

Tecnolengua Lingmotif at EmoInt-2017: A lexicon-based approach*

Antonio Moreno-Ortiz

University of Málaga

Spain

amo@uma.es

Abstract

In this paper we describe Tecnolengua Group's participation in the shared task on emotion intensity at WASSA 2017. We used the Lingmotif tool and a new, complementary tool, Lingmotif Learn, which we developed for this occasion. We based our intensity predictions for the four test datasets entirely on Lingmotif's TSS (text sentiment score) feature. We also developed mechanisms for dealing with the idiosyncrasies of Twitter text. Results were comparatively poor, but the experience meant a good opportunity for us to identify issues in our score calculation for short texts, a genre for which the Lingmotif tool was not originally designed.

1 Introduction

For this shared task on emotion intensity we have used the Lingmotif (Moreno-Ortiz, 2017a) sentiment analysis software. This tool is not specifically built to classify texts, although it offers this feature. It is designed more as a general text analysis tool with a focus on sentiment analysis. It offers several text metrics and displays a detailed view of the analysis results, where specific text segments are marked and annotated with their valence and other data.

For sentiment analysis, it relies on its rich lexical sources rather than on sophisticated machine learning algorithms. We undertook this shared task as an evaluation of the performance of our tool for short texts,¹ and as a good opportunity to

*This research was supported by Spain's MINECO through the funding of project Lingmotif2 (FFI2016-78141-P).

¹We use the term *short text* to refer specifically to under 140 characters, such as those used in Twitter and other social networks.

learn about the linguistic features and issues that such texts raise in a strictly lexicon-based sentiment analysis tool. It also meant a first attempt to use Lingmotif's sentiment data as features in classification and regression algorithms.

1.1 Task Description and datasets

Unlike most shared tasks on sentiment analysis, the EmoInt Shared Task at WASSA-2017 (Mohammad and Bravo-Marquez, 2017b) focused on sentiment intensity rather than classification. Several annotated Twitter datasets were provided for system training, development and testing. Tweets were classified as belonging in one of three negative emotions (*anger*, *fear*, and *sadness*) and one positive emotion (*joy*).

The training datasets were labeled for sentiment intensity. The annotation system to obtain these datasets is described in Mohammad and Bravo-Marquez (2017a). Basically, they polled the Twitter API to extract tweets that contained representative words for each of the four emotions, which they selected using Roget's Thesaurus. They collected over 7,000 tweets, differentiating between those that contained the query term in hashtag form and those that included them in non-hashtag form. Then they crowdsourced the annotation for this dataset using a Best-Worst Scaling system, whose details we will not reproduce here.

In our experience with the datasets, we believe this procedure offers very reliable results, although we have come across a number of questionable annotations and some obvious errors.²

1.2 Lexicon-based Sentiment Analysis

Within Sentiment Analysis it is common to distinguish corpus-based approaches from lexicon-

²The authors themselves (Mohammad and Bravo-Marquez, 2017a) warn about the cognitive load that is placed on the respondents during the annotation process.

based approaches. Generally speaking, lexicon-based approaches are preferred for sentence-level classification (Andreevskaia and Bergler, 2007), whereas corpus-based, statistical approaches are preferred for document-level classification. Of course, these methods can be combined (for example, Riloff et al. (2006)).

Using sentiment dictionaries has a long tradition in the field. WordNet (Fellbaum, 1998) has been a recurrent source of lexical information (Kim and Hovy, 2004; Hu and Liu, 2004; Andreevskaia and Bergler, 2006) either directly as a source of lexical information or for sentiment lexicon construction. Other common lexicons used in English sentiment analysis research include The General Inquirer (Stone and Hunt, 1963), MPQA (Wilson et al., 2005), and Bing Liu’s Opinion Lexicon (Hu and Liu, 2004). Yet other researchers have used a combination of existing lexicons or created their own (Hatzivassiloglou and McKeown, 1997; Turney, 2002). The use of lexicons has sometimes been straightforward, where the mere presence of a sentiment word determines a given polarity. However, negation and intensification can alter the valence or polarity of that word.³ Modification of sentiment in context has also been widely recognized and dealt with by some researchers (Kennedy and Inkpen, 2006; Polanyi and Zaenen, 2006; Choi and Cardie, 2008; Taboada et al., 2011).

