IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

An Idiom Based Approach To Sentiment Analysis In Turkish

[1]Özgün KOŞANER, [2]Harun EROĞLU
[12]Department of Linguistics, Dokuz Eylül University
İzmir, Turkey

Abstract: In this study, a model was developed to perform sentiment analysis of frequently used Turkish idioms. The main purpose of the study is to analyze the sentiments of idioms categorized as positive, negative, and neutral, thus examining these linguistically and culturally rich expressions using natural language processing (NLP) techniques. The dataset consists of 147 idioms randomly selected from Püsküllüoğlu's Turkish Idioms Dictionary. Sentiment labels were assigned to these idioms manually, and a model was created and trained using TF-IDF and Random Forest algorithms. The model's accuracy was determined to be 61%, with high performance observed particularly in the positive and negative sentiment classes. This study serves as a first step towards addressing the lack of corpora for sentiment labeling of idioms in Turkish. It is expected that this foundational work will be enhanced with larger datasets and more advanced models in the future, thereby expanding its scope.

***Index Terms* -** Idiom, Sentiment Analysis, Sentiment Labeling, Natural Language Processing, Corpus Linguistics**.**

I. INTRODUCTION

Corpus Linguistics is the tagging, compilation and analysis of spoken and written texts that describe the structure and use of languages. The generated corpora enable the application of advanced natural language processing (NLP) techniques such as sentiment analysis. Sentiment analysis has become a very important field in NLP and has achieved high accuracy rates with many studies in the past and present. In fact, while the average percentage of accuracy that indicates that a study is acceptable for certain applications in Computer Science is at least 60% [1], today, even in unlabeled data, programs such as VADER can reach over 70% accuracy, provided that the inputs are high. On the other hand, although it is easily possible to exceed even these percentages, the data set to be studied should have thousands of inputs if possible; only in this way will the sentiment analysis model utilized be able to work with high accuracy.

Sentiment analysis studies are generally applied on written texts, with the purpose as diverse as analyzing the emotion of comments on a product on sale on an e-commerce site or analyzing review comments on a movie. In doing so, the analysis can be carried out on a pre-created data set, or one can create one's own data set. The generated datasets are trained on a predetermined sentiment analysis model such as BERT, Word2vec or Scikit-learn, thus adding to the model.

The words, emotions of which are annotated in the datasets are labeled according to the emotion and meaning they evoke in natural language. For example, in the sentence "I love her.", the word "love" is labeled as "positive" in the dataset since it reflects a positive emotion, and in the sentence "I don't love her.", the word "don't love" is labeled as "negative" in the dataset since it evokes a negative emotion. On the other hand, the emotions evoked by complex expressions that are specific to language and culture, such as idioms and proverbs, can also be done by consulting natural language speakers depending on the scope of the study. However, it would be more appropriate to apply such methods in studies where not only negative and positive but also complex emotions such as fear, happiness and sadness are labeled.

In particular, for English, there are already very advanced studies and lists of idioms (e.g. Potential Idiomatic Expression - PIE). For Turkish, on the other hand, although many sentiment studies have been conducted, it is not possible to say the same for idioms. Therefore, the aim of this study is to label idioms in order to improve sentiment analysis in Turkish. Frequently used idioms in Turkish are labeled in 3 different categories as "positive", "negative" and "neutral/objective" and since the scope of the study is not very large, the relevant model is trained on Scikit-learn, TF-IDF and Random Forest instead of more complicated programs such as Word2vec etc. In addition, since the dataset consists of 147 inputs, the goal was to achieve an accuracy of up to 60%. Therefore, the aim with this small-scale dataset is to increase the number of inputs in the long run, which will directly increase the accuracy rate.

The software used in the study, how the sentiment categories of idioms were determined, and the outputs and relevant tables will be discussed in detail in the methodology, and findings and discussion sections.

II. Literature Review

Idioms are one of the richest elements of language, adding depth to natural language in a culturally specific way. Linguists and other experts have developed many classifications and methods to understand this complex nature of idioms. These include understanding the linguistic structure of idioms as well as examining their cultural origins.

