Procesamiento del Lenguaje Natural, Revista nº 58, marzo de 2017, pp. 29-36 recibido 11-11-2016 revisado 12-01-2017 aceptado 06-02-2017

On Evaluating the Contribution of Text Normalisation Techniques to Sentiment Analysis on Informal Web 2.0 Texts*

Evaluación de la Contribución de la Normalización al Análisis de Sentimiento en Textos Informales de la Web 2.0

Alejandro Mosquera
University of Alicante
Alicante, Spain
amosquera@dlsi.ua.es

Yoan Gutiérrez
University of Matanzas
Matanzas, Cuba
yoan.gutierrez@umcc.cu

Paloma Moreda
University of Alicante
Alicante, Spain
moreda@dlsi.ua.es

Abstract: The writing style used in social media usually contains informal elements that can lower the performance of Natural Language Processing applications. For this reason, text normalisation techniques have drawn a lot of attention recently when dealing with informal content. However, not all the texts present the same level of informality and may not require additional pre-processing steps. Therefore, in this paper we explore the results of applying lexical normalisation applied to a sentiment analysis classification task on Web 2.0 texts, shows more than a 2.6 % improvement over average F1 for the most informal data.
Keywords: Informality, normalisation, sentiment analysis, opinion mining

Resumen: El tipo de lenguage empleado en las redes sociales suele incluir elementos informales que pueden afectar el rendimiento de las herramientas de procesamiento del lenguaje natural. El uso de técnicas de normalización léxica es una de las opciones que se han estado usando a la hora de tratar contenidos de la Web 2.0. Sin embargo, no todos los textos requieren dicho pre-procesamiento ya que pueden exhibir diferentes niveles de informalidad. En este trabajo exploramos el impacto de aplicar normalización léxica evaluando los resultados de un sistema de análisis del sentimiento antes y después de la normalización. Los resultados de nuestra investigación muestran una mejora de mas del 2.6 % sobre el F1 para los textos mas informales.
Palabras clave: Informalidad, normalización, minería de opiniones

1 Introduction

Nowadays, Web 2.0 applications are some of the most popular forms of communication between Internet users. Blogs, social networks or short text messaging platforms have become a very important participation channel where users publish their comments and opinions. This valuable source of information contains insights about user opinions and sentiments regarding almost any topic. These can determine the reputation of public companies or figures, mine opinion patterns and measure the popularity of news and events.

Sentiment analysis (SA) is the sub-field of Natural Language Processing (NLP) that extracts and identifies subjective information. A basic task in SA deals with measuring the overall polarity orientation of a document about some topic. When SA is applied to social media comments it can be used to increase the effectiveness of marketing campaigns, discover new market threats and opportunities or react faster to customer issues.

However, the language used in social media websites and applications can contain a

* This work has been partially funded by the European Commission under the 7th Framework Programme for Research and Technological Development through the SAM (FP7-611312) project, by the Spanish Government through the ATTOS (TIN2012-38536-C03-03) and LEGOLANGUAGE (TIN2012-31224) projects, and by the University of Alicante through the project "Explotación y tratamiento de la información disponible en Internet para la anotación y generación de textos adaptados al usuario" (GRE13-15) and "Tratamiento inteligente de la información para la ayuda a la toma de decisiones" (GRE12-44). We also thank our anonymous reviewers for their insightful comments and valuable feedback

ISSN 1135-5948

variable amount of informal elements such as lexical variants, slang or non-standard punctuation (Thurlow, 2003) that can make any NLP task challenging. For this reason, these texts can benefit from a pre-processing step that understands these informal features and replaces them by their formal equivalent (Wang and Ng, 2013).

The use of lexical normalisation to enhance NLP processing is not a new topic and it has been the subject of recurrent research applied to short and noisy texts such as SMS (Aw et al., 2006). The similarities shared by SMS and more recent genres such as microblogs (Han and Baldwin, 2011) have helped to develop similar approaches. Moreover, not all Web 2.0 genres have the same level of informality. For example, micro-blog posts have a character limit that favors contractions and ellipsis while blog entries or product reviews are usually larger and more elaborated (Santini, 2006).

