Turk J Elec Eng & Comp Sci
() : –
© TÜBİTAK
doi:10.3906/elk-

A polarity calculation approach for lexicon-based Turkish sentiment analysis

Gökhan YURTALAN[1], **Murat KOYUNCU**[2*], **Çiğdem TURHAN** [2]
[1]HAVELSAN A.Ş., Ankara, Turkey
[2]Faculty of Engineering, Atilim University, Ankara, Turkey

Received: .201 • **Accepted/Published Online:** .201 • **Final Version:** ..201

Abstract: Sentiment Analysis attempts to resolve the senses or emotions that a writer or speaker intends to send across to the people about an object or event. It generally uses natural language processing and/or artificial intelligence techniques for processing electronic documents and mining the opinion specified in the content. In recent years, there have been many successful sentiment analysis studies for the English language which consider many words and word groups that set emotion polarities arising from the English grammar structure, and then use datasets to test their performance. However, there are only a limited number of studies for the Turkish language, and these studies have lower performance results compared to those studies for English. The reasons for this can be incorrect translation of datasets from English to Turkish and ignoring the special grammar structures in the latter. In this study, special Turkish words and linguistic constructs which affect the polarity of a sentence are determined with the aid of a Turkish linguist, and an appropriate lexicon-based polarity determination and calculation approach is introduced for this language. The proposed methodology is tested using different datasets collected from Twitter, and the test results show that the proposed system achieves better accuracy than the previously developed lexical-based sentiment analysis systems for Turkish. The authors conclude that especially the analysis of the word groups increases the overall performance of the system significantly.

Key words: Sentiment analysis, lexicon-based, Turkish language, opinion mining

1. Introduction

Recently, natural language processing and artificial intelligence techniques have been emerging as a solution for automatic sentiment analysis in different studies, whose general aim is to determine the intended emotion specified in documents, sentences, clauses or words that include observations and assessments concerning any aspects related to a given product, person or subject. Although emotions are specified in a comprehensive way using a variety of words, the senses associated with them may only be coarsely defined as either positive, negative or neutral [1]. In recent years, sentiment analysis has gained importance among individuals, institutions and organizations who wish to receive feedback about their products and services. This type of analysis, which cannot be easily done by humans, can be achieved with the development of efficient sentiment analysis tools.

Although it is possible to see other classification techniques in the literature, machine learning and lexicon-based classification are the most leading techniques from a technical point of view for sentiment analysis [2, 3]. It is possible to find studies combining lexicon-based and machine learning-based approaches as well [4]. Specifically, for the Turkish language, the majority of the research on sentiment analysis is based on machine learning methods as can be seen in the studies given in the next section. Only a few studies exist that use pure

*Correspondence: mkoyuncu@atilim.edu.tr

lexicon-based classification for Turkish [5, 6] with accuracy rates of 76% and 75%, respectively. Obtained success rates are lower compared to the studies on English and other languages because of the complex structure of the Turkish language. This complex structure makes natural language processing and sentiment analysis tasks difficult for this language [7]. Due to their own unique features and rules, English and Turkish have very different grammar structures. For this reason, structures specific to one cannot be used in the other in such studies. Therefore, working with Turkish linguists is a requirement to develop a successful system because of these inherent differences. On the other hand, some studies on sentiment analysis use translation from Turkish to English or vice versa [5, 6]. One approach is to translate Turkish dataset into English and use an algorithm developed for English. Another approach is to translate polarity words from English sources to Turkish and use them for sentiment analysis. Although translation algorithms have improved their success rates in recent years, fully accurate translations can barely be ever achieved.

In this study, we aim to propose an improved lexicon-based (LB) method to evaluate the polarity of Turkish social media data which achieves a better accuracy rate than the previous LB-based studies by taking into account the special features of the Turkish language in more detail. Sentiment analysis on Turkish tweets is performed where the dataset is collected from instant tweets with Twitter API. The polarity value is calculated for the entire length of the tweet using polarity values of words/word groups/idioms/proverbs and, ternary classification of the tweet is achieved as positive, negative or neutral. The contributions of this study are as follows:

- A new domain-independent LB polarity determination and calculation algorithm focusing on the word groups and using a specially built lexicon is proposed for the Turkish language which achieves better accuracy than the previously developed LB sentiment analysis systems;
- With the help of a Turkish linguist, many Turkish linguistic constructs, which are effective in setting the polarity values, are considered in the implementation.

The rest of the paper is organized as follows: In the next section, the related work on sentiment analysis are presented. In Section 3, the system architecture is explained in detail. Later, the test results and evaluation are explained in Section 4. Finally, the derived conclusions and future work are given in Section 5.

2. Related work

Sentiment analysis has been an active research area and different algorithms have been developed in various studies in recent years. For various languages, especially for English, numerous studies have been performed using different machine learning (ML) techniques, such as Naive Bayes (NB) and Support Vector Machine (SVM) with high success rates [8–13]. In some of these studies, the accuracy rate exceeds 90%. Besides ML-based studies, LB sentiment analysis is also one of the techniques used actively among researchers. For example, Turney achieved up to 84% accuracy [14]; whereas, Moreo, Romero, Castro and Zurita achieved 89% accuracy as the average of five different news datasets [15] and Fernandez-Gavilanes, Alvarez-Lopez, Juncal-Martinez, Costa-Montenegro and Gonzales-Castano came up with 74.8% accuracy in their studies using LB sentiment analysis [16]. More information about the techniques used in sentiment analysis can be found in [2] and [3].