According to common definitions, idioms are linguistic expressions whose meaning is structurally unpredictable [2]. The main reason for this definition of idioms is that idioms are culture-specific. In addition, as we will discuss later, there are also neutral idioms whose meaning is neither positive nor negative, in other words, although their meaning does not change, they manipulate the perception of the receiver according to the context. For example, while the idiom "to be speechless" clearly means "to be overjoyed and excited" (positive emotion), and "to find absurd" means "to regard (something) as nonsense" (negative emotion), the idiom "to be speechless" represents a neutral emotion because its meaning can change depending on the context. Thus, when someone who is expected to speak but does not speak starts to speak, the idiom will be closer to a positive emotion, whereas when someone who normally speaks little and is expected to do so speaks a lot in an inappropriate environment, the idiom will be closer to a negative emotion.

Linguistically, there are 3 different idiom types: Grammatical, morphological and semantic. These are, respectively, the degree of recognition, the formal complexity of the structure, and the figurative qualities of the meaning. Although these principles can be applied to idioms themselves and other constructions containing idioms, the distinguishing features are not always clear. For this reason, as Langlotz pointed out in his 2006 study, the important thing is not to define idioms precisely, but rather to establish the degree of idiomaticity [3]. In light of this, the following definition of idiomatic constructions can be made: An idiom is a compromised structure that consists of at least two lexical units and has the complex structure of a phrase that may contain structural idiosyncrasies [3].

Idioms and idiomaticity are actually quite complicated, and following some definitions given in this framework, sub-types of idioms can also be mentioned. In other words, since there is no clear definition of idioms by experts, it would be appropriate to examine and exemplify idioms in terms of sub-types.

Makkai mentioned that idioms can basically be analyzed in 2 subdivisions. These are lexical and semantic idioms [4]. Lexical idioms consist of categories such as "phrasal verbs" and "phrasal compounds". Examples of these are "make up" and "black mail" respectively. In addition, there are categories such as "exaggerations" and "greetings" under the umbrella of semantic verbs, and the example of "He won't even lift a finger." can be given for an exaggeration, while the example of "How do you do?" can be given for a greeting.

All of Makkai's work has paved the way for much more advanced idiomatization based on semantics, syntax and function. The time series monthly data is collected on stock prices for sample firms and relative macroeconomic variables for the period of 5 years. The data collection period is ranging from January 2010 to Dec 2014. Monthly prices of KSE -100 Index is taken from yahoo finance [5].

We have already mentioned that there are different approaches to the definition of idioms and how they should be classified. It would not be wrong to say that idioms, which can be considered as a field of study in linguistics, are in a closed development period. Although today, the Turkish Language Institution (TDK) mostly utilizes the collections of proverbs and idioms, many works have been published on idioms in Turkish since the past. Already in the first quarter of the Republican period, names such as M. Esat İleri, M. Nihat Özön and Ö. Asım Aksoy made serious developments on behalf of the phenomenon of meaning in idioms, and Mehmet Ali Ağakay's Dictionary of Metaphors in Turkish in 1949 took this to the next level [5].

More recent research on idioms include Alper Yıldırım's (2015) New Dictionary of Proverbs and Idioms and Ali Püsküllüoğlu's Dictionary of Turkish Idioms [6]. Yıldırım's work is a dictionary that can be productive even for those who are educated at primary school level with its full explanations and examples, while Püsküllüoğlu's work draws attention with the fact that it contains the meanings and examples of a great number of idioms, and being one of the most frequently referenced sources in the literature [7].

2.1Sentiment Analysis

Sentiment analysis is a relatively new application that is carried out with NLP techniques through various programs in the computer environment. From the past to the present, many corpora have been created for sentiment analysis and these have become applications in fields such as AI. Especially corpus studies for English have made sentiment analysis algorithms for English quite comprehensive. One of them is Potential Idiomatic Expression (PIE). In the corpus in which this study was conducted, situations such as simile, personification, exaggeration, and irony were labeled and an algorithm that works with an accuracy rate of approximately 90% was developed [8].

The studies are not limited to emotion classifications, but have even been used to measure the emotion of individuals in global events such as COVID-19 [9].

2.2Sentiment Analysis in Turkish

Although some sentiment analyses have already been conducted for Turkish for satisfaction or other emotions by analyzing user comments, a comprehensive sentiment dictionary has not been created in a general framework until recent years. The study published by Kaya et al. in 2012 is one of the first representatives of sentiment dictionary construction for Turkish [10]. In the study, a four-supervised machine learning algorithm consisting of Naive Bayes, Maximum Entropy, SVM and n-Gram Language Model was developed for sentiment classification of Turkish political columns. The study has an accuracy rate of 65% to 77%, which can be considered as an efficient study for a difficult language to analyze such as Turkish [10].

