



(ISSN: 2587-0238)

Şimşek, R. & Erol, T. (2023). A Corpus-Based Suggestion for Words to be Taught at C1 Level In Teaching Turkish as a Foreign Language , *International Journal of Education Technology and Scientific Researches*, 8(21), 414-427.

DOI: <http://dx.doi.org/10.35826/ijetsar.555>

Article Type (Makale Türü): Research Article

---

## A CORPUS-BASED SUGGESTION FOR WORDS TO BE TAUGHT AT C1 LEVEL IN TEACHING TURKISH AS A FOREIGN LANGUAGE

**Ramazan ŞİMŞEK**

Associate Professor, Nevşehir Hacı Bektaş Veli Üniversitesi, Nevşehir, [ramazansimsek@nevsehir.edu.tr](mailto:ramazansimsek@nevsehir.edu.tr)  
ORCID: 0000-0002-8295-8903

**Teymur EROL**

Assist. Prof, Muş Alparslan Üniversitesi, Muş, [t.erol@alparslan.edu.tr](mailto:t.erol@alparslan.edu.tr)  
ORCID: 0000-0003-1738-0858

Received: 09.10.2022

Accepted: 14.02.2023

Published: 05.03.2023

### ABSTRACT

The planning of inputs such as grammatical structure, pronunciation and vocabulary in the foreign language education process directly affects the efficiency. Vocabulary is at the centre of the process in terms of the functional use of basic language skills of individuals learning the target language. Individuals who learn the target language through lists shaped according to levels will achieve high productivity with less effort. In the international literature, vocabulary studies are carried out through frequency lists. These frequency lists are formed through data repository called corpus, which are created with written and spoken texts with various types and structural features of the target language. In this study, it is aimed to determine the most frequently used noun, adjective and verb type words for C1 level in teaching Turkish as a foreign language. The data set of the research consists of the textbooks "Gazi, Yedi İklim, Türkçeye Yolculuk, Yeni Hitit, Yeni İstanbul" used at the C1 level in the field of teaching Turkish as a foreign language, and the reading materials Yenisey and Türkçe Okuyorum, which are used at the C1 level. In this context, the relevant materials were adapted to the database of the software used in the field of corpus linguistics and transferred to the digital environment. After the UTF code conversion, optical character reading errors were eliminated and C1 Level YDTÖ Materials Corpus consisting of 326,591 words was created. Words for which frequency lists were created by means of C1 Level YDTÖ Materials Corpus were transformed into base (lemma) by removing inflectional morpheme. Considering the frequency intensities, the most frequently used words in the type of noun, adjective and verb were determined as a result of the concordance analysis by considering the minimum f(10) value. Through this list, an important contribution has been made to those who learn and teach Turkish as a foreign language; 250 nouns, 250 adjectives and 250 verbs that are most frequently used in teaching Turkish as a foreign language at an advanced level have been suggested. This study is not a final study, but with the aim of being a reference for researchers and publishers, it has prioritized to use the advantages of software and language education fields.

**Keywords:** Turkish, Corpus, Turkish as a foreign language, advanced level, course materials

---

## INTRODUCTION

The Republic of Turkey has made radical changes and transformations in the field of education in the first quarter of the 21<sup>st</sup> century. These transformations have touched every aspect of education and necessitated a change in the vision of teaching and administrative staff. Teaching Turkish as a foreign language has also gained a high momentum around this transformation, and the number of foreign students coming to Turkey has exceeded hundreds of thousands.

The concept of education, which is among the most important indicators of development, is at the focal point of international competition. Especially in higher education institutions, student exchange programs through internationalization, international projects and the number and qualifications of accepted foreign students come to the fore. In this context, the number of foreign students in higher education institutions in Turkey has increased to over two hundred thousand, as a result of the fast increase seen especially in the last ten years. As the target language in which this specified number of international students will receive education is Turkish, the development in the field of teaching Turkish as a foreign language is remarkable. After the 2000s, the number of studies, materials and textbooks in the field of teaching Turkish as a foreign language has increased significantly. Thus, different contents have been presented for the audience learning Turkish as a foreign language, and the number of alternatives has increased to make basic language skills functional and for learners to acquire vocabulary.

In foreign language education, the functions of materials and textbooks are not designed solely on the basis of progress through the curriculum or the relationship of the various stages of the process with each other. The functional use of the language and the acquisition of basic vocabulary are among the primary objectives of language education. The most important factor that directs communication in language actions and bears the semantic load is the word (Şimşek, 2020). It has been explained to foreigners that there are problems in various studies related to the proficiency of Turkish teachers (Memiş, 2021). Vocabulary is very important for basic reading comprehension skills. Making sense of language inputs according to the levels depends on the acquisition of the basic vocabulary and word frequencies.