As for Turkish, sentiment analysis has only started to gain interest in recent years. For example, Kaya, Fidan and Toroslu, who work on sentiment analysis with multiple machine learning methods for the purpose of performing comparisons, use the news domain [17]. They have obtained higher accuracy with supervised techniques and achieved an accuracy of 77% in binary classification of political news. Vural, Cambazoglu, Senkul and Tokgoz present a lexicon based sentiment analysis (unsupervised) framework which uses the translation of SentiStrength lexicon to Turkish [5]. They employ an approach based on summing lexicon scores of sentiment oriented words in related text. The accuracy of their framework is reported as 76.0% for movie reviews obtained from a popular social media site. Çetin and Amasyalı carried out several experiments to compare different term weighting methods for sentiment analysis on Turkish data [18]. Consequently, they found that the supervised term-weighting method which includes terms' distribution of classes is more successful, achieving approximately 62% accuracy. Balahur, Turchi, Steinberger, Perea-Ortega, Jacquet, Küçük, Zavarella and El Ghali translate English data into Turkish and analyze those using ML algorithms, and achieve an accuracy of 60% [19]. In that paper, the authors argue that there is insignificant difference between human translations and machine translations of datasets. Yıldırım, Çetin, Eryiğit and Temel were able to increase their accuracy to 79% by adding layers of natural language processing to ML methods [20]. There are other studies on Turkish sentiment analysis using ML techniques as well [21–24]. In addition, Türkmenoğlu and Tantuğ compare Lexicon based and Machine Learning-based sentiment analysis methods on Turkish social media [6]. They form a lexicon by translating an English opinion lexicon to Turkish and calculate the overall score by summing up sentiment scores of terms in the lexicon. They construct a baseline Machine Learning (ML) based method using different feature sets. They have applied both approaches for binary (positive/negative) classification and achieved an accuracy of 75.2% and 85% in binary classiffication of Twitter data with LB and ML, respectively. Similarly, Akgül, Ertano and Diri also have attempted to compare LB and n-gram methods for sentiment analysis on Twitter data and found that LB method has outperformed the n-gram method with 73.2% and 70% accuracy, respectively [25]. Furthermore, Dehkharghani, Yanikoglu, Saygin, and Oflazer propose a comprehensive sentiment analysis system for Turkish in which they cover different levels of sentiment analysis such as aspect, sentence, and document levels as well as some linguistic issues such as conjunction and intensification in Turkish sentiment analysis [26]. Their system is evaluated on Turkish movie reviews and the obtained accuracies range from 60% to 79% in ternary and binary classification tasks at different levels of analysis. Additional studies have been conducted to aid the current sentiment analysis research in Turkish such as the studies by Parlar and Özel [27] who propose a new feature selection method for sentiment analysis, Sağlam, Sever and Genç [28] who aim to develop a Turkish sentiment lexicon, and Omurca, Ekinci and Türkmen [29] who have developed an annotated corpus to be used in this domain. Finally, some of the studies in this domain have focused on aspect-based sentiment analysis in Turkish [30, 31].

3. System architecture

The developed system is a lexicon-based sentiment analysis tool for Turkish tweets. The lexicon of the system consists of positive and negative word roots, part-of-speech (POS) tags and the polarity values of 1181 data items. In addition, 398 data items including idioms and proverbs are also used to detect emotions in the data analysis phase. The overview of the system is shown in the Figure. Tweets are collected from Twitter by the Crawler module, forming the datasets to be processed. After this stage, every tweet passes through a set of processes. In the end, each tweet is classified either as positive, negative or neutral. A detailed description of each module is given in the following sections supported by examples.

3.1. Crawler

In this study, Twitter is used as the primary data source since it is more favorable compared to other social media services for sentiment analysis because the postings have a maximum limit of 140 characters within which

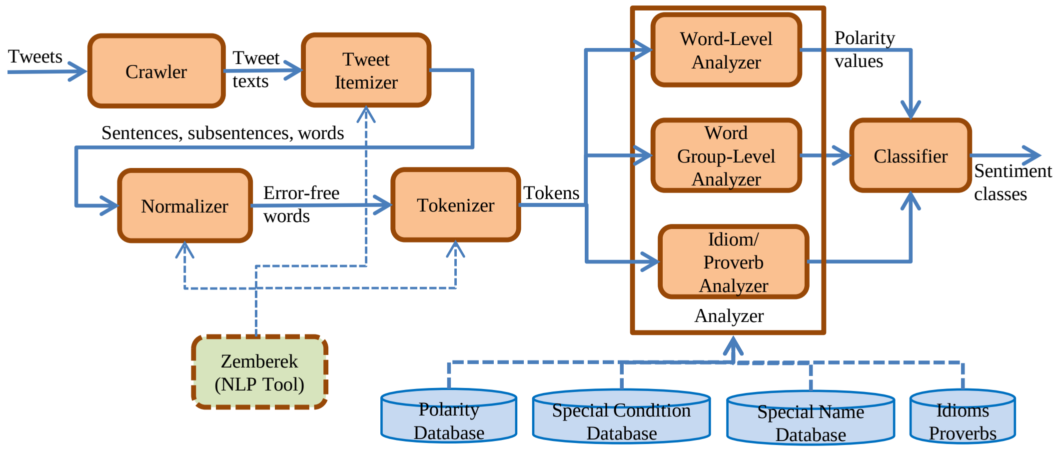

Figure. System architecture.

limits users to specify their feelings. In other words, they have to express their emotions in a concise manner. Twitter REST API v1.1 is utilized to collect data from Twitter. It allows queries against the indices of recent or popular tweets while behaving similarly to, but not completely as, the search feature available in the Twitter mobile or web clients. It requires user authentication and the responses are available in JSON format. In this study, GET search/tweets query option is used in the Crawler to filter and retrieve related tweets on a subject. Twitter supports more than 1100 emojis such as love symbols, country flags, hand signals and smileys. Also, Twitter announced that shared photographs, gifs, videos, URLs, and @username forms will not be counted within the 140 character limit. With this approach, users can reflect their opinions using text and additional components. Visually meaningful symbols on the screen are not systematically mapped to unique symbols or strings when tweets are converted into text as given below for a tweet.