In addition, a model that assigns positive and negative labels to texts in Turkish blogs in order to provide an overview of products and services and a mood study in this framework can also be mentioned. This model also uses semi-supervised learning based on the Naive Bayes approach. The training set consists of Turkish words indicating emotion. Accuracy obtained in this study in different situations is up to 84% [11].

Corpora for Turkish, which is difficult to analyze from a morphological point of view, are also limited. Nevertheless, in addition to SentiTurkNet, other important corpora for Turkish have been created. One of them is Zemberek, introduced by Akın in 2007 [12]. This NLP tool is an important model for morphological analysis, synonymy, etc.

The most recent corpus study is the Turkish National Corpus (TUD). One of the largest corpora for Turkish, TUD is a general purpose corpus with 50 million words. Thanks to this corpus, users can perform queries on topics such as author gender, text type, and media [13].

Another corpus is the METU Turkish Corpus. Consisting of 2 million words, the corpus covers 10 different text types, each consisting of 2 thousand words [14].

Finally, SentiTurkNet is one of the prominent sentiment tagging studies. This study includes a general corpus that shows the positive, negative and neutral levels of each unit in the Turkish wordNet and can be applied to other languages [15].

Apart from these, there are other corpora such as TurCo and BOUN Corpus, but these are corpora without an online interface [13].

III. Method and Material

3.1Method

In line with the aim of the study, in order to improve the mood analysis in Turkish, frequently used idioms in Turkish were labeled in 3 different poles as "positive", "negative" and "neutral". The programs used in the study, how the mood poles of the idioms were determined, the software applications used, and the findings and their tables will be shared in detail.

3.2Material

In this study, Püsküllüoğlu's Dictionary of Turkish Idioms was utilized as it is one of the most frequently used dictionaries prepared for idioms in Turkish as well as being the most up-to-date one.

This dictionary contains nearly 14 thousand idioms. Compared to other dictionaries published in Turkish, the dictionary contains more units and each idiom is exemplified. This made it easier to determine their meanings and connotations when assigning poles such as positive and negative to idioms.

3.3Sample

As mentioned before, idioms that are not agreed upon have more than one category, and for this study, certain restrictions were imposed in order to select the idioms to be included in the limited data set to be created. Among the idioms in Püskülüoğlu's Dictionary of Turkish Idioms, only the semantic ones were evaluated and classified as positive, negative, neutral among the sub-categories defined by Makkai for idioms. In order not to manipulate the model to be trained, these were randomly selected to ChatGPT by random sampling method. In addition, the selected phrases

Before the labeling was done manually, the existing definitions were compared through the internet-based TDK dictionary. The 147 idioms that make up the database of the study were categorized as positive (37), negative (55) and neutral (55). Ten percent of the included idioms are presented with their labels in the table below.

Table 3.1. 10% of the Idioms Composing the Database

Bahtı açık olmak	Having good fortune	Positive
Göz kamaştırmak	Dazzle	Positive
Havalara uçmak	Be on cloud nine	Positive
Köşeyi dönmek	Striking it rich	Positive
Perileri bağdaşmak	Come to terms	Positive
Başının etini yemek	Niggle at	Negative
Çamur atmak	Point the bone at someone	Negative
Eceline susamak	Get a death wish	Negative
Hapı yutmak	Be done for	Negative
Kazık yemek	Get ripped off	Negative
Akıl vermek	Give advice	Neutral
Çağı geçmek	Be out of date	Neutral
Ikınıp sıkınmak	Grunt and strain	Neutral
İpe dizmek	String	Neutral
Kulağını çınlatmak	Make someone’s ear ring	Neutral

Since the related idioms are few in number, instead of presenting them on an additional page, all of them are included in the code under the heading "coding stages" so that they can be examined within the code.

3.4.Coding

The phrases tagged in Python were trained and run on TF-IDF and Random Forest for mood analysis and it was observed that the targeted 60% accuracy rate was achieved despite having a very limited data set. Before describing the full code, the steps can be summarized step by step as follows:

i. As the first stage of the coding process, the pandas library was loaded, sklearn modules were called for operations such as adding/removing features and calculating correctness, and re modules were called for regular expressions.
ii. Idioms were written according to their categories.
iii. The process of separating the idioms in the dataset according to their categories was performed.
iv. Minority classes were replicated.
v. A balanced data set was created.
vi. Text cleaning functions were written.
vii. The cleaning process was performed.
viii. Feature extraction was performed with TF-IDF method.
ix. Separation of training and test sets was performed.
x. The model was trained with Random Forest.
xi. A prediction function was written on the test set.
xii. Performance evaluation function was written.
xiii. Functions to print the output to the screen have been written.