Because of these genre differences not all the Web 2.0 texts would experience the same benefits after a normalisation step. For this reason, in this paper we analyse the effects of replacing informal lexical variants with their canonical version on social media texts. This has been applied to a SA classification task with aim to enhance polarity detection results. In order to do this, we have carried out several polarity classification experiments using English texts with different degrees of informality and evaluated the impact of normalisation in the results. We have also explored the use of informality analysis as a way of measuring the need of pre-processing on each case.

This article is organised as follows: In Section 2 the state of the art is reviewed. Section 3 describes the informality analysis process. The SA systems used in the experiments are explained in Section 4. In Section 5, the text normalisation step is introduced. The corpora used for all the experiments are detailed in Section 6. Section 7 contains the obtained results and their analysis. Finally, our main conclusions and future work are drawn in Section 8.

2 Related Work

Both industry and academic researchers have increased their interests on measuring user sentiments from social media. After the initial works of Pang, Lee and Vaithyanathan (2002) several applications of opinion mining have been developed focused on microblogs (Barbosa and Feng, 2010; Bifet and Frank, 2010) using both machine learning (Turney, 2002) and lexicon-based approaches (Taboada et al., 2011). The real-time nature of tweets provides a large amount of metadata that can be used as a training corpus for opinion mining systems (Pak and Paroubek, 2010) without requiring annotated corpora (Wiebe, Wilson, and Cardie, 2005).

Whilst normalisation is a common pre-processing step in several areas of NLP (Sproat et al., 2001; Adda et al., 1997) the rise of social media has expanded the concept and meaning of this process. Lexical normalisation techniques (Liu et al., 2011; Han, Cook, and Baldwin, 2013) based on the substitution of out of vocabulary (OOV) words have been used in opinion mining systems before (Mukherjee et al., 2012; Gutiérrez et al., 2013; Sidorov et al., 2013) but this process is usually presented as an intermediate transformation step without explicitly detailing the contribution of normalisation to the classification results. In a more recent analysis of the improvements of using text normalisation applied to SA tasks (Mosquera and Moreda, 2013) it has been shown that normalisation can have positive effects on informal genres. On the other hand, there are different genres within the Web 2.0 and they do not have the same level of informality (Mosquera and Moreda, 2012c), so the enhancements obtained after normalisation can be more or less relevant depending on that level.

Regarding the analysis of the formality/informality of documents most of the prior research tried to measure text formality using readability indexes, and the concept of lexical density (Fang and Cao, 2009). There were attempts to create a formality score by using the frequency of deictic words, that are expected to increase with the informality of a text and, conversely, the frequency of non-deictic words should increase with text formality (Heylighen and Dewaele, 1999). While this score can be used to detect deep formality this approach cannot quantify stylistic or grammatical deviations. Regarding approaches measuring informality, the work of Mosquera and Moreda (2012a) uses multi-dimensional analysis in order to determine the informality level of Web 2.0 texts. This method not only shows information about what texts are more informal but it also allows the comparison of texts from other corpora or genres by using a set of dimensions (Mosquera and Moreda, 2012c).

For this reason, in this paper we study the cases where Web 2.0 texts benefit from using normalisation techniques. We apply this analysis to a common NLP task such as SA, and evaluate when this pre-processing step is necessary and can really enhance the classification results. In order to to this, informality analysis is used to score and rank the SA corpora by their informality level before and after the normalisation step.

3 Informality Analysis

Because Web 2.0 texts have specific informal features not usually present in more formal genres we have applied informality analysis using the SMILE (Mosquera and Moreda, 2012b) tool, a framework for classifying texts by their informality level based on four dimensions: Complexity, Emotiveness, Expressiveness and Incorrectness. These dimensions are based on aggregated text features such as the presence of slang and offensive words, incorrect capitalisations and punctuation marks, frequency of character repetitions, readability measures, frequency of emoticons or the frequency of SMS-style contractions.

