the representation and management of the lexical resources. The underlying design principle is to isolate lexical knowledge from processing as much as possible, so that the processors can use the data directly from the database. The idea behind this design is that all lexical sources can be edited at any time by any member of the team, which is facilitated by a PHP interface specifically developed to this end (GDB). This kind of flexibility would not be possible with the monolithic design typical of earlier proof-of-concept systems.

4.1 Lexical resources

Sentitext relies on three major sources: the individual words dictionary (*words*), the multiword expressions dictionary (*mwords*), and the context rules set (*crules*), which is our implementation of Contextual Valence Shifters.

The individual words dictionary currently contains over 9,400 items, all of which are labeled for valence. The acquisition process for this dictionary was inspired by the bootstrapping method recurrently found in the literature (e.g., Riloff and Wiebe 2003; Gamon and Aue 2005). Lexical items in both dictionaries in our database were assigned one of the following valences: −2, −1, 0, 1, 2. Since the words dictionary contains only sentiment-carrying items, no 0-valence word is present.

The most similar sentiment analysis system to ours (Taboada et al. 2011) uses a scale from −5 to 5, which makes sense for a number of graded sets of near synonyms such as those given as examples by the authors (273). In our opinion, however, as more values are allowed, it becomes increasingly difficult to decide on a specific one while maintaining a reasonable degree of objectivity and agreement among different (human) acquirers, especially when there is no obvious graded set of related words, which is very often the case.

There are two ways in which the original valence of a word or phrase can be modified by the immediately surrounding context: the valence can change in degree (intensification or downtoning), or it may be inverted. Negation is the simplest case of valence inversion.

The idea of Contextual Valence Shifters (CVS) was first introduced by Polanyi and Zaenen (2006), and implemented for English by Andreevskaia and Bergler (2007) in their CLaC System, and by Taboada et al. (2011) in their Semantic Orientation CALculator (SO-CAL). To our knowledge, apart from Brooke et al.'s (2009) adaptation of the SO-CAL system, Sentitext is the only sentiment analysis system to implement CVS for Spanish natively.

4.2 Global sentiment value

An important variable concerning sentiment analysis is the degree of granularity that the system aims to achieve. Most work on the field has focused on the *Thumbs up or thumbs down* approach, i.e., producing a positive or negative rating. Turney's (2002) work, from which the name derives, is no doubt the most representative. A further step involves an attempt to compute not just a binary classification of documents, but a numerical rating on a scale. The *rating inference* problem was first posed by Pang and Lee (2005), and the approach is usually referred to as *seeing stars* in reference to this work.

Sentitext provides results as a number of metrics in the form of an XML file, which is then used to generate the reports and graphical representations of the data. The crucial bit of information is the Global Sentiment Value (GSV), a numerical score (on a 0–10 scale) for the sentiment of the input text. Other data include the total number of words, total number of lexical words (i.e., content, non-grammatical words), number of neutral words, etc.

To arrive at the global value, a number of scores are computed beforehand, the most important of which is what we call Affect Intensity, which modulates the GSV to reflect the percentage of sentiment-conveying words the text contains.

Before we explain how this score is obtained, it is worth stressing the fact that we do not count words (whether positive, negative, or neutral), but *text segments* that correspond to lexical units (i.e., meaning units from a lexicological perspective, or "units of meaning" in the sense explained in Section 2.2).

As we mentioned before, items in our dictionaries are marked for valence with values in the range −2 to 2. Intensification context rules can add up to three marks, for a maximum score of 5 (negative or positive) for any given segment.

The simplest way to compute a global value for sentiment would be to add negative values on the one hand and positive values on the other, and then establishing a result by simple subtraction. However, as others have noted (e.g., Taboada et al. 2011), things are more complicated than that. Our Affect Intensity measure is an attempt to capture the impact that different proportions of sentiment-carrying segments have in a text. We define Affect Intensity simply as the percentage of sentiment-carrying segments. Affect Intensity is not used directly in computing the global value for the text; however, an intermediate step consists of adjusting the upper and lower limits (initially −5 and 5). The Adjusted Limit equals the initial limit unless the Affect Intensity is greater than 25 (i.e., over 25% of the text's lexical items are sentiment-carrying. Obviously, using this figure is arbitrary, and has been arrived at simply by trial and error. The Adjusted Limit is obtained by dividing the Affect Intensity by 5 (since there are 5 possible negative and positive valence values).