Word frequency can be defined as the rate of use of a word (Aydın, 2015). In the introduction to his work "The Teacher's Word Book" published in 1921, Thorndike states that the term "frequency" can be defined as the answer to the question of "How often is the word used?" (as cited in Coxhead, 2000: 217).

When the use of textbooks for vocabulary development is considered, two interrelated features of words can be emphasized in texts for vocabulary development. The first feature is repetition (Matsuoka and Hirsh, 2010). Word repetition is a positive condition for teaching vocabulary (Nation, 2001). Research on the effect of word repetition on vocabulary learning has focused on three variables: the number of repetitions, the range of repetitions and the type of repetitions. A certain number of repetitions of a word does not guarantee its learning (Huckin and Coady, 1999). However, there is a strong correlation between vocabulary and reading

comprehension (Matsuoka and Hirsh, 2010: 56; Webb, 2007). Therefore, the words that should be taught first should be selected depending on their use, functionality and frequency in the target language. In a study conducted by Hwang and Nation (1995), it was stated that individuals learning the target language could recognize 84% of the words in various types of texts if they learned the 2,000 most frequently used words.

The lists to be created about the basic vocabulary of the target language will help learners to use basic language skills effectively by breaking their prejudices towards the process. Frequency lists related to the vocabulary also help the functional use of the target language. In recent years, vocabulary studies have been carried out with corpus linguistics applications using more comprehensive databases. This term, which is translated to Turkish as "derlem" and "bütüncü", is a word of Latin origin and means "body" and the plural of the word "corpora" is formed with suffixes used to make plural in Latin (Şimşek and Gün, 2021). Corpus is defined as data sets that can represent the relevant language segment, can be read by machine, contain written or spoken texts created by using special criteria and can describe the language for various research (McEnery and Hardie 2011: 32; Weisser, 2016: 13; Gries, 2009: 7; Dash, 2008: 27). For advanced foreign language learners, it is very important to determine the most frequently used word lists in the target language on a corpus basis in terms of the breadth of the data set and the functionality of the output. Corpus linguistics applications used through the combination of large data sets and software play important roles in many areas such as semantics, pragmatics and collocations. Corpus linguistics applications have come to the fore in teaching Western languages and determining frequency lists according to levels since the second half of the 20<sup>th</sup> century in terms of foreign language education because the needs of the learners from different levels will differ; the vocabulary need of a learner who learns the target language at a beginner level and the vocabulary need of a learner who learns it at an advanced level will be different and naturally the target word list and vocabulary to be taught to learners from different levels will be different (Şimşek, 2021). This study is an important attempt to fill an important gap in the literature by developing a corpus-driven frequency list for the vocabulary that individuals learning Turkish as a foreign language at the C1 level should primarily acquire.

### **Problem Statement**

- 1- What are the most frequently used nouns for individuals learning Turkish as a foreign language at C1 level?
- 2- What are the most frequently used adjectives for individuals learning Turkish as a foreign language at C1 level?
- 3- What are the most frequently used verbs for individuals learning Turkish as a foreign language at C1 level?

### **Purpose**

Determining the vocabulary according to the levels in foreign language education and presenting it to the target audience positively affects the efficiency of the process. In this study, it is aimed to determine the most

frequently used verbs, adjectives and nouns for individuals who learn Turkish as a foreign language at an advanced level and present them to the target audience.

## **METHOD**

In the current study, a corpus-based approach was adopted and a special-purpose corpus was created. Wolfgang Teubert and Anna Čermáková (2004) defined special-purpose corpora as corpora prepared for the purpose of researching a special case rather than the purpose of working on the standard language with a holistic approach. Special-purpose corpora which are smaller than general reference corpora are known to be designed for various research purposes. Special corpora may be a source for a researcher in a study with a narrow scope rather than in general linguistic concerns.

## **Data Analysis**

The documents in the data set of the study were transferred to the digital environment by going through OCR processes and UTF code transformations and the vocabulary of the course materials was analyzed with WordSmith Tools 7.0 and AntConc software used in corpus-based studies and *Turkish C1 Corpus* was created. In the next stage, the most frequently used words were determined, and the most frequently used words were tagged as adjective, noun and verb. The data set labelled with the word types was listed according to frequency. When the relationship between word frequency and vocabulary learning is evaluated in the international literature, it is seen that the vocabulary elements that are planned to be imparted to learners should have a frequency value of at least 10 and above (Pigada and Schmitt, 2006: 18). The words in the *Turkish C1 Corpus* data set, such as nouns, adjectives and verbs, were also analyzed according to this frequency value and lists consisting of 250 different nouns, adjectives and verbs were created.