One disadvantage on relying solely on a sentiment lexicon is that different domains may greatly alter the valence of words, a fact well recognized in the literature (Aue and Gamon, 2005; Pang and Lee, 2008; Choi et al., 2009). A number of solutions have been proposed to these, mostly using ad hoc dictionaries, sometimes created automatically from a domain-specific corpus (Tai and Kao, 2013; Lu et al., 2011).

Our approach to using a lexicon takes some ideas from the aforementioned approaches. We describe it in the next section.

2 The Lingmotif SA tool

The Tecnolengua group started work on lexicon-based sentiment analysis with the development of Sentitext, a linguistically-motivated sentiment analysis system for Spanish, and evolved within

³The use of the terms *valence* and *polarity* is used inconsistently in the literature. We use *polarity* to refer to the binary distinction positive/negative sentiment, and *valence* to a value of intensity on a scale.

the Lingmotif project to integrate English, French, Italian, and German.⁴

Lingmotif is based on the same principles as Sentitext: a reliance on wide-coverage lexical resources rather than a complex set of algorithms. It utilizes a number of lexical sources and analyzes context, by means of sentiment shifters, in order to identify sentiment-laden text segments and produce a number of scores that qualify a text from a SA perspective, as well as other various text analytics.

Analysis is produced by the identification of words and phrases that are stored in its lexicon. The overall score for a text is computed as a function of the accumulated negative, positive and neutral scores. Specific domains can be accounted for by applying user-provided dictionaries, which can be imported from CSV files, and used along with the application’s core dictionary.

Lingmotif was not designed as a sentiment classifier, but as a user-focused text analysis tool. It offers a visual representation of the *sentiment profile* of texts, which allows users to compare the profile of multiple documents side by side, and can process ordered document series. Such features are useful in discourse analysis tasks, where sentiment changes are relevant, whether within or across texts, such as political speeches and narratives, or to track the evolution in sentiment towards a given topic (in news, for example). It uses a simple, easy-to-use GUI that allows users to select input and options, and launch the analysis. Details of the GUI’s capabilities can be found in Moreno-Ortiz (2017b).

Results are generated as an HTML/Javascript document, which is saved locally to a predefined location and automatically sent to the user’s default browser for immediate display. Internally, the application generates results as an XML document containing all the relevant data; this XML document is then parsed against one of several available XSL templates, and transformed into the final HTML.

2.1 Lexical data

Lingmotif’s main asset is its comprehensive lexical sources. For each language, Lingmotif uses the following resources:

⁴The current version of Lingmotif supports English and Spanish. Version 1.2 will include initial support for French. Italian and German will be added in future versions.

- A wide-coverage core sentiment lexicon that contains both unigrams and multiword expressions, from bigrams to 6-grams.
- A set of context rules, where sentiment shifters are defined using a template approach.
- Optionally, a plugin lexicon can be used to account for domain-specific sentiment expression.

A part of speech tagger and lemmatizer are also used. Lingmotif's lexicons are still under development. For this shared task, version 1.2 was used.⁵

2.2 The Lingmotif lexicon

Lexicon entries have the structure `<form>`, `<part-of-speech>`, `<valence>`, where valence is an integer from -5 to 5, 0 being neutral. All single-word entries have a non-zero valence. Unigrams can be entered as literals or as lemmas (expressed by angled brackets), in which case they will be inflected during import and expanded into their possible forms. Examples are: `<safe>`, JJ, 2; `<fallacy>`, ALL, -3; `insolent`, ALL, -3⁶

Multi-word expressions are a big asset of Lingmotif. No other sentiment lexicon, to our knowledge, contains a significant amount of, or any at all, multi-word expressions. Avoiding MWEs has practical advantages; first, it obviously makes lexicon construction much simpler, as it does the identification process of sentiment words during analysis, thus facilitating bag-of-words approaches. However, it also ignores the fact that idiomaticity plays a huge role in the expression of sentiment. While it is true that many MWEs contain individual words of the same polarity as the overall expression, for example "turn a blind eye", "raise the alarm", "smear campaign", many do not contain any sentiment words at all ("raise the bar", "silver lining", "lose ground", "peanut gallery"), or even words with the opposite polarity ("smile at danger", "penny wise and pound foolish"). Finally, many zero-valence MWEs do contain individual

words with some valence: "vanity bag", "proper fraction", "fancy dress".