- #Arçelik'in reklami harika olmus [?]! Yerli olarak üretilen her şey bni çok gururlandiriyor. Başımızın üstünde yerin var @kocholding [?] / *Arçelik's ad is great [?]!. Everything produced domestically makes me very proud. You have a special place in our minds @kocholding [?]*

Twitter API provides developers with the ability to select the tweet writer's language. In our application, Turkish is selected as the language; however, using the Twitter profile feature, a user can select the Turkish language, and, yet post tweets in a different language. To solve this problem, Google Language Detection API is used to make sure that the text is genuinely Turkish, and discard statements written in other languages.

Note that we only process text in the Twitter and not symbols, smileys, emojis, etc. Although some of these symbols may be important to analyze the sense in a tweet, they are out of the scope of this study. Twitter API provides different features for tweeting purposes, such as a list including hashtags, mentions and medias which are specific to Twitter. At this stage, specific characters and URLs that, in terms of sentiment bear no meaning are removed from the text. However, hashtags and mentions can be essential in a tweet in terms of meaning; as such they are kept within the message and handled as a proper name, and the remaining items are sent to the tweet itemizer module as seen in the following text:

- #Arçelik'in reklami harika olmus! Yerli olarak üretilen her şey bni çok gururlandiriyor. Başımızın üstünde yerin var @kocholding / *Arçelik's ad is great! Anything produced domestically makes me very proud. You have a special place in our minds @kocholding*

3.2. Tweet itemizer

In this module, the tweet which is being processed turns into a text and is parsed into its components. Sentences, phrases, words and suffixes in a tweet are separated and formatted as lists. For example, the tweet text given above is converted to a list of sentences with the help of the Zemberek SentenceBoundaryDetector method [32], and then each sentence is processed individually.

3.3. Normalizer

In the social media, users tend to make mistakes, intentionally or unintentionally, when writing words. Typical mistakes include absence or repetitions of letters, abbreviations and conjoined words. In order to eliminate these kinds of mistakes, Zemberek's spell checking module is applied to every word within each sentence and incorrect ones are then replaced with their correct forms. This process is referred to as 'normalization' and some normalization examples are as follows:

Letter repetition is encountered frequently in the social media by which users emphasize a certain sentiment, such as:

- onu çooook özlüyorummm / *I missssss him sooooooo much*

With the spell-checking method, the sentence is converted into the following:

- onu çok özlüyorum / I miss him so much

Similarly, the uncertainty regarding missing letters as the residual letters also has to be eliminated. In the tweet example given above, the word 'bni' is replaced with 'beni'.

Some letters used in Turkish such as 'ğ', 'I', 'ş', 'ç', 'ö' and 'ü' may not necessarily be defined and supported by certain mobile brands. Even if they are, users have to spend more effort to type them, e.g., having to press a key several times to find the correct Turkish letter. For this reason, they may simply choose to do without them when texting. This problem is also resolved in the normalization process using the related method in Zemberek. The following example shows the correct Turkish letter to be used on the right-hand side:

- reklami =>reklamı

Since Zemberek is an open source NLP tool, we have extended it in the following ways to increase its success for spellchecking:

- We have inserted 800 new items into its XML database. With these insertions, for example, not only "çook", but also "çooook" is converted correctly to "çok". Some other examples which can be corrected by modified spell-checking method are "haaaayırr", "hiççççççç", "tşk", "cnm", "gidiyooon", hiiişşşşşşttt, soole, napıooonn, dimi, işallahh.

- We have also modified the spell-checking method of Zemberek, since it sometimes returns unrelated words in the first order. After detecting the correct form of words using XML database, we also check the root of the corrected words with the root finder of Zemberek. After these two operations, the suggestion list of Zemberek is formed and we use the first one from the list.

3.4. Tokenizer

In this phase, tokens are created with the help of the NLP tool where the token definition includes the original word, POS (Part of Speech) tag, polarity value, root, suffixes and the special condition code of each word. The POS shows the linguistic category of the words as noun, verb, adjective, adverb, pronoun, preposition, conjunction and interjection. The POS tags are especially important for determining the word groups. Stop words which are not of importance in sentiment analysis are eliminated in this phase.