A further variable needs some explaining. Our approach to computing the GSV is similar to Polanyi and Zaenen's (2006) original method, in which equal weight is given to positive and negative segments, but it differs in that we place more weight on extreme values. This is motivated by the fact that it is relatively uncommon to come across such values (e.g. "extremely wonderful"), so when they do appear, it is a clear marker of positive sentiment. Other implementations of Contextual Valence Shifters (Taboada et al. 2011) have put more weight only on negative segments when modified by valence shifters (up to 50% more weight), operating under the so-called "positive bias" assumption (Kennedy and Inkpen 2006), i.e., negative words and expressions appear more rarely than positive ones, and therefore have a stronger cognitive impact, which should be reflected in the final sentiment score.

In our implementation, equal weight is placed on positive and negative values. However, we do not simply assign more weight to both extremes of the scale (−5 and 5), we place more weight on each increasingly toward both ends of the scale.

The resulting method for obtaining the Global Sentiment Value for a text is defined as:

$$GSV \; \frac{(\sum_{\mathrm{i}-1}^{5} 2.5 \cdot i \cdot N_i \mid \sum_{\mathrm{i}-1}^{5} 2.5 \cdot i \cdot P_{\mathrm{i}})}{5 \cdot (LS \quad NS)}$$

where N_i is the number of each of the negative valences found, and P_i is the equivalent for positive values. The sum of both sets is then multiplied by the Affect Intensity. *LS* is the number of lexical segments and *NS* is the number of neutral ones. Although not expressed in the equation, the number of possible scale points (5) needs to be added to the resulting score, which, as mentioned before, is on a 0–10 scale.

5. Context rules and contextual valence shifters: The use of corpora to identify and modulate valence assignment in text

It is important to understand the way our context rules work in order to appreciate how closely they interact with the other lexical data sources, especially the multiword dictionary. Simply accounting for negative and positive words and phrases found in a text would not be enough. There are two ways in which their valence can be modified by the immediately surrounding context: the valence can change in degree (intensification or downtoning), or it may be inverted altogether. Negation is the simplest case of valence inversion.

Our CVS system is implemented in what we call Context Rules, which are expressed as data structures that are parsed against the lemmatized input text. Table 1 describes its components.

Table 1. Nouns researched in our corpus, with valence assignment and classified according to Parrott (2001)

Data attribute	Description
Unit Form	Freeling-compliant morpho-syntactic definition of the item being modified (e.g.: "AQ").
Unit Sign	Polarity of the item being modified (e.g. "+").
CVS Definition	Modifier definition (e.g.: "muy").
CVS Position	position of the modifier (e.g. "L" for left).
CVS Span	Maximum number of words where the modifier can be found from the modified item.
Result	Valence result of the modification. This result can be expressed as either an operator or a set valence. Operators are one of the following: – INV (valence/polarity INVersion) – INTn (valence INTensification of *n*) – DOWn (valence DOWntoning of *n*).

The *n* argument in the last two operators is the degree by which the operator is to be applied. The result can also be a set valence, in which case it looks like any valence expressed in the dictionaries.

This system allows us to describe fairly elaborate context rules. For instance, having multiword modifiers such as those in (1) and (2) below. A context rule for type (1) constructions would cause the polarity of the negative adjective to be inverted, whereas a rule for type (2) constructions would intensify the valence of the negative adjective. In many senses, our use of context rules is very similar to those grammar patterns devised by Hunston and Francis referred to in Section 2.2, with the added advantage that we count on an existing lexicon of individual words marked with an assigned valence, which allows us to express things like "negative adjective". On the other hand, all resources are annotated using the same scheme, the one imposed by our morpho-syntactic analyser, Freeling.