4 Sentiment Analysis

In order to carry out the SA experiments we have used a 3-class (positive, negative and neutral) unsupervised SA classification system based on WordNet (WN)(Fellbaum, 1998) and additional resources. In order to enrich the WN resource, it has been linked with different lexical resources such as WordNet Domains (WND) (Magnini and Cavaglia, 2000) a lexical resource containing the domains of the synsets in WordNet, SUMO (Niles and Pease, 2003) an ontology relating the concepts in WordNet, WordNet Affect (WNA) an extension of WN where different synsets are annotated with one of the six basic emotions proposed by Ekman (1999), SentiWordNet (Esuli and Sebastiani, 2006) a lexical resource where each synset is annotated with polarity, Semantic Classes (SC)(Izquierdo-Bevià, Suàrez, and Rigau, 2007) a set of Base Level Concepts (BLC) based on WN. RST (Gutiérrez et al., 2010) is a method able to disambiguate the senses of the words contained in a sentence by obtaining the Relevant Semantic Trees from different resources. For SA, RST makes use of the polarity information from SentiWordNet that is contained in ISR-WN (Gutiérrez et al., 2010). In order to measure the association between concepts in each sentence according to a multidimensional perspective, RST uses the Association Ratio (AR) measure (Vázquez, Montoyo, and Kozareva, 2007). The purpose is to include the Multidimensional Semantic Analysis into the Opinion Analysis using RSTs (Gutiérrez, Vázquez, and Montoyo, 2011) with WNDs and SCs.

5 Text Normalisation

We have used TENOR (Mosquera, Lloret, and Moreda, 2012), a multilingual lexical normalisation tool for English and Spanish texts in order to transform noisy and informal words into their canonical form. After this step they can be easily processed by NLP tools and applications.

In order to do this, OOV words are detected with a dictionary lookup. TENOR uses a custom-made lexicon built over the expanded Aspell dictionary and then augmented with domain-specific knowledge from the Spell Checking Oriented Word Lists (SCOWL)[1] package.

The OOV words are matched against a phone lattice using the double metaphone algorithm (Philips, 2000) to obtain a list of substitution candidates. With the Gestalt pattern matching algorithm (Ratcliff and Metzener, 1988) a string similarity score is calculated between the OOV word and its candidate list.

Nevertheless, there are acronyms and abbreviated forms that can not be detected properly with phonetic indexing techniques *(lol - laugh out loud)*. For this reason, TENOR uses an exception dictionary with common Internet abbreviations and slang collected from online sources[2].

Moreover, a number transliteration lookup table and several heuristics such as word-lengthening compression, emoticon translation and simple case restoration are applied to improve the normalisation results. Finally, TENOR uses a trigram language model in order to enhance the clean candidate selection.

[1] http://wordlist.sourceforge.net/

[2] http://en.wiktionary.org

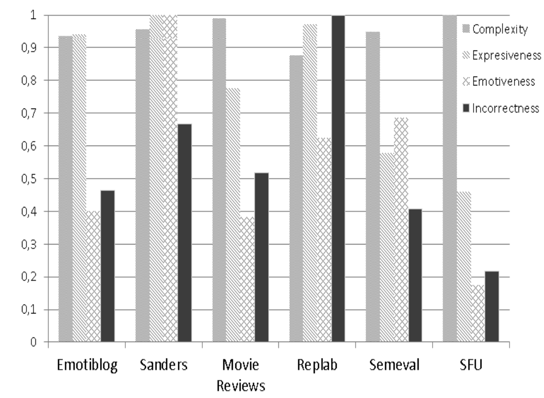

Figure 1: Distribution of informality dimensions in the corpora

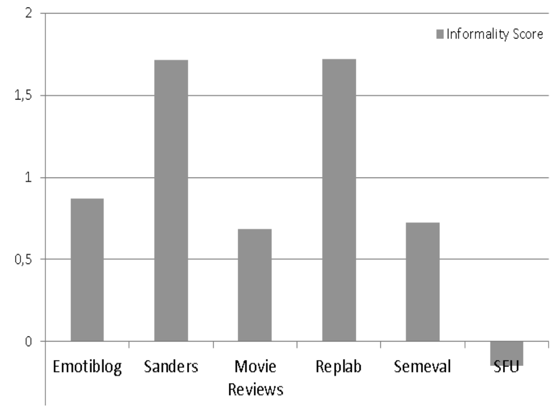

Figure 2: Normalised informality scores

6 Datasets

The polarity classification system has been evaluated using annotated English and Spanish texts from different Web 2.0 genres:

Microblog publications (Sanders): 5513 Twitter messages in English [3].

Blog posts: The Kyoto sub-set of the EmotiBlog corpus [4] corpus comprising 1173 English texts.

Movie reviews: Polarity dataset from movie reviews (Pang and Lee, 2005) containing 10662 sentences.