While creating the data set of the study, manual entries and corrections were made for the wrongly scanned words in the OCR processes. In this context:

- “ı” characters were wrongly coded characters and thus they were re-coded and corrected to “i”.
- “n” characters were wrongly coded characters and thus they re-coded and corrected to “r”.
- “m” characters were wrongly coded characters and thus they were re-coded and corrected to “ru”.
- “ni” characters were wrongly coded characters and thus they were re-coded and corrected to “m”.
- “H” characters were wrongly coded characters and thus they were re-coded and corrected to “Fl”.
- “8” characters were wrongly coded characters and thus they were re-coded and corrected to “B”.
- “ü” characters were wrongly coded characters and thus they were re-coded and corrected to “ıı”.
- “ğ” characters were wrongly coded characters and thus they were re-coded and corrected to “ai”.
- “l” characters were wrongly coded characters and thus they were re-coded and corrected to “h”.
- “?” characters were wrongly coded characters and thus they were re-coded and corrected to “7”.
- “v” characters were wrongly coded characters and thus they were re-coded and corrected to “r”.
- “a” characters were wrongly coded characters and thus they were re-coded and corrected to “ci”.

- “S” characters were wrongly coded characters and thus they were re-coded and corrected to “5”.
- “Çi” characters were wrongly coded characters and thus they were re-coded and corrected to “O”.
- “O” characters were wrongly coded characters and thus they were re-coded and corrected to “0”.

## FINDINGS

### Findings on the Turkish C1 Corpus

The materials used in the field of teaching Turkish as a foreign language (YDTÖ) constitute the data set of the *Turkish C1 Corpus* in the current study. Reading texts in C1 level textbooks such as Altın Köprü, Gazi, Yeni İstanbul, Türkçeye Yolculuk, Yeni Hitit, Yedi İklim and reading texts in Yenisey C1+ and Dilmer Türkçe Okuyorum 4 sets were analyzed on a corpus basis. In the study, the data set, which passed through OCR scanning and code conversion, was adapted to the database of the program to be used for analysis. The corpus of C1 level course materials and the data of the materials that make up the corpus are shown in Table 1.

In the study, the number of tokens in the corpus of the C1 level course materials was determined to be 326,591. The number of types in the corpus of the course materials was determined to be 61,974.

**Table 1.** Findings Regarding the Data Set of the Turkish C1 Corpus

Data Set	Total Number of Tokens	Number of Types	Type/Token Ratio
Altın Köprü	26,356	10,198	38.69
Gazi	27,989	10,162	36.31
Yeni Hitit	54,267	17,776	32.76
Türkçeye Yolculuk	38,275	12,235	31.97
Yedi İklim	44,732	13,089	29.26
Yeni İstanbul	48,786	14,400	29.52
Yenisey C+	70,109	22,791	32.51
Türkçe Okuyorum 4	16,077	6,525	40.59
Türkçe C1 Derlemi	326,591	61,974	18.98

When Table 1 is examined, it is seen that there are sub-corpus consisting of eight different textbooks in the data set of the Turkish C1 Corpus. Considering the data richness of the sub-corpus, the fewest data in terms of the number of words were obtained from the book set named Türkçe Okuyorum 4. On the other hand, the data obtained from Yenisey C+ material constitutes the centre of the corpus. Yenisey C+ is in the position of the sub-corpus, which is the richest among the book sets in terms of the number of different words it contains. When TTR levels were analyzed, the data set with the highest variety of vocabulary was found to be Türkçe Okuyorum 4. The TTR level of the examined books in terms of the number of different words is 13,397.

In the next stage of the study, the most frequently used nouns were determined by using the data set of the Turkish C1 Corpus. The first most frequently used 250 nouns are shown in Table 2 and Table 3.