The code in question is as presented below:

```
pip install pandas scikit-learn
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, accuracy_score import re
from sklearn.utils import resample


positive_idioms = ["bahtı açık olmak", "baş tacı etmek", "başı göğe ermek", "can ciğer kuzu sarması", "canına minnet", "çiçeği burnunda", "dört elle sarılmak", "esen kalmak", "etekleri uzamak", "el üstünde tutmak", "gül gibi bakmak", "göklere çıkarmak", "göğüs germek", "göz kamaştırmak", "göğsü kabarmak", "gözleri parlamak", "havalara uçmak", "iki dirhem bir çekirdek", "kapağı atmak", "kısmeti açılmak", "köşeyi dönmek", "mürüvvetini görmek", "omuzda taşımak", "perileri bağdaşmak", "rayına oturtmak", "rüya gibi", "rüzgar gelecek delikleri tıkamak", "sağlam kazığa bağlamak", "sevincinden ağzı kulaklarına varmak", "sekiz köşe olmak", "sevinçten uçmak", "sırt sırta vermek", "tatlıya bağlamak", "üstüne titremek", "yüzüne kan gelmek", "yüreği ferahlamak", "yüzü gülmek"]


negative_idioms = ["abuk sabuk konuşmak", "başının etini yemek", "başına kakmak", "başı çatlamak", "belaya çatmak", "beyin yıkamak", "boğazı düğümlenmek", "boyunun ölçüsünü almak", "buluttan ne kapmak", "burun kıvırmak", "can çekişmek", "canından bezmek", "cayırtı koparmak", "cehennemin dibine gitmek", "cin tutmak", "curcunaya çevirmek", "çıkmaza girmek", "çamur atmak", "çile çekmek", "dalavere çevirmek", "dara düşmek", "deli etmek", "diş geçirememek", "eceline susamak", "elini kana bulamak", "hapı yutmak", "hor kullanmak", "hurdaya çevirmek", "hafife almak", "içi kan ağlamak", "içini kemirmek", "içi içini yemek", "inim inim inlemek", "ipiyle kuyuya inilmez", "kabak tadı vermek", "kaçacak delik aramak", "kara kara düşünmek", "karizmayı çizdirmek", "kasıp kavurmak", "kazık yemek", "kimi kimsesi olmamak", "koynunda yılan beslemek", "kötü yola sapmak", "kuyusunu kazmak", "küçük düşmek", "küplere binmek", "nevri dönmek", "ödü kopmak", "öküz gibi bakmak", "pestilini çıkarmak", "şeytana uymak", "tüy dikmek", "yıldırımları üstüne çekmek", "zokayı yutmak"]


neutral_idioms = ["abayı yakmak", "abanoz kesilmek", "abdesti gelmek", "abluka altına almak", "akıl vermek", "bel bağlamak", "bin dereden su getirmek", "boyun eğmek", "büyük söz söylemek", "bülbül gibi konuşmak", "çağı geçmek", "cephe almak", "çenesi düşmek", "dalıp gitmek", "deli divane olmak", "dili tutulmak", "el ele vermek", "ezilip büzülmek", "gazel okumak", "geri basmak", "gıkı çıkmamak", "gır gır geçmek", "girecek delik aramak", "göbek bağlamak", "gönül eğlendirmek", "göz kulak olmak", "hop oturup hop kalkmak", "ışığını cıcığını çıkarmak", "ıkınıp sıkınmak", "ısıtıp ısıtıp önüne koymak", "ırgat pazarına döndürmek", "icabına bakmak", "içi geçmek", "ikili oynamak", "iğne atsan yere düşmez", "ipe dizmek", "iş başa düşmek", "işkembesini şişirmek", "jilet gibi", "kabasını almak", "kaşık atmak", "karnı zil çalmak", "kırkı çıkmak", "kıtlıktan çıkmış gibi yemek", "köşesine çekilmek", "köstek vurmak", "kulağını çınlatmak", "kuş uçurmamak", "laçka olmak", "merdiven dayamak", "silip süpürmek", "soluk aldırmamak", "şakakları ağarmak", "tabanları yağlamak"]

data = {'Idiom': positive_idioms + negative_idioms + neutral_idioms, 'Sentiment': ['Positive'] * len(positive_idioms) + ['Negative'] * len(negative_idioms) + ['Neutral'] * len(neutral_idioms)}

df = pd.DataFrame(data)
df_positive = df[df['Sentiment'] == 'Positive'] df_negative = df[df['Sentiment'] == 'Negative'] df_neutral = df[df['Sentiment'] == 'Neutral']
```

```
df_neutral_upsampled = resample(df_neutral, replace=True, n_samples=len(df_positive),
random_state=42) df_negative_upsampled = resample(df_negative, replace=True,
n_samples=len(df_positive), random_state=42)

df_balanced = pd.concat([df_positive, df_neutral_upsampled, df_negative_upsampled])

def temizle(text):
    text = re.sub(r'\d+', '', text)  # Sayıları kaldır
    text = re.sub(r'\W', ' ', text) # Özel karakterleri kaldır text = re.sub(r'\s+', ' ', text) # Ekstra
    boşlukları kaldır text = text.lower()  # Küçük harfe çevir
    return text

df_balanced['cleaned_idiom'] = df_balanced['Idiom'].apply(temizle)

vectorizer=TfidfVectorizer(max_features=5000,ngram_range=(1,2))X
= vectorizer.fit_transform(df_balanced['cleaned_idiom'])
y = df_balanced['Sentiment']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

rf_model_balanced = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model_balanced.fit(X_train, y_train) y_test_pred_balanced = rf_model_balanced.predict(X_test)
test_accuracy_balanced = accuracy_score(y_test, y_test_pred_balanced) classification_rep_balanced
= classification_report(y_test, y_test_pred_balanced)

print(f"Doğruluk Oranı: {test_accuracy_balanced}") print("Sınıflandırma Raporu:")
print(classification_rep_balanced)
```

