

## **A Corpus-Based Proposal for the Vocabulary to be Taught at A1 Level in the Teaching of Turkish as a Foreign Language**

**Ramazan ŞİMŞEK<sup>1</sup>**

Nevşehir Hacı Bektaş Veli Üniversitesi

**Mesut GÜN<sup>2</sup>**

Mersin Üniversitesi

### **Abstract**

While the significance of corpus-based research in the teaching of foreign language and native language is increasing, the need for such studies is also increasing accordingly. The priority of the vocabulary taught in language education are identified through corpus-based studies in line with the international criteria set in the relevant field. Corpora and sub-corpora consisting of many layers, written and oral, improve the efficiency of the teaching/learning process for both teachers and learners. Corpus-based studies aiming to create word lists to be taught at the basic level are significant in this context to meet the needs of learners and teachers. In this study, 7 different A1 level textbooks and about 114-thousand-word data set used in teaching Turkish as a foreign language were examined based on the relevant corpus, and the type/token ratios of the sub-corpus were identified. The most frequently used words in A1 corpus in the field of Turkish teaching as a foreign language are classified according to their types. Frequency lists of the words that are considered to be functional and useful in the teaching of Turkish as a foreign language for both teachers and learners have been prepared; 250 most frequently used nouns, adjectives and verbs; 750 different words, which are considered to be a fairly comprehensive number for the basic level, are presented as suggestions within the scope of this study.

**Keywords:** Corpus, Vocabulary Load, Coursebooks, Teaching Turkish to Foreigners as a Foreign Language

**DOI:** 10.29329/epasr.2020.334.19

---

<sup>1</sup> Lecturer Dr, TÖMER, Nevşehir Hacı Bektaş Veli Üniversitesi, Nevşehir, ORCID ID: 0000-0002-8295-8903, **Correspondence:** simsekramazan1989@gmail.com

<sup>2</sup> Assoc. Prof. Dr., Faculty of Education, Department of Turkish Education, Mersin Üniversitesi, Mersin, Turkey  
e-mail: mesutgun07@gmail.com ORCID ID: 0000-0001-9663-1066

## Introduction

Language is the greatest communication tool of mankind, who has communicated directly or indirectly in history. Language is the agreement and people produce with words or signs to let one another know what they think and hear (Türkçe Sözlük, 2005: 526). Language is a multifaceted and advanced system that allows thoughts, feelings and desires in society to be passed on others using the shared elements and rules in terms of sound and meaning (Aksan, 1977: 55).

In language actions, the most important factors that drive communication and attract the semantic load is the word or lexical item. A word can be defined as a base possessing a meaning or task ready to be processed with inflexional suffixes (Baş, 2011). As humankind is a social being, s/he expresses what s/he wants through words (Yilmaz and Doğan, 2014). There is a close relationship between the effective use of comprehension and expression skills and the richness of vocabulary knowledge (Karatay, 2007: 143). Word and sentence knowledge are absolutely important for reading at a very basic level (Grabe, 1991: 380) because there is a strong relationship between vocabulary knowledge and reading comprehension (Matsuoka and Hirsh, 2010: 56). A word that acts as a bridge between receiving and transmitting skills and transmits a semantic load makes basic language skills work.

The main purpose of language teaching is to develop listening, speaking, reading and writing skills in learners, which are called four basic language skills. In line with these goals, as in every educational environment, various teaching materials are used to facilitate the process and ensure permanent learning in the field of language teaching. Teaching material is a teaching-learning aid specifically designed for students so that they can learn and for the teacher so that they can provide effective teaching (Demirel, 2005: 7). Textbooks are key for both teachers and students. Planning, controlling the educational process and transferring social and cultural codes to the target audience is achieved through the contents, texts and words presented in textbooks. In this context, the texts to be included in textbooks must have some general characteristics, as well as some certain standards in terms of their vocabulary content.

Vocabulary load is the main factor determining the quality of texts according to the level and the level of semantic load. According to Aksan (2015), vocabulary load is what covers not only the words of the target language, but also its idioms, stereotypes, proverbs, terms, and various narrative patterns. What is described with this term also reflects the material and spiritual culture, worldview, living conditions and experiences of the nation speaking that language (Aksan, 2014: 13)

Frequency studies play an important role in the teaching of a language both as a mother tongue and as a foreign language. Frequency is defined as "the number of occurrences of the same linguistic phenomenon or unit in a speech or article in a certain length " (Vardar, 2002). Word

Frequency can be defined as the rate of frequent use of a word (Aydin, 2015). Thorndike also published a book namely “the teacher's Word Book” in 1921 and he pointed out that frequency is the answer to the question “how often is the word used?” (Cited in Armut. Coxhead, 2000: 217). When considered in terms of education and training processes, the most commonly used material is the textbook and the vocabulary load stands out as an evaluation criterion when evaluating their content and quality.

In foreign language teaching “vocabulary is a complement of the four basic language skills and language knowledge, and it plays a fundamental role in the development of these areas.”(Kurudayıoğlu and Dölek, 2019: 30). Considering from the perspective of those learning Turkish as a foreign language, “the vocabulary needs of learners are different from those whose native language is Turkish, and it requires more effort.” (Karatay, 2020: 256). As with native language education, the basic vocabulary used in foreign language education and factors such as the motivation of students to learn, the ability to achieve, are important for the functional use of the language. Lists that will be created about the basic vocabulary of the target language will help a student use the language functionally. In this context, the needs of learners will differ from an individual who learns the target language at an initial level and to an individual who learns it at an advanced level, and the target word list and the vocabulary load will also be among these differences.

When the relationship between literary texts and life is examined, it is seen that the reality of life in literary products and the state of humanity, is established in the inner world of authors. In this regard, literary texts have a significant share in understanding different aspects of man and society. In literary texts written in a foreign language, it is also possible to find the reflection of the world of that language, the life that is foreign to us (Polat, 2012). In the light of the evaluations, textbooks prepared for foreign language education are expected to use text content in line with the level, needs and purposes of the individuals learning that language. At this stage, the needs of instructors and learners should be monitored, and the content presented in the course should be designed based on international standards.

