

A corpus Study on Narrative Texts in Turkish Language*

İlker Aydınⁱ
Ordu University

Abstract

This study aiming to determine the frequency of use of words in the narrative texts in which the best examples of standard language use are displayed ranks 268 stories compiled from the prominent names of Turkish story-writing. Frequency analyses were made on a corpus of approximately 450 thousand words and the frequency of use of words were tried to be listed according to their lexical categories. The words ranking at a certain percentage (about 2%-4%) were scored. It was observed that general words and grammatically functional words had higher frequencies, and the words were used with their antonyms. In terms of Turkish, categorization of words according to their positions in the sentence and their affixations, possible categorization of words in more than one lexical category, and variable meanings which homonyms gain in context were found as the most significant problems in frequency analysis and corpus formation.

Keywords: Turkish, corpus, narrative text, vocabulary teaching.

DOI: 10.29329/ijpe.2019.203.15

* This paper was extended from the oral declaration presented at the 4th International Contemporary Educational Research Congress, "Education, Youth and Future" October 4-7 2018, in Muğla

ⁱ **İlker Aydın**, Assoc. Prof. Dr., Ordu University, Faculty of Education.

Correspondence: ilkaydin67@hotmail.com

INTRODUCTION

It is an important problem that how many words are there in the language and how many of them should be learned or taught. It's almost impossible to say exactly how many words there are in a natural language. This is counted as approximately one million for English (McCarten, 2007). In the *Turkish Dictionary*, it is emphasized that there is a total of 616,767 words, including words, idioms, terms and names (<http://www.tdk.gov.tr>).

Counting words is hard work and this depends primarily on what we understand from the word. For example, words derived from the same root, such as *koş-* 'run', *koşma* 'running', *koşu* 'running', *koşucu* 'runner', *koştur-* 'rush' should be counted as one word or 5 words? Is the verb *koş-* 'run', in the same sense as *maraton koş-* 'run a marathon', *şart koş-* 'stipulate', *işin peşinde koş-* 'pursue' or *atı arabaya koş-* 'harness to'? The same complexity is in question for the homonymic words such as *yüz* ('swim' IMP, 'hundred', 'face'), *yaz* ('write' IMP, 'summer'), *kaz* ('dig' IMP, 'goose'). New words are being added to the language all the time; especially via the internet and information technologies new words such as *yonga* 'chip', *genel ağ* 'internet', *blog* 'blog', *blogcu blogger*, *özçekim* 'selfie', *ağda gezinmek* 'surf the web', *ağ yöneticisi* 'webmaster' etc. are added to Turkish. Despite such difficulties, researchers have tried to estimate how many words native speakers know in order to assess the number of words learners need to learn. Estimates for native speakers vary between 12,000 and 20,000 depending on their level of education. One estimate is that a native speaker graduated from university knows for about 20,000 words (Goulden, Nation, and Read, 1990).

It is difficult to determine which and how many words should be taught to students not only in native language but also in foreign language teaching. According to McCarten (2007), it is possible to get along in English with fewer than 20,000 words and the way of deciding the number of words learners need is to count how many different words are used in an average spoken or written text. Because some high-frequency words are repeated, learners can understand a large proportion of texts with a relatively small vocabulary. McCarten (2007) cites that learners who know the most frequent 2,000 words should be able to understand almost 80 percent of the words in an average text, and a knowledge of 5,000 words increases learners' understanding to 88.7 percent (Francis and Kucera 1982), and for spoken language, the news is even better since about 1,800 words form over 80 percent of the spoken corpus (McCarthy 2004; O'Keeffe, McCarthy, and Carter 2007). It seems important to identify what are the most frequent 2,000 to 5,000 vocabulary items are in target language and to give them priority in teaching.

When the subject is approached from the Turkish, as it is understood from the Turkish Teaching Program, Turkish textbooks, school dictionaries and children's books, the development of the vocabulary cannot actualise incrementally, conspiratorially, and regularly. It is not certain which words will be taught as a priority (Keklik, 2011).

Corpus and vocabulary

In the current *TDK Turkish Dictionary*, the concept of corpus, which corresponds to the term 'derlem', is defined as a combined cluster of the samples compiled from various kinds of language usage area to be used in the grammar and theoretical linguistic studies, in a way to be read by the computer (<http://www.tdk.gov.tr>). The term corpus is derived from the Latin word corpus that means "body". Within the domain of modern corpus linguistics, the term 'corpus' refers to "a large collection of linguistic data, either written texts or a transcription of recorded speech, which can be used as a starting point of linguistic description or as a means of verifying hypotheses about a language" (Crystal 1995). A corpus is a collection of texts, written or spoken, usually stored in a computer database. Written texts in corpus might be drawn from books, newspapers, or magazines that have been scanned or downloaded electronically. Other written corpus might contain works of literature, or all the writings of one author (e.g., Peyami Safa). Such corpus helps us to see how language is used in contemporary society, how our use of language has changed over time, and how language is used in different situations. Spoken corpus, on the other hand, contains transcripts of spoken language. Such

transcripts may be of ordinary conversations recorded in people's homes and workplaces, or of phone calls, business meetings, radio broadcasts, or TV shows. Such written or spoken corpus shows us how a language is used in real life and in many different contexts (McCarthy, 2007).