**Table 2.** Most Frequently Used Nouns in the Turkish C1 Corpus

No	Word	Freq	No	Word	Freq	No	Word	Freq
1	human	1286	43	student	324	85	difference	219
2	time	1070	44	moment	324	86	life	218
3	place	1013	45	name	320	87	spouse	217
4	day	882	46	expression	320	88	religion	216
5	work	727	47	book	319	89	month	215
6	language	712	48	love	316	90	story	214
7	house	701	49	question	315	91	sports	214
8	year	700	50	culture	311	92	continuation	210
9	world	685	51	part	310	93	period	204
10	thing	672	52	history	302	94	creature	203
11	person	647	53	answer	301	95	species	199
12	oneself	633	54	order	300	96	writing	199
13	life	594	55	man	299	97	section	197
14	subject	557	56	science	297	98	construction	196
15	text	532	57	game	287	99	result	196
16	in	529	58	right	284	100	television	194
17	reason	518	59	number	279	101	dance	193
18	child	506	60	music	272	102	cause	192
19	country	464	61	water	269	103	product	192
20	state	455	62	word	268	104	opposition	192
21	money	449	63	lesson	258	105	folk	191
22	hand	431	64	mother	257	106	city	191
23	letter	416	65	long live	253	107	movement	190
24	unit	400	66	woman	252	108	cinema	189
25	friend	396	67	art	250	109	goal	189
26	path	393	68	population	247	110	sound	187
27	sentence	392	69	reality	244	111	necessity	186
28	Turk	389	70	watch	241	112	thought	184
29	information	389	71	special	241	113	male	182
30	film	388	72	effect	233	114	attention	180
31	meaning	379	73	space	230	115	interest	179
32	side	362	74	down	230	116	bottom	179
33	shape	357	75	family	229	117	addition	175
34	eye	353	76	problem	228	118	teacher	171
35	saying	349	77	word	226	119	university	171
36	head	344	78	sign	226	120	acceptance	169
37	dish	341	79	duration	224	121	research	168
38	Turkey	331	80	see	221	122	need	167
39	Turkish	330	81	illness	226	123	week	167
40	age	341	82	father	224	124	program	165
41	sample	336	83	face	221	125	decision	164
42	middle	327	84	event	220			

It was determined that the words *human, time, place, day, work, language, house, year, world, thing* are among the 10 most frequently used words in the Turkish C1 Corpus. When the list is examined, it can be said that the 100 most frequently used nouns consist of daily life, language and education themed content. In Table 3, nouns in the range of 126-250 are listed.

**Table 3.** Most Frequently Used (126-250) Nouns in the Turkish C1 Corpus

No	Word	Freq	No	Word	Freq	No	Word	Freq
126.	news	164	168.	idiom	132	210.	intelligence	105
127.	out	163	169.	environment	132	211.	doctor	105
128.	writer	163	170.	preference	132	212.	point	103
129.	work of art	159	171.	imagination	132	213.	credit	102
130.	car	158	172.	plant	132	214.	young	102
131.	stone	158	173.	strength	132	215.	right	102
132.	future	158	174.	direction	130	216.	partner	102
133.	girl	156	175.	shape	130	217.	morning	101
134.	condition	155	176.	series	129	218.	son	101
135.	society	153	177.	verb	126	219.	dream	101
136.	past	153	178.	computer	126	220.	heart	101
137.	post	152	179.	foundation	126	221.	century	100
138.	communication	152	180.	repetition	123	222.	clothing	100
139.	education	152	181.	branch	123	223.	behaviour	100
140.	emotion	152	182.	class	123	224.	fear	99
141.	ratio	152	183.	telephone	122	225.	bringing	99
142.	system	149	184.	owner	122	226.	shopping	99
143.	brain	149	185.	government	121	227.	fashion	98
144.	exam	149	186.	group	121	228.	death	97
145.	animal	149	187.	usage	120	229.	light	97
146.	opinion	147	188.	value	118	230.	flirtation	96
147.	spirit	145	189.	rule	118	231.	war	96
148.	relationship	145	190.	end	116	232.	character	96
149.	air	144	191.	festival	116	233.	paragraph	95
150.	income	144	192.	angle	116	234.	bus	94
151.	action	143	193.	plenty	114	235.	journey	94
152.	coming	141	194.	happiness	114	236.	trace	94
153.	olympics	140	195.	health	113	237.	song	94
154.	evening	140	196.	foot	113	238.	space	93
155.	flour	140	197.	importance	112	239.	bread	92
156.	energy	140	198.	sea	111	240.	vehicle	91
157.	source	139	199.	specialist	110	241.	god	91
158.	mister	139	200.	upper	110	242.	tradition	91
159.	narration	137	201.	idea	110	243.	mouth	90
160.	poem	137	202.	europe	110	244.	flower	90
161.	night	135	203.	gold	109	245.	coffee	90
162.	horse	135	204.	success	109	246.	curiosity	90
163.	principle	134	205.	shade	108	247.	newspaper	90
164.	teacher	134	206.	sun	108	248.	love	90
165.	school	134	207.	door	107	249.	tourist	89
166.	note	134	208.	hour	107	250.	preparation	88
167.	mathematics	133	209.	garden	106			

In the Turkish C1 Corpus, there are words that refer to different conceptual fields such as *computer, mathematics, newspaper, teacher, success* as well as interrelated words such as *news, author* and *work of art* among the most frequently used nouns in the range of 126-250.