(1) no tener nada de (be not at all) + negative adjective
Ese no tiene nada de tonto/estúpido/…
'he is not at all dumb/stupid…'

(2) (ser) un completo (be a complete) + negative adjective
Es un completo inútil
he's a complete idiot

What is interesting about this representation scheme is that it gives us greater flexibility than simply having a repository of multiword expressions. Without context rules, it would be very difficult to represent (and successfully process for sentiment analysis) these types of expressions, in which part of them is defined by the existence of a semantic prosody feature that triggers a certain polarity (e.g., adjectives denoting a negative quality).

As we mentioned in Section 2.2, semantic prosody is a strongly collocational phenomenon (Louw 2000, 50) and it depends on a word being frequently associated with others that carry some sort of evaluative meaning, and needs to be distinguished from connotation (or in sentiment analysis terms, from valence (see Section 2.1), in which the semantic associations that we make with a word are irrespective of co-occurrence factors. In our sentiment analysis system, these collocational phenomena are treated in two different ways: as multiword expressions with a particular valence assigned to them, or as context rules, when a word collocates with semantic set or a particular colligational category (i.e. adverbs of negation, or emphasisers, see Section 4.2).

We have relied heavily on corpora for the acquisition process of all lexical sources, but this is especially true of context rules, since no such specific resource was available.[8] Our very pragmatic approach was to use the original seed set for the semi-automatic acquisition of individual words, expanded to include other parts of speech and synonyms, and manually analyse the contexts in which they appeared, obtaining recurrent morphosyntactic patterns in which they were observed to modify their original valence.

Our seed set was based on Parrott's (2001) classification of emotions. Table 2 displays a sample of nouns we employed to identify context rules, e.g., *violencia extrema* (intensification), and phrases, e.g., *no tener queja* (inversion). The numbers in square brackets are the valence the nouns have in the individual words dictionary.

This simple approach allowed us to identify a large number of recurrent valence-modifying patterns, which we wouldn't have thought of otherwise. Some valence shifters are fairly obvious, such us negation by the adverb *no*, but others could hardly be obtained by introspection alone. We proceeded by part of speech, studying left and right contexts with different spans, obtaining long lists of modifiers. Table 3 below shows some examples for nouns.

8. For the individual words dictionary, we developed an application that took as input the OpenOffice Spanish thesaurus and a seed set of emotion-laden antonyms. The application interacted with the lexicographer offering candidate words in a cyclic looping over the full thesaurus. Most phrases where obtained from diverse sources and semi-automatically given the Freeling format. For these resources we basically used the corpus to check on hard cases.

Table 2. Nouns researched in our corpus, with valence assignment and classified according to Parrott (2001)

Emotion	Assigned nouns and valences
rage	violencia [−2], ataque [−2]
neglect	ignorancia [−2], deterioro [−2]
nervousness	amenaza [−2], preocupación [−2]
pride	dignidad [2], éxito [2]
irritation	enfado [−2], queja [−2], provocación [−1]
cheerfulness	alegría [2], felicidad [2]
lust	ambición [−1]
suffering	dolor [−2], malestar [−1]
sadness	tristeza [−2], desolación [−2]
affection	solidaridad [2], empatía [1]
contentment	tranquilidad [2], satisfacción [1]
exasperation	aburrimiento [−1], cansancio [−1]
horror	pánico [−2], devastación [−2]
disgust	corrupción [−2], desprecio [−2]
shame	culpa [−2], equivocación [−1]
relief	alivio [2], seguridad [2]
optimism	confianza [1], esperanza [1]
envy	envidia [−2], celos [−2]
disappointed	decepción [−2], fracaso [−2]
sympathy	ayuda [1], compasión [1]
longing	ilusión [2], ensueño [1]