A corpus stored in a database can tell us about (McCarten, 2007):

Frequency: Which words and expressions are most frequent and which ones are rare;

Differences in speaking and writing: Which vocabulary is the most spoken and which one is the most written;

Contexts of use: The situations in which people use certain vocabulary;

Collocation: Which words are often used together;

Grammatical patterns: How words and grammatical rules are combined to form patterns;

Strategic use of vocabulary: Which words and expressions are used to organize and manage discourse.

Corpus is collected for different purposes. Learner's dictionaries, grammar reference materials, vocabulary learning materials, and course books all benefit from the information in corpus. Materials developed through the corpus can be more reliable and can illustrate language as it is really used. Briefly, a corpus is a large collection of texts that we can analyze using computer software. It is not a theory of language learning or a teaching methodology, but it does influence our way of thinking about language and the kinds of texts and examples we use in language teaching (McCarthy, 2007).

Words are more than just symbols or signs of something in the real world. They carry with them special connotations evolving from the individual, personal experience of each speaker as well as from the society of which he/she is a part. And frequency of a word in a language means the number of occurrence of a linguistic item in a written or spoken form (Richards & Theodore, 2001). Frequency is to know the degree of probability of encountering the word in speech or in print. It is noticeable that frequency also differs in speech and in writing (Elyas & Alfaki, 2014).

Purpose of the Study

The aim of this study is to create a corpus based on narrative texts in which the most beautiful examples of the language are exhibited and to make some comments on the vocabulary of Turkish language by identifying the words with the highest frequency of use.

METHOD

This study, which aims to identify the words with a certain rate of usage over a corpus based on narrative texts, is designed and conducted according to qualitative research paradigm. Qualitative research is defined as a way where qualitative data gathering methods such as observation, interview and document analysis are used (Karasar, 2010).

Research Pattern

This study based on corpus linguistics, has a descriptive pattern. A descriptive design seeks to describe the current status of a variable or phenomenon. The researcher does not begin with a hypothesis, but typically develops one after the data is collected. Data collection is mostly observational in nature.

Data Collection

The data that is subject to the study, which was carried out as a result of a long struggle, was obtained from a corpus of approximately 450 thousand words, including 268 stories of well-known Turkish storytellers such as Ömer Seyfettin, R. Halit Karay, M. Şevket Esendal, S. Faik Abasıyanık, Orhan Kemal, Kemal Tahir, Adalet Ağaoğlu, Pınar Kür, Ferit Edgü, Firuzan, etc.

Data Analysis

In the analysis of the data, it was benefited from the software of analysis named as “word counter 2.10-41” and from the web sites which can make text analysis such as “https://www.browserling.com” and “http://www.csgnetwork.com”. Words can be considered as the smallest independent elements in language and communication. Based on their use and functions, words are categorized into several types or parts of speech. The parts of speech, also known as word classes, explain how a word is used in a sentence. In the Turkish language, words can be classified under 8 major word types or parts of speech namely, nouns, pronouns, adjectives, verbs, adverbs, conjunctions, prepositions, and interjections. In this study, the classification of words is based on *Turkish Dictionary* (TDK, 2018). The data from data collection tools were analyzed using descriptive statistics. According to parts of speech, frequency percentage tables are arranged. The words ranking at a certain percentage (about 2%-4%) is scored. The interjection was not included in this study.

FINDINGS AND COMMENTS

In this section, according to corpus analysis, word frequency and percentage values are given. On the basis of their frequencies, the first few words are interpreted from each table.

According to software, the first 200 words with the highest frequency of use are given in Table 1:

Table 1: Top 200 words with the highest frequency of use

Rank	Word	Frequency	%				
1	bir	15696	3,46	26	onu	877	0,19
2	ve	4790	1,06	27	içinde	872	0,19
3	bu	4779	1,05	28	zaman	872	0,19
4	de	3729	0,82	29	gün	865	0,19
5	o	3293	0,73	30	var	846	0,19
6	da	3191	0,70	31	onun	830	0,18
7	gibi	2826	0,62	32	en	787	0,17
8	ne	2646	0,58	33	mı	776	0,17
9	sonra	2070	0,46	34	bana	776	0,17
10	daha	1892	0,42	35	beni	769	0,17
11	için	1779	0,39	36	değil	760	0,17
12	kadar	1694	0,37	37	doğru	750	0,17
13	dedi	1691	0,37	38	böyle	741	0,16
14	her	1580	0,35	39	artık	738	0,16
15	diye	1500	0,33	40	şimdi	736	0,16
16	ben	1465	0,32	41	ile	733	0,16
17	çok	1431	0,32	42	ya	722	0,16
18	ama	1396	0,31	43	vardı	668	0,15
19	iki	1239	0,27	44	başka	667	0,15
20	ki	1201	0,27	45	fakat	654	0,14
21	hiç	1052	0,23	46	benim	649	0,14
22	şey	976	0,22	47	biraz	649	0,14
23	bile	935	0,21	48	ona	624	0,14
24	bütün	927	0,20	49	nasıl	621	0,14
25	mi	923	0,20	50	sen	621	0,14