### **Corpus and Customized Corpus**

It is known that studies on the use of vocabulary have been carried out more often in recent years. “Corpus”, which is a Latin-rooted term, means body, and the plural form of the word “corpora” is formed with Latin plural suffixes. McEnery and Hardie (2011: 32) defined “corpus” as "a whole of text that has the highest level of representative power, that can be read by a machine, and has certain boundaries. Weisser (2016: 13) defined “corpus” as a collection of oral or written texts selected for use in a linguistic evaluation, based on specific design criteria shaped by the purpose and scope of particular linguistic research. Gries (2009: 7) defined “corpus” as a collection of machine-readable

(written or oral) texts produced in a natural communication environment. These texts are collected to be representative and balanced in terms of a certain language variable, language item or literary type, and to conduct linguistic studies. Elena (2001: 55) defined corpus as a computer-processed set of actual texts that are suitable for automatic or semi-automatic processing and review. If a definition for the corpus-based on the evaluations is made; it seems that the emphasis is on transferring texts collected for a general or special purpose to computer environment and processing them electronically for evaluations such as text type, word structure, frequency, context analysis.

Corpora have arc data set volumes according to their purpose. In terms of data size and the area of research it addresses, macro corpora are expressed as “general corpus”. Corpora are divided into two as general and private. According to Tahiroğlu (2010), general corpora include all sub-varieties in one language. General corpora with a variety of data set from very different fields try to describe the macrostructures in the language. Special corpora are specific to the fields such as law, medicine, literature, etc. and newspapers, novels, magazines, academic articles (Tahiroğlu, 2010). Depending on the selected genre, the representation power of general corpora increases or decreases, while the representation power of special corpora are limited by the level of words that they contain (McEnery and Andrew, 2004).

The most advanced form of frequency studies is undoubtedly computer-generated corpora. Corpora performed with natural language processing software are large-volume studies. The larger the volume of content processed in the corpus is, the more reliable the resulting frequency is. In this regard, corpora-related studies are a reference in language teaching (Karadağ, 2019a). However, specialized studies are also carried out for a specific field and discipline through special corpora. With special corpora, researchers are directed to the focus, missing out many variables that big data has. This is also done by combining data sets for the research area.

Among the main data sources of the corpora are written products. Because of the language characteristics, it provides important data to portray the overall structure of that language. On the other hand, written language could be measured as a research object and easily accessible to researchers. In this sense, the reliability of the data gathered from relevant corpus also means that the description made is also reliable (Özkan, 2013). In the design of the corpus, for example, the type of text contained in the content, the number of texts, the selection of specific texts, the selection of samples from available texts, the size of the sample, etc. are all the measures regarding whether the design was made purposefully or not (Sampson et al. 2005: 174). In terms of written resources in the language teaching process, textbooks are among the basic materials in meeting needs and achieving goals, and they have the same significance in the teaching of Turkish as well. Textbooks used in teaching Turkish as a foreign language may differ in terms of frequency lists. When the relevant

literature is considered, it is observed that frequency lists in teaching Turkish as a foreign language are usually designed away from the corpus-based approach.

Vocabulary use in the corpus-based researches conducted at the beginning of 20. Century in Europe is the basis of frequency studies. Corpus-based studies have also been one of the research interests in Turkey in recent years. Özkan (2010) claimed that the positive outputs of processing language data through the computer are significantly obvious in the fields of linguistics; computerized language studies that have expanded its field, and this has been obvious in corpus, linguistics lexicology, morphology, phonology, machine translation, semantics, etc. In this study, the vocabulary load in A1(basic) level textbooks used in teaching Turkish as a foreign language will be examined based on a relevant corpus, and thus A1 level corpus will be created. 250 most frequently used nouns, adjectives and verbs will be listed according to their parts of speech using the relevant corpus.

## **Method**

### **Research Design**

This study adopted a corpus-based approach, and a specific purpose corpus was created. Wolfgang Teubert and Anna Čermáková (2004: 119) defined specific-purpose corpora as a collection prepared by those who created the corpora for their research on a special situation, rather than to make research on a standard language adopting a holistic approach. Specific-purpose corpora, which are smaller in size than general-reference corpora, are known to have been designed for various research purposes. Rather than general linguistic concerns, they may be used by the researchers in narrow-scope researches. A specific-purpose corpus can also be created by gathering texts from a general-purpose reference corpus in parallel with the purpose of the research (Yazıcı, 2018). In this study, 7 different textbooks used in teaching Turkish as a foreign language were digitized, and thus A1 level corpus was created. The following are the textbooks that make up the A1 level corpus:

- A1 level coursebook of Altay, a set for teaching Turkish as a foreign language
- A1 level coursebook of Dedam, a set for teaching Turkish as a foreign language,
- A1 level coursebook of Dilmer, a set for teaching Turkish as a foreign language,
- A1 level coursebook of Gazi, a set for teaching Turkish as a foreign language,
- A1 level coursebook of İstanbul, a set for teaching Turkish as a foreign language,
- A1 level coursebook of Türkçeye Yolculuk, a set for teaching Turkish as a foreign language,
- A1 level coursebook of Yedi İklim, a set for teaching Turkish as a foreign language

## Analysis of Data

A1-level textbooks used in teaching Turkish as a foreign language examined within the scope of the research were transferred to a computer environment, books were digitized with OCR software and a data set was created by converting through UTF-8 code. The obtained data set was prepared for analysis using WordSmith Tools 7.0 software, frequency lists were created and sorted according to parts of speech (noun, verb, adjective,). In this process, inflectional suffixes were removed from words and thus word roots were obtained. For example:” F(1) “was assigned to the word “book” as the frequency value by eliminating the case suffixes in the words “kitaba, kitabı, kitaptan". Thus, the lemmatisation and tag stages in the corpus linguistics were completed. In the final stage, 250 nouns, 250 adjectives, 250 verbs obtained from the A1 level textbooks corpora for basic Turkish learners are presented in tables.

