International Journal of Linguistics, Literature and Translation
ISSN: 2617-0299 (Online); ISSN: 2708-0099 (Print)
DOI: 10.32996/ijllt
Journal Homepage: www.al-kindipublisher.com/index.php/ijllt

IJLLT
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

A Survey of Some Italian Literature Works using Sentiment Analysis

Luca Pavan
Institute of Foreign Languages, Vilnius University, Vilnius, Lithuania; Language Studies Centre, Faculty of Creative Industries, Vilnius Tech, Vilnius, Lithuania; Department of Foreign Languages, Literary and Translation Studies, Vytautas Magnus University, Kaunas, Lithuania
Corresponding Author: Luca Pavan, **E-mail**: pavan@panservice.it

| ABSTRACT
The article describes the use of software, written by the author, to analyze some Italian literature works from the last century by using sentiment analysis. The software is lexicon-based, with a sentiment Italian dictionary including about 30,000 positive and negative words. To set the sentiment of each word, a scale of points ranging from 0 to 100 was introduced. To check the overall sentiment in literary works, the software makes use of a new parameter, the index of positivity. It is found that sentiment analysis is an efficient way to detect the tendency of opinions (positive or negative) in literature works. The results of this analysis coincide with the critics on the classification of authors' tendencies. The index of positivity could be useful also for bookstores: the customers, checking the index, would expect in advance, which is the overall tendency in a literary work. In this case, it would be easier to select a certain kind of work, according to the taste or the wish of customers.

| KEYWORDS
Computational linguistics, Sentiment analysis, Italian literature.

| ARTICLE DOI: 10.32996/ijllt.2022.5.1.16

1. Introduction

In the last two decades, sentiment analysis has become very popular in computational linguistics; Sentiment analysis studies, among other things, the people's opinions, sentiments and emotions (Liu, 2012). There are different ways to perform sentiment analysis, from lexicon-based analysis to artificial intelligence and machine learning (Dharani Devi, Kamalakkannan, 2020). In the end, the results help to understand the sentiment tendency (positive, negative, or neutral) or to classify opinions and emotions in a text without the need to read it. This kind of analysis is popular among companies of e-commerce to understand the taste of customers as fast as possible. It is also popular to analyze people's opinions on social networks, such as politics. However, sentiment analysis can be useful to study the sentiment's tendency in any kind of written text. Though, the analysis results do not always show the truth: for example, often sentiment analysis is not able to detect the irony in a text. Such limitation can be a problem when analysing the text is a literature work. However, in most cases, it is still possible to analyze literature works with a certain amount of precision. In the last years, there were even attempts to extrapolate emotions (joy, anger etc.) from a text (Kim, Klinger, 2018, pp. 3-5). In the case of the Italian language, dictionaries of emotions are still in progress. However, large dictionaries with classic tendencies (positive, negative or neutral) can quite easily be enlarged and included in some software to perform the analysis.

This article describes the software of sentiment analysis and the results of the comparative research of selected Italian literature authors made by using that software. There was no method for selecting the works for analysis except the timing and popularity – several among the most popular Italian literature works from the last one hundred years.

2. Method

This section describes a lexicon-based software to perform sentiment analysis, called Psychoword. The software is written in C language. C is not reputed as the best language to write such kind of software, but it has a lot of documentation, making it not too

difficult to carry out this project. The dictionary was created manually, and it is based on an existing dictionary of positive and negative sentiment words (Porcu, 2016). About 3,000 words were added by the author. At present, the dictionary includes about 30,000 words. The present version of Psychoword works with single words, but it also could process sentences - this feature could be used in future versions. The words include nouns, adjectives (with superlatives), adverbs and verbs. Each word contains a label showing the *level of positivity* (LP) using a scale of points ranging from 0 to 100. Rather than having a bipolar scale (Taboada et al., 2011) author preferred a unipolar one, taking into consideration only the positivity: a word with 100 points is totally positive, a word with 50 points is ambiguous, a word with less than 50 points has a predominance of negativity, and a word with 0 points is totally negative (it has no positivity). The words included in the dictionary were labelled either 0 or 100 points, but for some words, the sentiment of which is more complex, the score is different. For example, the word *discreto* was labelled with 75 points and *Malinconico* with 15 points. To analyze the overall tendency of a text, Psychoword uses the arithmetic mean of all the LPs, which is equal to an *index of positivity* (IP). Because the mean between the minimum and the maximum LP is 50 when the IP (which is normalized from 0 to 100) is less than 50, the overall tendency of the test is negative. Values of IP above 50 shows an overall positive tendency. However, Psychoword can change the values of the LPs according to *valence shifters* (Taboada et al., p. 269). In the software, 11 valence shifters, usually adverbs, can influence the results of the analysis. The most important valence shifters are *non* and *mai*: if they are found in a sentence, the polarities of the next sentiment words are reversed. Other valence shifters which augment or diminish the score of LPs of a certain amount are *abbastanza*, *appena*, *assai*, *meno*, *più*, *molto*, *poco*, *troppo*, *tanto*.