51	büyük	617	0,14	101	biri	329	0,07
52	öyle	610	0,13	102	başını	316	0,07
53	adam	607	0,13	103	kız	311	0,07
54	kadın	606	0,13	104	az	310	0,07
55	yok	596	0,13	105	birden	310	0,07
56	yine	591	0,13	106	ağır	305	0,07
57	uzun	590	0,13	107	şeyler	304	0,07
58	hep	568	0,13	108	gene	304	0,07
59	kendi	564	0,12	109	beyaz	304	0,07
60	belki	531	0,12	110	altında	302	0,07
61	küçük	528	0,12	111	mu	299	0,07
62	sanki	509	0,11	112	üstüne	298	0,07
63	dedim	504	0,11	113	tam	298	0,07
64	olan	481	0,11	114	baktı	296	0,07
65	şu	481	0,11	115	yavaş	295	0,07
66	oldu	481	0,11	116	senin	294	0,06
67	gece	475	0,10	117	yoktu	286	0,06
68	iyi	474	0,10	118	Ali	283	0,06
69	eski	470	0,10	119	sana	280	0,06
70	güzel	460	0,10	120	gelen	279	0,06
71	yeni	452	0,10	121	babam	279	0,06
72	pek	449	0,10	122	yer	278	0,06
73	hemen	434	0,10	123	olsun	276	0,06
74	hem	410	0,09	124	eve	276	0,06
75	ilk	401	0,09	125	olarak	273	0,06
76	üç	397	0,09	126	su	272	0,06
77	genç	395	0,09	127	karşı	271	0,06
78	yalnız	393	0,09	128	aynı	268	0,06
79	bizim	386	0,09	129	gözlerini	267	0,06
80	tek	386	0,09	130	akşam	266	0,06
81	beş	386	0,09	131	içine	264	0,06
82	önce	377	0,08	132	seni	264	0,06
83	bey	377	0,08	133	yerde	262	0,06
84	gözleri	376	0,08	134	kendini	261	0,06
85	olduğunu	376	0,08	135	sabah	258	0,06
86	çocuk	370	0,08	136	kim	258	0,06
87	hiçbir	362	0,08	137	tekrar	258	0,06
88	işte	360	0,08	138	diyor	257	0,06
89	arkadaş	357	0,08	139	yeniden	256	0,06
90	son	354	0,08	140	önünde	255	0,06
91	geldi	350	0,08	141	annem	254	0,06
92	on	349	0,08	142	üzerine	254	0,06
93	başladı	347	0,08	143	kez	253	0,06
94	bunu	345	0,08	144	olsa	252	0,06
95	arasında	343	0,08	145	türlü	250	0,06
96	biz	340	0,08	146	dört	250	0,06
97	birkaç	338	0,07	147	insan	250	0,06
98	olduğu	333	0,07	148	çünkü	249	0,05
99	olur	330	0,07	149	olacak	248	0,05
100	yere	330	0,07	150	ev	247	0,05

151	zaten	243	0,05	176	üstünde	212	0,05
152	sesi	242	0,05	177	evin	206	0,05
153	gitti	240	0,05	178	yıl	204	0,05
154	iş	239	0,05	179	onları	201	0,04
155	göz	238	0,05	180	beri	201	0,04
156	birlikte	235	0,05	181	etti	201	0,04
157	elini	234	0,05	182	yoksa	201	0,04
158	saat	230	0,05	183	alıp	200	0,04
159	para	229	0,05	184	kendine	199	0,04
160	yol	229	0,05	185	işte	198	0,04
161	üzerinde	228	0,05	186	burada	197	0,04
162	yanına	227	0,05	187	dolu	196	0,04
163	bize	225	0,05	188	adammın	195	0,04
164	birer	224	0,05	189	aldı	194	0,04
165	belli	224	0,05	190	çıkıtı	194	0,04
166	ince	223	0,05	191	geçen	192	0,04
167	orada	222	0,05	192	sarı	192	0,04
168	yüz	221	0,05	193	ise	191	0,04
169	demek	218	0,05	194	önüne	191	0,04
170	geri	218	0,05	195	koca	190	0,04
171	başına	217	0,05	196	an	190	0,04
172	ses	217	0,05	197	değildi	190	0,04
173	kara	213	0,05	198	şöyle	189	0,04
174	el	213	0,05	199	açık	189	0,04
175	yana	213	0,05	200	onlar	188	0,04