Microblog publications (RepLab 2013): 70139 polarity-annotated English and Spanish tweets from the RepLab2013 testing dataset (Amigó et al., 2013).

Microblog publications (Semeval 2013): 6434 polarity-annotated English tweets from the Semeval 2013 sentiment analysis training dataset (Nakov et al., 2013).

Online reviews (SFU): The SFU corpus (Taboada and Grieve, 2004) contains 400 online reviews in English for several product categories.

We have obtained the distribution of informality dimensions (see Figure 1) and normalised informality scores for each corpus (see Figure 2) by using informality analysis. These results can be aggregated in three main groups: Very informal (Sanders and RepLab tweets), Informal (Semeval tweets and movie reviews) and Formal (SFU reviews).

[3]http://www.sananalytics.com/lab/twitter-sentiment/sanders-twitter-0.2.zip

[4]http://gplsi.dlsi.ua.es/gplsi11/allresourcespanel

7 Results

A ten-fold cross-validation evaluation of SA classification has been conducted on the previously described linguistic resources before and after the normalisation step using TENOR. The results on Table 1 show how Sanders and Emotiblog texts obtained more than a 4 % and 3.5 % F1 improvement respectively on polarity classification by using the WN-Domain approach. All the F1 scores obtained during the experiments have been checked for statistical significance at 0.95 % confidence level. The aforementioned cases where normalisation contributes the most to SA have a high classification confidence.

Regarding the Semantic Class method, F1 results are in overall lower and they seem to improve after normalisation where texts are very informal (e.g. Sanders, Replab) only. Enhancements in F1 after using TENOR are slightly higher with the WN-domain approach, especially on the corpora with medium informality level. Moreover, if we take into account the average values by informality level (see Table 2) we can appreciate that these differences are just 1.5 percentual points higher for Semantic Class when analysing the most informal texts, obtaining in overall similar results.

To reduce the dependency of the results on the two unsupervised SA approaches, we have also repeated the same experiments using a supervised SA system (Mosquera and Moreda, 2013) with a subset of the original corpora (Sanders, Emotiblog and Movie reviews). Interestingly, the improvements of normalisation on Sanders are considerably higher (6,45 %) while the results on less informal corpora such as Emotiblog show a substantial decrease on F1 (-5,27 %). This supervised SA model performs better with medium-informality corpora and needs a normalisation step in order to obtain comparable scores for the most informal texts.

After the experimental results, we can conclude that text normalisation consistently helps to improve SA classification systems on the most informal Web 2.0 genres and can be useful on some of the less informal texts.

7.1 Discussion

During the evaluation we found several cases of sentences that were correctly labeled without any pre-processing by the SA systems but generated FPs/FNs after being processed by TENOR. We have manually classified these into 4 main categories:

Boosted/Reduced polarity: Normalisation can reduce the polarity of sentences by removing character repetitions. Some interjections can have a higher score when they include repeated characters *Ohhhhh* (disappointment) vs *Oh* (surprise). e.g. Before normalisation: Ohh I think he did and there's so much more which won't be covered now (Negative) And after normalisation: Oh I think he did and there's so much more which won't be covered now (Neutral)

OOV words are usually ignored by dictionary-based SA systems but after normalisation these are now processed. This is usually the desired effect and improves the results. *Hasta luego :) (See you later)* vs *Hasta luego estoy feliz. (See you later I am happy)* It goes from a detected neutral polarity to a high-positive one after the normalisation of the emoticon with the textual equivalent. But in the case of very short sentences it can cause FPs by boosting the polarity when there is not enough context: *Que frio (so cold)* vs *Que frío.* (Neutral to Negative polarity change after normalisation).

Different language: As we have not performed any language detection during the process, the presence on OOV words in another language will result in a poor quality normalisation that will affect negatively the sentiment detection: e.g. *es una version frida superficial , preciosista y sin ningun contenido (a cold and superficial version, without any content)* vs *eyes una viewers and due frida superficial, preciosista why sin knowing and continued.* (Negative polarity before normalisation but Neutral after).