Polarity values of word roots are defined in a lexicon which is shown as the polarity database in the Figure. This lexicon which contains 672 negative and 509 positive roots is prepared with a special effort. First, around 4000 tweets are investigated to determine the mostly used words which are effective for sentiment analyses. The initial lexicon is constructed after this study. Then, we have used several documents of Turkish Language Association (http://www.tdk.gov.tr/) to determine the words having positive and negative meanings, synonyms and antonyms of existing words and accordingly extended the lexicon. The polarity values are assigned manually. We have also received help from a Turkish linguist during this process. The final lexicon includes 285 adjectives (e.g., "cazip"), 518 nouns (e.g., "felç"), 365 verbs (e.g., "coş", "kutla") and 13 interjections (e.g., "imdat"). Notice that the polarity database is constructed independently from the datasets given in this paper. The polarity value of a word root is defined in the lexicon and can take one of the values given in Table 1. If a word root together with its POS tag is not defined in the lexicon, then 0 (zero) is assigned as its polarity value. In Table 2, the tokenizing results of the first sentence of the example tweet are listed and the polarity values of the roots of the words are shown under the column *polarity value*. The condition column specifies whether the word has a special condition and needs further processing, as explained in the next section.

Table 1. Polarity values of word roots.

Polarity Name	Polarity Value
Very negative	-3
Negative	-2
Slightly negative	-1
Neutral	0
Slightly positive	1
Positive	2
Very positive	3

Table 2. Polarity values of word roots.

Word	POS	Polarity value	Root	Suffixes	Condition
Arçelik'in	Noun	0	Arçelik	-in	none
reklamı	Noun	0	reklam	-ı	none
harika	Adjective	2	harika		none
olmuş	Verb	0	ol	-muş	none

At this stage, another important process is to determine entity names such as person, institution or organization, because a whole name or a part of a name may indicate moods and lead to improper evaluations. For example, in the 'Dost Eli Konya Gıda Bankası Yardımlaşma ve Dayanışma Derneği'/ *'Dost Eli Konya Food Bank Association for Fraternity and Solidarity'*, the words 'yardımlaşma' (fraternity) and 'dayanışma' (solidarity) have positive nuance and positive polarity values in the database. However, these words are used in the name of an organization and, therefore, should be discarded and not included in the sentiment classification. Another benefit of extraction of entity names is to prevent wrong determination of bigrams and trigrams which are constructed in the next stage. The special name database (shown in the Figure) which includes about 1200 names with their acronyms is used for this process.

3.5. Analyzer

The analyzer module includes three submodules, namely the word-level analyzer, word group-level analyzer, and idiom/proverb analyzer. This design aims to handle in a systematic way the special conditions in Turkish affecting the sense of tweets.

3.5.1. Word-level analyzer

As Turkish is an agglutinative language, word roots are derived by taking on many suffixes, thereby may form different meanings and conveying various emotions. Every suffix can change the word root's tense, meaning and POS thereby changing its polarity [7]; thus, the roots and suffixes have to be examined individually. The word-level analyzer module handles special conditions at the word level. Here, a special condition defines the changes in the meaning of word roots when combined with specific suffixes.

In this scope, we have to consider the suffixes which are used as negators in the Turkish language because they can change the meaning of words from positive to negative or vice versa. One of them is the '-me' negator. Depending on the previous vowel, it may turn into '-ma' to provide vowel harmony. If used in the present continuous tense, it turns into either '-mı', '-mi', '-mu' or '- mü' [33]. The suffix '-me' is used after the verb root or body and reverses the meaning. Likewise, the '-maz' and '-mez' suffixes can be added to a verb root or body considering vowel harmony. Then, the word becomes an adjective and has a negative meaning. Another special condition to be handled is the suffixes '-lı' or '-li' and '-sız' or '-siz' which indicate presence or absence of a quality, respectively, as shown below:

- Haysiyet (root) + li (suffix) / honorable
- Haysiyet (root) + siz (suffix) / dishonorable

Although the root is defined as a positive word in the database, its polarity is converted to negative due to the negative suffix '-siz'. Table 3 shows how the polarities are affected when a root is combined with a suffix reversing the meaning of the attached root.

Table 3. Word-level special conditions.

Root+Suffix	**New polarity**
Positive root + negative suffix	Negative
Negative root + negative suffix	Positive
Neutral root + negative suffix	Negative

3.5.2. Word group-level analyzer

In the word-group level analysis phase, the n-grams are not used in the conventional sense where they are constructed from the adjacent words of the sentence, but rather from the contiguous noun phrases, adjective phrases, verb phrases, etc. which form meaningful word groups. These meaningful word groups are sometimes very critical for sentiment analysis. For example, consider the tweet “Arçelik’ten akilli ev uygulamasi: Home-Whiz”. Here, “akıllı ev uygulaması (smart home application)” is a noun phrase which does not specify negative or positive meaning. This tweet is labelled as a neutral tweet in the database. If we use only word-level analyzer the tweet is labelled as a positive tweet since “akıllı” has a positive polarity value in the lexicon. When we use word-level analyzer, “akıllı ev uygulaması” is determined as a noun phrase and zero polarity value is assigned to it. In this way, a correct classification is obtained.

The determination of these phrases relies on the case markings showing the constituents of the phrase, the order of the constituents as well as syntactic rules specific to Turkish. One of these rules used in the determination of noun phrases is that noun phrases in Turkish consist of a head and one or more optional modifiers which always precede the head where some of the constituents can be marked with specific suffixes. In the noun phrase given below, the genitive case suffix marks the start of the noun phrase which includes all the adjacent modifiers until the head with a possessive marker is found.

Uçağın	kırık	pervanesi	/ The broken propeller of the plane.
Plane+GEN	broken	propeller+POSS	

As such, using various morphological and syntactic rules of the Turkish grammar, different types of noun, adjective and verb phrases are determined and the whole phrase is included as one of the items in a bigram. Subsequently, trigrams are formed using the common words of the bigrams as shown in the example below:

(Uçağın kırık pervanesi) (problem)	(problem) (yarattı)
Bigram-1	**Bigram-2**
Trigram	
The broken propeller of the plane created a problem.	