Table 3. Valence modifying patterns for nouns in Spanish

Modification type	Noun polarity	Intensification examples	Inversion examples
Premodification by verb or deverbal noun	Positive	*asegurar, defender, duplicar, facilitar*	*criticar, bloquear contradecir, destruir*
Premodification by verb or deverbal noun	Negative	*agravar, alentar, exacerbar, reavivar*	*abordar, considerar, desdramatizar*
Premodification by adjective	Positive	*absoluto, evidente, imperecedero*	*aparente, cierto, pobre, presunto, solo*
Premodification by adjective	Negative	*completo, eterno, grave, insoluble*	*insignificante, ningún, mínimo*
Postmodification by adjective	Positive	*absoluto, continuado, generalizado*	*limitado, pasajero, posible, relativo*
Postmodification by adjective	Negative	*abierto, acuciante, adicional, candente*	*secundario, simple, inferior, tenue*

Work on context rules is hard, and our efforts on-going. Furthermore, since our interests are totally focused on implementation concerns, we have run into practical issues when it comes to applying context rules: some rules overlap with others, and it becomes increasingly hard for our current algorithms to apply them successfully as the number of rules escalates. It is, however, a straightforward approach that is able to account for a large number of cases with a fairly reduced number of rules.

6. Conclusion

In this paper we have briefly summarised the current state of the art in the analysis of evaluative language, both from a linguistic and a NLP perspective. We have offered a description of the approaches that have been proposed and the results obtained, both theoretically and in practical applications, to focus on the use of text corpora as a common resource employed by practitioners of both disciplines, albeit with different strategies and interests.

As first-hand experience, we have also presented some of the work carried out by our research team, Tecnolengua, which has concentrated on the construction of a domain-neutral, linguistically-motivated sentiment analysis tool that makes extensive use of lexical resources of various types. We have also discussed our implementation and acquisition process of *contextual valence shifters*, which determine the high success rate of our system. The inclusion of multiword expressions ("extended units of meaning" in linguistic terms) in our database is a key factor in the improved performance of Sentitext, as they play an integral role in the creation and application of context rules, and improve precision by avoiding "false positives". They also help to reduce ambiguity by blocking a number of highly polysemous words that are components of these multiword units, thus improving valence assignment of a word or multiword construction. We hope we have managed to show that both linguistic and NLP perspectives coalesce: the study of form-function interaction for the study of evaluative language and its applications can only be tackled successfully by studying language in context.

References

Andreevskaia, Alina, and Sabine Bergler. 2007. "CLaC and CLaC-NB: Knowledge-based and Corpus-based Approaches to Sentiment Tagging." In *Proceedings of the 4th International Workshop on Semantic Evaluations*, 117–120. Prague, Czech Republic: Association for Computational Linguistics.

Atserias, Jordi, Bernardino Casas, Elisabet Cornelles, Meritxell González, Lluis Padró, and Muntsa Padró. 2006. "FreeLing 1.3: Syntactic and Semantic Services in an Open-source NLP Library." In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, 2281–2286. Genova, Italy.

Aue, Anthony, and Michael Gamon. 2005. "Customizing Sentiment Classifiers to New Domains: A Case Study." In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*. Borovets, Bulgaria. Available at: http://research.microsoft.com/pubs/65430/new_domain_sentiment.pdf

Bakhtin, M. M. 1953. "The Problem of Speech Genres." In *Speech Genres and Other Late Essays*, ed. by C. Emerson, and M. Holquist, 60–102. Austin: University of Texas Press.

Balahur, Alexandra, Zornitsa Kozareva, and Andrés Montoyo. 2009. "Determining the Polarity and Source of Opinions Expressed in Political Debates." In *Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing '09* 468–480. Berlin: Springer-Verlag.

Barnbrook, J., and John Sinclair. 1995. "Parsing Cobuild Entries." In *The Languages of Definition: The Formalisation of Dictionary Definitions for Natural Language Processing*, ed. by J. Sinclair, M. Hoelter, and C. Peters, 13–58. Luxemburg: European Commission.

Bednarek, Monika. 2006. *Evaluation in Media Discourse: Analysis of a Newspaper Corpus*. Research in Corpus and Discourse. London: Continuum.