The explanations of the relevant codes are briefly mentioned here and the entire code is included. Details such as the accuracy of the model will be discussed in the findings and discussion section.

IV. RESULTS AND DISCUSSION

In this study, a model is developed to perform sentiment analysis of Turkish idioms. The dataset used includes 148 randomly selected idioms from Püsküllüoğlu's work and these idioms are labeled as positive, negative and neutral. Model training and testing phases were performed using TF-IDF feature extraction method and Random Forest algorithm.

The accuracy rate obtained by the model on the test set and the results in the classification report are presented in the table below.

Table 4.1: Accuracy Rate and Classification report

Accuracy	%6086956521739131		
	Precision	Recall	F1-Score
Positive Idioms	%50	%88	%64
Negative Idioms	%71	%71	%71
Neutral Idioms	%100	%25	%40

These results show that the model performs relatively well in the positive and negative emotion classes; however, the performance is lower in the neutral class. The overall accuracy of the model is around 60%, which can be considered as a reasonable performance in the targeted context.

When the accuracy and classification results of the model are analyzed, several important findings stand out. First, the higher performance of the model in positive and negative emotion classes is due to the fact that these classes have distinctive features.

On the other hand, the fact that the neutral class has wider and more diverse features can be considered as one of the details that reduce the performance of the model in this class. The limitation and partial imbalance in the data set is an important factor affecting the performance of the model. In this study, the method of replication of minority classes was used to overcome the data imbalance. However, this method may not always yield the desired results. Therefore, using larger and more balanced data sets for future studies will improve the overall performance of the model.

Furthermore, different feature extraction techniques (e.g. word embeddings) and more complex models (e.g. deep learning based models) can be used to improve the performance of the model. The TF-IDF and Random Forest method used in this study provides a basic approach and has some limitations compared to more advanced techniques.

In this study, a model for sentiment analysis of Turkish idioms is developed and evaluated. The dataset used is 147 idioms selected from Püsküllüoğlu's Turkish Idioms Dictionary and labeled as positive, negative and neutral. While labeling, the definitions of the idioms were additionally compared with TDK's web-based Dictionary of Proverbs and Sayings. Model training and testing phases were performed using TF-IDF feature extraction method and Random Forest algorithm.