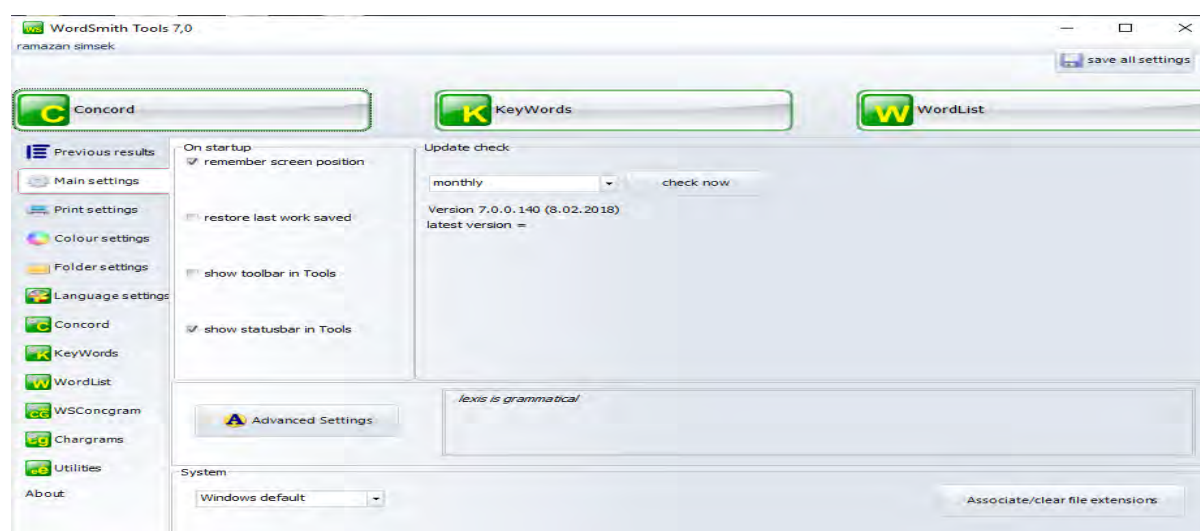


Figure 1. The interface of WordSmith Tools Software

## Findings

### Vocabulary Use of the Corpus of A1 Level Book Sets for Teaching Turkish to Foreigners as a Foreign language

A1 level coursebooks of Altay, Gazi, Istanbul, Türkçeye Yolculuk, Yedi İklim, Dedam, Dilmer sets for teaching Turkish to foreigners were analysed based on the corpus. In this context, the vocabulary load (token) of the textbook corpus was found to be 113957 words. The number of different words of the coursebook corpus was found to be 13958.

Table 1. Findings regarding the Vocabulary Load of YDTÖ A1 Level Coursebooks

Data Set	Number of Words in total (Token)	Number of Different Words (Type)	Coefficient of Word Treasure (Type/Token)
Yedi İklim	19677	5084	0.26
Türkçeye Yolculuk	19308	4351	0.22
Dilmer	19061	3032	0.16
Altay	15882	3882	0.24

İstanbul	15813	3757	0.24
Gazi	13604	3843	0.28
Dedam	10612	2454	0.23
Corpus of A1 Level Coursebooks	113957	13958	0,12

In the study, it was found that the number of words (tokens) of Yedi İklim A1 level textbooks was 19677, the ratio of type and token was 0.26, and the number of different words was 5084. The number of the corpus-based vocabulary of A1 level coursebook of Türkçeye Yolculuk was found to be 19308, the number of different words was 4351, and the type/token ratio was found to be 0.22. In A1 level textbook of Dilmer, it was found that the total number of words (tokens) was 19061, the number of different words(type) is 3032, and the type/token ratio was 0.16.

It was found that the number of words (tokens) of Altay A1 level textbooks was 15882, the number of different words was 3882 and the rate of type/token was 0.24. The number of words (tokens) of Istanbul A1 level textbooks was 158813, the number of different words was 3757, and the ratio of type/token was 0.24. The number of words (tokens) of Gazi A1 level textbooks was 13604, the number of different words was 3843, and the ratio of type/token was 0.28. The number of words (tokens) of Dedam A1 level textbooks was 10612, the number of different words was 2454, and the ratio of type/token was 0.23. It was determined that the total number of words (tokens) in the corpus consisting of all A1 level textbooks was 113957, the number of different words(type) was 13958, and the ratio of type/token was 0.12.

At the next stage of the research, 250 nouns, the most frequently used in teaching Turkish as a foreign language, were identified.

**Table 2.** The most frequently used 1-125 nouns in the YDTO A1 level textbook

No	Word	Freq	No	Word	Freq	No	Word	Freq
1	Saat	700	43	Kahvaltı	183	85	Dünya	117
2	Ev	629	44	Sabah	182	86	Renk	113
3	Gün	574	45	Kelime	178	87	Sokak	112
4	Ünite	462	46	Et	178	88	Televizyon	112
5	Kitap	443	47	Otobüs	178	89	Mustafa	112
6	Yemek	426	48	Kardeş	176	90	Yurt	109
7	Yaz	395	49	Adım	175	91	Boşluk	108
8	Zaman	376	50	Yer	172	92	Kişi	107
9	Türkçe	369	51	Bilgi	172	93	İşaret	106
10	Arkadaş	350	52	Resim	171	94	Sınav	105
11	Okul	344	53	Dün	170	95	Kendi	103
12	Ders	339	54	Tatil	166	96	Amir	103
13	Cevap	326	55	Tamam	164	97	Hayat	102
14	Soru	312	56	Ek	159	98	Bahçe	102
15	Oku	311	57	Kaç	154	99	Çiçek	102
16	İş	297	58	Yaş	148	100	Sema	101
17	Hafta	296	59	Üniversite	148	101	Kahve	101
18	Ad	296	60	Su	147	102	Türk	100
19	Merhaba	288	61	Kilo	146	103	Yol	99

20	Öğrenci	280	62	Yıl	146	104	Pazar	99
21	Sınıf	278	63	Kız	145	105	Bayram	97
22	Hayır	276	64	Dil	145	106	Market	97
23	Öğretmen	271	65	Hava	143	107	Çanta	97
24	Anne	263	66	Son	140	108	Şey	95
25	Telefon	254	67	Şehir	138	109	Kalem	95
26	Çocuk	251	68	Insan	133	110	Efendi	93
27	Akşam	245	69	Ay	132	111	Uçak	93
28	Oda	239	70	Bilgisayar	132	112	Yan	92
29	Araba	228	71	Eş	132	113	Sağ	92
30	Türkiye	224	72	Deniz	132	114	Yarın	92
31	Hanım	212	73	Ara	131	115	Dinle	91
32	Cümle	207	74	Meslek	130	116	Öğle	91
33	Baba	205	75	Çay	125	117	Tarih	90
34	Masa	195	76	Film	124	118	An	88
35	Metin	194	77	Spor	122	119	Kat	87
36	Lira	194	78	Para	122	120	Dolap	87
37	Alışveriş	194	79	Ülke	122	121	Öğren	86
38	Bey	192	80	Diyalog	121	122	Fotoğraf	86
39	Din	191	81	Sinema	120	123	Kedi	86
40	Doktor	191	82	Örnek	119	124	Da	86
41	Aile	190	83	Müşteri	118	125	Durum	86
42	Yabancı	183	84	Park	118			