This is the reason why MWEs in Lingmotif can have a 0 valence; the aim is to block detection of individual words which are part of a MWE and whose valence may or may not be the same as that in the MWE. Other zero-valence MWEs are included because they are valence shifters used in the CVS system, mostly intensifiers such as "kind of", "a fair bit of", "through and through".

In version 1.2 multiword expressions can also contain variables that act as placeholders for any word, such as `<fall>_into_2_hands`, which will match any sequence of any form of the lemma "fall" followed by "into", then 0 to 2 words (e.g., "the" "his", "the wrong"), then "hands". This allows flexible representation and identification of variable MWEs and collocations.

Version 1.2 of the English Lingmotif lexicon contains 13,250 unigram lemmas (which expand to 21,300 forms), 12,300 MWE lemmas (which expand to 37,700 forms), and 720 context rules (sentiment shifters).

As for its origin, the Lingmotif lexicon was initially compiled from a lexicographic perspective, aiming at comprehensiveness. The core single-word lexicon was jumpstarted using existing sentiment lexicons, namely, the Harvard General Inquirer (Stone and Hunt, 1963), MPQA (Wilson et al., 2005), and Bing Liu's Opinion Lexicon (Hu and Liu, 2004). These resources were expanded by using a thesaurus and derivational generation rules. The lexicon has been subsequently refined manually using corpus analysis techniques as well by qualitative techniques.

2.3 Sentiment shifters

A sentiment word or expression can change its valence in context. It can be intensified or down-toned, by means of quantifiers, for example, or its valence may be inverted altogether (negation being the most obvious case), thus altering the polarity.

Lingmotif implements a contextual valence shifter (CVS) system based on the matching of a number of context rules that define how a sentiment item changes its polarity in context. Such approach has been used by Polanyi and Zaenen (2006), Kennedy and Inkpen (2006), and Taboada et al. (2011), among others. In our implementation, we use simple addition or subtraction of in-

⁵At the time of editing this document version 1.0 can be downloaded from the Tecnolengua website (<http://tecnolengua.uma.es/lingmotif>). Version 1.2 will be made available during 2017.

⁶The "ALL" notation simplifies acquisition and avoids matching problems derived from bad part-of-speech tagging at run-time.

Inversion
NN, -, avoid*, LR, 5, INVO
JJ, +-, not, L, 2, INVO
Intensification
NN, -, avoid*, LR, 5, INVO
JJ, +-, not, L, 2, INVO
Downtoning
NN, -, avoid*, LR, 5, INVO
JJ, +-, not, L, 2, INVO

Table 1: Sentiment shifters

tegers to modify the original valence, as specified by a set of patterns in which certain features are matched, namely, the part of speech and polarity of the sentiment word, the form, location (left or right), and span (in number of words) of the shifter, and the result of the rule application. Version 1.2 contains over 700 such rules for English. These are some examples:

When a context rule is matched, the resulting text segment is marked as a single unit and assigned the calculated valence, as specified by the rule. New in version 1.2 is multiple rule matching, where results are aggregated. Thus the sequences "really interesting" and "really really interesting" produce different results. This is an experimental feature that we have yet to improve, as it can produce some unexpected results.

2.4 Lingmotif Learn

For this task we created a new tool, still under development, tentatively called "Lingmotif Learn". This is a GUI-enabled convenience tool that manages datasets and uses the Python-based scikit-learn (Pedregosa et al., 2011) machine learning toolkit. This tool facilitates loading and preprocessing of datasets, getting the text run through the Lingmotif SA engine, and feeding the resulting data into one of several machine learning algorithms.

It makes it easy to compare the performance of different combinations of the available Lingmotif data as features and classification/regression algorithms. After the optimal features and algorithm have been selected, the model is trained and saved; then it can be loaded to classify the development and test datasets.