Consider the word-level parsing result given in Table 2, the first bigram can be constructed as ‘#Arçelik’in reklamı’. However, since the suffix ‘-in’ is a genitive case suffix, it forms a noun phrase together with the possessive suffix ‘ı’ of the next word. Therefore, the noun phrase ‘#Arçelik’in reklamı’ is taken as a single item and the bigram is created by adding the next word as follows:

Bigram:	#Arçelik’in reklamı	harika
Polarity values:	0	2
In English:	*Arçelik’s ad*	*great*

The next stage is creating a trigram by combining two bigrams (‘#Arçelik’in reklamı harika’ and ‘harika olmuş’) which takes the following form:

Trigram:	#Arçelik’in reklamı	harika	olmuş
Polarity values:	0	2	0
In English:	*Arçelik’s ad*	*great*	*is*

After the trigrams are determined, their polarity values are calculated by adding the polarity values of the components. As a result, the following is obtained:

Polarity value of trigram$= 0 + 2 + 0 = 2$

As in the word-level stage, word groups also have special conditions. A negative noun sentence is an example of a special condition to be handled. Negative noun sentences are used to describe the absence of a quality or asset [33]. The typical words used in these sentences are 'değil', 'hiç', 'yok', 'hayır', 'ne…ne(de)', 'ama', 'aksi halde', 'yoksa'. The words 'ne…ne (de)', 'ama', 'fakat', 'lakin' are conjunctions which can make combination of word groups. Thus, each special condition arising in Turkish must be taken into account one by one in the application. Some examples of special conditions handled by the application are given below:

Çok / az (very, a lot / little, less): The words representing abundance and scarcity can shape polarity values. For example, the word 'gururlandırıyor' has a positive polarity value. When it is coupled with the word 'çok', its polarity value is changed to very positive which represents intensification in the polarity (increment the previously calculated polarity value by 1). Conversely, if coupled with the word 'az' it will take slightly positive value, which represents a reduction in the polarity (decrement the previously calculated polarity value by 1).

Ama / fakat / lakin / ancak / oysaki (but / yet / nevertheless / however / although): These words are typically used as conjunctions between two phrases which are supposed to have opposite polarities. The general rule applied to these conjunctions is to intensify the whole polarity considering the polarity of the second phrase. Conjunctions listed above causes the overall polarity to be incremented or decremented by 1 depending on the positivity or negativity of the second part of a sentence. In the given example below, polarities of two sentences are calculated separately, and then -1 is added because of the conjunction considering the negative polarity of the second phrase.

Trigrams:	Akşam yemeğimiz harika görünüyor	ancak	benim miğdem rahatsız	
	trigram	conjunction	trigram	
Polarity values:	2	-1	-2	= -1
In English:	Our dinner looks delicious	but	my stomach hurts	

Ne… ne(de) (neither..nor): Although this conjunction combines words with positive meanings, it reverses the meaning, converting them to negative. While forming trigrams in sentences, if there is the 'ne…ne(de)' conjunction, its word group should be represented by a trigram. As seen below, the first bigram has not been completed yet and has been formed using only two items because the 'ne…ne(de)' conjunction is included in the second trigram. Although the sentence without 'ne…ne(de)' has a positive meaning, the conjunction word changes the meaning and the overall polarity value is calculated as slightly negative (-1). Because of the conjunction, the existing polarity value (+2) is multiplied by 0 to make it neutral and, then, -1 is added to make it negative.

Trigrams:	Bu çalışmalar	ne doğru ne de kullanılabilir	bir yaklaşım sergiliyor	
	bigram	trigram	trigram	
Polarity values:	0	2*0-1	0	= -1
In English:	*These studies present neither an accurate nor a useful approach.*			

Değil / yok (not): These words reveal the absence or negativity in the concept of their preceding noun. While forming trigrams within the application, these words are added to the trigrams as the last item. As seen in the example below, when the word has a negative meaning and is used with a word indicating negativity (in this case, 'değil'), the result is calculated as slightly positive. In this example, 'kötü bir çocuk' has a (-2) polarity value. Because of the word ‘değil’, its polarity value is converted to neutral, multiplied by 0, and then +1 is added to make it slightly positive.

Trigrams:	Sen geçmişte	kötü bir çocuk değildin	
	bigram	bigram	
Polarity values:	0	-2*0+1	= +1
In English:	*You were not a bad kid in the past.*		

3.5.3. Idiom/proverb analyzer

At the word-level analyzer, the polarity of each word is taken from the polarity database based on the root of the word. Then, the initial polarity values are evaluated considering the suffixes and modified as explained in Section 3.5.1. At the word group-level analyzer, the polarity values produced by the word-level analyzer are used as inputs. Depending on the bigrams and trigrams determined, the polarity values are processed as explained in Section 3.5.2 and delivered to the idiom/proverb analyzer as the sum of the polarities obtained from word groups (bigrams and trigrams). The idiom/proverb-level analysis reveals the idioms and proverbs used in the tweets since the polarity of an idiom may be different than the polarity of its parts [7]. If an idiom or proverb cannot be detected, the polarity value taken from the word group-level analyzer is directly used as the polarity of the sentence. Otherwise, the whole polarity defined for the idiom or proverb is taken from the database and used as the polarity of the sentence.