**Table 3.** The most frequently used 126-250 nouns in the YDTO A1 level textbook

No	Word	Freq	No	Word	Freq	No	Word	Freq
126	Balık	86	168	Kart	64	210	Teyze	49
127	Satıcı	85	169	Oyun	63	211	Seyahat	49
128	Salon	85	170	Diş	63	212	At	49
129	Futbol	85	171	Adres	63	213	Yolcu	49
130	Mevsim	84	172	Ifade	62	214	Süt	49
131	Bilet	84	173	Çorba	62	215	Not	48
132	Neden	83	174	Bardak	61	216	Metro	48
133	Hastane	82	175	Çevre	60	217	Alt	48
134	Otel	81	176	Koltuk	60	218	Dans	48
135	Bölüm	81	177	İnce	60	219	Kıyafet	48
136	Cadde	81	178	Sol	59	220	Üst	47
137	Garson	81	179	Hemşire	59	221	Makine	47
138	Müzik	80	180	Mağaza	59	222	Sözcük	46
139	Elbise	78	181	Gelecek	58	223	Konu	46
140	Kütüphane	77	182	Sağlık	58	224	Erkek	46
141	Yolculuk	76	183	Kıta	57	225	Çık	45
142	Doğum	76	184	Sekreter	56	226	Pantolon	45
143	Kapı	76	185	Cuma	56	227	Durak	45
144	İsim	75	186	Fiil	55	228	Kutu	45
145	Göz	75	187	Ödev	55	229	Kar	44
146	Sıra	75	188	Gül	55	230	Ofis	44
147	Mutfak	74	189	Temel	54	231	Dışarı	44
148	Yatak	74	190	Adam	53	232	Pencere	44
149	Manav	74	191	İnternet	53	233	Lise	43
150	Baş	74	192	Hayvan	53	234	Un	43
151	Çalış	73	193	Domates	53	235	Eşya	43
152	Gece	73	194	Ekmek	53	236	Karşı	42
153	Kafe	72	195	Oğul	53	237	Apartman	42
154	Tekrar	72	196	Defter	53	238	Pazartesi	42



155	Numara	70	197	Meyve	52	239	Peynir	42
156	Lokanta	70	198	Elma	52	240	Kredi	42
157	Yardım	70	199	Gazete	52	241	Gözlük	42
158	Can	70	200	Müze	51	242	Maç	41
159	Ağaç	70	201	İki	51	243	Karın	41
160	Köpek	69	202	Kol	51	244	Konser	41
161	Ayakkabı	69	203	Banka	51	245	Mayıs	41
162	Yüz	68	204	Sandalye	51	246	Günaydın	41
163	El	68	205	Hâl	50	247	Kış	41
164	Merkez	68	206	Amca	50	248	Mektup	41
165	Kurs	66	207	Tiyatro	50	249	Teyze	41
166	Dakika	66	208	Saç	50	250	Sayı	41
167	Geçmiş	64	209	Havuz	50			

A1-level textbooks used in teaching Turkish as a foreign language were examined based on the corpus and 250 nouns, which are the most frequently used according to their frequency values, were identified. As a result of the examination, basic nouns such as *saat, ev, kitap, okul, soru, cevap, masa, oda, araba, telefon, anne, baba* stood out with their frequency values. In terms of basic foreign language education, nouns stand out with their frequency values in the context of daily life, school, basic needs.

At the next stage of the study, 250 verbs, which are most frequently used in teaching Turkish as a foreign language, were identified.

**Table 4.** The most frequently used 1-125 verbs in the YDTO A1 level textbook

No	Word	Freq	No	Word	Freq	No	Word	Freq
1	Git	784	43	Öğren	66	85	Çal	23
2	Yap	747	44	Gir	65	86	Yağ	22
3	Gel	708	45	Ağrı	63	87	Götür	21
4	Iste	587	46	Seyret	60	88	Kutla	21
5	Et	541	47	Dinle	59	89	Göster	21
6	Ol	477	48	Ne	58	90	Eğlen	21
7	Al	457	49	Aç	57	91	Öde	21
8	Ye	315	50	Beğen	56	92	Dinlen	21
9	Sev	262	51	Tanı	56	93	At	20
10	Oku	254	52	Kul	55	94	Ağla	19
11	Ver	241	53	Dol	51	95	Değerlen	18
12	Çık	205	54	Anla	50	96	Özle	18
13	De	182	55	Koy	49	97	Yıka	18
14	Çalış	180	56	Yat	47	98	Incele	17
15	Bak	167	57	An	44	99	Üşü	17
16	Gör	159	58	Ara	44	100	Kazan	17
17	Otur	153	59	Gez	43	101	Tut	16
18	Bil	125	60	Kork	38	102	Görüş	16
19	İç	123	61	Uyan	37	103	Yık	16
20	Kal	119	62	Uy	36	104	Kapat	16
21	Konuş	116	63	Buluş	35	105	Evlen	15
22	Oyna	115	64	Düşün	35	106	Bırak	15
23	Yaz	114	65	Çek	34	107	Bulun	14
24	Söyle	110	66	İzle	33	108	Doğ	14
25	Geç	96	67	Kur	32	109	Çağır	13

26	Dön	93	68	Giy	31	110	Düş	13
27	Başla	91	69	Koş	30	111	Tanış	12
28	Gül	89	70	Yok	29	112	Güzel	12
29	Yürü	86	71	In	29	113	Geçir	12
30	Sor	85	72	Unut	28	114	Yıl	12
31	Bekle	82	73	Piş	28	115	Kaç	11
32	Kalk	82	74	Getir	28	116	Sıklı	11
33	Nere	82	75	Dur	27	117	Kilo Al	11
34	Yor	80	76	Yüz	26	118	Mutlu	11
35	Uyu	76	77	Kız	25	119	Dök	11
36	Yaşa	75	78	Sür	25	120	Ayrıl	11
37	Yaş	75	79	Hoşlan	24	121	Oluş	10
38	Kullan	72	80	Dile	24	122	Ulaş	10
39	Anlat	68	81	Seç	24	123	Kaydet	10
40	Bit	67	82	Din	24	124	Ekle	10
41	Bin	67	83	Sat	23	125	Boya	10
42	Bul	66	84	Kes	23			