Table 2 lists the features available for each text after the Lingmotif analysis.

As we will discuss in section 4 below, for this shared task we used only TSS as a predictor. TSS attempts to summarize the overall sentiment of a

ID	Name	Description
1	tss	Text Sentiment Score
2	tsi	Text Sentiment Intensity
3	lex_items	Number of lexical Items
4	pos_score	Positive score
5	neg_score	Negative score
6	pos_items	Number of positive items
7	neg_items	Number of negative items
8	crules	Number of sentiment shifters
9	V0	Number of items with valence 0
10	V-1	Number of items with valence -1
11	V-2	Number of items with valence -2
12	V-3	Number of items with valence -3
13	V-4	Number of items with valence -4
14	V-5	Number of items with valence -5
15	V1	Number of items with valence 1
16	V2	Number of items with valence 2
17	V3	Number of items with valence 3
18	V4	Number of items with valence 4
19	V5	Number of items with valence 5
20	x_marks	Number of exclamation marks
21	q_marks	Number of question marks
22	pmarks	Number of punctuation marks
23	handles	Number of user handles
24	hashtags	Number of hashtags
25	urls	Number of exclamation URLs

Table 2: Set of available features

text on a 0-100 scale. It is arrived at by calculating a *sentiment weight*, which is dependent on text length, and is encapsulated in the TSI feature, which, in turn, is calculated by combining the *pos_score*, *neg_score*, and *lex_items* features. A more detailed description of these scores can be found in Moreno-Ortiz (2017a).

3 Dealing with social media text

It is only recently that we have begun experimenting with social media content analysis. Our focus so far has been on longer texts (user reviews, political debates and speeches, narratives). We undertook this task as a challenge that would give us a first glimpse of the potentiality of our system to analyze tweets and other social media short texts, which certainly show certain specific characteristics, such as the intensive use of emoticons and emojis, hashtags, repetitions, etc. As a first approach to this type of texts, we adapted our system as described below.

3.1 Emoticons and emojis

Emoticons are a well known source of emotion expression, and very common in social media in general and Twitter in particular. Even though the relationship between emoticons and the sentiment conveyed in the overall message is not always unambiguous (Wang and Castanon, 2015),

they clearly play an important role in the expression of sentiment, and, relevant to this task, they have been found to have a strong impact in the intensity of the emotions expressed in the message. Accordingly, they have recurrently been used as features for machine learning classifiers in sentiment analysis tasks, even from the first efforts to classify Twitter data, e.g., [Go et al. \(2009\)](#).

Further, the generalization of emoji keyboards in mobile devices in the recent years has no doubt contributed to the proliferation of emojis. If (text) emoticons display certain ambiguity, the sentiment conveyed by emojis is obviously more sophisticated, as is its relation to the text.

This shared task gave us an opportunity to improve on the management of emoticons and emojis we have used so far in Lingmotif. In the current version (1.0), emoticons are dealt with during preprocessing and are converted to a placeholder lexical item with a certain polarity. Emojis are simply ignored.

For this task we implemented support for emojis by including them in the lexicon just like any other sentiment word. Currently, the list of emojis is limited to 126 positive and negative items, which were selected as these and other English and Spanish Twitter datasets. All these emojis are more or less consistent in their usage in terms of their polarity. Emojis denoting surprise, and others which exhibit a high degree of variability in their denoted polarity were not included. At this stage, all emojis in our lexicon have the same level of intensity, i.e., 3/-3 (medium). This is of course far from ideal, and our intention is to provide better intensity ratings, for which we intend to use [Novak et al. \(2015\)](#)'s results, which provide reliable polarity and intensity data for 970 emojis in 13 European Languages.

3.2 Treatment of hashtags

Hashtags have been shown to be excellent cues of the sentiment conveyed in tweets ([Mohammad, 2012](#); [Mohammad and Kiritchenko, 2015](#)). Making sense of hashtags is not an easy task, however, since users can be extremely creative in their use. Efforts have been made to process and normalize their content, some of them quite sophisticated ([Declerck and Lendvai, 2015](#)).