Entity removal: One of the preprocessing steps of TENOR is the removal of all the OOV entities such as URLs, Twitter hashtags and mentions. This eases the processing for NLP tools but sometimes these tags contain semantic information that can be relevant for SA, e.g.: *Office 2014 #Mac #sucks #hate* (Negative) vs *Office 2014* (Neutral after normalisation) *#NWTrue Blood can't wait til the new one comes on tomorrow* (Positive) vs *Blood can't wait tile the new one comes on tomorrow.* (Negative after normalisation)

Incorrect normalisation: We have not found many cases where incorrect normalisation caused a FP or a FN but these may happen when some of the words are incorrectly treated as OOV if they are not present in the normalisation dictionary. One of the limitations of TENOR is the absence of superlatives and diminutives for Spanish as these are not treated as IV when ending with the suffixes -ato or -ito. e.g. *Toma ya! Esto si que es un piropazo!!!!!* (Neutral) vs *Toma ya! Esto si que es un propicio!* (Positive after normalisation)

A more granular analysis of these effects by informality level can be observed in Table 3.

Level	B/R	Lang.	Entity	Badnorm
Informal	7.50	13,16	55,05	24,29
Medium	2.54	7,35	37,13	52.98
Formal	0.00	5,30	1.90	92,80

Table 3: Percentage of misclassification after normalisation by informality type: Boosted/Reduced polarity(B/R), Different language(Lang.), Entity removal(Entity) and Incorrect normalisation(Badnorm)

8 Conclusion

In this paper we have evaluated the use of lexical normalisation techniques with aim to enhance SA classification by replacing informal lexical variants with their canonical version. Our experiments with Web 2.0 English texts consistently show higher average F1 over the original data. However, after using informality analysis we have discovered that these improvements are higher for the most informal texts (2.6 % avg. F1), while in the less informal and formal corpora normalisation only

System	Corpus	Precision	Recall	F1	Diff.	F1 Confidence (+-)
Mosquera-Moreda2013	Sanders	81,20	80,00	80,60		2.020
Mosquera-Moreda2013	Sanders Norm	**85,80**	**85,80**	**85,80**	6,45 %	2.860
Mosquera-Moreda2013	Emotiblog	84,70	86,10	85,40		2.853
Mosquera-Moreda2013	Emotiblog Norm	79,00	82,89	80,90	-5,27 %	2.249
Mosquera-Moreda2013	Reviews	66,70	66,70	66,70		2.861
Mosquera-Moreda2013	Reviews Norm	**67,10**	**67,10**	**67,01**	0,01 %	2.854
WN-Domain	Sanders	59,49	41,95	49,20		0.354
WN-Domain	Sanders Norm	**61,10**	**44,04**	**51,19**	4,04 %	0.352
WN-Domain	Emotiblog	53,46	44,92	48,82		0.354
WN-Domain	Emotiblog Norm	**53,80**	**47,67**	**50,55**	3,54 %	0.353
WN-Domain	Replab	35,42	35,63	35,52		0.894
WN-Domain	Replab Norm	35,34	**36,38**	**35,85**	0,93 %	0.949
WN-Domain	Semeval	37,79	37,89	37,84		0.948
WN-Domain	Semeval Norm	36,55	36,44	36,49	-3,57 %	0.892
WN-Domain	SFU	53,83	49,98	51,80		0.949
WN-Domain	SFU Norm	**54,37**	**50,22**	**52,20**	0,77 %	0.948
WN-Domain	Reviews	55,90	45,10	49,90		1.043
WN-Domain	Reviews Norm	55,70	45,00	49,80	-0,20 %	1.319
WN-Semantic Class	Replab	34,66	34,97	34,80		1.317
WN-Semantic Class	Replab Norm	**34,80**	**35,91**	**35,35**	1,58 %	0.921
WN-Semantic Class	Semeval	36,11	36,55	36,32		1.319
WN-Semantic Class	Semeval Norm	**36,12**	36,55	36,32	0,00 %	1.313
WN-Semantic Class	SFU	52,26	50,48	51,40		1.185
WN-Semantic Class	SFU Norm	**52,60**	**51,01**	**51,80**	0,78 %	1.175
WN-Semantic Class	Sanders	57,41	49,20	52,99		1.176
WN-Semantic Class	Sanders Norm	**58,07**	**51,85**	**54,79**	3,40 %	1.175
WN-Semantic Class	Emotiblog	53,80	53,69	53,75		4.896
WN-Semantic Class	Emotiblog Norm	53,25	53,49	53,37	-0,71 %	4.898
WN-Semantic Class	Reviews	55,00	48,20	51,40		4.895
WN-Semantic Class	Reviews Norm	55,00	**48,50**	**51,50**	0,19 %	4.896