The software Psychoword works with a text file as input and three text files as output. The output files include the results of the analysis. In one output file, the input text is rewritten, showing the labels of sentiment words: this allows the user to find the sentiment words easily. Another output file shows which line of the input text to find sentiment words and other valuable information: the value of IP, the number of positive and negative words, the most used positive word and the most used negative word. Finally, the third output file shows the list of sentiment words found in the input text sorted by their frequency, like in Fig. 1 (from Italo Svevo's *Corto viaggio sentimentale*). In this case, it is possible to check which are the most popular sentiment words in a text. The author also built an English version of the software Psychoword (with valence shifters), based on the dictionary of Hu and Liu (2004, https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html). However, the sentiment analysis results can change by some amount, according to the user dictionary. More tests in different languages will be performed in future to observe the differences among dictionaries.

Bene[100]	29
Vero[100]	27
Dolore[0]	23
Amico[100]	21
Bella[100]	20
Vera[100]	18
Difficile[0]	17
Male[0]	17
Amore[100]	15
Sogno[100]	15

Fig. 1 – Some sentiment words with label and frequency sorted in descending order

3. Results and discussion

To perform a sentiment analysis with the software Psychoword the author has chosen several Italian literature works of different authors from the last one hundred years. In total there are 16 authors and the genre of works are either poetry or novels. The works (with authors in alphabetical order) are: *Una donna* (S. Aleramo, 1906), *Castelli di rabbia* (A. Baricco, 1991), *Il barone rampante* (I. Calvino, 1957), *Donne* (A. Camilleri, 2016), *L'edera* (G. Deledda, 1908), *Un uomo* (O. Fallaci, 1979), *L'amica geniale* (E. Ferrante, 2011), *Le occasioni* (E. Montale, 1939), *1934* (A. Moravia, 1982), *Il codice di Perelà* (A. Palazzeschi, 1911), *La spiaggia* (C. Pavese, 1942), *Sei personaggi in cerca d'autore* (L. Pirandello, 1921), *Oboe sommerso* (S. Quasimodo, 1932), *Il contesto* (L. Sciascia, 1971), *Corto viaggio sentimentale* (I. Svevo, 1928), *Sentimento del tempo* (G. Ungaretti, 1933).
Psychoword performs the analysis using four parameters:

IP – Index of positivity

PT – Percentage of text used in the analysis

PW – Number of positive words

NW – Number of negative words

The main results of the analysis of the above-listed literature work using Psychoword are shown in Fig. 2.

Author	IP	PT	PW	NW
Camilleri	57.94	4.59	1102	807
Svevo	55.77	5.48	958	746
Pavese	54.72	4.51	446	405
Baricco	53.50	4.19	1192	1029
Ferrante	52.93	4.85	2576	2135
Aleramo	51.79	6.45	2241	2066
Moravia	51.58	4.53	1933	1900
Ungaretti	49.77	4.75	182	192
Palazzeschi	48.72	4.71	1168	1226
Pirandello	48.71	3.88	457	451
Calvino	48.16	3.68	1234	1356
Sciascia	44.26	5.20	619	887
Quasimodo	42.14	5.04	94	111
Deledda	40.74	5.74	1246	2043
Fallaci	40.73	5.00	3851	6190
Montale	38.98	2.55	98	164

Fig. 2 – Results of sentiment analysis of several authors sorted according to the score of IPs

The IP is the main parameter of sentiment analysis. To check the accuracy of Psychoword results, the same software was tested on some positive Amazon's reviews. The value of the IP, as expected, was very high (above 90). However, in the case of literary works, the IP normally does not reach very high or low values. In fact, in such works, the authors try to keep a certain balance: a work with an IP too high or too low would probably result boring for the readers because of lack of variety. However, the IP clearly shows which are the works with a positive, neutral, or negative tendency. All the values above 50 show positivity, for example, in authors like Camilleri and Ferrante. If IP is 50, there is neutrality. Authors like Montale and Quasimodo, with an IP below 50, show a certain amount of negativity.

The second parameter of Psychoword (PT) is simply calculated by taking the number of sentiment words of the total number of words. Usually, PT is a small number, but it could be useful for comparing works. If PT is too close to 0, the analysis would be not very reliable.

Psychoword also shows the number of positive (PW) and negative words (NW). This can be useful, especially to evaluate the presence of valence shifters. If the value of PW is higher than NW, it demonstrates a predominance of positivity. The authors of the works which have an IP below 50 have a negative tendency in their style. For example, the analysis of Montale's and Quasimodo's works coincides with critics. The poetry of Montale seems to include metaphysical negativity, similarly to Giacomo Leopardi's works (Montale - De Rogatis, 2018, intr.). In Quasimodo's work, it is possible to find the negativity coming from hermeticism's style (Mauro – Quasimodo, 2020, intr.). Also, in Fallaci's work, the negative IP does not contradict the critics: in this book, there is, first of all, tragedy and death (Procacci – Fallaci, 2010, intr.). In the works of other authors with negative IP, like Palazzeschi and Pirandello, the software can hardly detect the irony. However, the works of these authors also have a negative view of the world of the bourgeoise and its contradictions (Palazzeschi, ed. 2015, intr.; Pirandello – Zavanella, ed. 2013, intr.).
Among authors who show a positive tendency, it can be noticed that some of them wrote their works more recently. In fact, the literature of the time period starting after the II World War often is not characterized by a negative view of the world. Therefore, authors like Camilleri, Baricco, Ferrante and Moravia have a positive IP.

