ICDEL Journal,
Vol.2, No.1 (2017)

Can Sentiment in our Words be Quantified? An Introduction to Lingmotif, a Sentiment Analysis Software Tool and its Educational Application

Javier Fernández-Cruz

Universidad de Málaga, Spain

Pontificia Universidad Católica del Ecuador, Sede Esmeraldas

Email for correspondence: fernandezcruz@uma.es

Receipt date: May 8th, 2016

Approval date: June 28th, 2016

How to cite this article (APA Norms)

Fernández-Cruz, J. (2017). Can Sentiment in our Words Be Quantified? An Introduction to Lingmotif, a Sentiment Analysis Software Tool and its Educational Application. International Congress on the Didactics of the English Language Journal, Vol. 2, No.1. ISSN 2550-7036. Retrieved from http://revistas.pucese.edu.ec/ICDEL/index

Abstract

In this paper, we present a concise introduction to Sentiment Analysis. We introduce a multilingual lexicon-based SA application, Lingmotif, developed by the Tecnolengua team at the Universidad de Málaga (Spain). Lingmotif is a lexicon-based, multi-platform desktop application for Sentiment Analysis. This software detects sentiment-laden words and multi-word expressions and returns a detailed analysis of its sentiment. In this paper, we provide a description of the tool's interface along with a brief proposal for its application in education and English Language Teaching.

***Keywords**:* Sentiment analysis; opinion mining; Natural Language Processing; education; discourse analysis; digital humanities

Introduction

Sentiment Analysis (SA), sometimes referred as Opinion Mining, is the discipline within Natural Language Processing and Computational Linguistics involved with the automatic treatment of opinion and subjectivity in texts. The development of SA permits the classification of sentiment in texts and its sections whether positive or negative. SA applications receive any type of electronic text as an input and returns automatically a final overall sentiment score along with its corresponding text analytics. SA software developments are currently applied as a tool in many end-user fields of knowledge, whether professional, commercial or academic.

Pang and Lee (2008) retract the foundation of SA as a discipline to 2001 but, according to the authors, research on opinion and evaluative language and emotion talk can be traced back to the 1970s (Bednarek 2008) as research on *connotation* (Lyons, 1977), *affect* (Besnier, 1993) and *attitude* (Halliday, 1994). On the other hand, Artificial Intelligence approaches to *subjective understanding* on political discourse and ideology can be traced back to Carbonell (1979).

In relation with SA, all the information we can find in a text of any kind can be classified into facts and opinions. Facts are considered as the objective expressions that may describe any entity and its sub-attributes, while opinions are subjective expressions that describe any subject's emotion, sentiments and assessment about any entity in real life (Liu, 2010, 2015). Mostly, opinions are expressed on target entities which may contain different hierarchical sub components. Its top rank subdivision is the *object*, which is defined as any entity in the real world (*i.e.* a person, a product, and idea…). Even if there are more complex patterns of sentiment, such as the five-variable scheme by Liu (2012) or the model described by Pawar, *et. al* (2016).

For our purpose, this can be simplified in a *parts* and *features* scheme as in Moreno-Ortiz (2016b). *Parts* deal with the different physical or abstract subdivisions of an object, while features deal with their variables. To illustrate this, an orange (as an *object*) has certain sentiment-carrying subdivisions. The *parts* of an orange would be i.e. *peel, pulp, juice, seeds* and their *features* would include *colour, aroma, taste, size,* etc. In this opinion scheme, as a matter of illustration, an orange with *big seeds* carries a more negative semantic orientation if we compare it with an orange with *small seeds* (Pawar et al., 2016**;** Liu, 2012; Moreno-Ortiz, 2016b).

A teacher-friendly SA application: Lingmotif

In 2012, the Tecnolengua group at the University of Málaga developed Sentitext, a lexicon-based sentiment analysis system for Spanish (Moreno-Ortiz, et. al, 2010) which set the grounds for current software. Lingmotif, in its current presentation**,** is a bilingual Spanish-English lexicon-based SA tool. As a desktop suite, it is developed entirely in Python for Windows, MacOS and Linux. Its detailed lexicon was acquired semi automatically and, subsequently, revised by hand. As for results, it leads to a strong commitment to both coverage and quality. In addition, Lingmotif permits using custom lexicons when dealing with specialised domains. Future efforts will involve the development of Lingmotif 2 in 2017, which will include analysis in French, German and Italian (Moreno-Ortiz, 2016ab, 2017ab).

As a lexicon-based system, the automatic calculation of semantic orientation in Lingmotif relies on three major linguistic data sources for each language: an individual words dictionary, a multiword expressions (MWE) dictionary and a context rules set. Lexical items in both dictionaries in our database were assigned a valence that marked their orientation and degree (from -2 to 2). To illustrate this with figures, Lingmotif's lexicon includes 77,000 entries (word forms) and nearly 500 context rules for English. Lexical items in both dictionaries in our database were assigned a valence marking their orientation and degree (from -2 to 2) (Moreno-Ortiz, 2017ab).