Table 1: Polarity classification results by corpus

Level	DOM	Sem.CLS	DOM.NRM	Sem.CLS.NRM	Diff.DOM	Diff.Sem.CLS
Informal	42,36	43,89	**43,52**	**45,07**	2,74 %	2,68 %
Medium	45,52	47,16	45,62	47,06	0,21 %	-0,20 %
Formal	51,80	51,40	**52,20**	**51,80**	0,77 %	0,78 %

Table 2: Polarity classification results (F1) by informality level using WordNet Domain (DOM) and WordNet Semantic Class (CLS) methods before and after normalisation (NRM)

shows a positive impact in some cases. To understand these possible negative effects introduced by normalisation we conducted a case by case analysis and identified the four main scenarios where normalisation can not only fail to improve but also worsen SA results. Finally, we can conclude that normalisation improves SA techniques and informality analysis can be used to determine which texts could benefit from this pre-processing step.

References

Adda, G., M. Adda-decker, J. luc Gauvain, and L. Lamel. 1997. Text normalization and speech recognition in french. In *Proc. ESCA Eurospeech 1997*, pages 2711–2714.

Amigó, E., J. Carrillo de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, T. Martín, E. Meij, M. de Rijke, and D. Spina. 2013. Overview of RepLab 2013: Evaluating Online Reputation Monitoring Systems. In *Proceedings of the Fourth International Conference of the CLEF initiative*, pages 333–352.

Aw, A., M. Zhang, J. Xiao, and J. Su. 2006. A phrase-based statistical model for sms text normalization. *Proceedings of the COLING/ACL*, pages 33–40.

Barbosa, L. and J. Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 36–44, Stroudsburg, PA, USA. Association for Computational Linguistics.

Bifet, A. and E. Frank. 2010. Sentiment knowledge discovery in twitter streaming data. In *Proceedings of the 13th international conference on Discovery science*, DS'10, pages 1–15, Berlin, Heidelberg. Springer-Verlag.

Ekman, P., 1999. *Basic Emotions*, pages 45–60. John Wiley and Sons, Ltd.

Esuli, A. and F. Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2016*, pages 417–422.

Fang, A. C. and J. Cao. 2009. Adjective density as a text formality characteristic for automatic text classification: A study based on the british national corpus. In O. Kwong, editor, *PACLIC*, pages 130–139. City University of Hong Kong Press.

Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books.

Gutiérrez, Y., A. González, R. Pérez, J. I. Abreu, A. Fernández Orquín, A. Mosquera, A. Montoyo, R. Muñoz, and F. Cámara. 2013. UMCC_DLSI-(SA): Using a ranking algorithm and informal features to solve Sentiment Analysis in Twitter. *Semeval 2013, Proceedings of the 7th International Workshop on Semantic Evaluations*.

Gutiérrez, Y., A. F. Orquín, A. Montoyo, and S. Vázquez. 2010. Integración de recursos semánticos basados en WordNet. *Procesamiento del Lenguaje Natural*, 45:161–168.

Gutiérrez, Y., S. Vázquez, and A. Montoyo. 2011. Sentiment classification using semantic features extracted from WordNet-based resources. In *Proceedings of the 2Nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA 2011, pages 139–145, Stroudsburg, PA, USA. Association for Computational Linguistics.

Han, B. and T. Baldwin. 2011. Lexical normalisation of short text messages: Makn sens a #twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 368–378, Portland, Oregon, USA, June. Association for Computational Linguistics.

Han, B., P. Cook, and T. Baldwin. 2013. Lexical normalization for social media text. *ACM Trans. Intell. Syst. Technol.*, 4(1):5:1–5:27, February.

Heylighen, F. and J.-M. Dewaele. 1999. Formality of language: definition, measurement and behavioral determinants. Technical report, Free University of Brussels.

Izquierdo-Bevià, R., A. Suàrez, and G. Rigau. 2007. A proposal of automatic selection of coarse-grained semantic classes for WSD. *Procesamiento del Lenguaje Natural*, 39:189–196.

Liu, F., F. Weng, B. Wang, and Y. Liu. 2011. Insertion, deletion, or substitution? normalizing text messages without precategorization nor supervision. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 71–76, Portland, Oregon, USA, June. Association for Computational Linguistics.