Psychoword is also able to show the most used positive and negative words in a text. This feature allows one to compare authors' works and find out which sentiment words, among the most used, are in common. The list of these words is shown in Fig. 3.

MP – Most used positive word

MN – Most used negative word

Author	MP	MN
Camilleri	*Amore*	*Morte*
Svevo	*Bene*	*Dolore*
Ferrante	*Bene*	*Male*
Pavese	*Amici*	*Male*
Ungaretti	*Sogno*	*Morte*
Moravia	*Amore*	*Morte*
Aleramo	*Amore*	*Dolore*
Baricco	*Bene*	*Odio*
Calvino	*Bene*	*Paura*
Palazzeschi	*Bene*	*Morte*
Sciascia	*Bene*	*Errore*
Pirandello	*Vero*	*Dramma*
Deledda	*Bene*	*Morto*
Montale	*Dolce*	*Morte*
Fallaci	*Bene*	*Morte*
Quasimodo	*Amore*	*Morti*

Fig. 3 – Most used negative and positive words according to the authors

According to this list, there are many sentiment words in common. In total, without duplicates, only 15 words are left. The most used word is *bene*, followed by *morte*, *amore* and *dolore*. For a better understanding, it is possible to put the list in a cloud, in which the frequency is proportional to the size of words (Fig. 4).

Fig. 4 – Cloud showing the most used negative and positive word

Finally, with Psychoword, it is possible to analyze all the authors together, getting a list of sentiment words sorted according to their frequency (Fig. 5). If performed on a bigger corpus of authors, this kind of analysis could reveal, for example, which sentiment words are popular in a certain period. Furthermore, it would be possible to get the overall index of positivity in a particular period.

Bene[100]	808
Amore[100]	528
Vero[100]	458
Morte[0]	438
Male[0]	372
Paura[0]	308
Bella[100]	260
Morto[0]	224
Amico[100]	218
Sorriso[100]	205

Fig. 5 – Some sentiment words found in a group of authors with label and frequency (sorted in descending order)

To know the index of positivity could be useful also for customers in a bookstore. Even if sentiment analysis sometimes cannot be reliable (it fails to detect the positivity or negativity of certain emotions), it is generally possible to know the overall sentiment of work, helping the readers understand in advance which kind of product is close to their taste. In the near future, with the creation of more specific dictionaries, it will also be possible to classify the emotions, getting useful information about literary works.

4. Conclusions

The article described Psychoword - a software to perform sentiment analysis of texts, along with results of sentiment analysis of several Italian literature works. It was found that sentiment analysis is reliable using a lexicon-based dictionary. The results of the analysis by Psychoword are like the opinions of critics. Also, they demonstrate that sentiment analysis can help compare authors and observe which lexical features they have in common. A scale from 0 to 100 is proposed to get an index of positivity, using negative and positive polarities. This index could be used by bookstores to show the sentiment of books, giving the customers a greater help to understand which kind of product they are buying. Finally, even if the new tendencies of studies on sentiment analysis include the field of emotions, it was shown that relying only on positivity and negativity is enough to increase the understanding of literary works.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

[1] Dharani D, G., and Kamalakkannan, S. (2020). Literature Review on Sentiment Analysis in Social Media: Open Challenges toward Applications. *International Journal of Advanced Science and Technology*, *29*(7), 1462-1471.
[2] Fallaci, O. (2010). *Un uomo*. Procacci, D. (ed.). Milano: Rizzoli.
[3] Kim, E., and Klinger, R. (2018). A Survey on Sentiment and Emotion Analysis for Computational Literary Studies. *ArXiv, preprint abs/1808.03137*.
[4] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. San Rafael, US: Morgan & Claypool Publishers.
[5] Montale, E. (2018). *Le occasioni*. De Rogatis, T. (ed.). Milano: Mondadori.
[6] Palazzeschi, A. (2015). *Il codice di Perelà*. Milano: Mondadori.
[7] Pirandello, L. (2013). *Sei personaggi in cerca d'autore*. Zavanella, A. (ed.). Firenze: Edimedia.
[8] Porcu, V. (2016). *Text mining e Sentiment Analysis con R*. Roma: Streetlib.
[9] Quasimodo, S. (2020). *Tutte le poesie*. Mauro, C. (ed.). Milano: Mondadori.
[10] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*(2):267-307.