**Table 5.** The most frequently used 1-125 verbs in the YDTO A1 level textbook

No	Word	Freq	No	Word	Freq	No	Word	Freq
126	Bahset	10	168	Öğret	6	210	Kay	3
127	Topla	10	169	Fark Et	5	211	Yakala	3
128	Yak	10	170	Hangi	5	212	Süpür	3
129	Yarar	9	171	Yanı	5	213	Aşçı	3
130	Kıl	9	172	Kesil	5	214	Üşü	3
131	Sil	9	173	Doyur	5	215	Çiz	3
132	Tak	9	174	Hatırla	5	216	Yetiş	3
133	Duy	9	175	Kaybet	5	217	Azal	3
134	Bayıl	9	176	Öp	5	218	Paylaş	3
135	Tuşla	9	177	Yaklaş	5	219	Tatıl	3
136	Taşı	9	178	Çarp	5	220	Art	3
137	Acık	9	179	Şaşır	5	221	Kaçır	3
138	Uç	9	180	Bağla	5	222	Uğra	3
139	Temiz	9	181	Hastalan	5	223	Inan	3
140	Um	9	182	Düzenle	5	224	Zayıfla	3
141	Dolaş	8	183	Kop	5	225	Emin	3
142	Değiş	8	184	Doy	4	226	Koru	3
143	Yapış	8	185	Açıkla	4	227	Uzun	3
144	Karşıla	8	186	Dokun	4	228	Sarar	3
145	Anlaş	8	187	Bulan	4	229	Er	3
146	Atla	8	188	Hak Et	4	230	Uzan	3
147	Sevin	8	189	İsmarla	4	231	Kısa	2
148	Üzül	8	190	Bık	4	232	Çevir	2
149	Ak	8	191	Sün	4	233	Yüksel	2
150	Öksür	8	192	Kirala	4	234	Çıkar	2
151	Kapan	8	193	Acı	4	235	Kat	2
152	Ilerle	7	194	Tara	4	236	Kokla	2
153	Hisset	7	195	Patla	4	237	Ov	2
154	Katıl	7	196	Giyin	4	238	İlgilen	2
155	Kır	7	197	Zor	4	239	Duygulan	2
156	Yet	7	198	Benze	4	240	Üre	2
157	Sun	7	199	Ör	4	241	Alış	2
158	Bas	6	200	Affet	4	242	Tanıt	2
159	Boz	6	201	Okşa	3	243	Sin	2
160	Sakla	6	202	Hapşır	3	244	Kuru	2
161	Değ	6	203	İyileş	3	245	Tın	2

162	Gönder	6	204	Tart	3	246	Sus	2
163	San	6	205	Tık	3	247	Say	2
164	Kişi	6	206	Kısal	3	248	Sön	2
165	Karış	6	207	Taşın	3	249	Öl	2
166	Öner	6	208	Utan	3	250	Aran	2
167	Dik	6	209	Karşılaş	3			

A1-level textbooks used in teaching Turkish as a foreign language were examined based on the corpus and 250 most frequently used verbs were identified according to their frequency values. It was found that the verbs such as *git, iste, ye al, ol, çalış, bak, gör* stood out with their frequency values. Frequently used verbs were found to stand out by considering the needs of the learner and the functional use of a foreign language.

At the next stage of the study, 250 adjectives, which are the most frequently used in teaching Turkish as a foreign language, were identified.

**Table 6.** The most frequently Used 1-125 adjectives in the YDTO A1 level textbook

No	Word	Freq	No	Word	Freq	No	Word	Freq
1	İyi	370	43	Pahalı	40	85	Renk	20
2	Aşağı	342	44	Mavi	39	86	Yağlı	20
3	Güzel	301	45	Hız	39	87	Özel	20
4	Doğru	187	46	Neşe	38	88	Bozuk	19
5	Büyük	176	47	Kendi	38	89	İlginç	19
6	Geç	159	48	Basit	37	90	Kat	19
7	Uygun	132	49	Zor	37	91	Tatlı	18
8	Biraz	120	50	Rahat	36	92	Hazır	18
9	Hoş	118	51	Ucuz	36	93	Kuru	17
10	Yeni	115	52	Aynı	36	94	Yardım	17
11	Yanlış	112	53	Bol	35	95	Heyecan	17
12	Küçük	90	54	Eski	35	96	Kalın	17
13	Günlük	81	55	Az	34	97	Güneşli	17
14	İlgi	78	56	Sosyal	33	98	Yarım	17
15	Uzun	77	57	Evli	33	99	Sevgi	16
16	Hasta	73	58	Hayırlı	32	100	Eğlence	16
17	Görev	70	59	Fazla	32	101	Karşılık	16
18	Örnek	65	60	Çalışkan	32	102	Kahverengi	16
19	İlk	65	61	Taze	31	103	Karlı	16
20	Kırmızı	59	62	Genel	29	104	Kötü	15
21	Başka	59	63	Olumsuz	29	105	Zayıf	15
22	Sıcak	58	64	Yüksek	28	106	Boy	15
23	Çeyrek	56	65	Yalnız	28	107	Akıllı	14
24	Yakın	56	66	Yeşil	28	108	Sakin	14
25	Bütün	55	67	Acil	27	109	Tarihî	14
26	Mutlu	53	68	Tam	26	110	Turistik	14
27	Yorgun	51	69	Genç	25	111	Karlı	13
28	Soğuk	51	70	Öyle	25	112	Sivri	13
29	Beyaz	51	71	Yabancı	25	113	Dolu	13
30	Siyah	50	72	Kapalı	25	114	Sesli	13
31	Erken	50	73	Yukarı	25	115	Sağlık	13
32	Uzak	48	74	Kutu	23	116	Özgür	12
33	Ünlü	48	75	Yavaş	23	117	Çabuk	12
34	Kalabalık	48	76	Tek	23	118	Sisli	12
35	Açık	47	77	Lezzetli	23	119	Çift	12

36	Boş	47	78	Olumlu	23	120	Temiz	12
37	Kolay	46	79	Mutsuz	22	121	Benzer	11
38	Kısa	45	80	İç	22	122	Şişman	11
39	Buçuk	45	81	Yaş	21	123	Hafif	11
40	Farklı	45	82	Üst	21	124	Komik	11
41	Önemli	44	83	Sarı	21	125	Çoğul	11
42	Son	43	84	Geniş	20			

