

# Use of Natural Language Processing Methods in Teaching Turkish Proverbs and Idioms

Ertürk ERDAĞI

School Principal, Republic of Türkiye Ministry of National Education, İstanbul, Türkiye

**Abstract**—In this study, a series of studies are proposed for easy learning of proverbs and idioms in the language. In Turkish, proverbs and idioms are structures that are used both in the academic environment and in their daily lives, especially by 10-year-old students who have entered the abstract thinking stage. Since this structure contains abstract expressions, it seems difficult to learn at first. In the study, 2396 proverbs and 11209 idioms in the online dictionary of the Turkish Language Association were used. A pre-test was conducted to measure the knowledge level of 20 students selected as the study group. The structure of idioms and proverbs was analyzed using Natural Language Processing methods. With the analysis, difficulty groups were divided according to information such as word count, n-gram analysis, frequency level, and the student was asked questions from the online question pool for the tutorial and the test during the process. Generative artificial intelligence enables semantic analysis of texts containing idioms and proverbs. Following the studies, a test was applied to the students and the efficiency of the process was tried to be measured. As a result, students' idiom knowledge increased by 51.8% and proverb knowledge increased by 59.40%.

**Keywords**—Idiom; proverb; natural language processing; word frequency; n-gram analysis; contextual analysis

## I. INTRODUCTION

Proverbs and idioms are linguistic treasures that are passed down from generation to generation by combining the accumulation and experience gained over centuries with the observations and life practices of the people. These expressions are an important part of cultural heritage. Social values, ethical teachings and deep knowledge about various aspects of life are reflected in these expressions. Proverbs and idioms are the richest and most colorful elements of language. They are frequently used in everyday life and strengthen communication.

Proverbs have often originated anonymously. They are expressed in a short and concise way and usually convey a universal truth. They are often based on experience and reflect the common values of society [1]. Proverbs allow to express a lot of things with few words in language. The desired expressions are conveyed in a powerful way. Proverbs are a didactic element. With these words, advice is given, a warning is given and guidance is given. Proverbs provide individuals with clues about concepts such as right and wrong, good and bad, beautiful and ugly. At the same time, proverbs also point to the aesthetic and artistic aspect of language. Through figures of speech, metaphors and rhythmic structures, proverbs create impressive and memorable words.

Idioms are used to describe a particular situation or emotion. Idioms are fixed expressions or groups of words, usually consisting of more than one word. Idioms usually carry a meaning different from the literal meaning of words. In other words, they are used figuratively. Idioms reveal the richness of the language and increase the power of expression [2]. Idioms, which are frequently used in daily conversations, strengthen the expression and attract the attention of the listener. Idioms reveal the subtleties and depths of language. Because idioms often gain meaning through cultural and social contexts.

In Turkish culture, proverbs and idioms are of great importance not only for the aesthetic and artistic dimension of language but also for the protection of social and cultural values. Proverbs and idioms reflect the common memory and experiences of the society in which they exist from the past to the present. These expressions have a wide range of usage in daily life, literature, education and media. Proverbs and idioms contain important elements for the social structure, traditions and worldview of Turkish society [3].

In literature, proverbs and idioms are used as expressions that strengthen and enrich expression. Many writers and poets in Turkish literature use proverbs and idioms in their works. Proverbs and idioms add depth and meaning to the texts. With this use, the reader's interest in the text increases and the stories told are presented more impressively.

Proverbs and idioms play an important role in language teaching. Proverbs and idioms can be used for students to develop their language skills and discover the aesthetic aspects of language [4]. In addition, through these expressions, students learn cultural values and gain knowledge about social consciousness [5]. Proverbs and idioms can develop vocabulary in language. In this way, different features of the language are recognized and the power of expression is increased. Especially proverbs and idioms, in which figurative expression is intense, enable the learning of this situation, which is indispensable in the structure of the language. Learning figurative expression is developed through the use of idioms and proverb patterns [6].

## II. RELATED WORK

Baptista and Reis identified the proverbs in Portuguese. While identifying proverbs in the text, they used abbreviations and derivatives as well as their original form. They conducted a study on a new corpus using information about the form and diversity of proverbs as well as their frequency in the corpus. It was thought that teaching the Portuguese language would also be productive [7]. Development was made through the database

named WordNet.PT [8]. This database was created by developing WordNet [9].

Ghosh and Srivastava stated in their study that complex analogical evaluations come to a certain level with large language models, but are not fully useful for structures such as proverbs containing abstract expressions. For a proverb, the topic has performed a prediction process for similar content discovery. A study was conducted for English proverbs and a data set consisting of 250 proverbs was obtained. Mood scoring was made for each proverb. When compared with BERT [10], a similarity rate of 25% was obtained [11].

Goren and Strapparava performed word-level metaphor detection using the zero-shot model in GPT 3.5 [12]. A data set consisting of 891 English proverbs was used in the study. It has been tried to explain the metaphor determination and meaning relationship of the words in proverbs. A satisfactory performance was achieved with word-level metaphor detection [13].

Özbal et al. used proverbs for metaphor definition and interpretation. They thought that a collection of proverbs could be useful for different fields of study. They created a data set called PROMETHEUS consisting of English proverbs and their Italian equivalents. In addition to the metaphor structure at the word level, the general metaphor degree and meaning of the proverb were obtained by questions posed to a study group. There are 761 sentences and 13642 words in the dataset. They worked with a linguist for the Italian equivalents. Many sources were used here, and then the results obtained were cross-examined or verified by different people. Tokenization and POS tagging were done during the preprocessing stage. The structure was added based on the similarity in meaning of Italian and English proverbs [14].