Bednarek, Monika. 2008. *Emotion Talk Across Corpora*. New York: Palgrave Macmillan. DOI: 10.1057/9780230285712

Biber, D. 2006. "Stance in Spoken and Written University Registers." *Journal of English for Academic Purposes* 5: 97–116. DOI: 10.1016/j.jeap.2006.05.001

Biber, Douglas, and Edward Finegan. 1998. "Adverbial Stance Types in English." *Discourse Processes* 11: 1–34. DOI: 10.1080/01638538809544689

Biber, Douglas, Stig Johansson, Geoffrey N. Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Boldrini, Ester, Alexandra Balahur, Patricio Martínez-Barco, and Andrés Montoyo. 2009. "EmotiBlog: An Annotation Scheme for Emotion Detection and Analysis in Non-traditional Textual Genres." In *Proceedings of the 2009 International Conference on Data Mining*, 491–497. Las Vegas, USA: CSREA Press.

Brooke, Julian, M. Tofiloski, and Maite Taboada. 2009. "Cross-Linguistic Sentiment Analysis: From English to Spanish." In *Proceedings of RANLP 2009, Recent Advances in Natural Language Processing*, 50–54. Borovets, Bulgaria.

Bublitz, W. 2003. "Emotive Prosody: How Attitudinal Frames Help Construct Context." In *Anglistentag 2002, Bayreuth – Proceedings*, ed. by E. Mengel, H. Schmid, and M. Steppat, 381–391. Trier: Wissenschaftlicher Verlag Trier.

Buhler, K. 1934. *Speachtheorie. Die Darstellungsfunktion Der Sprache*. Jena: Fisher.

Caffi, Claudia, and Richard W. Janney. 1994. "Towards a Pragmatics of Emotive Communication." *Journal of Pragmatics* 22: 325–373. DOI: 10.1016/0378-2166(94)90115-5

Carter, Ronald. 2004. *Language and Creativity*. London: Routledge.

Conrad, Susan, and Douglas Biber. 2000. "Beyond Exchange: Appraisal Systems in English." In *Evaluation in Texts. Authorial Stance and the Construction of Discourse*, ed. by Susan Hunston, and Geoffrey Thompson, 56–73. Oxford: Oxford University Press.

Cruz, Fermín, José A. Troyano, Fernando Enríquez, and Javier Ortega. 2008. "Clasificación de documentos basada en la opinión: Experimentos con un corpus de críticas de cine en español. [A classification of documents based on opinion: Experiments with a corpus of movie reviews in Spanish]." *Procesamiento del Lenguaje Natural* 41: 73–80.

Dirven, René. 1997. "Emotions as Cause and the Cause of Emotions". In *The Language of Emotions*, ed. by Susanne Niemeier, and René Dirven, 55–86. Amsterdam: John Benjamins. DOI: 10.1075/z.85

Englebretson, Robert. 2007. *Stancetaking in Discourse*. Amsterdam: John Benjamins. DOI: 10.1075/pbns.164

Fairclough, Norman. 2001. *Language and Power*. (*2nd Revised Edition*). Harlow: Longman.

Firth, John R. 1957. "A Synopsis of Linguistic Theory, 1930–1955." *Studies in Linguistic Analysis. Special Volume, Philological Society*. 1–32.

Francis, Gill, Susan Hunston, and Elizabeth Manning. 1996. *Collins Cobuild Grammar Patterns 1: Verbs*. London: Harper Collins.

Francis, Gill, Susan Hunston, and Elizabeth Manning. 1998. *Collins Cobuild Grammar Patterns 2: Nouns and Adjectives*. London: Harper Collins.

Gamon, Michael, and Anthony Aue. 2005. "Automatic Identification of Sentiment Vocabulary: Exploiting Low Association with Known Sentiment Terms." In *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, 57–64. Ann Arbor, Michigan: Association for Computational Linguistics.

Goodwin, Charles, and M. Goodwin. 2000. "Emotion Within Situated Activity". In *Linguistic Anthropology: A Reader*, ed. by A. Duranti, 239–257. Malden, MA: Blackwell.