Within the context of the study, the most frequently used adjectives were determined from the data set of the corpus of C1 level course materials. The first most frequently used 250 adjectives are presented in Table 4 and Table 5.

**Table 4.** Most Frequently Used Adjectives in the Turkish C1 Corpus

No	Word	Freq	No	Word	Freq	No	Word	Freq
1	good	496	44	easy	102	87	upper	49
2	correct	454	45	natural	97	88	healthy	49
3	lower	449	46	certain	94	89	interesting	49
4	first	432	47	successful	92	90	hidden	48
5	new	356	48	young	91	91	emotional	48
6	relevant	346	49	traditional	89	92	valuable	47
7	large	344	50	fast	87	93	slim	47
8	wrong	323	51	orderly	83	94	intense	47
9	important	307	52	called	79	95	noisy	46
10	same	269	53	cultural	78	96	ready	45
11	different	267	54	effective	77	97	warm	45
12	other	256	55	rich	76	98	green	44
13	final	240	56	clever	75	99	blessed	44
14	beautiful	230	57	approximate	75	100	adequate	42
15	long	224	58	scientific	75	101	crowded	41
16	more	223	59	negative	74	102	combined	40
17	whole	220	60	separate	74	103	material	40
18	small	217	61	elderly	73	104	social	40
19	suitable	209	62	wide	72	105	previous	40
20	less	196	63	red	70	106	excessive	38
21	old	190	64	natural	68	107	definite	38
22	some	172	65	dependent	67	108	full	38
23	single	170	66	rightful	67	109	cold	38
24	coming	167	67	international	66	110	upper	38
25	necessary	165	68	strong	66	111	authentic	37
26	short	161	69	general	64	112	colourful	35
27	such	150	70	positive	64	113	sweet	35
28	complete	142	71	white	63	114	clean	34
29	special	138	72	meaningful	62	115	physical	34
30	famous	133	73	empty	61	116	pair of	34
31	various	128	74	far	61	117	technical	33
32	social	125	75	artificial	60	118	expensive	33
33	high	125	76	in love	57	119	figurative	33
34	so	123	77	economical	55	120	personal	33
35	daily	119	78	slow	55	121	historical	32
36	difficult	115	79	alone	55	122	national	32
37	open	115	80	helping	55	123	unfounded	32
38	bad	112	81	early	55	124	abundant	31
39	possible	111	82	heavy	54	125	counted	31
40	foreigner	108	83	similar	54			
41	happy	107	84	modern	53			
42	comfortable	106	85	ill	53			
43	near	103	86	diverse	51			

It is understood from Table 4 that the words *good, correct, lower, first, new, relevant, large, wrong, important, same* are among the 10 most frequently used adjectives in the corpus of C1 level course materials. When the list is examined, it can be said that the 100 most frequently used adjectives have a content that takes into account advanced communication needs. Adjectives in the range of 126-250 most frequently used adjectives are listed in Table 5.