Rassi et al. used the regular syntactic structure of proverbs. They worked on automatically detecting proverbs in Brazilian Portuguese. They used the finite state automaton structure to search within word combinations. The fact that proverbs use certain word combinations and contain metaphors despite their narrow lexical structure shows the difficulty of studying. They worked on a data set by collecting proverbs and their derivatives. They achieved a success rate of 60.15% [15].

Zongjin et al. helped students understand the meaning of Chinese proverbs and apply them. They have offered an online platform using Natural Language Processing for students whose native language is not Chinese. It is possible to search on the desired word using keywords. There are 12 questions about the literal and semantic meanings of proverbs. The target participants of this study are Malay and Indian students. Participants were given 30 minutes to answer this development task. They were allowed to use search functions and consult online resources to research and learn the structure and semantic meaning of proverbs. Thanks to this search, the metaphor word structure in proverbs will be learned structurally [16].

### III. METHOD

#### A. Data Collection

Turkish proverbs and idioms are in the online dictionary of the Turkish Language Association. In this dictionary, you can

see whether an expression is an idiom or a proverb, its meaning and an example sentence. The data here was obtained by web scraping method using Python programming language. Web scraping analyzes the HTML structure in the source code of a web page and collects the desired data [17]. There are various libraries available to do this process and automatically retrieve the data. The most common of these is the BeautifulSoup library. With this library, data extraction processes are made fast and automatic [18]. Idioms and proverbs were extracted from the Turkish Language Association's idioms and proverbs section using Python programming language and BeautifulSoup library. The data was stored in \*csv file type. In total, 2396 proverbs and 11209 idioms were obtained.

#### B. Preprocessing

Preprocessing consists of the steps of organizing the data before processing, removing missing data, transforming it if necessary or deleting unnecessary data [19]. A preprocessing stage was applied for the deconstructions and unnecessary data that may occur after the data extraction. In preprocessing, the number of phrases and words were first checked. They had to contain at least two words. Since the meaning and example sentence part of the data is presented in a single section, this part needs to be separated.



Fig. 1. Turkish Language association dictionary of proverbs and idioms [20].

Fig. 1 shows the online Turkish Language Association Dictionary of Proverbs and Idioms. In blue color, the idiom or proverb is presented. Under this section, the meaning of the idiom or proverb is shown first, followed by an example sentence. Since the meaning and the example sentence will be evaluated separately in the study, the information here has been separated.

An example data from the data set is presented in Table I. The first field of the data set contains the Turkish text version of the idiom or proverb, and the second field contains the idiom or proverb information. The third column contains the meaning of the idiom or proverb, and the fourth column contains the example sentence presented in the dictionary. In some idioms and proverbs, since sample sentences are not provided in the dictionary, these sections were taken as blank, and then sample

sentences were produced for the empty fields. The production of sentences was carried out with the help of generative artificial intelligence. For realistic data generation, generative models are used to increase the data size or to fill gaps in the data [21].

TABLE I. SAMPLE DATA FROM THE DATA SET

Idiom / Proverb	Type	Meaning	Example Sentence
gözdağı vermek	Idiom	sonradan verilecek bir ceza ile korkutmak, yıldırım, tehdit etmek, caydırmaya çalışmak	Sarhoş ağabeyi, parası pulu ile gözdağı vermeye kalktı onlara.
abesle iştigal etmek	Idiom	yersiz, yararsız işlerle vakit öldürmek	Yazarlarımızın çoğu, yalnızca kendi ürünlerinin ne amaçla üretildiğini sayıp dökerek bir anlamda abesle iştigal ediyorlar.
acele ile menzil alınmaz	Proverb	ivmekle daha çabuk sonuç alınır sanılmamalıdır	Acele ile menzil alınmaz. Telaşlanıp sabırsız davranmakla, daha çabuk sonuç alacağımız, başarı kazanacağımız sanılmamalıdır.
bağrına basmak	Idiom	kucaklamak	İzmir'den kalkıp Mısır'a kadar beni görmeye, beni okşamaya, beni bağrına basıp sevmeye gelirdi.

C. Analyze

Fig. 2 shows the number of idioms and proverbs in the data set. There are 2396 proverbs and 11209 idioms in the data set. These idioms and proverbs were checked for the possibility of repeatable data during the preprocessing stage. Missing data were detected only for the sample sentence part, and it was observed that these were not included in the dictionary.

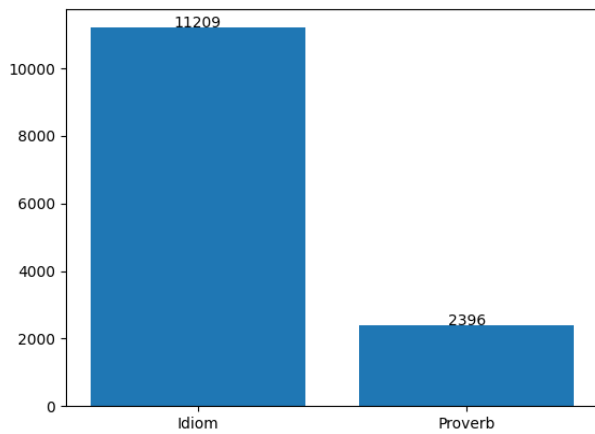


Fig. 2. Number of proverbs and idioms in the data set.

Fig. 3 shows the graph of the word numbers of idioms. This graph shows that the vast majority of idioms consist of two words. In Turkish, idioms generally consist of two words and contain verbs that are not structurally inflected in the infinitive form [22]. The conjugation process on the verb is applied

according to the subject and tense. Since there are many two-word idioms, these idioms were used in the pre-test and post-test.