In Table 1, the values obtained are directly reflected without going to any classification. When Table 1 is examined, it is seen that in terms of frequency of occurrence, among the first 15 words, functional words such as conjunctions (*ve* ‘and’, *de/da* ‘also’, *ne* ‘neither’), prepositions (*gibi* ‘as’, *için* ‘for’, *kadar* ‘until’), adverbs (*sonra* ‘after’, *daha* ‘more’) and multifunctional words such as *bir* ‘one’, *bu* ‘this’ are mainly observed. Since functional words form a limited list in the language, their frequency of use is higher. The language item *dedi* ‘s/he said’ is an expression of direct/indirect speech based on dialogue in the narrative texts. Therefore, it has been the first finite verb with the highest usage frequency. On the other hand, nouns with high frequency of use are general words, and adjectives are ranked with their opposites. According to Holliday & Hassan (1976) “general noun is itself a borderline case between a lexical item (member of an open set) and a grammatical item (member of closed system). The class of general noun is a small set of nouns having generalised reference within the major noun classes, those such as ‘human noun’, ‘place noun’, ‘fact noun’ and the like.”

According to software, the first 50 nouns with the highest frequency of use are given in Table 2:

Table 2: Top 50 nouns with the highest frequency of use

Rank	Noun	Frequency	%	Frequency Distribution (in the first 50 noun)
1	bir	15696	3,46	
2	şey	1463	0,32	şey (976), şeyler (304), şeyi (183)
3	iki	1239	0,27	
4	göz	881	0,19	gözleri (376), gözlerini (267), göz (238)
5	zaman	872	0,19	
6	yer	870	0,19	yere (330), yer (278), yerde (262)
7	gün	865	0,19	
8	adam	802	0,18	adam (607), adamın (195)
9	ev	729	0,16	eve (276), ev (247), evin (206)
10	kadın	606	0,13	
11	baş	533	0,12	başını (316), başına (217), baş (0)
12	gece	475	0,10	
13	ses	459	0,10	sesi (242), ses (217)
14	el	447	0,10	elini (234), el (213)
15	ön	446	0,10	önünde (255), önüne (191), ön (0)
16	yan	440	0,10	yanına (227), yana (213), yan (0)
17	üç	397	0,09	
18	yüz	394	0,09	yüz (221), yüzüne (173)
19	beş	386	0,09	
20	bey	377	0,08	
21	çocuk	370	0,08	
22	arkadaş	357	0,08	
23	on	349	0,08	
24	kız	311	0,07	
25	beyaz	304	0,07	
26	Ali	283	0,06	
27	babam	279	0,06	
28	su	272	0,06	
29	karşı	271	0,06	
30	akşam	266	0,06	
31	içine	264	0,06	
32	sabah	258	0,06	
33	annem	254	0,06	
34	insan	250	0,06	
35	dört	250	0,06	
36	iş	239	0,05	
37	saat	230	0,05	
38	para	229	0,05	
39	yol	229	0,05	
40	kara	213	0,05	
41	yıl	204	0,05	
42	sarı	192	0,04	
43	koca	190	0,04	
44	an	190	0,04	
45	siyah	188	0,04	
46	hanım	184	0,04	
47	ay	181	0,04	
48	erkek	181	0,04	
49	İstanbul	178	0,04	
50	taş	172	0,04	

The noun refers to words that are used to name persons, things, animals, places, ideas, or events. When Table 2 is examined, it is seen that the word *bir* ‘one’ which is used as definite numeral adjective and indefinite numeral adjective is as the word with the highest frequency of use. The word *şey* ‘thing’ in the second order is a multifunctional word and indicates uncertainty. In the narrative texts, the words *bir* ‘one’ and *şey* ‘thing’ are statements of uncertainty that facilitate narration. The word *iki* ‘two’ is seen the most commonly used numerical adjective. The terms *zaman* ‘time’ and *gün* ‘day’ are mainly associated with structures that function as adverb (*o zaman* ‘then, at the time’, *o gün* ‘that day’) rather than noun. The word *yer* ‘place’ completes mainly the meaning of the predicate with the interest of space, by taking the dative suffix (*yere* ‘to the place’) or locative suffix (*yerde* ‘in the place’) and it is a general word. The word *göz* ‘eye’ which is in the fourth order is the most important organ of human being. It is much used in the narrative and descriptive texts. The word *ev* ‘house’ is the ninth order and, it is used as the most important shelter for people. On the other hand, the general words *kadın* ‘woman’ and *adam* ‘man’ are taken place in the first ten ranks, with a close frequency to each other. As to other family members, the frequency of the words *baba* ‘father’, *anne* ‘mother’ and *insan* ‘human’ is close to each other. As to proper nouns, names of city and person, it is noticed that the first proper noun is *Ali* and the first city name is *Istanbul*. It is also seen that the first color name is *beyaz* ‘white’ and the second organ name is *el* ‘hand’.