As a first approach, we introduced in Lingmotif a simple system to process hashtags. Our strategy consisted of trying to match substrings in the hash-

tag against our single-word lexicon, either as the whole string (minus the hash symbol) or in Camel-Case. Simple as it is, this system turned out to be able to decode the content of a significant proportion of hashtags,

4 Analysis and results

We approached the task by running a Lingmotif analysis of each emotion dataset as a single document. Since training datasets were provided already sorted by emotion intensity, this was straightforward and could give us a rough idea of the performance. We used Lingmotif's "Sentiment Profile" feature to quickly check if there was a viable correlation. The Sentiment Profile is a line graph whose data points are obtained by breaking the input text into segments of varying lengths (dependent on the text's overall length), and computing the valence for each segment by averaging the valences of the lexical words and phrases (after the sentiment shifters system discussed above has been applied) contained in the segment. [Figure 1](#) shows the sentiment profile obtained for the anger training dataset.

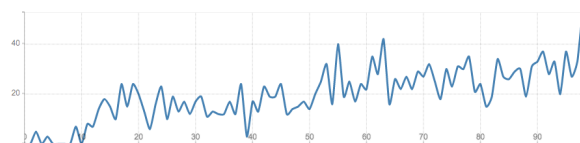


Figure 1: Sentiment Profile for the "anger" training dataset

Higher scores in this graph indicate more positive sentiment. Tweets in the dataset were sorted in decreasing order of intensity, so, as we are using a negative emotion, a higher TSS indicates a lower intensity, and therefore a correlation between TSS and the scores in the dataset. This gave us the impression that average to good performance could be achieved simply by using the TSS data of each individual tweet. Our approach to building the statistical model then consisted of using a simple linear regression (best fit with least squares), using Lingmotif's Text Sentiment Score (TSS) as the independent variable.

For the analysis, we decided to include the emotion word as part of the text to be analyzed. This would ensure that at least one word of the same polarity was included in every tweet. This turned out not to be a good solution, as we will discuss in [section 5](#) below.

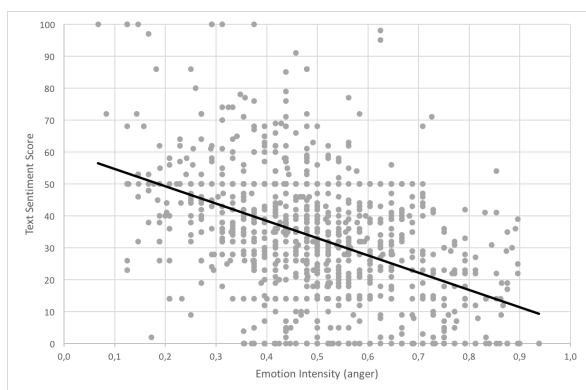


Figure 2: Lingmotif TSS vs. intensity in the anger training dataset

Dataset	Pearson	Spearman
Anger	0.324	0.324
Fear	0.466	0.454
Sadness	0.436	0.449
Joy	0.408	0.393
Average	0.409	0.405

Table 3: Official results.

This first impression, however, turned out no to be too accurate. Figure 2 shows the scatter plot of the anger training dataset in terms of intensity vs Lingmotif’s TSS. As the figure shows, it is a relatively poor predictor. The final results obtained are detailed in Table 3.

5 Discussion

We believe these comparatively poor results were due to the fact that Lingmotif’s TSS is not well suited to extremely short texts. Even though identification of sentiment words (or hashtags) and expressions is fairly good, thanks to the wide coverage of the Lingmotif Lexicon and sentiment shifters, the intensity reflected by TSS does not seem to finely reflect the intensity as perceived by human annotators of tweets.

As expected, there were also a number of analysis errors, many of them related to the nature of social media text. An analysis of the annotated text, which Lingmotif produces, allowed us to discover certain recurrent problems:

- Unaccounted/bad shifters: ”**zero** tolerance for honesty her alliance”
- Overreaching of shifters: ””Why are people that do [**not have iPhones so bitter**] about iPhones????”

- Bad spelling and/or grammar: ”These guys **dcan not get nothing right**”
- Irony and sarcasm: ”**thanks** or saying My wife and I were getting our iPhones today and then losing both of them with no eta **thanks**”
- Complex wording: ”You will never find someone who loved you like I did. And that my love, will be my revenge.”