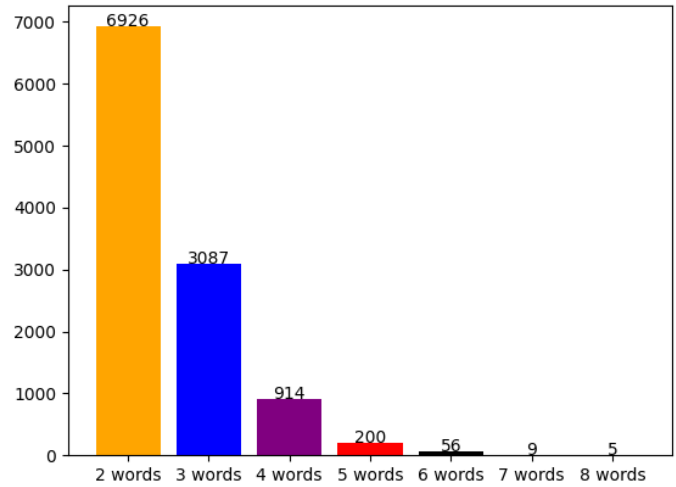


Fig. 3. Idioms word counts.

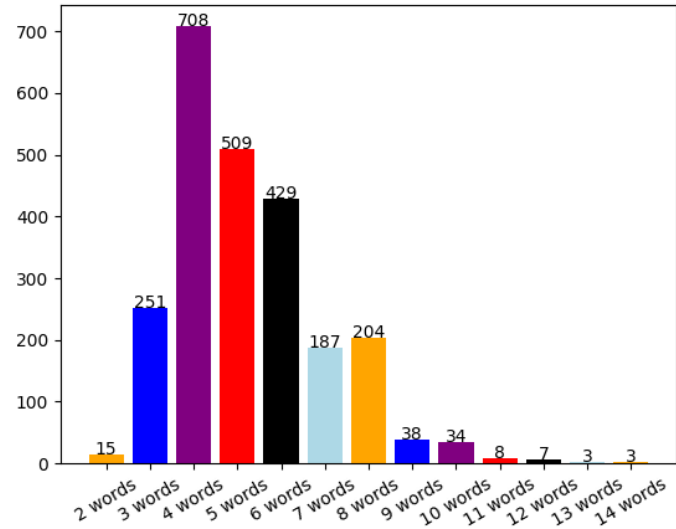


Fig. 4. Proverbs word counts.

Fig. 4 shows the graph of the word numbers of proverbs. In this graph, it can be seen that the four-five-six-word versions of proverbs are in majority. In Turkish, proverbs usually contain one or more verbs [23]. Since they are sentences containing subject, predicate and object, unlike idioms, they consist of sentences with more than two words. For the pre-test and post-test, proverbs containing mostly four-five-six words were used.

Word frequency processing was applied to the idioms in the dataset. The words most used in idioms in Turkish are important. Therefore, understanding the contextual analysis of the words with the highest frequency will be effective in learning many idioms [24]. For the pre-test and post-test, idioms containing the ten words with the highest frequency seen in Fig. 5 were used.

The results of the word frequency process for proverbs are presented in Fig. 6. Although the frequency results are predominantly stop words, these words are the determining

factor in many proverbs [25]. Therefore, stop words are not excluded from the scope. The proverbs containing the ten words with the highest frequency were used in the pre-test and post-test.

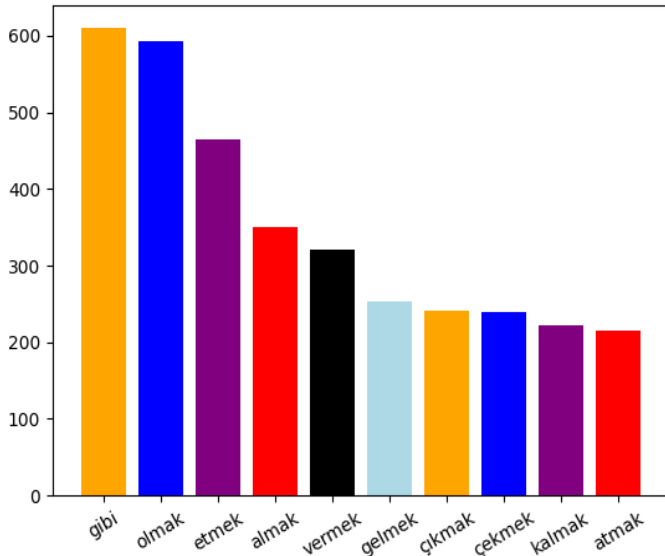


Fig. 5. Top 10 words with the highest word frequency in idioms.

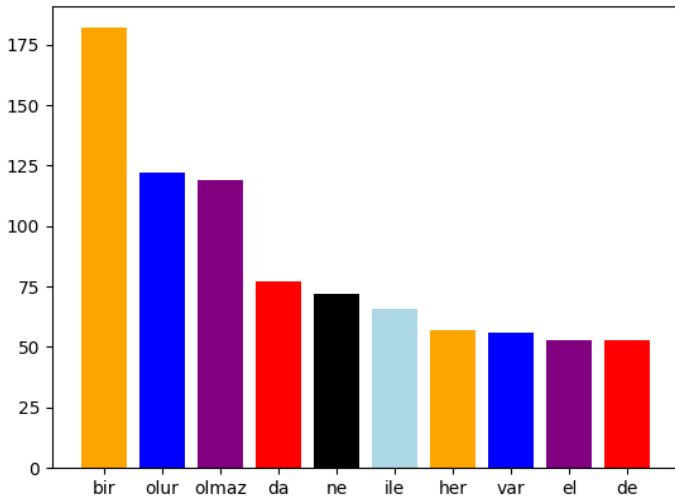


Fig. 6. Top 10 words with the highest word frequency in proverbs.