Halliday, Michael A. K. 1994. *An Introduction to Functional Grammar (2nd Ed.)*. London: Edward Arnold.

Harkins, Jean, and Anna Wierzbicka. 2001. *Emotions in Cross-linguistic Perspective*. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110880168

Hatzivassiloglou, Vasileios, and Kathleen R. McKeown. 1997. "Predicting the Semantic Orientation of Adjectives." In *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics*, 174–181. Madrid, Spain: Association for Computational Linguistics.

Huddleston, Rodney D., and Geoffrey K. Pullum (eds). 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hunston, Susan. 1994. "Evaluation and the Planes of Discourse: Status and Value in Persuasive Texts." In *Advances in Written Text Analysis*, ed. by Michael Coulthard, 191–218. London: Routledge.

Hunston, Susan, and Gill Francis. 1999. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.

Hunston, Susan, and John Sinclair. 2000. "A Local Grammar of Evaluation." In *Evaluation in Texts. Authorial Stance and the Construction of Discourse*, ed. by Susan Hunston, and Geoffrey Thompson, 75–100. Oxford: Oxford University Press.

Hunston, Susan. 2011. *Corpus Approaches to Evaluation. Phraseology and Evaluative Language*. New York: Routledge.

Hunston, Susan, and Geoff Thompson (eds). 2000. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.

Hyland, Ken. 2009. "Corpus Informed Discourse Analysis: The Case of Academic Engagement." In *Academic Writing: At the Interface Between Corpus and Discourse*, ed. by Maggie Charles, Diane Pecorati, and Susan Hunston, 110–128. London: Continuum.

Janney, Richard W. 1996. *Speech and Affect. Emotive Uses of English*. Munich: R. W. Janney.

Julian, P. 2009. *An Exploration of Strategies to Convey Evaluation in the "Notebook" Texts*. University of Birmingham: Unpublished PhD Thesis.

Kennedy, Alistair, and Diana Inkpen. 2006. "Sentiment Classification of Movie Reviews Using Contextual Valence Shifters." *Computational Intelligence* 22 (2): 110–125. DOI: 10.1111/j.1467-8640.2006.00277.x

Kövecses, Zoltán. 2000. *Metaphor and Emotion. Language, Culture and Body in Human Feeling*. Cambridge: Cambridge University Press.

Leech, Geoffrey N. 1994. *Semantics*. Harmondsworth: Penguin Books Ltd.

Lemke, Jay L. 1998. "Resources for Attitudinal Meaning: Evaluative Orientations in Text Semantics." *Functions of Language* 5 (1): 33–56. DOI: 10.1075/fol.5.1.03lem

Louw, William E. 1993. "Irony in the Text or Insincerity in the Writer?" In *Text and Technology*, ed. by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 157–176. Amsterdam: John Benjamins.

Louw, William E. 2000. "Contextual Prosodic Theory: Bringing Semantic Prosodies to Life." In *Words in Context: A Tribute to John Sinclair on His Retirement*, ed. by C. Heffer, H. Sautson, and G. Fox, 48–94. Birmingham: University of Birmingham.

Lyons, John. 1977. *Semantics*. Vol. 2. Cambridge: Cambridge University Press.

Martin James R. 2000. "Beyond Exchange: Appraisal Systems in English." In *Evaluation in Texts. Authorial Stance and the Construction of Discourse*, ed. by Susan Hunston, and Geoffrey Thompson, 142–175. Oxford: Oxford University Press.

Martin James R., and Peter R. R. White. 2005. *The Language of Evaluation: Appraisal in English*. Hampshire: Palgrave Macmillan.

Moreno-Ortiz, Antonio, Francisco Pineda Castillo, and Rodrigo Hidalgo García. 2010. "Análisis de valoraciones de usuario de hoteles con Sentitext: Un sistema de análisis de sentimiento independiente del dominio [An analysis of hotel users reviews with Sentitext: A domain-neutral analysis of sentiment]." *Procesamiento de Lenguaje Natural* 45: 31–39.