**Table 5.** Most Frequently Used Adjectives (126-250) in the Turkish C1 Corpus

No	Word	Freq	No	Word	Freq	No	Word	Freq
126.	cool	31	169.	diverse	22	212.	genius	17
127.	spherical	31	170.	quality	21	213.	geographical	17
128.	normal	31	171.	antique	21	214.	environmental	17
129.	originating	31	172.	parallel	21	215.	adverse	17
130.	serious	30	173.	academic	21	216.	opposite	17
131.	hard	30	174.	logical	21	217.	passionate	17
132.	named	30	175.	superior	21	218.	chronic	17
133.	simple	30	176.	careless	21	219.	dangerous	17
134.	half	30	177.	classical	21	220.	enjoyable	17
135.	main	30	178.	straight	21	221.	dynamic	17
136.	official	30	179.	aimless	20	222.	nervous	16
137.	pleasing	29	180.	light	20	223.	related	16
138.	closed	29	181.	tiny	20	224.	unworldly	16
139.	mutual	29	182.	technological	20	225.	unhappy	16
140.	visual	28	183.	vital	20	226.	pure	16
141.	calm	28	184.	religious	20	227.	folded	16
142.	deep	27	185.	chemical	20	228.	planned	16
143.	with children	26	186.	tired	20	229.	real	16
144.	blue	26	187.	insecure	20	230.	contemporary	16
145.	grey	26	188.	homeless	19	231.	free	16
146.	front	26	189.	psychological	19	232.	popular	15
147.	quick	26	190.	mad	19	233.	limited	15
148.	national	26	191.	formal	19	234.	useful	15
149.	lively	25	192.	perfect	19	235.	documented	15
150.	yellow	25	193.	local	19	236.	renowned	15
151.	victorious	25	194.	paid	19	237.	strong	15
152.	common	25	195.	unique	18	238.	ecstatic	15
153.	collective	25	196.	legal	18	239.	urgent	15
154.	funny	25	197.	medical	18	240.	numbered	15
155.	spiritual	25	198.	harmful	18	241.	aneroid	15
156.	low	24	199.	built	18	242.	cheap	15
157.	thick	24	200.	intellectual	18	243.	central	15
158.	hard-working	24	201.	literary	18	244.	added	15
159.	guilty	24	202.	narrow	18	245.	rude	15
160.	superstitious	24	203.	equal	18	246.	comprehensive	14
161.	detailed	23	204.	beneficial	17	247.	permanent	14
162.	verbal	23	205.	weak	17	248.	biological	14
163.	complex	23	206.	smart	17	249.	minute	14
164.	content	23	207.	inadequate	17	250.	fresh	14
165.	example	23	208.	primitive	17			
166.	conscious	22	209.	broken	17			
167.	black	22	210.	stupid	17			
168.	wonderful	22	211.	virtual	17			

In the corpus of C1 level course materials, adjectives such as *simple, half, light, primitive, fresh*, documented are between the 126<sup>th</sup> and 250<sup>th</sup> most frequent adjectives. The adjectives in question undertake similar functions in the functional use of the language. It can be said that descriptive adjectives about phenomena and events come to the fore.

At the last stage of the study, the most frequently used verbs were determined from the data set of the C1 level course materials corpus. The first 250 most frequently used verbs are shown in Table 6 and Table 7.

**Table 6.** Most Frequently Used Verbs in the Turkish C1 Corpus

No	Word	Freq	No	Word	Freq	No	Word	Freq
1	be	1943	44	change	105	87	dress up	48
2	do	1121	45	stop	102	88	watch	47
3	make	943	46	participate	101	89	get angry	47
4	give	684	47	win	100	90	affect	47
5	say	577	48	listen	97	91	mention	46
6	come	535	49	increase	92	92	appear	46
7	take	511	50	clarify	90	93	remove	45
8	want	430	51	hold	89	94	take away	45
9	see	413	52	recognize	89	95	examine	44
10	tell	405	53	feel	88	96	include	44
11	go	384	54	leave	85	97	satisfy	44
12	stay	360	55	believe	83	98	choose	44
13	know	345	56	improve	83	99	complete	42
14	use	315	57	wait	80	100	pay attention	42
15	exit	291	58	play	80	101	walk	42
16	think	289	59	pull	79	102	fill	42
17	start	274	60	fall	79	103	indicate	41
18	work	270	61	arrive	77	104	be born	41
19	eat	219	62	forget	76	105	feed	41
20	explain	218	63	drink	73	106	like	40
21	live	184	64	put	72	107	reproduce	40
22	read	183	65	drive	71	108	follow	40
23	pass	183	66	realize	71	109	decrease	40
24	find	182	67	throw	71	110	wait	40
25	look	181	68	carry	69	111	spoil	40
26	need	176	69	suppose	68	112	stand up	39
27	show	173	70	convey	67	113	reflect	39
28	write	168	71	burn	66	114	behave	38
29	understand	167	72	sit	66	115	adapt	38
30	bring	156	73	add	62	116	share	38
31	learn	149	74	start	64	117	remember	38
32	love	138	75	sleep	60	118	pay	38
33	finish	133	76	answer	60	119	tire	37
34	speak	133	77	smile	58	120	organize	36
35	hear	120	78	die	58	121	resemble	35
36	seek	119	79	present	57	122	cry	35
37	express	116	80	dread	57	123	solve	35
38	ask	116	81	resent	56	124	convert	35
39	provide	113	82	force	53	125	discuss	34
40	establish	113	83	mark	53			
41	prepare	113	84	determine	53			
42	enter	109	85	turn	52			
43	utilize	108	86	lose	49			

In Table 6, verbs that are used frequently in daily life such as *be, do, make, give* are at the top. In Table 8, verbs in the range of 126-250 most frequently used verbs are listed. When the list is examined, it can be said that the verbs related to the basic needs of daily life come to the fore.