In this stage, the roots of words within the sentence are compared and matched with the database containing the roots of idioms and proverbs. The reason for using the roots is that word suffixes may vary according to the person who forms the sentence. This way, matching can be done more easily by eliminating the suffixes. In the example tweet, the last sentence is an idiom and should be evaluated as a whole as given below.

Sentence:	Başımızın üstünde yerin var	@kocholding	
	idiom		
Polarity values:	2	0	= +2
In English:	*You have a special place in our minds @kocholding.*		

3.6. Classifier

When the analyzer module completes its processing, the classifier is executed and the tweet’s total polarity value is calculated using Equation (1),

$$polarityValue = \sum_{i=1}^{n} T_i \tag{1}$$

where n represents the number of sentences in a tweet, and Ti represents the polarity value of the sentence i. The polarity values of the sentences are added to obtain the final value. The polarity value of the example tweet is calculated as shown in Table 4.

The final polarity of the whole tweet is decided according to the final polarity value. For the example given in Table 4, the calculated final polarity value is greater than zero; therefore, this tweet is classified as a positive tweet.

Table 4. Calculation of the final polarity value.

Sentences	**Polarity Values**
#Arçelik'in reklamı harika olmuş!	+2
Başında Türk geçen her şey beni çok gururlandırıyor.	+3
Başımızın üstünde yerin var @kocholding.	+2
Polarity of the whole tweet	**+7**

4. Test and evaluation

4.1. Test results

For testing purposes, three datasets are created in different categories as shown in Table 5. Initially, to create the datasets the collected tweets are analyzed by three experts in Turkish language to classify the tweets as positive, negative and neutral, since we have decided to work on ternary classification. Each expert has classified the tweets individually. If a tweet is classified into the same class by the three experts, then it is included in the dataset. Notice that this classification does not discard the tweets with ambiguity in terms of sentiment analysis. Here, the idea is that a tweet should have a clear sense for a human. As such, our datasets reflect real tweets from Twitter including complex and ambiguous tweets as well as simple ones.

Chi-square test was applied to evaluate the results of the system as shown in (Sağlam, Sever and Genç, 2016). p-values for the three datasets were found as 0.998, 0.999 and 0.995 respectively, which prove that the results were statistically meaningful.

Table 5. Datasets

Dataset	**Topic**	**Category**	**Total Tweet**	**Polar Type**	**Ground Truth**	**%**
Dataset-1	Aziz Sancar	Science	300	Positive	184	61.3
				Negative	55	18.3
				Neutral	61	20.3
Dataset-2	Beşiktaş	Sports	364	Positive	149	40.9
				Negative	42	11.5
				Neutral	173	47.5
Dataset-3	Arçelik	Brand	537	Positive	142	26.4
				Negative	36	6.7
				Neutral	359	66.9

We present the results as accuracy, precision, recall and F1-score to evaluate the performance of our system. We prefer to use the accuracy measure since it is the mostly used metric in Turkish related studies [5, 6, 17]. F1-score is also calculated since our datasets are not well-balanced. For three datasets, the obtained rates are detailed in Table 6. As seen in the table, the average accuracy for all datasets is above 87%, with the highest accuracy of 88.2% obtained for the dataset on the topic of 'Aziz Sancar'. The lowest accuracy values are obtained for the neutral class. Since it is the hardest class to be distinguished between positive and negative, this result seems logical. However, the accuracy values for positives is lower than those for negatives. Although the system classifies negative tweets more successfully, we could not obtain the same success for positives. Here, the problem is that the system has difficulty distinguishing positives and neutrals, and needs more improvement to solve this problem. The averages of F1-scores are between 78.3 and 82.1 for different datasets and the first dataset achieves the highest score, similar to accuracy.

In addition, we have activated and deactivated different modules of the system to see the contribution of

Table 6. Performance test results

Dataset	Polar Type	Accuracy (%)	Average (%)	Precision (%)	Recall (%)	F1-score (%)	Average (%)
Dataset-1	Positive	85.3	88.2	98.6	77.2	86.6	82.1
	Negative	96.7		97.9	83.6	90.2	
	Neutral	82.7		54.1	96.7	69.4	
Dataset-2	Positive	85.7	87.7	92.9	70.5	80.2	81.1
	Negative	95.9		88.6	73.8	80.5	
	Neutral	81.6		74.5	93.1	82.8	
Dataset-3	Positive	85.1	87.8	68.7	80.3	74.0	78.3
	Negative	96.6		75.0	75.0	75.0	
	Neutral	81.8		89.0	83.0	85.9	

modules to the performance of sentiment analysis and obtained test results are submitted in Table 7. At the baseline, we calculate polarities only taking into account the polarity values taken from the lexicon and achieve 62.8% accuracy for the first dataset. As the next test case, the normalizer and word-level analyzer modules are included and around 10% improvement is achieved for the first dataset as seen in the table. The contribution of the word group-level analysis is evaluated in another test, where the word-level analyzer module is also activated in addition to the previous modules. This module has provided about 15% improvement on the first dataset compared to the previous test case. In another test case, the idiom/proverb analyzer is activated, but the word group-level analyzer is deactivated. Using the idiom/proverb analyzer directly on top of the word-level analyzer has improved the performance in a limited amount, for example, 0.7 for the first dataset. In the last row of Table 7, the performance results of the full system are shown. Consequently, each module of the system has a reasonable impact on improving the overall accuracy of the system, but by far, the most significant improvement occurs when the word-group level analyzer is activated which increases the accuracy by more than 12% for all the datasets.