According to software, the first 15 pronouns with the highest frequency of use are given in Table 3:

Table 3: Top 15 pronouns with the highest frequency of use

Rank	Pronoun	Frequency	%	Frequency Distribution (in the first 15 pronouns)
1	o	5792	1,28	o (3293), onu (877), onun (830), ona (624), ondan (168)
2	bu	5454	1,20	bu (4779), bunu (345), bunun (168), buna (162)
3	ben	3659	0,81	ben (1465), bana (776), beni (769), benim (649)
4	kendi	1566	0,35	kendi (564), kendini (261), kendine (199), kendisine (150), kendisi (144), kendisini (132), kendimi (116)
5	sen	1459	0,32	sen (621), senin (294), sana (280), seni (264)
6	biz	951	0,21	bizim (386), biz (340), bize (225)
7	onlar	646	0,14	onları (201), onlar (188), onların (127), onlara (130)
8	şu	481	0,11	
9	kim	376	0,08	kim (258), kimi (118)
10	bunlar	332	0,07	bunları (182), bunlar (150)
11	biri	320	0,07	
12	siz	177	0,04	
13	herkes	169	0,04	
14	kimse	164	0,04	
15	hepsi	156	0,03	

A pronoun is a part of a speech which functions as a replacement for a noun. Considering the pronoun analysis, *o* ‘s/he, it’ and *bu* ‘this’ which are in the first two ranks are also used as ‘demonstrative adjective’ and ‘demonstrative pronoun’. Therefore, the frequency of the words *o* ‘s/he, it’ and *bu* ‘this’ is very close to each other. In addition, the narrative texts tend to use *o* ‘s/he, it’ rather than *ben* ‘I’. Thus, the first-person singular pronoun *ben* ‘I’ has been at the third order, yet it is logically expected to be in the first order. The reflexive pronoun *kendi* ‘him/herself’ with the highest frequency of use is used in place of all the personnel pronouns by taking suffixes. Besides, the personal pronouns attract attention mainly with their suffixed forms. Indefinite pronouns such as *herkes* ‘everyone’, *kimse* ‘nobody’, *hepsi* ‘all’ usually do not take suffixes and their frequency of use is low.

According to software, the first 50 adjectives with the highest frequency of use are given in Table 4:

Table 4: Top 50 adjectives with the highest frequency of use

Rank	Adjective	Frequency	%	Frequency Distribution (in the first 50 adjectives)
1	bu	4779	1,05	
2	her	1580	0,35	
3	var	1468	0,32	var (846), vardı (622)
4	çok	1431	0,32	
5	bütün	927	0,20	
6	yok	882	0,19	yok (596), yoktu (286)
7	doğru	750	0,17	
8	böyle	741	0,16	
9	başka	667	0,15	
10	biraz	649	0,14	
11	büyük	617	0,14	
12	öyle	610	0,13	
13	uzun	590	0,13	
14	küçük	528	0,12	
15	şu	481	0,11	
16	iyi	474	0,10	
17	eski	470	0,10	
18	güzel	460	0,10	
19	yeni	452	0,10	
20	pek	449	0,10	
21	ilk	401	0,09	
22	genç	395	0,09	
23	hiçbir	362	0,08	
24	son	354	0,08	
25	birkaç	338	0,07	
26	az	310	0,07	
27	ağır	305	0,07	
28	tam	298	0,07	
29	yavaş	295	0,07	
30	aynı	268	0,06	
31	türlü	250	0,06	
32	belli	224	0,05	
33	birer	224	0,05	
34	ince	223	0,05	
35	dolu	196	0,04	
36	açık	189	0,04	
37	fazla	177	0,04	
38	derin	175	0,04	
39	deli	164	0,04	
40	tatlı	158	0,03	
41	sıcak	150	0,03	
42	yaşlı	148	0,03	
43	kötü	137	0,03	
44	boş	137	0,03	
45	kısa	137	0,03	
46	garip	135	0,03	
47	yarı	130	0,03	
48	hafif	128	0,03	
49	sık	127	0,03	
50	sağ	121	0,03	

Adjective as a part of a speech is used to describe and determine a noun or a pronoun. Adjectives can specify the quality, the size, and the number of nouns or pronouns. In Turkish language, adjectives are multifunctional words of noun origin. When an adjective comes before the noun, it qualifies the noun in the function of adjective and when it is used alone, it takes over all the tasks that the noun can undertake in the sentence. When used with the predicate, its function is adverb.

Güzel kız ‘beautiful girl’ (adjective)

Güzel dışarı çıktı. ‘Beautiful vent out.’ (noun)

Kız *güzel* konuştu. ‘Girl spoke *beautifully*.’ (adverb)

When Table 4 is examined, it is seen that the demonstrative adjective *bu* ‘this’ is at the top of the list. This adjective serves also as the demonstrative pronoun.