**Table 7.** The most frequently Used 126-250 nouns in the YDTO A1 level textbook

No	Word	Freq	No	Word	Freq	No	Word	Freq
126	Mübarek	10	168	Karışık	7	210	Elektronik	4
127	Üstün	10	169	Peşin	7	211	Düşünceli	4
128	Bahçe	10	170	Böyle	7	212	Utangaç	4
129	Üzgün	10	171	Sevimli	7	213	Ağaçlı	4
130	Sinirli	10	172	Ağır	7	214	İndirimli	4
131	Tembel	10	173	Zehir	7	215	Razı	4
132	İnce	10	174	Rüzgârlı	7	216	Belirsiz	4
133	Çeşit	10	175	Dalgalı	7	217	Islak	4
134	Ilık	10	176	Tekil	7	218	Boyalı	4
135	Serin	10	177	Belli	6	219	Güleryüzlü	4
136	Gürültülü	10	178	Sade	6	220	Çiçek	4
137	Dar	10	179	Düz	6	221	Anlamlı	4
138	Pembe	10	180	Manzaralı	6	222	Sobalı	4
139	Düzenli	10	181	Sıkıcı	6	223	Kaloriferli	4
140	Kırık	10	182	Klasik	6	224	Zararlı	4
141	Bekâr	9	183	Zıt	6	225	Evcil	4
142	Millî	9	184	Fena	6	226	Korkunç	4
143	Belirli	9	185	Kilolu	6	227	Geveze	4
144	Resimli	9	186	Dayanıkl	6	228	Dinî	4
145	Resmi	9	187	Türlü	6	229	Zevkli	4
146	Harika	9	188	Sulu	5	230	Kızgın	4
147	Zengin	9	189	Burslu	5	231	Konuşkan	4
148	Tuzlu	9	190	Tok	5	232	Kibar	4
149	Sarışın	8	191	Sabırlı	5	233	Sınırlı	4
150	Yaklaşık	8	192	Faydalı	5	234	Etlî	4
151	Romantik	8	193	Devamlı	5	235	Değerli	4
152	Değişik	8	194	Mezun	5	236	Yetişkin	4
153	Güçlü	8	195	Medeni	5	237	Ön	4
154	Çirkin	8	196	Şaşkın	5	238	Emin	4
155	Dikkat	8	197	Normal	5	239	Kıvrıkcık	4
156	Tehlike	8	198	Kıskanç	5	240	Elim	4
157	Mor	8	199	Acele	5	241	Parçalı	4
158	Şekerli	8	200	Önce	5	242	Antik	4
159	Dümdüz	7	201	Buharlı	5	243	Turuncu	4
160	Keyif	7	202	Yiğit	5	244	Üzer	4
161	Konforlu	7	203	Hayali	5	245	Hareketli	4
162	Müstakil	7	204	Saçlı	5	246	Uç	4
163	Meşgul	7	205	Gözlüklü	5	247	Davetli	3
164	Gözde	7	206	Yoğun	5	248	Sayıllı	3
165	Kirli	7	207	Resmî	5	249	Acılı	3
166	Fakir	7	208	Serbest	4	250	Yerli	3
167	Başarılı	7	209	Sonra	4			

A1-level textbooks used in teaching Turkish as a foreign language were examined based on the corpus and 250 adjectives, which are the most frequently used according to frequency values,

were identified. It was found that the adjectives such as *doğru, yanlış, iyi, aşağı, yukarı, güzel, çirkin, tatlı, tuzlu* stood out with their frequency values. Besides, descriptive and quantity adjectives stand out in the corpus with their frequency values. In terms of learners' needs, it was found that priority is given to the adjectives that individuals will use most often in daily life at the basic level.

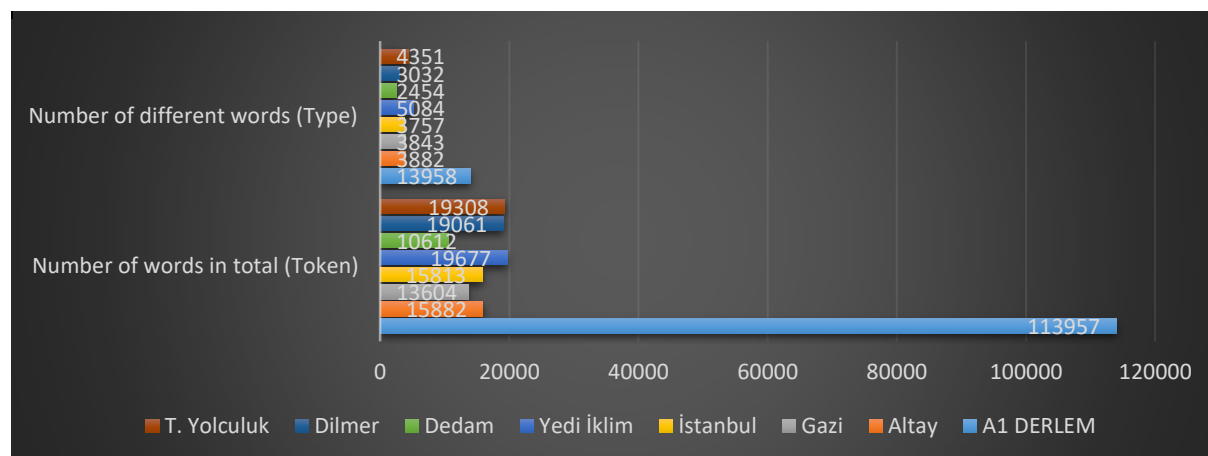
### **Discussion, Conclusion and Suggestions**

Corpus-based frequency lists are fundamental references, especially in Foreign Language Teaching. The use of corpus-based frequency lists or frequency dictionaries in the teaching of the language for general-purpose is highly functional. The common approach is to give priority to the teaching of the most frequently used units in the language. This teaching process varies based on the levels in foreign language education. There are studies on the frequency at varying levels (Kara ve Ulutaş, 2018; Şimşek, 2020a; Şimşek 2020b; Karadağ, 2019b; Karadağ, 2019a; Arslan, 2014).

In recent years, an increase in the number of learning materials has been observed, along with the increase in the demand for learning Turkish as a foreign language. These materials are classified based on their levels, and they offer different contents for the target learners at each level. The fact that the words to be taught at A1 level are identified according to international standards and needs to meet the basic needs of the foreign language learners and the ability to express themselves at the basic level directly affects the success of the teaching process.