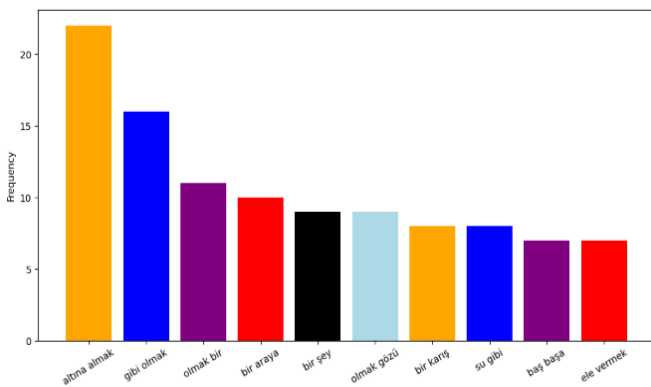


Fig. 7. The 2-gram phrase with the highest frequency for idioms.

An N-gram is an N-character part of a text [26]. 2-gram analysis was conducted due to the idiom structure. The result of the 2-gram analysis for idioms is presented in the graph in Fig. 7. Binary word groups following each other in idioms are important. The fact that there are two-word groups for idioms made this result more important. The word groups obtained here provided information for the most important idiom groups and were used for the pre-test and post-test.

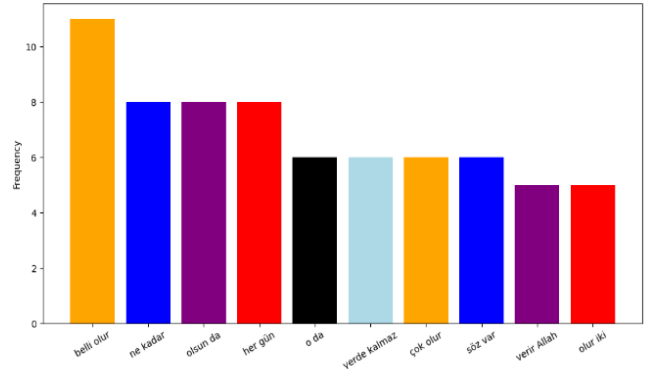


Fig. 8. The 2-gram phrase with the highest frequency for proverbs.

The ten-word groups with the highest frequency obtained by 2-gram analysis for proverbs are presented in the graph in Fig. 8. The word groups provided information for the most important proverbs and were used for the pre-test and post-test.

#### D. Pre-Test

This study aims to improve the learning process of idioms and proverbs. Therefore, it is necessary to compare prior knowledge and subsequent knowledge in a sample group. A group of 20 students aged 10-11 years was selected to conduct the study. This group of students was given a test containing 10 idioms and 10 proverbs. In this test, first of all, the students' level of knowledge about idioms and proverbs was measured. A multiple-choice test was prepared to measure this knowledge. Students were asked to match the meanings of 10 idioms with two words, three words and 10 proverbs with three words, four words and five words. The number of words in idioms and proverbs was chosen because of the regions where there is density.

A web application was developed with C# programming language using ASP.NET Core technology. Students were asked to log in to the system with the code given to each student. In this way, which student answered which question will be recorded. As can be seen in Fig. 9, in the web application, the idiom and proverb were written at the top, then four options were presented and the correct meaning was asked to be found. Only one of the four options is correct. Students answered for 10 idioms and 10 proverbs.

In Table II, the correct answers given by the students for the idiom test are indicated with + for each idiom. In Table III, the correct answers for the proverb test are indicated with a + sign.

Fig. 10 shows the students' correct answers for the 20-question test consisting of idioms and proverbs. When the graph is analyzed, it is seen that the students obtained results close to each other, with an average value of 5.4 for idioms and 5.05 for proverbs.

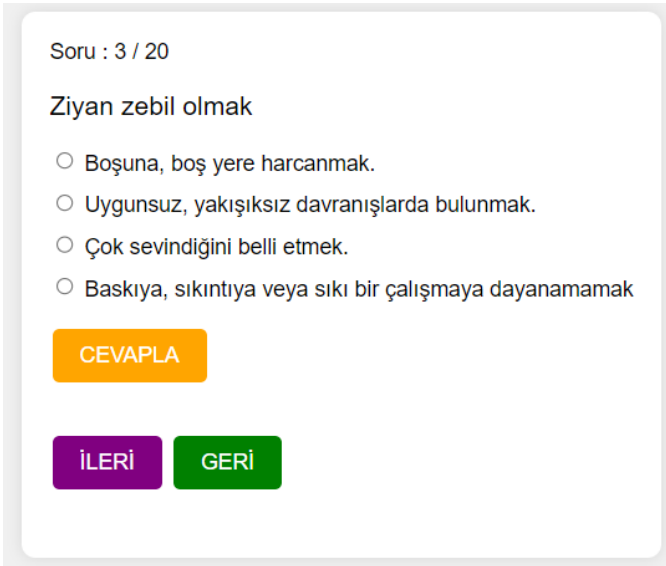


Fig. 9. Web application screenshot for pre-test and post-test.

TABLE II. PRE-TEST DATA FOR IDIOM KNOWLEDGE

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	Score
S1	+	+		+		+			+	+	6
S2		+	+		+		+	+	+		6
S3	+		+	+				+		+	5
S4		+	+		+	+	+		+		6
S5		+	+		+		+	+	+		6
S6	+		+	+	+		+		+	+	7
S7		+		+	+			+			4
S8	+	+	+			+		+			5
S9	+			+		+		+		+	5
S10	+		+		+		+		+		5
S11		+	+		+		+	+		+	6
S12	+	+		+		+			+	+	6
S13	+		+		+	+		+		+	6
S14		+		+		+	+			+	5
S15		+			+			+		+	4
S16	+			+		+		+	+	+	6
S17			+		+		+		+		4
S18	+	+	+		+			+		+	6
S19	+		+				+		+	+	5
S20		+		+	+	+			+		5

TABLE III. PRE-TEST DATA FOR PROVERB KNOWLEDGE

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Score
S1	+			+	+			+	+		5
S2		+	+	+			+		+	+	6
S3	+	+		+	+	+		+	+		7
S4		+		+	+			+		+	5