Table 7. Contributions of modules

Modules	Dataset-1 (%)	Dataset-2 (%)	Dataset-3 (%)
Baseline (only lexicon)	62.8	62.4	66.1
+ normalize + word-level analyzer	72.1	74.2	72.6
+ normalize + word-level analyzer + word group-level analyzer	87.5	86.1	87.7
+ normalize + word-level analyzer + idiom/proverb analyzer	72.8	75.8	72.7
Full system	88.2	87.7	87.8

4.2. Evaluation

In the previous studies on Turkish sentiment analysis, the highest accuracy with ML has been achieved by Meral and Diri [22] with 90% (ternary classification) and Türkmenoğlu and Tantuğ [6] with 85% (binary classification) on Twitter datasets, wheras in the few studies that implemented the LB approach, the highest accuracy was again obtained by Türkmenoğlu and Tantuğ [6] with 75.2% on the Twitter dataset and 79% on the Movie dataset for binary classification. Consequently, the 88.2% accuracy achieved in this study for ternary classification is significantly better than the previously developed LB systems and not far from to the ML implementations.

Interestingly, after comparing machine learning and lexical based methods for Turkish, Türkmenoğlu and Tantuğ have concluded that LB sentiment analysis is more preferable due to its unsupervised and domain free nature [6]. In addition, Ravi and Ravi state that machine learning systems yield maximum accuracy while lexicon-based systems provide better generality [2]. Therefore, LB sentiment analysis has some advantages and may be preferred in some cases.

In their study for the LB approach, Türkmenoğlu and Tantuğ have developed a lexicon which was translated from English [6]. In addition, the level of morphological analysis and multiword extraction implemented in that study are not as detailed as our analysis which covers more constructs specific to Turkish, especially at the word-group and idiom/proverb levels. Vural et al., on the other hand, present a lexicon-based sentiment analysis framework which uses the translation of the SentiStrength lexicon to Turkish which may cause losses of meaning, thereby resulting in incorrect polarity of the translated words [5]. Because our system takes into account the special features of the Turkish language in more detail with a lexicon specially formed for Turkish and introduces a novel approach to focus on the word-groups rather than the words in a sentence, the results of our system show more than 9% increase in accuracy when compared to the results of the previous LB studies. Notice that the test results of the studies given above are not directly comparable with our results as the datasets are different. However, the authors believe that the LB polarity calculation approach and the lexicon structure introduced in this study are more suitable for the Turkish language, consequently achieving better results.

5. Conclusion

In this paper, a polarity determination and calculation method for lexicon-based sentiment analysis is developed and tested on Turkish tweets. The datasets are examined at three levels; namely, word, word group and idiom/proverb levels in order to reach the correct sense behind each tweet. At the word level, following the normalization step, the words are parsed into root and suffixes. Within the scope of the analysis, the interaction of the roots and suffixes according to the negators, presence - absence of suffixes and POS tags are determined. At the word group-level stage, word groups are analyzed with the representation of special circumstances present in the Turkish language. Lastly, at the idiom/proverb level the idioms and proverbs are examined as a whole. The developed system is tested on three datasets with the help of Twitter API, and an accuracy of 88.2% is observed which is the highest accuracy level achieved compared to the previous LB sentiment analysis on Turkish texts. Furthermore, we have also tested the contribution of the word group-level and idiom/proverb analyses. The results show that the success metric values increase considerably, proving that the lexicon-based analysis of Turkish texts at word, word group and idiom/proverb levels improves the system performance with the word-group level module being the most effective.

The authors believe that LB sentiment analysis is important for Turkish sentiment analysis since it is more effective in the implementation of specific rules to the language, analysis of the words that signify different meanings according to the used sentences, and determination of the relationship of meanings between the words. The lexicon is extremely important in LB analysis and it may affect the overall results substantially. Therefore, in such studies, it is important to pay careful attention to the creation of the lexicon.

As future work, in order to improve the performance of the proposed method, the lexicon can be expanded to cover more domains or existing polarity databases such as SentiTurkNet [7] can be used. In addition, the identification of the words that have different meanings in different contexts and their integration into the analysis stage can be considered as another future attempt. The various symbols, emojis, etc. supported by Twitter can be incorporated into the analysis to achieve better results. Furthermore, the proposed LB method for Turkish texts can be utilized for datasets retrieved from other social media platforms such as news sites, LinkedIn and Youtube reviews. Finally, this method can be integrated with ML-based methods to improve the overall performance of sentiment analysis systems.

Acknowledgment

We thank Assoc.Prof.Dr. Murat Küçük for his valuable support related with the Turkish language.

References

[1] Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology 2010; 61: 2544–2558.

[2] Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems 2015; 89: 14–46.

[3] Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal 2014; 5: 1093–1113.

[4] Mudinas A, Zhang D, Levene M. 2012. Combining lexicon and learning based approaches for concept-level sentiment analysis. In: the First Int. Workshop on Issues of Sentiment Discovery and Opinion Mining; 12-16 August 2012; Beijing, China: ACM. pp.1-8.

[5] Vural AG, Cambazoglu BB, Senkul P, Tokgoz ZO. A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish. Computer and Information Sciences 2012; 3: 437–445.

[6] Türkmenoğlu C, Tantuğ AC. Sentiment analysis in Turkish media. In: Workshop on Issues of Sentiment Discovery and Opinion Mining; 25 June 2014; Beijing, China: IMLS. pp.1-11.

[7] Dehkharghani R, Saygin Y, Yanikoglu B, Oflazer K. SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. Language Resources and Evaluation 2016; 50: 667–685.

