THE ANALYSIS OF ARKADAŞ TÜRKÇE SÖZLÜK (ARKADAŞ TURKISH DICTIONARY) AND THE SUGGESTED MODIFICATIONS FOR ITS LEARNER'S VERSION

Sinan Çakır[1]

Adıyaman University

Abstract: This study aims to draw attention to two of the problems that Turkish lexicography faces today. One of these problems is that there are not any Turkish to Turkish dictionaries that have been prepared for the people who are learning Turkish as a second or foreign language. Another problem that faces Turkish lexicography is the deficiency of a corpus-based approach in dictionary building process. In this study, the layout of a dictionary that was prepared for the native speakers of Turkish was analyzed, and some modifications were suggested for the learner's version of this dictionary. The importance of corpus based approach in dictionary building process is the other point that is emphasized throughout the study. The study consists of two parts. In the first part, the modifications that can be made on a Turkish-to-Turkish dictionary were determined through the analyses of two English-to-English dictionaries, one of which was prepared for the native speakers of English and the other for the foreign learners of English. The possible modifications that can be done in the front-back matters and the A-Z entries of the Turkish Dictionary were presented in a list. In the second part,

the entry structures of three headwords presented in the Turkish dictionary were modified to be used in the suggested "learner's version" of this dictionary. In the study, the data obtained from the web were used as a mini-corpus to show the advantages of having a corpus based approach in lexicography. This mini corpus was used in the modification of the entry structures of the target headwords.

Keywords: *Turkish lexicography, dictionary for foreigners, corpus linguistics*

ARKADAŞ TÜRKÇE SÖZLÜK'ÜN İNCELENMESİ VE ÖĞRENCİ VERSİYONU İÇİN ÖNERİLEN DEĞİŞİKLİKLER

Öz: Bu çalışma, Türkiyedeki sözlükçülük çalışmalarının karşı karşıya olduğu iki sorunu ele almaktadır. Bunlardan birincisi, bugüne değin Türkçeyi öğrenen yabancılar için hazırlanmış bir sözlüğün mevcut olmamasıdır. Bir diğer sorun da sözlükçülük çalışmalarında derleme dayalı bir yaklaşımın takip edilmemesidir. Bu çalışmada, Anadili Türkçe olan kullanıcılar için hazırlanmış bir Türkçe-Türkçe sözlük incelenmiş ve bu sözlüğün yabancı kullanıcılar için hazırlanabilecek "öğrenci versiyonu" için uygun olacak bazı değişiklikler önerilmiştir. Çalışma boyunca, sözlük hazırlama sürecinde derleme dayalı bir yaklaşımın takip edilmesinin gerekliliği vurgulanmaktadır. Çalışma iki bölümden oluşmaktadır. Birinci bölümde, biri anadili İngilizce olan kullanıcılar için, diğeri de bu dili öğrenen yabancılar için hazırlanmış iki İngilizce-İngilizce sözlük incelenmiş ve seçilen bir Türkçe sözlük üzerinde yapılabilecek değişiklikler saptanmıştır. Seçilen Türkçe sözlüğün ön ve arka kısımlarında ve A-Z girdilerinde yapılabilecek değişiklikler liste halinde sunulmuştur. İkinci bölümde, Türkçe sözlük içerisinde sunulan üç maddebaşı sözcüğün yapısı bu sözlüğün yabancılar için hazırlanacak versiyonunda kullanılabilecek şekilde değiştirilmiştir. Çalışmada, derleme dayalı sözlük geliştirmenin faydalarını vurgulamak üzere internetten elde edilen verilerle bir mini derlem oluşturulmuştur. Seçilmiş olan üç maddebaşı sözcükte değişiklikler yaparken bu derlemden yararlanılmıştır.

Anahtar sözcükler: *Türk sözlükçülüğü, yabancılar için hazırlanan sözlük, derlem dilbilim*

1. INTRODUCTION

Turkish lexicography faces some important problems today. One of such problems is that the number of the dictionaries produced for different target groups are very limited. For instance, there are not any dictionaries produced for the learners who are learning / acquiring Turkish as a second or foreign language. As Jackson (2002) states, the foreign learners have special needs and there should be dictionaries which are specially developed to meet their needs (p. 129). The foreign learners of Turkish, however, have to use the dictionaries which were produced for the native speakers of Turkish, and such dictionaries do not take their needs into account. For instance, majority of such dictionaries do not provide information on cultural issues and grammatical points. As Jackson further points out, the foreign learners of a language might need more information on grammatical and cultural points compared to the native speakers of that language, and the monolingual learners' dictionaries should present comprehensive information on such points (2002, pp. 135-140).

The foreign learners of Turkish may have difficulty in figuring out the word stress or the division of the syllables as well. They may also mispronunce the words since they are not fully acquainted with the Turkish articulatory system. However, majority of the Turkish dictionaries do not provide information on such points and there is a need for a dictionary that is specifically designed for the foreign learners of Turkish.

Another problem that faces Turkish lexicography is the deficiency of a corpus-based approach. Corpus linguistics studies have been carried out in the Western World for a few decades and the linguists who are working on this field have already established extensive corpora in those languages. The dictionaries that are produced by them lean on the data that are taken from these corpora. As Meijs (1996) states:

Over the past ten to fifteen years, the discipline of lexicography has changed almost beyond recognition. This change is due to the technological revolution which has computerized the lexicographers'

working environment to a very high degree and which has permitted a veritable quantum leap in the amount and variety of resources that can be brought to bear on the lexicographical process. The most important of these resources are computerized corpora of real, mostly written, but now increasingly also spoken, running text (p. 99).