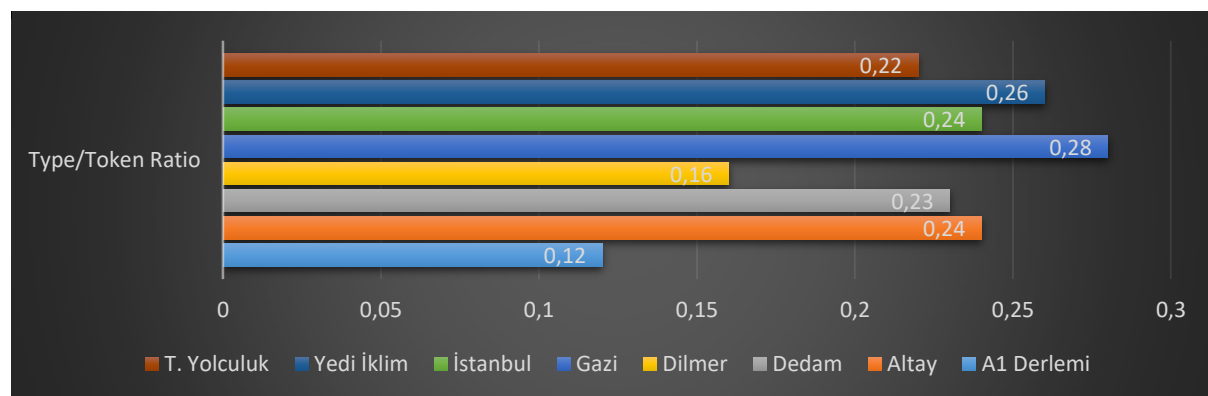
In this study, A1 level coursebooks of Altay, Gazi, İstanbul, Türkçeye Yolculuk, Dedam, Dilmer, Yedi İklim, which are the most commonly used sets for teaching Turkish to foreigners were examined based on the corpus. The total number of words in textbooks and different words were identified, and the most frequently used nouns, adjectives, verbs in these textbooks were also identified and listed based on their frequency values.

A1 level coursebooks of Altay, Gazi, İstanbul, Türkçeye Yolculuk, Dedam, Dilmer, Yedi İklim sets for teaching Turkish to foreigners were examined based on corpus in this research. In this context, the vocabulary load (token) of the A1 Level Textbook corpus was found to be 113957. The number of different words (type) included in the A1 Level Textbook corpus was found to be 13958. The ratio of the total number of words in A1 corpus to the number of different words was found to be 0.12.



**Figure 2.** Data Regarding the Corpus-based Vocabulary Load in YDTÖ A1 Level Coursebooks

The vocabulary load (token) of A1 level coursebook of Altay book set for teaching Turkish to foreigners, which is one of the sub-corpus of A1 level corpus, was found to be 15882 words, the number of different words was found to be 3882. The ratio of the total number of words to the number of different words was found to be 0,24. The vocabulary load of A1 level coursebook of Gazi book set for teaching Turkish to foreigners was found to be 13604, the number of different words was found to be 3843. The ratio of the number of total word to the number of the different word was found to be 0,28.



**Figure 3.** Details regarding the Type/token Ratio of YDTÖ A1 Level Coursebooks

The vocabulary load of A1 level coursebook of İstanbul book set for teaching Turkish to foreigners was found to be 13813 words, and the number of different words was found to be 3757. The ratio of the total number of words to the number of different words was found to be 0.24.

The vocabulary load of A1 level coursebook of Türkçeye Yolculuk book set for teaching Turkish to foreigners was found to be 19308 words, and the number of different words was found to be 4351. The ratio of the total number of words to the number of different words was found to be

0.22. A1 level coursebook of Türkçeye Yolculuk book set for teaching Turkish to foreigners is noted as the richest with its vocabulary load.

At the first stage of the research, a corpus of YDTO A1 level textbooks was created; the vocabulary loads of sub-corpora was examined. The data-rich one of the sub-corpora was found to be that of A1 level coursebook of Yedi İklim set with 19677 words. A1 level coursebook of Türkçeye Yolculuk book set was found to be the second among the sub-corpora with 19308 words in terms of data richness. Another sub-corpus, A1 level coursebook of Dilmer book set ranked third with 19061 words.

In terms of type/token (all words/different words) ratio, the A1 corpus was found to have a value of 0.12. Of the sub-corpora, A1 level textbook of Gazi YDTÖ had the highest value with 0.28 in terms of type/token ratio; another sub-corpus, A1 level textbook of Dilmer YDTÖ had the lowest value with 0.16. Türkçeye Yolculuk was found to have 0.22, Yedi İklim had a value of 0.26, İstanbul YDTÖ had a value of 0.24, Dedam YDTÖ had a value of 0.23, and Altay YDTÖ had a value of 0.24. Low type/token ratio in A1 corpus is assumed to be related to the content aiming at meeting basic language teaching and basic communication needs at a basic level. After evaluating the data set properties of the A1 corpus, the focus was on the frequency lists created based on types using this corpus.

As Özdemir (2013: 2052) highlighted, there should be complete compatibility between the "language " of the frequency lists and textbooks used in teaching Turkish as a foreign language and the language level of the taught students. To achieve this, authors of course books should first consider at what level the material that they have prepared will be used and the vocabulary treasure of the student groups at that stage. A1 corpus, created within the scope of this research, contains books used in teaching Turkish as a foreign language. Therefore, it could be claimed that the word lists created with this study are important both for learners and book publishers.

As true for all languages, the first 1000 words most frequently used in a language meet about 80% of the products made with that language (Aksan 2000). This is similar in the studies conducted on the content compiled from a book or a social group. In the study conducted for English language (Nation 1990, cited in Lipinski 2010) states that the 4,000-5,000 words most frequently used make up 95% of a written text, and the 1,000 words most frequently used make up 85% of speaking production. It seems that the words that stand out with their frequency serve as a bridge in written or oral communication processes. In this context, 7 A1 level textbooks used in teaching Turkish as a foreign language were structured as sub-corpora; 250 words in the most frequently used noun, adjective and verb forms were identified within the A1 corpus. In frequency, the first 250 words were evaluated according to their frequency values and thus a total of 750 different words were identified.