Bu ayakkabı güzel. ‘This shoe is beautiful.’ (demonstrative adjective)

Bu sana yakıştı. ‘This is good for you.’ (demonstrative pronoun)

The same situation is valid for the word *şu* ‘that’ which has taken place in the fifteenth order:

Şu çocuk kazandı. ‘That boy won.’ (demonstrative adjective)

Şunu beğendim. ‘I liked *that*.’ (demonstrative pronoun)

The most striking point in the adjectives is that they are used with their opposites such as *var/yok* ‘present/absent’, *büyük/küçük* ‘big/little’, *eski/yeni* ‘old/new’, *ilk/son* ‘first/last’ and their frequency of occurrence is similar. The words *her* ‘every’, *çok* ‘much’, *öyle* ‘so’, *böyle* ‘such’ whose frequency of use is very high are also used as adverb in the sentence.

According to software, the first 20 verbs with the highest frequency of use are given in Table 5:

Table 5: Top 20 verbs with the highest frequency of use

Rank	Verb	Frequency	%	Frequency Distribution (in the first 20 verbs)
1	ol-	4107	0,91	olan (481), oldu (481), olduğunu (376), olduğu (333), olur (330), olsun (276), olarak (273), olsa (252), olacak (248), olmuştu (143), olurdu (139), olmuş (136), oluyor (131), olmaz (131), olmak (129), oluyor (131), olup (117)
2	de-	2966	0,65	dedi (1691), dedim (504), diyor (257), demek (218), diyordu (174), derdi (122)
3	gel-	1299	0,29	geldi (350), gelen (279), geliyor (167), gelip (166), gelir (137), gelmiş (112), geliyordu (111)
4	et-	736	0,16	etti (201), eden (159), ediyordu (134), etmek (122), eder (120)
5	bak-	728	0,16	baktı (296), bak (171), bakıyordu (127), baktım (118)
6	al-	394	0,09	alıp (200), aldı (194)
7	git-	383	0,08	gitti (240), gidip (143)
8	geç-	375	0,08	geçen (192), geçti (183)
9	başla-	347	0,08	başladı (347)
10	dön-	273	0,06	döndü (154), dönüp (119)
11	dur-	261	0,06	duran (131), durdu (130)
12	kal-	258	0,06	kaldı (148), kalmış (110)
13	gör-	220	0,05	gördü (111), görünce (109)
14	çık-	194	0,04	çıktı (194)
15	bil-	152	0,03	bilir (152)
16	sor-	127	0,03	sordu (127)
17	kalk-	123	0,03	kalktı (123)
18	gir-	116	0,03	girdi (116)
19	ver-	114	0,03	verdi (114)
20	iste-	109	0,02	ister (109)

Verb is the most important part of a speech. A sentence is not existed without verb. Simply put, this is a word that shows an action (physical or mental) or state of being of the subject in a sentence. When Table 5 is examined, it is seen that the verb *ol-* ‘to be’ which is often used as an auxiliary verb has taken in the first order. It was followed by the verb *de-* ‘to say’. When the verb *gel-* ‘to come’ is placed in the third order, the auxiliary verb *et-* ‘to do’, which forms a compound verb by merging with the noun, is placed in the fourth order. As seen in the table, the verb *ol-* ‘to be’ is conjugated in almost all modes and tenses. The fact that the verb *de-* ‘to say’ takes in second order is due to the nature of the narrative texts. The verbs are mostly conjugated in the past tense mode with the 3rd person singular suffix. It is also observed that the use of participles is quite high depending on the descriptions.

According to software, the first 30 adverbs with the highest frequency of use are given in Table 6:

Table 6: Top 30 adverbs with the highest frequency of use

Rank	Adverb	Frequency	%
1	sonra	2070	0,46
2	daha	1892	0,42
3	hiç	1052	0,23
4	içinde	872	0,19
5	en	787	0,17
6	artık	738	0,16
7	şimdi	736	0,16
8	nasıl	621	0,14
9	yine	591	0,13
10	hep	568	0,13
11	belki	531	0,12
12	hemen	434	0,10
13	önce	377	0,08
14	birden	310	0,07
15	üstüne	298	0,07
16	yeniden	256	0,06
17	üzerine	254	0,06
18	zaten	243	0,05
19	birlikte	235	0,05
20	üzerinde	228	0,05
21	orada	222	0,05
22	geri	218	0,05
23	üstünde	212	0,05
24	burada	197	0,04
25	şöyle	189	0,04
26	neden	186	0,04
27	beraber	167	0,04
28	iyice	162	0,04
29	sadece	160	0,04
30	birdenbire	154	0,03

Just like adjectives, adverbs are also used to describe words, but the difference is that adverbs describe adjectives, verbs, or another adverb. If a word is associated with the verb and qualifies the verb, it is adverb; it is associated with the noun and qualifies the noun, it is adjective.