Magnini, B. and G. Cavaglia. 2000. Integrating subject field codes into WordNet. In *LREC*. European Language Resources Association.

Mosquera, A., E. Lloret, and P. Moreda. 2012. Towards facilitating the accessibility of web 2.0 texts through text normalisation. In *Proceedings of the LREC workshop: Natural Language Processing for Improving Textual Accessibility (NLP4ITA) ; Istanbul, Turkey.*, pages 9–14.

Mosquera, A. and P. Moreda. 2012a. A qualitative analysis of informality levels in web 2.0 texts: The facebook case study. In *Proceedings of the LREC workshop: @NLP can u tag #user_generated_content ?!; Istanbul, Turkey.*, pages 23–29.

Mosquera, A. and P. Moreda. 2012b. Smile: An informality classification tool for helping to assess quality and credibility in web 2.0 texts. In *Proceedings of the ICWSM workshop: Real-Time Analysis and Mining of Social Streams (RAMSS)*.

Mosquera, A. and P. Moreda. 2012c. The study of informality as a framework for evaluating the normalisation of web 2.0 texts. In *Proceedings of 17th International conference on Applications of Natural Language Processing to Information Systems (NLDB 2012)*. Springer.

Mosquera, A. and P. Moreda. 2013. Improving web 2.0 opinion mining systems using text normalisation techniques. *Recent Advances in Natural Language Processing (RANLP)*.

Mukherjee, S., A. Malu, B. A.R., and P. Bhattacharyya. 2012. Twisent: a multistage system for analyzing sentiment in twitter. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, CIKM '12, pages 2531–2534, New York, NY, USA. ACM.

Nakov, P., Z. Kozareva, A. Ritter, S. Rosenthal, V. Stoyanov, and T. Wilson. 2013. Semeval-2013 task 2: Sentiment analysis in twitter.

Niles, I. and A. Pease. 2003. Mapping WordNet to the SUMO ontology. In *Proceedings of the ieee international knowledge engineering conference*, pages 23–26.

Pak, A. and P. Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010)*, Valletta, Malta, May. European Language Resources Association (ELRA).

Pang, B. and L. Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the ACL*.

Pang, B., L. Lee, and S. Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

Philips, L. 2000. The double metaphone search algorithm. *C/C++ Users Journal*, 18:38–43, June.

Ratcliff, J. W. and D. E. Metzener. 1988. Pattern matching: The gestalt approach. *Dr. Dobb's Journal*, 13(7):46–72, July.

Santini, M. 2006. Web pages, text types, and linguistic features: Some issues. *ICAME Journal*, 30.

Sidorov, G., S. Miranda-Jiménez, F. Viveros-Jiménez, A. Gelbukh, N. Castro-Sánchez, F. Velásquez, I. Díaz-Rangel, S. Suárez-Guerra, A. Treviño, and J. Gordon. 2013. Empirical study of machine learning based approach for opinion mining in tweets. In *Proceedings of the 11th Mexican international conference on Advances in Artificial Intelligence - Volume Part I*, MICAI'12, pages 1–14, Berlin, Heidelberg. Springer-Verlag.

Sproat, R., A. W. Black, S. F. Chen, S. Kumar, M. Ostendorf, and C. Richards. 2001. Normalization of non-standard words. *Computer Speech and Language*, 15(3):287–333.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede. 2011. Lexicon-based methods for sentiment analysis. *Comput. Linguist.*, 37(2):267–307, June.

Taboada, M. and J. Grieve. 2004. Analyzing appraisal automatically. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, pages 158–161.

Thurlow, C. 2003. Generation txt? the sociolinguistics of young people's text-messaging. *Discourse Analysis Online*, 1(1).

Turney, P. D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA. Association for Computational Linguistics.

Vázquez, S., A. Montoyo, and Z. Kozareva. 2007. Word sense disambiguation using extended relevant domains resource. In H. R. Arabnia, M. Q. Yang, and J. Y. Yang, editors, *IC-AI*, pages 823–828. CSREA Press.

Wang, P. and H. T. Ng. 2013. A beam-search decoder for normalization of social media text with application to machine translation. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 471–481, Atlanta, Georgia, June. Association for Computational Linguistics.

Wiebe, J., T. Wilson, and C. Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39(2-3):165–210.