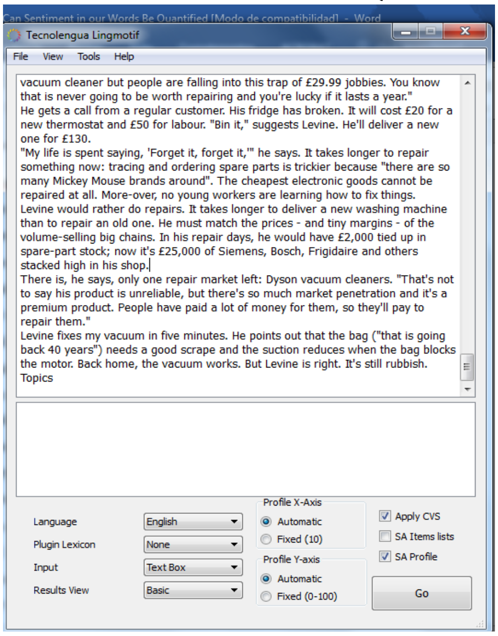

Figure 1. The interface of Lingmotif 1.0 – Windows (Moreno-Ortiz, 2016a)

Lingmotif's process of calculating semantic orientation can be briefly outlined as follows. Input text, whether copied and pasted or uploaded by batches of txt files, is pre-processed, tokenized, lemmatized, and part-of-speech tagged. Multiword expressions are identified and tagged, too. The analysis process can be briefly outlined as follows:

1. All lexical words and MWE's are looked up in the system's sentiment lexicons and matches are assigned their corresponding valence.
2. Context rules are searched for every lexical word or MWE. Matching segments are assigned the valence resulting from the application of the context rule. These items must be considered as "fuzzy quantifiers", as they are able to capture the changes that words such as *slightly*, *extremely* or *very* produce in their neighbouring context.
3. The proportion of positive and negative segments vs neutral segments, denominated *Affect Intensity*, is calculated.
4. Current system of valences is set on a [-5 - +5] range, being [–2, -5] negative valences and the [+2, +5] range valued as positive. As of output results, Lingmotif classifies "segments" rather than words. Any of the following items is considered to be a segment: a lexical word, a MWE or context rules conglomerate (Moreno-Ortiz, 2017b). Segments have weight in its final score calculation; while function words and other common stopwords have no weight whatsoever.

5. For the final results, positive and negative values are added after factoring in text size for each value. Two different output figures are returned in a scale ranging from 0 to 100 points: Text Sentiment Score (TSS) and Text Sentiment Intensity (TSI). The first score (TSS) quantifies the global valence of the analysed text and TSI returns the density of non-neutral lexical items.

Conclusions

A practical application: Sentiment Analysis in education

To conclude, we want to illustrate this document providing examples of SA (Table 1) in general education and ESOL, which are only two of the many possibilities Lingmotif has to offer. The emergence of SA has developed in parallel with the introduction of ICT in education. The combination of both offers promising new grounds for improving learning processes in traditional and online education. As we know, students with positive attitudes are generally more motivated in learning settings. Thanks to SA, we can measure the sentiment of students' interactions with educational software and online learning (Siemens & Baker, 2012; Wen, Yang and Rosé, 2014).

Table 1

Possible applications of Sentiment Analysis software in education

General educational purposes	ESOL
• Analysing the emotional part of our students written work.	• Working with connotation and denotation in the Language Lab.
• Automatic processing of students end-term evaluations and self-evaluations.	• Working on style in the School's Writing Lab.
• Analysing emotions in textbooks and classroom materials.	• Working on style in Spanish to English translators.
• School counsellors: detecting conflicts, motivation or self-esteem issues in students.	• Learning new vocabulary from a text by guessing its meaning from educational text sentiment analysis.
	• Contrasting different texts dealing with positive and negative vocabulary.

References

Bednarek, M. (2008). *Evaluation in media discourse: analysis of a newspaper corpus*. London: Continuum.

Besnier, N. (1993). *Reported speech and affect on Nukulaelae Atoll*. *Responsibility and Evidence in Oral Discourse*, 161–181.

Carbonell, J. G. (1979). *Subjective Understanding: Computer Models of Belief Systems.* Yale University, New Haven.

Halliday, M. A. K. (1985). *An introduction to functional grammar*. London ; Baltimore, Md., USA: E. Arnold.

Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis*. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.

Pawar, A. B., Jawale, M. A., & Kyatanavar, D. N. (2016). *Fundamentals of Sentiment Analysis: Concepts and Methodology*. In W. Pedrycz & S.-M. Chen (Eds.), *Sentiment Analysis and Ontology Engineering* (Vol. 639, pp. 25–48). Cham: Springer International Publishing.

Liu, B. (2012). *Sentiment analysis and opinion mining*. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Lyons, J. (1977). *Semantics (vols I & II)*. Cambridge, UK: CUP.

Moreno-Ortiz, A. (2017a). *Lingmotif: A User-focused Sentiment Analysis Tool*. Procesamiento de Lenguaje Natural, 58.

Moreno-Ortiz, A. (2017b). *Lingmotif: Sentiment Analysis for the Digital Humanities*. Presented at the Proceedings of the EACL, Valencia, Spain.

Moreno-Ortiz, A. (2016a). *Lingmotif* (Version 1.0) [Linux]. Málaga (Spain): Universidad de Málaga. Retrieved from http://tecnolengua.uma.es/lingmotif

Moreno-Ortiz, A. (2016b). *Extracting Domain-Specific Features for Sentiment Analysis Using Simple NLP Techniques: Running Shoes Reviews*. Presented at the Congreso Internacional de Lingüística de Corpus 2016, Málaga.

Siemens, G., & d Baker, R. S. (2012, April). *Learning analytics and educational data mining: towards communication and collaboration*. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 252-254). ACM.

Wen, M., Yang, D., & Rosé, C. P. (2014). *Linguistic reflections of student engagement in massive open online courses*. In the Eighth International Conference on Weblogs and Social Media. Palo Alto, California.