S5	+				+		+	+			4
S6			+	+			+		+	+	5
S7	+						+		+		3
S8	+			+			+		+		4
S9	+		+		+	+			+	+	5
S10		+	+			+				+	4
S11	+			+		+	+	+			5
S12			+	+	+	+			+	+	6
S13	+	+				+	+		+	+	6
S14			+			+		+	+	+	5
S15		+	+	+	+		+	+			6
S16			+		+		+	+		+	5
S17	+	+		+		+		+		+	6
S18			+			+		+	+		4
S19		+	+	+		+	+				5
S20		+		+			+	+			4

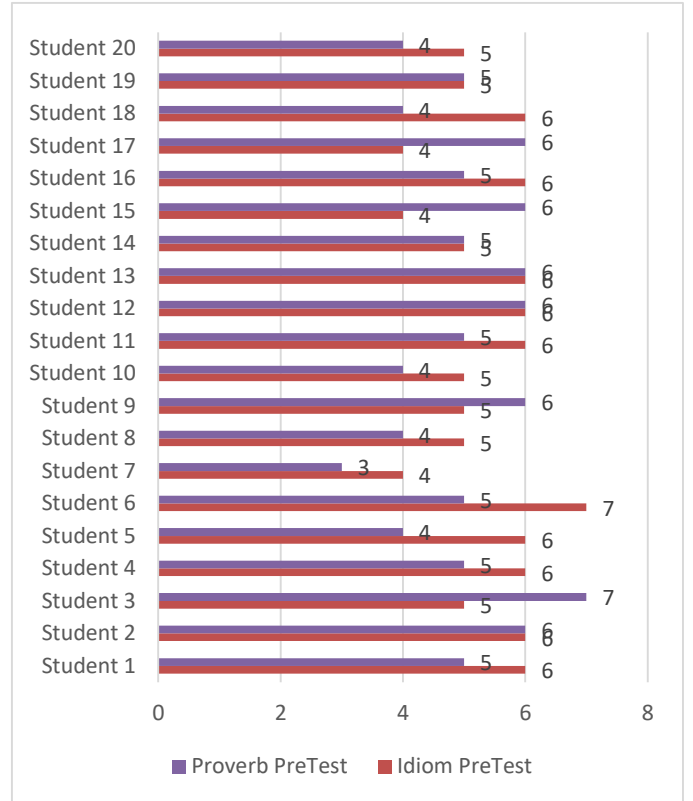


Fig. 10. Pre-test results for 20 students.

### E. Text Generation

The aim of the study is to create a method that will increase the number of correct answers given by students. For this situation, a story was created for each student using generative artificial intelligence. 10 idioms were used in the story. It is aimed that students can guess idioms using previous and next sentences, and they can do this through semantic analysis. An example text produced is shown in Fig. 11.



#### F. Post Test

Table IV shows the post-test results for idioms. Table V shows the post-test results for idioms. In this section, those shown with a + sign indicate the idioms that the students answered correctly in the pre-test. Those marked with an X sign indicate the questions that were not answered correctly in the pre-test but were answered correctly in the post-test. The empty parts indicate the questions that could not be answered correctly in both the pre-test and the post-test.

Bir varmış, bir yokmuş. Anadolu'nun küçük ve şirin bir köyünde, Adil adında bir genç yaşarmış. Adil, doğuştan şanssızmış, köyde **adi çıkmış dokuza, inmez sekize**. Her yaptığı işte bir aksilik çıkar, her sözünde bir yanlış anlaşılma olurmuş. Bu yüzden köyde herkes ona biraz mesafeli durmuş. Bir gün, köyün meydanında bir panayır kurulmuş. Adil de panayıra görmek ve biraz eğlenmek istemiş. Meydana vardığında, göz alıcı standlarla, rengarenk ışıklarla süslenmiş tezgahlar görmüş. Ancak, Adil fark etmiş ki, bazı tezgahlar sadece **göz boyamak** için yapılmış; içlerinde satılacak doğru dürüst bir şey yokmuş. Adil, bir tezgaha yaklaşmış ve satıcıyla biraz **lakırtı etmiş**. Satıcı, bir taraftan Adil'e türlü çeşit ürünler övüp, bir taraftan da kendi yalanlarını saklamaya çalışıyormuş. Ancak Adil'in **ağzının tadı bozulmuş**, çünkü sattıkları ürünler gördükleri kadar iyi değilmiş. Üstelik, tezgahın sahibi onu ikna etmeye çalışırken, **ağzını hayra açmamış**, hep kötü şeylerden bahsedip durmuş. O sırada, Adil'in en yakın arkadaşı Mehmet gelmiş yanına. Mehmet, Adil'e "Haydi başka bir yere gidelim, burada **ölçüyü kaçırdılar**," demiş. Adil de bu öneriyi kabul etmiş ve birlikte meydanın öteki ucuna yürümüşler. Ancak, şanssız Adil bu sefer de **belaya çarpmış**. Panayırdaki kavga çıkmış ve Adil istemeden olayların ortasında kalmış. Kavgayı ayırmaya çalışırken, biri ona çelme takmış ve yere düşmüş. O anda, **boğazından geçmemek** üzere olan bir çığlık kopmuş, ama kimse duymamış. Kavga sona erdikten sonra, Adil yerden kalkmış ve derin bir nefes almış. Mehmet yanına gelip onu teselli etmeye çalışmış. "Merak etme Adil, bu da geçer," demiş. Adil biraz rahatlamış, ama bir yandan da **irtihal eden** yaşlı komşusunu ve onun ona söylediği bilgece sözleri hatırlamış. Yaşlı komşusu, "Hayatta her şey geçicidir, yeter ki sabret," dermiş. Adil ve Mehmet, panayırdan ayrılıp köyün kenarındaki sessiz bir alana gitmişler. Burada oturup köyün üzerindeki yıldızları seyretmişler. Adil, başına gelenleri düşündükçe, bu köyde **açıkta kalmak** istemediğini anlamış. Ne kadar zorluk yaşarsa yaşasın, dostlarıyla birlikte olduğu sürece hayatın tadını çıkarabileceğini fark etmiş. O gece, Adil ve Mehmet uzun uzun sohbet etmişler. Gelecek planları yapmışlar, hayaller kurmuşlar. Adil, en sonunda içini rahatlatan bir karar vermiş. Hayatı boyunca ölçüyü kaçırmamak, her daim dürüst ve açık sözlü olmak istiyormuş. Ertesi sabah, köyde yeni bir gün başlarken, Adil de yeni bir başlangıç yapmaya karar vermiş. Ve böylece, Adil'in hikayesi de tıpkı diğer tüm hikayeler gibi, küçük ama önemli dersler barındırarak devam etmiş. Köyde adı çıkan, şanssız Adil, hayatın içinde yolunu bulmaya devam etmiş. Çünkü bilirmiş ki, hayat her zaman olduğu gibi, inişli çıkışlı bir yolculuktan ibarettir.