Arkadaşım, İngilizce'yi hızlı öğrendi. (My friend learned English *fast*.) (adverb)

Arkadaşım, hızlı arabaları çok seviyor. (My friend loves *fast* cars.) (adjective)

When Table 6 is examined, it is seen that the word with the highest frequency of use is the word *sonra* 'after' which is adverb of time. It is an adverb that describes when the action of a verb is carried out. Since the narrative texts are texts formed within the event frame, it is common for the adverb *sonra* 'after' to be in the first place. The word *daha* 'more' in the second order and the word *en* 'very' in the fifth order are adverbs of degree that express intensity of the action. They create adverb group such as *daha sonra* 'later', *en çok* 'at most', *daha erken* 'earlier', *en geç* 'the latest'. The word *hiç* 'any' in the third order strengthens the meaning of the action in the negative sentences and indicates uncertain, any time in question sentences. The word *içinde* 'in/within' in the fourth order completes the meaning of the action with the interest of time and space.

According to software, the first 20 prepositions with the highest frequency of use are given in Table 7:

Table 7: Top 20 prepositions with the highest frequency of use

Rank	Preposition	Frequency	%	Frequency Distribution (in the first 20 preposition)
1	<i>gibi</i>	2826	0,62	
2	<i>mi</i>	1998	0,44	<i>mi</i> (923), <i>mı</i> (776), <i>mu</i> (299)
3	<i>için</i>	1779	0,39	
4	<i>kadar</i>	1694	0,37	
5	<i>diye</i>	1500	0,33	
6	<i>değil</i>	950	0,21	<i>değil</i> (760), <i>değildi</i> (190)
7	<i>ile</i>	733	0,16	
8	<i>sanki</i>	509	0,11	
9	<i>yalnız</i>	393	0,09	
10	<i>tek</i>	386	0,09	
11	<i>işte</i>	360	0,08	
12	<i>beri</i>	212	0,05	
13	<i>evet</i>	227	0,05	
14	<i>ancak</i>	183	0,04	
15	<i>sadece</i>	160	0,04	
16	<i>hayır</i>	123	0,03	
17	<i>üzere</i>	107	0,02	
18	<i>acaba</i>	98	0,02	
19	<i>göre</i>	94	0,02	
20	<i>rağmen</i>	83	0,02	

A preposition shows the relationship of a noun or pronoun to another word. They can indicate time, place, or relationship. When Table 7 is examined, it is seen that the preposition with the highest frequency of use is the preposition *gibi* 'as'. It adds to the sentence the meanings of equality and

analogy. The question marker *mi* with its allomorph *mı* and *mu* is ranked in the second order. It reinforces the meaning, and establishes interest of time. The preposition *için* ‘for’ in the third order establishes a variety of meaning interests, in particular the purpose and the cause in the sentence. As to the preposition *kadar* ‘until’ in the fourth order establishes interests such as equality, similarity, approximation through the comparison. The preposition *diye* ‘so, that’ in the fifth order adds the sentence a meaning of purpose and cause. This language item is also related to the reported speech in the narrative texts. The word *değil* ‘isn’t’ is a preposition of negativity. It negates the predicate of nominal sentence. As to the word *evet* ‘yes’ it establishes interest of affirmation and verification.

According to software, the first 20 conjunctions with the highest frequency of use are given in Table 8:

Table 8: Top 20 conjunctions with the highest frequency of use

Rank	Conjunction	Frequency	%
1	ve	4790	1,06
2	de	3729	0,82
3	da	3191	0,70
4	ne	2646	0,58
5	ama	1396	0,31
6	ki	1201	0,27
7	bile	935	0,21
8	ile	733	0,16
9	ya	722	0,16
10	fakat	654	0,14
11	hem	410	0,09
12	gene	304	0,07
13	çünkü	249	0,05
14	yoksa	201	0,04
15	ise	191	0,04
16	ancak	183	0,04
17	hatta	182	0,04
18	hele	156	0,03
19	oysa	150	0,03
20	yani	136	0,03

A conjunction joins words, phrases, or clauses, and indicates the relationship between the elements joined. When Table 7 is examined, it is seen that the most common ones are *ve* ‘and’, *veya/ya da* ‘or’, and *ama/fakat* ‘but’. All these words have nuances and they all help to build up meaningful relationships within a sentence. In terms of frequency of use, the conjunction in the first order is the coordinating conjunction *ve* ‘and’. The conjunction in the second and third order is *de/da* ‘as well’. The language item *da* is allomorph of *de*. The conjunction *de/da* that strengthens the meaning of the word to which it is attached is defined as the conjunction of reinforcement. In the fourth order, the language item *ne* ‘what’ is a conjunction that enters different combinations such as *ne...ne* ‘neither...nor’, *ne...ne de* ‘neither...nor’, *ne var ki* ‘however’. The conjunction *ama/fakat* ‘but’ connects the sentences and judgments that have a contrast between them.