Moreno-Ortiz, Antonio, Chantal Pérez-Hernández, and Rodrigo Hidalgo-García. 2011. "Domain-neutral, Linguistically-motivated Sentiment Analysis: A Performance Evaluation." In *Actas Del XXVII Congreso De La Sociedad Española Para El Procesamiento Del Lenguaje Natural*, 847–856.

Padró, Lluís. 2011. "Analizadores multilingües en FreeLing [Multilingual analysers in FreeLing]." *Linguamatica* 3 (2): 13–20.

Painter, Claire. 2003. "Developing Attitude: An Ontogenetic Perspective on Appraisal." *Text* 23 (2): 183–209.

Pang, Bo, and Lillian Lee. 2004. "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarisation Based on Minimum Cuts." In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, 271–278. Barcelona, Spain: Association for Computational Linguistics.

Pang, Bo, and Lillian Lee. 2005. "Seeing Stars: Exploiting Class Relationships for Sentiment Categorisation with Respect to Rating Scales." In *Proceedings of ACL 2005*, 115–124. Stroudsburg, PA, USA: Association for Computational Linguistics.

Pang, Bo, and Lillian Lee. 2008. "Opinion Mining and Sentiment Analysis." *Foundations and Trends in Information Retrieval* 2 (1–2): 2–135.

Parrott, W. Gerrod. 2001. *Emotions in Social Psycology*. Philadelphia: Psychology Press.

Polanyi, Livia, and Annie Zaenen. 2006. "Contextual Valence Shifters." In *Computing Attitude and Affect in Text: Theory and Applications*, ed. by James G. Shanahan, and Yan Qu, 1–10. Dordrecht: Springer. DOI: 10.1007/1-4020-4102-0_1

Quirk, Randolph, Sidney Greenbaum, Geoffrey N. Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Harlow: Longman.

Riloff, Ellen, and Janyce Wiebe. 2003. "Learning Extraction Patterns for Subjective Expressions." In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP'03*. 105–112. Stroudsburg, PA, USA: Association for Computational Linguistics.

Rothery, Joan and Maree Stenglin. 2000. "Interpreting Literature: The Role of APPRAISAL." In *Researching Language in Schools and Functional Linguistics*, ed. by Len Unsworth, 222–244. London: Cassell.

Russell, Stuart, and Peter Norvig. 2010. *Artifical Intelligence: A Modern Approach*. Boston: Prentice Hall.

Selting, Margret. 1994. "Emphatic Speech Style – with Special Focus on the Prosodic Signalling of Heightened Emotive Involvement in Conversation." *Journal of Pragmatics* 22: 375–408. DOI: 10.1016/0378-2166(94)90116-3

Sinclair, John. 1991. *Corpus, Concordance, Collocation: Describing English Language*. Oxford: Oxford University Press.

Sinclair, John. 1996. "The Search for Units of Meaning." *Textus IX* (1): 75–106.

Sinclair, John. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Steward, Dominic. 2010. *Semantic Prosody: A Critical Evaluation*. New York: Routledge.

Stubbs, Michael. 2001. *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Taboada, Maite, Julian Brooks, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. "Lexicon-Based Methods for Sentiment Analysis." *Computational Linguistics* 37 (2): 267–307. DOI: 10.1162/COLI_a_00049

Thompson, Geoffrey, and Susan Hunston. 2000. "Evaluation: An Introduction." In *Evaluation in Texts. Authorial Stance and the Construction of Discourse*, ed. by Susan Hunston, and Geoffrey Thompson, 1–27. Oxford: Oxford University Press.

Turney, Peter D. 2002. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews." In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 417–424. Philadelphia, USA.

Van Dijk, Teun A. 1998. "Opinions and Ideology in the Press." In *Approaches to Media Discourse*, ed. by A. Bell, and P. Garret, 21–63. Oxford: Blackwell.

Weigand, Edda (ed). 2004. *Emotion in Dialogic Interaction. Advances in the Complex*. (Current Issues in Linguistic Theory, 248). Amsterdam: John Benjamins. DOI: 10.1075/cilt.248