Obviously, some of these issues are harder to fix than others. Irony and sarcasm are possibly the hardest cases to deal with automatically, and are very common in social media short texts.⁷ Others, however, are of a more practical nature and easier to tackle.

Since the EmoInt organizers allowed participants in the shared task to keep uploading results after the competition was over, we took this opportunity to tackle some of these issues. We started by removing the emotion tag from the tweets, which, in retrospect, we consider a bad decision. We then reduced the range of far-reaching sentiment shifters to avoid overreaching and adapt to the simpler syntactic structures found in tweets.

Another recurrent issue we found is repetitions of emojis. As explained in section ?? above, Lingmotif’s current TSS uses text length, in terms of number of lexical items to determine intensity. In ”regular” texts, for the same text length, the number of lexical items falls within consistent ranges. However, repetition of emojis as an intensification of emotion is very common in social media text, and, when emojis are treated as lexical items, as we have experimented here for the first time, we obtain some cases where the number of lexical items exceeds by far the average frequency in texts of that length. The result is that the tweet is treated by Lingmotif as a longer text, thus calculating the wrong intensity. We avoided this problem by controlling character repetition during preprocessing, and limiting it to three consecutive same emojis.

Even after this, we realized that our current thresholds for binning texts in terms of their length was too fine, and resulted in tweets falling in one one of three categories according to their length. We fixed this by defining fewer (broader) categories in the lower end of the range, thus making sure that all tweets fall within the same categories in terms of text length. The new text length

⁷Wallace et al. (2014) report that 10-15 percent of messages on reddit.com exhibit some form of irony or sarcasm.

Dataset	Pearson	Spearman
Anger	0.423	0.425
Fear	0.524	0.524
Sadness	0.490	0.470
Joy	0.492	0.490
Average	0.482	0.477

Table 4: Results after adjustments

threshold (25 lexical items) is based on the maximum number of lexical items found on the EmoInt datasets.⁸

After applying the above-mentioned adjustments and fixes, we ran the system again to measure their effect, if any. Results were significantly improved, as is reflected in Table 4.

It would have been interesting to experiment with multiple regression using other sentiment features provided by our system, something we were unable to do for this task due to time limitations. In particular, we feel that using the raw *pos_score* and *neg_score* features would have produced better results. Another possibility would be to use *pos_items* and *neg_items*. The difference being that the valence values assigned in the lexicon are ignored, and only polarity is taken into account.

6 Conclusions

This work has been extremely useful to us. We now have a clearer picture of what it means to deal with social media short texts, and the difficulties they pose. This task gave us the chance to adapt our analysis system in a number of ways, at least in terms of form (emojis, character repetitions, etc.).

From a linguistic perspective, we have also found clear evidence that dealing with short texts of the type commonly found in social media call for specific adaptations of our system than the merely superficial ones we have described in this paper. Not only are there a number of formal differences, but the message itself is expressed in extremely condensed ways.

Our most relevant conclusion is that Lingmotif’s present sentiment score may not be a good predictor because it does not encapsulate the features it is based on optimally, and we think better results would be achieved by combining such features (*pos_score*, *neg_score*, *lex_items*, and others) using more sophisticated statistical learning meth-

⁸After rounding it up. The actual highest number of lexical items found in the EmoInt datasets was 22. The average number of lexical items per tweet was 9.18.

ods, a path that we will explore in future developments.

References

- Alina Adreevskaia and Sabine Bergler. 2006. Mining wordnet for fuzzy sentiment: Sentiment tag extraction from wordnet glosses. In *11th Conference of the European Chapter of the Association for Computational Linguistics*. pages 209–216.
- Alina Andreevskaia 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*. Association for Computational Linguistics, Stroudsburg, PA, USA, SemEval ’07, pages 117–120.
- Anthony Aue and Michael Gamon. 2005. Customizing sentiment classifiers to new domains: A case study. Borovets, Bulgaria.
- Yejin Choi and Claire Cardie. 2008. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA, EMNLP ’08, page 793.
- Yoonjung Choi, Youngho Kim, and Sung-Hyon Myaeng. 2009. Domain-specific sentiment analysis using contextual feature generation. In *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. ACM, Hong Kong, China, pages 37–44.
- Thierry Declerck and Piroska Lendvai. 2015. Processing and normalizing hashtags. In Galia Angelova, Kalina Bontcheva, and Ruslan Mitko, editors, *Proceedings of RANLP 2015*. INCOMA Ltd, pages 104–110.
- Christiane Fellbaum, editor. 1998. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA; London.
- Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *Processing* pages 1–6.
- Vasileios Hatzivassiloglou 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*. Association for Computational Linguistics, Madrid, Spain, pages 174–181.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, Seattle, WA, USA, pages 168–177.