**Table 7.** Most Frequently Used Verbs (126-250) in the C1 Turkish Corpus

No	Word	Freq	No	Word	Freq	No	Word	Freq
126.	confirm	34	169.	spend	24	212.	get rid of	18
127.	fear	34	170.	notice	24	213.	spread	18
128.	send	33	171.	like	24	214.	get together	18
129.	wake up	33	172.	occlude	24	215.	inform	18
130.	travel	33	173.	discover	24	216.	get away	18
131.	supply	33	174.	cut	23	217.	display	18
132.	escape	33	175.	press	23	218.	land	17
133.	catch	32	176.	feel sorry	23	219.	research	17
134.	warn	32	177.	sell	23	220.	chop	17
135.	leave	32	178.	define	23	221.	withdraw	17
136.	come down	32	179.	teach	23	222.	get bored	17
137.	evaluate	31	180.	introduce	23	223.	miss	17
138.	protect	31	181.	raise	23	224.	hit	17
139.	dry	31	182.	update	23	225.	come over	17
140.	emerge	31	183.	get older	23	226.	hide	16
141.	steal	31	184.	settle down	23	227.	trust	16
142.	separate	31	185.	gaze at	23	228.	touch	16
143.	break	31	186.	call	23	229.	set fire	16
144.	shut	29	187.	comment	23	230.	become old	16
145.	seem	29	188.	originate	22	231.	turn	16
146.	wound	29	189.	liven up	22	232.	forbid	16
147.	get involved	29	190.	split	22	233.	live in	16
148.	succeed	29	191.	rain	22	234.	touch upon	16
149.	get better	28	192.	process	21	235.	remember	16
150.	tolerate	28	193.	experiment	21	236.	apply	15
151.	meet	28	194.	beautify	21	237.	hear	15
152.	listen	28	195.	get lost	21	238.	encounter	15
153.	come across	27	196.	get accustomed	20	239.	save	15
154.	lie down	27	197.	wish	20	240.	create	15
155.	give up	27	198.	defend	20	241.	freeze	15
156.	mind	26	199.	count	20	242.	market	15
157.	get surprised	26	200.	get interested	20	243.	get angry	14
158.	grow	26	201.	divorce	20	244.	be interrupted	14
159.	dive	26	202.	take a fixed form	20	245.	shout	14
160.	marry	25	203.	relax	20	246.	arrange	14
161.	bind	25	204.	embark	20	247.	stretch	14
162.	plan	25	205.	rank	20	248.	push	14
163.	consume	25	206.	design	20	249.	hug	14
164.	fulfil	25	207.	summarize	19	250.	remove	13
165.	interpret	25	208.	coat	19			
166.	run	24	209.	negotiate	19			
167.	wash	24	210.	draw	19			
168.	transfer	24	211.	conclude	18			

In the corpus of C1 level course materials, it is seen that the verbs *be, do, make, give, say, come, take, want, see, tell* are the 10 most frequently used verbs. It is also notable that verbs related to cognitive processes such as *say, think, read, write, speak, listen, tell* are among the 50 most frequently used verbs.

## **CONCLUSION and DISCUSSION**

As in all areas of education, textbooks in foreign language education are multifaceted tools that serve implicit and explicit functions. Pedagogical aspect, functionality and usefulness are the main features of any foreign language education material. Textbooks are an indispensable resource for educators in general. It is seen that the materials used in foreign language teaching of Turkish, especially textbooks, are the subject of some research. Gün, Akkaya and Kara (2014) evaluated the vocabulary of the materials used in teaching Turkish as a foreign language within the framework of basic language skills and drew attention to the frequency of words and quality of texts.

In the studies aimed at improving the vocabulary in terms of the functional use of the target language, the concept of frequency has been intensely emphasized. In studies based on vocabulary learning and remembering words (Baddeley, 1990; Bloom and Shuell, 1981; Dempster, 1987), it has been stated that target words should be repeated periodically in course materials. When the relationship between word frequency and vocabulary learning is evaluated, it is emphasized that the vocabulary elements that are planned to be imparted to learners should have a frequency value of at least 10 and above (Pigada and Schmitt, 2006). It has been stated that the first 1000 most frequently used words in a language account for approximately 80% of the productions made in that language, being true for all languages (Aksan, 2000). In this study, a special corpus was created for individuals learning Turkish as a foreign language at C1 level, based on C1 level course materials, frequency lists were analyzed, and a list of the most frequently used words including nouns, adjectives and verbs was created. As a result of the study, frequency lists were developed with a data-driven approach for individuals learning and teaching Turkish as a foreign language and presented to concerned parties.