The accuracy rate of the model on the test set was calculated as approximately 61% despite the limited data set and the classification report results were analyzed. While the model performs well in the positive and negative classes, the F-1 score is relatively low in the neutral class. This is due to the fact that neutral idioms have a wider and more diverse set of features.

There are many methods that can be applied to increase the accuracy capacity of the study. For example, the current dataset contains a limited number of phrases; increasing the number of phrases and creating a larger dataset with different sentiment tones will directly increase the generalization capability of the model. In addition, more advanced feature extraction techniques such as Word2Vec, BERT, and models such as LSTM will improve the semantic representation of phrases and the texts they form together. In addition, the developed model can be applied in different domains such as social media analysis, customer feedback, etc. to increase the effectiveness of the model in practice and thus the model can be applied more easily in real life. However, in order to obtain healthy outputs as a result of the application of these models and practices, expanding the scope of the data set should be a priority.

In conclusion, this study has taken an important step forward in realizing sentiment analysis of Turkish idioms. In fact, since there is no idiom-specifically labeled dataset for Turkish, the study provides an initial dataset for future mood analysis models. In addition to the relevant dataset, the results show that the model is functional at a certain level and that further research and development is needed in the field of Turkish NDI. Future studies and labeling can further improve the performance of the model and enable it to have a wider and more diverse range of applications.

.

References

[1] Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., ve Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS digital health. 2(2): 8.

[2] Kövecses, Z. (1996). The scope of metaphor. Applied Linguistics. 17(3): 326.

[3] Langlotz, A. (2006). Idiomatic Creativity. A cognitive-linguistic model of idiom representation and idiom variation in English. Amsterdam: John Benjamin.

[4] Makkai, A. (1972). Idiom structure in English. The Hague: Mouton.

[5] Tan, M. (2020). Türkiye’de Genel Atasözü ve Deyim Sözlüklerinde Anlam Verme Çalışmaları Toplu Bir Bakış. Türk Dünyası Dil ve Edebiyat Dergisi. (50): 257.

[6] Püsküllüoğlu, A. (2017). Türkçe Deyimler Sözlüğü. Ankara, Türkiye: Arkadaş Yayınevi.

[7] Yozgat, U. (2023). Bilgisayarlı Dilbilim Çerçevesinde Deyimlerin Bilişsel Yöneltimli Etiketlenmesi. [Unpublished PhD thesis]. Dokuz Eylül Üniversitesi.

[8] Adewumi, T., Chang, S. ve Ma, X. (2022). Potential Idiomatic Expression (PIE)-English: Corpus for Classes of Idioms. Proceedings of the 13th Conference on Language Resources and Evaluation. (2): 689.

[9] Gupta, V., Jain, N., Katariya, P., Kumar, A., Mohan, S., Ahmadian, A. ve Ferrara, M. (2021). An Emotion Care Model using Multimodal Textual Analysis on COVID-19. Chaos, Solitons & Fractals. (144): 1-9.

[10] Kaya, M., Fidan, G. ve Toroslu, I. H. (2012). Sentiment analysis of Turkish political news. 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology (ss. 174-180). Düzenleyen IEEE Computer Society. Washington. 4-7 Aralık 2012.

[11] Aytekin, C. (2013). An opinion mining task in Turkish language: A model for assigning opinions in Turkish blogs to the polarities. Journalism and Mass Communication. 3(3): 179-198.

[12] Akın, A. A., & Akın, M. D. (2007). Zemberek, an open source NLP framework for Turkic languages. Structure. (10): 1-5.

[13] Aksan, Y. ve Demirhan, U. (2013). Türkçe Ulusal Derlemi (TUD) arayüz özellikleri: Tanıtım ve uygulama. 27. Ulusal Dilbilim Kurultayı Bildirileri (ss. 280-281). Düzenleyen Hacettepe Üniversitesi. Ankara. 2-4 Mayıs 2013.

[14] Say, B., Zeyrek, D., Oflazer, K. ve Özge, U. (2002). Development of a corpus and a treebank for present-day written Turkish. Current research in Turkish linguistics: Proceedings of the 11th International Conference of Turkish Linguistics (ss. 4-5). Eastern Mediterranean University. Magusa. 2002.

[15] Dehkharghani, R., Şahin, M., Kılıç, Ö. ve Yıldırım, S. (2016). SentiTurkNet: A Turkish polarity lexicon for sentiment analysis. Computer Speech & Language. (40): 1-19.