- Alistair Kennedy and Diana Inkpen. 2006. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence* 22(2):110–125.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*. Association for Computational Linguistics, Geneva, Switzerland, page 1367.
- Yue Lu, Malu Castellanos, Umeshwar Dayal, and ChengXiang Zhai. 2011. Automatic construction of a context-aware sentiment lexicon: An optimization approach. In *Proceedings of the 20th International Conference on World Wide Web*. ACM, New York, NY, USA, WWW '11, pages 347–356.
- Saif Mohammad. 2012. #emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics, Montreal, Canada, pages 246–255.
- Saif Mohammad and Felipe Bravo-Marquez. 2017a. Emotion intensities in tweets. In *Proceedings of the sixth joint conference on lexical and computational semantics (*Sem)*. Vancouver, Canada.
- Saif Mohammad and Felipe Bravo-Marquez. 2017b. Wassa-2017 shared task on emotion intensity. In *Proceedings of the EMNLP 2017 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media*. Copenhagen, Denmark.
- Saif M. Mohammad and Svetlana Kiritchenko. 2015. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence* 31(2):301–326.
- Antonio Moreno-Ortiz. 2017a. Lingmotif: A user-focused sentiment analysis tool. *Procesamiento del Lenguaje Natural* 58(0):133–140.
- Antonio Moreno-Ortiz. 2017b. Lingmotif: Sentiment analysis for the digital humanities. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Valencia, Spain, pages 73–76.
- Petra Kralj Novak, Jasmina Smailovi, Borut Sluban, and Igor Mozeti. 2015. Sentiment of emojis. *PLOS ONE* 10(12):e0144296.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(12):1–135.
- Fabian Pedregosa, Gal Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and douard Duchesnay. 2011. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12:2825–2830.
- Livia Polanyi and Annie Zaenen. 2006. Contextual valence shifters. In *Computing Attitude and Affect in Text: Theory and Applications*, Springer, Dordrecht, The Netherlands, volume 20 of *The Information Retrieval Series*, pages 1–10. Shanahan, James G., Quyan, Wiebe, Janyce edition.
- Ellen Riloff, Siddharth Patwardhan, and Janyce Wiebe. 2006. Feature subsumption for opinion analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Stroudsburg, PA, USA, EMNLP '06, pages 440–448.
- Philip J. Stone and Earl B. Hunt. 1963. A computer approach to content analysis: Studies using the general inquirer system. In *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference*. ACM, New York, NY, USA, AFIPS '63 (Spring), pages 241–256.
- Maite Taboada, Julian Brooks, M. Tofiloski, K. Voll, and M. Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics* 37(2):267–307.
- Yen-Jen Tai and Hung-Yu Kao. 2013. Automatic domain-specific sentiment lexicon generation with label propagation. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services*. ACM, New York, NY, USA, IIWAS '13, pages 53:53–53:62.
- Peter D. Turney. 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)*. Philadelphia, USA., pages 417–424.
- Byron C. Wallace, Do Kook Choe, Laura Kertz, and Eugene Charniak. 2014. *Humans require context to infer ironic intent (so computers probably do, too)*, Association for Computational Linguistics (ACL), volume 2, pages 512–516.
- Hao Wang and Jorge A. Castanon. 2015. Sentiment expression via emoticons on social media. In *2015 IEEE International Conference on Big Data (Big Data)*. IEEE Computer Society, Washington, DC, USA, BIG DATA '15, pages 2404–2408.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Stroudsburg, PA, USA, HLT '05, pages 347–354.