CONCLUSION AND RECOMMENDATION

Corpus research, methodology, and application are closely related and corpora are not only of use in linguistics but in a variety of fields that are concerned with language, e.g. language teaching, translation studies, social sciences, etc. But, counting words of a language is a hard work because of the words derived from the same root, homonymic words and compound words. At the same time, new words included in native language from other languages makes it more difficult. Exposure to different languages is an important factor in the addition of new words to a language. For example, via the internet and information technologies many words are added to Turkish. The parts of speech explain how a word is used in a sentence. Type of word in Turkish; is determined according to the place where the word is found in the sentence, the suffix it receives and the meaning it has in the sentence context. Since some words can sometimes be in more than one part of speech, in Turkish these multifunctional words, like homonymic words, cause problems in composing corpus.

Words with high frequency of use are mainly functional words and general words. According to Mahlberg (2005) "Two assumptions about general nouns are put forward: general nouns are nouns that are used frequently, and they are characterized by local textual functions that will be defined as functions that account for the integration of lexical items in patterns of texts." The part of speech indicates how the word functions in meaning as well as grammatically within the sentence. An individual word can function as more than one part of speech when used in different circumstances. Subjects and objects are often realized by noun or noun phrase, in this case the nouns have a central role and figure prominently in the narrative texts.

In the narrative texts, there are many antonyms as cohesive elements, such as *beyaz-kara/siyah* 'white-black', *kadın-erkek* 'woman-man', *anne-baba* 'mother-father', *sabah-akşam* 'morning-evening', *gelmek-gitmek* 'to come-to go', *üstünde-altında* 'above-under' etc. Since the repetition of the nouns leads to a boring text, pronouns are used very much in the narrative text. The pronouns are words that dominate the task rather than the meaning. The pronouns, *bu* 'this', *o* 's/he, it, that', *ben* 'I/me', taking suffix are used in the function of object and complement. This increases the frequency of use of these words. Otherwise, since the pronoun *kendi* 'him/herself' can replace all the personal pronouns by taking suffixes, its frequency of use is high. Indefinite pronouns such as *hepsi* 'all', *herkes* 'everybody', *kimse* 'anyone' are pronouns with the lowest frequency of use. These pronouns are remarkable by their appearance without suffixes in the narrative texts. In the narrative texts, if it is not a narration of the first-person singular, there are more third person singular and plural pronouns such as *o* 's/he', *bu* 'this' and *onlar* 'they/them'. All types of pronouns are priority words in the language teaching. Adjectives such as *bu* 'this', *her* 'every', *var* 'there is/are', *çok* 'many'; conjunctions *de/da* 'also', *ve* 'and' with the highest frequency of use are in the first places.

This study, based on narrative texts, should be also done for other text types and verbal language. In addition, in a such study, collocated words which tend to occur in the same textual context can be examined. Furthermore, the high frequency of functional words should be evaluated in terms of grammar teaching. And in the teaching of Turkish as both a mother tongue and as a foreign language, the obtained data should be used. Because vocabulary is central to language teaching and without sufficient vocabulary students cannot understand or express their own ideas. Wilkins (1972) states that "While without grammar very little can be conveyed without vocabulary nothing can be conveyed."

REFERENCES

- Crystal, D. (1995). *The Cambridge Encyclopaedia of the English Language*. Cambridge: Cambridge University Press.
- Elyas T, and Alfaki, I. (2014). Teaching Vocabulary: The Relationship between Techniques of Teaching and Strategies of Learning New Vocabulary Items. *English Language Teaching*, 7 (10), 40-56.

- Francis, W. N., and Kucera, H. (1982). *Frequency analysis of English Usage*. Boston: Houghton Mifflin Company.
- Goulden, R., Nation, N. and Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11 (4), 41-63.
- Halliday, M.A.K., Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Karasar, N. (2010). *Bilimsel Araştırma Yöntemi*. Ankara: Nobel.
- Keklik, S. (2011). Türkçede On Bir Yaşına Kadar Çocuklara Öğretilmesi Gereken, Birleşim Gücü Yüksek İlk Bin Kelime. *ODU Sosyal Bilimler Enstitüsü Sosyal Bilimler Araştırmaları Dergisi*, 2 (4), 80-95.
- McCarten J. (2007). *Teaching Vocabulary, Lessons from the Corpus Lessons for the Classroom*. New York: Cambridge University Press.
- McCarthy M. (2007). *Touchstone, From Corpus to Course Book*. New York: Cambridge University Press.
- Mahlberg M. (2005). *English General Nouns: A Corpus Theoretical Approach*. Philadelphia: John Benjamins Publishing.
- Richards, J. C., and Theodore, R. (2001). *Approaches and Methods in Language Teaching*. US: CUP.
- TDK, (2018). *Türkçe Sözlük*. Ankara: Türk Dil Kurumu Yayınları. (<http://www.tdk.gov.tr>).
- Wilkins, D. A. (1972). *Linguistics in Language Teaching*. Hodder & Stoughton Educational.