[8] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In: the Conference on Empirical Methods in Natural Language Processing (EMNLP); July 2002; Philadelphia, USA: ACL. pp. 79-86.

[9] Bai X. Predicting consumer sentiments from online text. Decision Support Systems 2011; 50: 732–742.

[10] Saleh MR, Martín-Valdivia MT, Montejo-Ráez A, Ureña-López LA. Experiments with SVM to classify opinions in different domains. Expert Systems with Applications 2011; 3: 14799–14804.

[11] Zhang Z, Ye Q, Zhang Z, Li Y. Sentiment classification of Internet restaurant reviews written in Cantonese. Expert Systems with Applications 2011; 38: 7674–7682.

[12] Lochter JV, Zanetti RF, Reller D. Short text opinion detection using ensemble of classifiers and semantic indexing. Expert Systems with Applications 2016; 62: 243-249.

[13] Kauer AU, Moreira VP. Using information retrieval for sentiment polarity prediction. Expert Systems with Applications 2016; 61: 282-289.

[14] Turney PD. 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; 07-12 July 2002; Philadelphia, Pennsylvania, USA: ACL. pp. 417-424.

[15] Moreo A, Romero M, Castro JL, Zurita JM. Lexicon-based comments-oriented news sentiment analyzer system. Expert Systems with Applications 2012; 39: 66–80.

[16] Fernandez-Gavilanes M, Alvarez-Lopez T, Juncal-Martinez J, Costa-Montenegro E, González-Castaño FJ. Unsupervised method for sentiment analysis in online texts. Expert Systems with Applications 2016; 58: 57-75.

[17] Kaya M, Fidan G, Toroslu IH. Sentiment analysis of Turkish political news. In: IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology; 4-7 Dec. 2012; Macau, China: IEEE. pp. 174-180.

[18] Çetin M, Amasyalı MF. Supervised and traditional term weighting methods for sentiment analysis. In: 21st Signal Processing and Communications Applications Conference (SIU); 24-26 April 2013; Haspolat, Turkey: IEEE. pp.1-4.

[19] Balahur A, Turchi M, Steinberger R, Perea-Ortega JM, Jacquet G et al. Resource creation and evaluation for multilingual sentiment analysis in social media texts. In: the 9th edition of the Language Resources and Evaluation Conference (LREC); 26-31 May 2014; Reykjavik, Iceland: ELRA. pp. 4265-4269.

[20] Yıldırım E, Çetin FS, Eryiğit G, Temel T. The impact of NLP on Turkish sentiment analysis. TBV Journal of Computer Science and Engineering; 2014; 7: 43-51.

[21] Akba F, Uçan A, Sezer EA, Sever H. Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews. In: the 8th European Conference on Data Mining; 15-17 July 2014; Lisbon, Portugal: pp.180-184.

[22] Meral M, Diri B. Sentiment analysis on Twitter. In: 22nd Signal Processing and Communications Applications Conference; 23-25 April 2014; Trabzon, Turkey: IEEE. pp.690-693.

[23] Çoban O, Özyer B, Özyer GT. Sentiment analysis for Turkish Twitter feeds. In: 23rd Signal Processing and Communications Applications Conference; 16-19 May 2015; Malatya, Turkey: IEEE. pp. 2388-2391.

[24] Türkmen H, Omurca İO. An empirical study for Turkish sentiment analysis by machine learning methods. In: International Conference on Advanced Technology and Sciences; 12-15 August 2014; Antalya, Turkey: pp. 589-592.

[25] Akgül ES, Ertano C, Diri B. Twitter verileri ile duygu analizi. Pamukkale University Journal of Engineering Sciences 2016; 22: 106-110.

[26] Dehkharghani R, Yanikoglu B, Saygin Y, Oflazer K. Sentiment analysis in Turkish at different granularity levels. Natural Language Engineering 2017; 23: 535-559.

[27] Parlar T, Özel SA. A new feature selection method for sentiment analysis of Turkish reviews. In: International Symposium on Innovations in Intelligent SysTems and Applications (INISTA); 2-5 August 2016; Sinaia, Romania.

[28] Sağlam F, Sever H, Genç B. Developing Turkish sentiment lexicon for sentiment analysis using online news media. In: IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), 29 Nov.-2 Dec. 2016; Agadir, Morocco: IEEE. pp.1-5.

[29] Omurca İO, Ekinci E, Türkmen H. An annotated corpus for Turkish sentiment analysis at sentence level. In: International Artificial Intelligence and Data Processing Symposium (IDAP); 16-17 Sept. 2017; Malatya, Turkey: IEEE. pp.1-5.

[30] Türkmen H, Omurca Sİ, Ekinci E. An aspect based sentiment analysis on Turkish hotel reviews. In: International Symposium on Engineering, Artificial Intelligence and Applications; 2015; Girne. pp. 9-15.

[31] Kama B, Öztürk M, Karagöz P, Toroslu IH, Kalender M. Analyzing implicit aspects and aspect dependent sentiment polarity for aspect-based sentiment analysis on informal Turkish texts. In: Proceedings of the 9th International Conference on Management of Digital EcoSystems; 07-10 November 2017; Bangkok, Thailand: ACM. pp.134-141.

[32] Akın AA, Akın MD. Zemberek, an open source NLP framework for Turkic Languages. Structure 2007; 10: 1-5.

[33] İlhan N, Kabadayı C. Türk Dilinde Olumsuzluk. İstanbul, Turkey: Kesit Yayınları, 2017 (in Turkish).