Fig. 11. Text produced using idioms and proverbs.

TABLE IV. POST TEST DATA FOR IDIOM KNOWLEDGE

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	Score
S1	+	+	X	+	X	+			+	+	8
S2	X	+	+	X	+		+	+	+	X	9
S3	+	X	+	+			X	+		+	7
S4	X	+	+	X	+	+	+	X	+		9
S5	X	+	+	X	+	X	+	+	+	X	10
S6	+	X	+	+	+	X	+		+	+	9
S7	X	+		+	+	X		+	X	X	8
S8	+	+	+	X		+		+	X	X	8
S9	+	X		+		+	X	+	X	+	8
S10	+	X	+		+	X	+	X	+	X	9
S11		+	+	X	+		+	+		+	7
S12	+	+	X	+	X	+		X	+	+	9
S13	+		+		+	+		+	X	+	7
S14		+		+		+	+	X		+	6
S15		+	X	X	+		X	+		+	7
S16	+	X	X	+		+		+	+	+	8
S17	X	X	+	X	+	X	+	X	+	X	10
S18	+	+	+	X	+			+	X	+	8
S19	+		+		X	X	+	X	+	+	8
S20	X	+	X	+	+	+	X		+	X	9

TABLE V. POST TEST DATA FOR PROVERB KNOWLEDGE

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Score
S1	+			+	+			+	+	X	6
S2	X	+	+	+	X		+	X	+	+	9
S3	+	+	X	+	+	+		+	+	X	9
S4	X	+		+	+	X	X	+		+	8
S5	+	X	X		+	X	+	+	X		8
S6		X	+	+		X	+	X	+	+	8
S7	+	X		X	X	X	+	X	+	X	9
S8	+	X		+	X		+	X	+	X	8
S9	+		+	X	+	+		X	+	+	8
S10		+	+		X	+		X		+	6
S11	+	X		+	X	+	+	+	X	X	9
S12		X	+	+	+	+	X		+	+	8
S13	+	+	X		X	+	+		+	+	8
S14	X		+	X		+	X	+	+	+	8
S15	X	+	+	+	+	X	+	+		X	9
S16	X	X	+		+	X	+	+	X	+	9
S17	+	+	X	+	X	+	X	+	X	+	10
S18	X		+	X	X	+	X	+	+		8
S19	X	+	+	+		+	+				6
S20	X	+		+		X	+	+	X		7

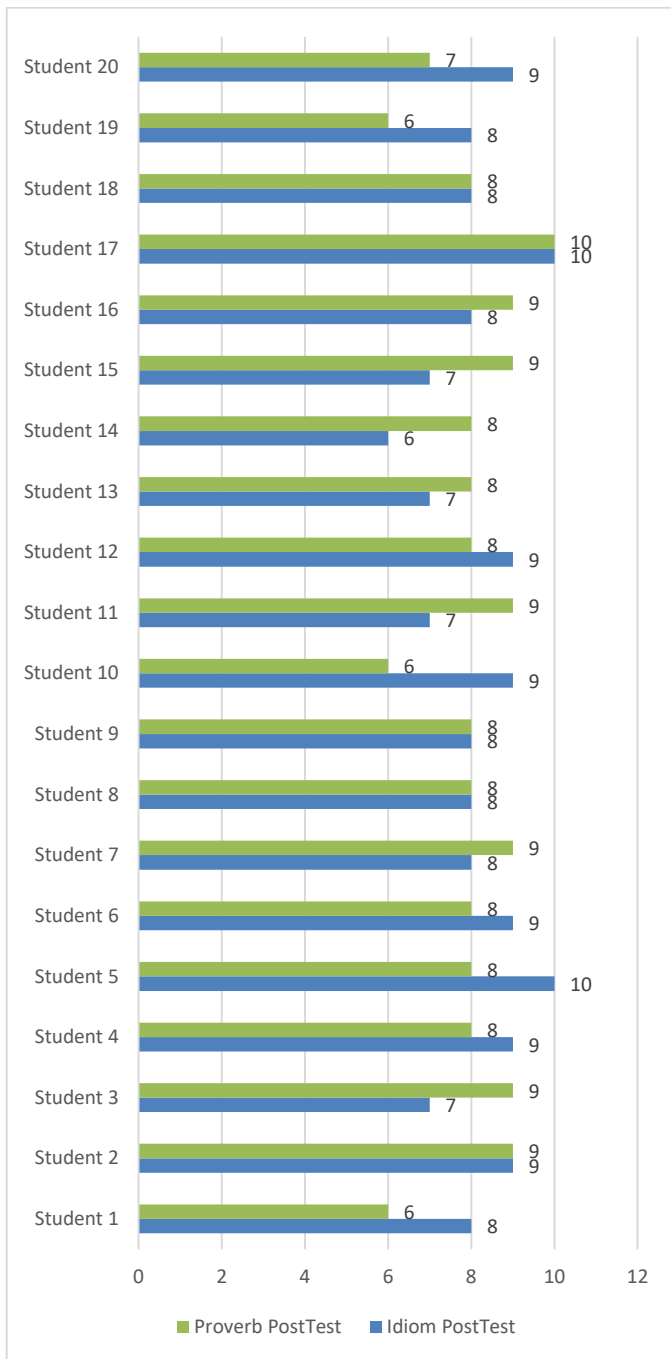


Fig. 12. Post-test results for 20 students.