In the corpus of C1 level course materials, it was determined that the words *human, time, place, day, work, language, house, year, world, thing* are among the top 10 most frequently used nouns. When the most frequently used adjectives were examined, it was seen that the words *good, correct, lower, first, new, relevant, large, wrong, important, same* are among the first 10 adjectives. When the verbs were examined, it was seen that the words *be, do, make, give, say, come, take, want, see, tell* are among the 10 most frequently used verbs. It is also notable that verbs related to cognitive processes such as *say, think, read, write, speak, listen, tell* are among the 50 most frequently used verbs. With these lists containing the 250 most frequently used adjectives, nouns and verbs in teaching materials of Turkish as a foreign language, the words that individuals learning Turkish will encounter most frequently are revealed. In addition, the functional use of basic language skills will be supported through the most frequently used nouns, adjectives and verbs in the materials. As Chomsky (2011) stated, when it is considered that unlimited meaning can be produced with a limited number of words, it is possible to develop the target language comprehension and usage functions of individuals learning Turkish as a foreign language

with this type of corpus-driven frequency lists. In this context, the word lists that are planned to be taught according to the levels in the field of teaching Turkish as a foreign language should be determined with a corpus-based approach. The word lists that are planned to be imparted to learners should be determined on the basis of the sections representing the target language and should be shaped by expert teams.

#### RECOMMENDATIONS

- In the field of teaching Turkish as a foreign language, target word lists should be determined with a corpus-based approach.
- Word lists should be categorized according to word types.
- Word lists prepared with an intuitive approach should be removed from use without delay.

#### ETHICAL TEXT

“This article complies with journal writing rules, publication principles, research and publication ethics, and journal ethics. Responsibility for any violations that may arise regarding the article belongs to the author(s).

Since the survey model was used in this study, it is a study that does not require an ethics committee document.

**Author(s) Contribution Rate:** In this study, the contribution rate of the first author was 51% and the contribution rate of the second author was 49%.

#### REFERENCES

- Aksan, D. (2000). *Her yönüyle dil ana çizgileriyle dilbilim*. TDK Yayınları
- Baddeley, A. (1990). *Human memory*. Lawrence Erlbaum Associates.
- Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, 74, 245-248.
- Chomsky, N. (2011). *Dil ve Zihin 3*. Edisyon, çev. Ahmet Kocaman, Bilgesu Yayınları.
- Dash, N. S. (2008). *Corpus linguistics: An introduction*. Pearson Education.
- Dempster, F. N. (1987). Effects of variable encoding and spaced presentation on vocabulary learning. *Journal of Educational Psychology*, 79, 162–170.
- Grabe, W. (1991). Current developments in second language reading research. *TESOL quarterly*, 25(3), 375-406.
- Gries, S. T. (2016). *Quantitative corpus linguistics with R: A practical introduction*. Taylor & Francis.
- Gün, M., Akkaya, A. ve Kara, Ö. T. (2014). Yabancılara Türkçe öğretimi ders kitaplarının Türkçe öğretim merkezlerinde görev yapan öğretim elemanları açısından değerlendirilmesi. *Turkish Studies*, 9/6, p. 116.
- Hwang, K., & Nation, I. S. P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. *Reading in a Foreign Language*, 6, 323–335.
- Hwang, K., & Nation, I. S. P. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23, 35–41.

- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 21, 181–193
- Matsuoka, W., & Hirsh, D. (2010). Vocabulary Learning through Reading: Does an ELT Course Book Provide Good Opportunities?. *Reading in a foreign language*, 22(1), 56-70.
- McEney, T., and Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: a case study. *Reading in a Foreign Language*, 18(1), 1-28.
- Şimşek, R. & Gün, M. (2021). A Corpus-based Research: The Vocabulary Content and Parts of Speech of A1 Level Textbooks Used in Teaching Turkish as a Foreign Language. *Journal of Language and Linguistic Studies*, 17.
- Teubert, W., & Čermáková, A. (2004). Directions in corpus linguistics. *Lexicology and corpus linguistics*, 113-166.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28, 46–65
- Weisser, M. (2016). *Practical corpus linguistics: An introduction to corpus-based language analysis* (Vol. 43). John Wiley & Sons.