When we consider the most frequently used verbs such as *git, iste, ye al, ol, çalış, bak, gör, çalış, konuş, oku, yaz* stood out in the A1 corpus with their frequency values. These verbs which are most frequently used in A1 corpus makes researchers think that the criteria for functional use of the language and meeting the needs of the learner were observed. 250 most frequently used nouns were identified in A1 corpus. In this context, nouns such as *saat, ev, kitap, okul, soru, cevap, masa, oda, araba, okul, öğrenci, sınav telefon, anne, baba, kardeş* stood out with their frequency values. It is seen that basic communication needs and the items to be used in such contexts were observed. The most frequently used 250 adjectives were identified in A1 corpus. The examination revealed that the adjectives such as *doğru, yanlış, iyi, aşağı, yukarı, güzel, çirkin*, stood out with their frequency values. Besides, descriptive and quantity adjectives stand out in the corpus with their frequency values.

In the corpus-based frequency study conducted by Aksan (2017), it could be claimed that A1 level corpus is competent in terms of the ability to represent A1 corpus, which seems to largely match with the most frequently used nouns, adjectives and verbs in Turkish. Frequency lists created based on the corpus in the form of the noun, adjective and verbs could be claimed to be functional in meeting the needs of learners and teachers and to be one of the steps taken to meet one of the significant needs. It can also be used as a reference source for textbook publishers.

### **Suggestions**

- The vocabulary loads of the teaching materials used in teaching Turkish as a foreign language should be organized based on the relevant corpus.
- These lists, whose frequency has been identified based on the corpus, can be beneficial in teaching processes.
- Similar corpus-based researches could be conducted.

### **References**

- Aksan, D. (2000). Her yönüyle dil ana çizgileriyle dilbilim. TDK Yayınları
- Aksan, D. (2014). *Anadilimizin söz denizinde* (3. Basım). Bilgi Yayınevi.
- Aksan, D. (2015). *Türkçenin sözvarlığı* (birinci basım). Bilgi Yayınevi.
- Aksan, D. (2017). *Türkçenin gücü* (15. basım). Bilgi Yayınevi.
- Aksan, Y., Aksan, M., Mersinli, Ü. ve Demirhan, U. U. (2017). *A frequency dictionary of Turkish*. London and New York: Routledge.
- Aksan, Y., Demirhan, U. (2015). Genel adlar: Sıklık listesinden işlevsel ulamlara. 28. *Ulusal Dilbilim Kurultayı Bildiri Kitabı*, 3-17. Ayrıntı Basım Yayın ve Matbaacılık.
- Baş, B. (2011). Söz varlığı ile ilgili çalışmalarda kullanılacak ölçütler. *Türklük Bilimi Araştırmaları*, 29(29), 27-61.



- Elena, T. B. (2001). Corpus linguistics at work. *Studies in Corpus Linguistics, John Benjamins Publishing Company*.
- 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.
- Kara, M , Ulutaş, M . (2018). Gazi Üniversitesi Vakfı Özel Anadolu Lisesi Öğrencilerinin Kelime Hazinesinin Derlem Tabanlı İncelenmesi. *Uluslararası Türkçe Edebiyat Kültür Eğitim (TEKE) Dergisi*, 7(3), 1902-1916.
- Karadağ, Ö. (2019a). Neden Bir Çocuk Dili Derlemine İhtiyacımız Var?. *Ana Dili Eğitimi Dergisi*, 7(3), 765-780.
- Karadağ, Ö. (2019b). Aynı Sınıf düzeyi için hazırlanan Türkçe ders kitaplarının ortak sözcük varlığı açısından karşılaştırılması. *Ana Dili Eğitimi Dergisi*, 7(4), 1130-1140.
- Karatay, H. (2007). Kelimesözcük öğretimi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 27(1), 141-153.
- Karatay, H. (2020). Türkçenin yabancı dil olarak öğretiminde söz varlığı ve geliştirilmesi. H. Karatay (Ed.), *Türkçenin Yabancı Dil Olarak Öğretimi El Kitabı* (s. 253-270). Pegem Akademi Yayınları.
- Lipinski, S. (2010). A Frequency Analysis of Vocabulary in Three First-Year Textbooks of German. *Die Unterrichtspraxis/Teaching German* 43 (2), 167-174.
- 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.
- McEnery, T., Andrew W. (2004), *Corpus Linguistics- An Introduction*, 2. basım, Edinburg Üniversitesi Yayını.
- McEnery, T., Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Nation, I. S. P. (1990) *Teaching and learning vocabulary*. Boston: Heinle & Heinle.
- Polat, T. (2012). Yazınsal metinler ve yabancı dil olarak Almanca öğretimi. *Alman Dili ve Edebiyatı Dergisi*, 8, 181-190.
- Özkan, B. (2010). Türkçenin öğretiminde sıfatların eşdizim sözlüğü: yöntem ve uygulama ve *Uluslararası Eğitim Araştırmaları Dergisi*, 1(2), 51-65.
- Özkan, B. (2013). Yöntem ve uygulama açısından “Türkiye Türkçesi söz varlığının derlem tabanlı sözlüğü”. *Bilig*, 66, 149-178.
- Sampson, Geoffrey and Diana McCarthy (2005). *Corpus Linguistics: Reading in a Widening Discipline*.
- Şimşek, R. (2020a). B1 düzeyi ders materyallerinin söz varlığının ve B1 düzeyi öğrencilerin yazılı söz varlığının çeşitli değişkenler açısından derlem tabanlı incelenmesi. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, (Ö7), 152-171.
- Şimşek, R. (2020b). Türkçeye Yolculuk B1 Düzeyi Ders Kitabının Söz Varlığı ve B1 Düzeyi Öğrencilerin Yazılı Söz Varlığının İncelenmesi. *Journal of History School*, 48, 3274-3293

- Teubert, W., Čermáková, A. (2004). Directions in corpus linguistics. *Lexicology and corpus linguistics*, 113-166.
- Vardar, B. (2002). Açıklamalı Dilbilim Terimleri Sözlüğü. Multilingual.
- Weisser, M. (2016). *Practical corpus linguistics: An introduction to corpus-based language analysis* (Vol. 43). John Wiley & Sons.
- Yılmaz, T. , Doğan, Y. (2014). 7. sınıf öğrencilerinin anlamını bilmedikleri kelime sözcükler ve Türkçe ders kitaplarındaki kelime sözcük çalışmaları bağlamında kelime sözcük öğretimi. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(25), 279-295.