The graph in Fig. 12 shows the post-test results. Here, one student answered both groups correctly and one student answered all questions in the idiom group correctly.

The pre-test results showed an average value of 5.4 for idioms and 5.05 for proverbs. With the post-test, the average value for idioms increased to 8.2, while the average value for proverbs increased to 8.05. When the graph in Fig. 13 is analyzed, an increase of 51.85% was achieved for idioms and 59.40% for proverbs.

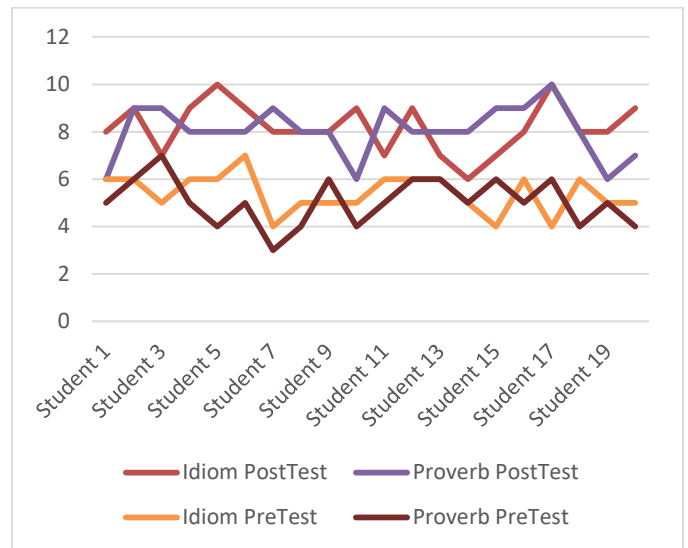


Fig. 13. Comparison of pre-test and post-test results.

#### IV. RESULTS AND DISCUSSION

In the study, Natural Language Processing methods were used to learn idioms and proverbs. The 2396 proverbs and 11209 idioms in the online dictionary of the Turkish Language Association were used. A group of 20 students aged 10-11 years was formed as the study group. Students were selected at this age because they are at an age-appropriate for abstract expression skill levels. The proverbs and idioms in the online dictionary were retrieved by the web scraping method. These data include idiom/proverb, meaning and sample sentence data. The retrieved data were first analyzed with Natural Language Processing methods. In particular, the difficulty levels of idioms and proverbs were determined with methods such as word count, word frequency, n-gram analysis. A pre-test was conducted to measure the efficiency of the method to be created for the students. For this process, a web application was created with C# programming language using ASP.NET Core technology. The answers given by the students were stored on SQL Server. The data here will be used for the post-test at the end of the study. In the method, students were randomly assigned 10 idioms and 10 proverbs according to their difficulty level. They were asked to find the meanings of idioms and proverbs with a multiple-choice test. After the results obtained, texts consisting of idioms and proverbs were presented to the students in line with the semantic analysis, word frequency and word count elements of idioms and proverbs. Here, a generative artificial intelligence was used and texts were produced in the form of a story. With the texts produced, students tried to discover the meanings of idioms and proverbs and to guess them with contextual analysis with the help of the preceding and following sentences. At the end of the whole process, the post-test application was presented to the students in multiple-choice format. After the answers, students' scores increased by 51.85% for idioms and 59.40% for proverbs. This shows the positive effect of students seeing idioms and proverbs within the scope of abstract meaning contextually in sentences, the preceding and following sentences in the story and the story flow.

## V. CONCLUSION AND FUTURE WORK

Proverbs and idioms consist of words and word groups containing metaphors. This negatively affects the learning process or causes the learning process to prolong. For the teaching process of proverbs and idioms in Turkish, words and word groups need to be analyzed. Natural Language Processing methods perform this process in both fast and effective ways. The results obtained from the study showed that the learning process was faster and more effective by revealing the relationship between words and word groups through analysis methods. The fact that the students increased by 51.85% on idioms and 59.40% on proverbs clearly proves this. Since the data set created in the study consists of data from the Turkish Language Association, an official authority, it is also important that it does not contain any errors. In the study, not only the meanings of proverbs and idioms are discussed, but also their usage areas are given with example sentences. This situation showed how semantic comprehension can be reflected in sample applications.

In future studies, text production by students to improve students' knowledge of idioms and proverbs will bring good results. Placing the proverb or idiom where it will be used by using the context of the previous sentence and the next sentence and processing the plot will further develop this skill. Exploring the use of different proverbs and idioms by researching other online resources will ensure good results.

## REFERENCES

- [1] J. Obelkevich, 'Proverbs and social history', in *Wise Words (RLE Folklore)*, Routledge, 2015, pp. 211–252.
- [2] H. Findlay and G. Carrol, 'Contributions of semantic richness to the processing of idioms', *The Mental Lexicon*, vol. 13, no. 3, pp. 311–332, 2018.
- [3] S. Sarıtaş, 'Türk Kültüründe Yüzle İlgili Deyim ve Atasözleri Üzerine Bir Çalışma', *SUSBED*, no. 28, Art. no. 28, Aug. 2012.
- [4] G. Göçen, G. Karabulut, N. Y. Memiş, and M. Darma, 'The Frequency of Use and Distribution of Reduplications, Idioms and Proverbs in Turkish Graded Readers For Foreigners by Levels', vol. 8, no. 2, 2020.
- [5] S. Altaylı, 'Atasözü ve Deyimler Arasındaki Farklar'.
- [6] R. Aksoy Arkan, 'Çeviride Kavramsallaştırma Atasözleri Deyim Örnekleri', *Journal of International Social Research*, vol. 14, pp. 142–150, Jan. 2021, doi: 10.17719/jisr.11527.
- [7] J. Baptista and S. Reis, 'Automatic Classification of Portuguese Proverbs', in *DROPS-IDN/v2/document/10.4230/OASICS.SLATE.2022.2*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi: 10.4230/OASICS.SLATE.2022.2.
- [8] P. Marrafa, 'Portuguese WordNet: general architecture and internal semantic relations', *DELTA*, vol. 18, pp. 131–146, 2002, doi: 10.1590/S0102-44502002000300008.
- [9] C. Fellbaum, 'WordNet', in *Theory and Applications of Ontology: Computer Applications*, R. Poli, M. Healy, and A. Kameas, Eds., Dordrecht: Springer Netherlands, 2010, pp. 231–243. doi: 10.1007/978-90-481-8847-5\_10.
- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds., Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.
- [11] S. Ghosh and S. Srivastava, 'ePiC: Employing Proverbs in Context as a Benchmark for Abstract Language Understanding', May 17, 2022, arXiv: arXiv:2109.06838. Accessed: Jul. 25, 2024. [Online]. Available: <https://arxiv.org/abs/2109.06838>
- [12] X. Liu et al., 'GPT understands, too', *AI Open*, Aug. 2023, doi: 10.1016/j.aiopen.2023.08.012.
- [13] G. Goren and C. Strapparava, 'Context Matters: Enhancing Metaphor Recognition in Proverbs', in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, and N. Xue, Eds., Torino, Italia: ELRA and ICCL, May 2024, pp. 3825–3830. Accessed: Jul. 25, 2024. [Online]. Available: <https://aclanthology.org/2024.lrec-main.338>
- [14] G. Özbal, C. Strapparava, and S. S. Tekiroğlu, 'PROMETHEUS: A Corpus of Proverbs Annotated with Metaphors', in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds., Portorož, Slovenia: European Language Resources Association (ELRA), May 2016, pp. 3787–3793. Accessed: Jul. 25, 2024. [Online]. Available: <https://aclanthology.org/L16-1600>
- [15] A. P. Rassi, J. Baptista, and O. Vale, 'Automatic Detection of Proverbs and their Variants', *OASICS, Volume 38, SLATE 2014*, vol. 38, pp. 235–249, 2014, doi: 10.4230/OASICS.SLATE.2014.235.
- [16] H. Zongjin, W. L. Yann, and A. Y. A. Aziz, 'Student-Oriented-Learning Strategy for Learning Chinese Numerical Proverbs Based on Natural Language Processing Online Database', *Conference Proceedings. Innovation in Language Learning 2022*, Nov. 2022, Accessed: Jul. 25, 2024. [Online]. Available: [https://conference.pixel-online.net/library\\_scheda.php?id\\_abs=5808](https://conference.pixel-online.net/library_scheda.php?id_abs=5808)
- [17] M. Anandarajan, C. Hill, and T. Nolan, 'Introduction to Text Analytics', in *Practical Text Analytics: Maximizing the Value of Text Data*, M. Anandarajan, C. Hill, and T. Nolan, Eds., Cham: Springer International Publishing, 2019, pp. 1–11. doi: 10.1007/978-3-319-95663-3\_1.
- [18] L. Richardson, 'Beautiful soup documentation'. April, 2007.
- [19] C. P. Chai, 'Comparison of text preprocessing methods', *Natural Language Engineering*, vol. 29, no. 3, pp. 509–553, May 2023, doi: 10.1017/S1351324922000213.
- [20] 'Türk Dil Kurumu | Sözlük'. Accessed: Jun. 23, 2024. [Online]. Available: <https://sozluk.gov.tr/>
- [21] Y. Li, Q. Pan, S. Wang, T. Yang, and E. Cambria, 'A Generative Model for category text generation', *Information Sciences*, vol. 450, pp. 301–315, Jun. 2018, doi: 10.1016/j.ins.2018.03.050.
- [22] M. Kurdayıoğlu and Ö. Karadağ, 'Kelime Hazinesi Çalışmaları Açısından Kelime Kavramı Üzerine Bir Değerlendirme', *GEFAD*, vol. 25, no. 2, Art. no. 2, Jun. 2005.
- [23] A. Güzel and Ö. Karadağ, 'Kelime Sıklığı Açısından Türk Atasözleri Üzerine Bir Değerlendirme', *MEÜFD*, vol. 9, no. 1, Art. no. 1, Mar. 2013, doi: 10.17860/efd.48788.
- [24] M. Brysbaert, M. Buchmeier, M. Conrad, A. M. Jacobs, J. Bölte, and A. Böhl, 'The Word Frequency Effect', *Experimental Psychology*, vol. 58, no. 5, pp. 412–424, Jul. 2011, doi: 10.1027/1618-3169/a000123.
- [25] C. C. Çakmakçı, 'Türk Atasözleri ve Deyimlerindeki Kelime Serveti Ve Kavram Geliştirme Sürecinde Kullanımları', *Zeitschrift Für Die Welt Der Türken/Journal Of World Of Turks*, vol. 10, no. 3, pp. 148–168, 2018.
- [26] W. B. Cavnar and J. M. Trenkle, 'N-Gram-Based Text Categorization'.