Similarly, Vitayapirak & Ratiroch (2006) point out that corpus-based approach to linguistic issues has become increasingly popular in the last decades. The compilation and the analysis of corpora stored in computerized databases provide a resource for the description of language at large. Their potential for describing specialized forms of language has also been recognized (p. 1). The first entirely corpus-based dictionary—COBUILD1—came out in 1987, it was on the basis of a corpus of around 20 million words of connected text (Ooi, 1998, pp. 36-37). Now all major British dictionary publishers use corpora of at least one hundred million words of text.

Sinclair defines corpus as: "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (in Atkins & Rundell, 2008, p. 54). Atkins and Rundell extend this definition as: "a lexicographic corpus must be a genuine – and inclusive- snapshot of a language, not a set of texts that have been specially chosen to advance someone's notion of what constitutes 'good' usage" (2008, p. 56). Hence, a reliable dictionary is the one whose generalizations about word behavior approximate closely to the ways in which people normally use language when engaging in real communicative acts.

Rather than leaning on a corpus, the dictionaries have been built by traditional methods in our country so far, and the lexicographers have determined the data that would be used in the dictionary by their own selection criteria (Özkan, 2013). However, the data for the dictionaries should not be composed of just the words that exist in the mental lexicon of their producers, but it should contain the words that are used in real life situations (Kocaman, 1998, 112). Tietze & Kurtböke (1996) underline this problem in Turkish lexicography as well. They claim that since there has been little study on the real-life usages of the words, the lexicographers have to find out all the usages of the

words by themselves (p. 69). Similarly, Atkins & Rundell (2008) state that consulting ones own mental lexicon, is a form of evidence, but it cannot form the basis of a reliable dictionary alone, since one individual's store of linguistic knowledge is inevitably incomplete and idiosyncratic (pp. 46-47). Thus, it is clear that dictionaries should rely on corpora to be able to reflect the languages more objectively. Çotuksöken (1996) also underlines the imporance of corpus in lexicography studies and states that the dictionaries should be based on consciously collected real life usages of the words (p. 87).

Fortunately, there are promising developments on this issue in the recent years. For instance, Göz (2003), Ölker (2011) and Çal (2015) carried out corpus-based studies to determine the most frequently used words in Turkish in different time periods. Their studies contained word pools of 975.141 words, 929.015 words and 549.366 words respectively. Such corpora are useful for determining the most frequently used words in Turkish that can be taught to the foreign learners of Turkish primarily. In addition to such studies, there are far larger corpus-based projects in Turkey as well. "ODTÜ Sözlü Türkçe Derlemi Projesi" (Spoken Turkish Corpus Project by METU), "ODTÜ- Sabancı Türkçe Ağaç Yapılı Derlem Projesi" (The Corpus Project of METU- Sabacı Turkish Tree Bank), Türkçe Ulusal Derlemi (Turkish National Corpus), TS Corpus Project are some of the projects that have been carried out to produce extensive corpora for Turkish (Sarıgül, 2011; Aksan & Aksan, 2009; Say et.al, 2002). By making use of such corpora, many projects have been carried out in the recent years. For instance, Özkan et al. (2014) reported the results of the TUBİTAK (The Institution of Scientific and Technological Researches in Turkey) projects that were carried out in accordance with the principles and methodology of corpus linguistics. The findings of the TUBITAK projects like "Türkiye Türkçesi Çevrim İçi Haber Metinlerinde Yeni Sözlerin Otomatik Çıkarımı" (The Automatic Inferences of the New Words in the News Reports in Turkish), "Türkiye Türkçesi Eşdizim Sözlüğü'nün Sayısallaştırılması" (Digitization of the Collocation Dictionary of Turkish), "Türkçe'nin Tarihsel Derlemini Hazırlamak" (Preparing the Historical Corpus of Turkish), "Derlem Tabanlı Çevrim İçi Türkçe Öğrenici Sözlüğü – Önadlar A Maddebaşı-" (The Corpus-based Online Turkish Learner's Dictionary – Adjectives A Headwords-) are presented in their book, which promise highly positive developments in this field. In the recent

years, carrying out such studies is on the rise and in the near future, it is highly possible that new dictionaries leaning on large corpora will be compiled.

1.1. THE PURPOSE OF THE STUDY

The purpose of this study is to suggest a dictionary structure which appeals to the needs of the foreigners. That is to say, the aim of this study is to analyze the layout of a dictionary that is prepared for the native speakers of Turkish, and to suggest some modifications for the learner's version of that dictionary. The importance of a corpus based approach in dictionary building process is the other point that is emphasized throughout the study.

2. METHODOLOGY

The study consisted of two main parts. In the first part, the modifications that could be made on a Turkish-to Turkish dictionary were determined through the analyses of two English-to-English dictionaries, one of which was prepared for the native speakers of English and the other for the foreign learners of English. The names of the dictionaries are *Concise Oxford Dictionary* (COD, hereafter) and *Oxford Advance Learner's Dictionary* (OALD, hereafter) respectively. One of the reasons for the selection of these dictionaries was that they were produced by the same institution (Oxford). Since both of them were produced by Oxford, they have certain similarities in the layout of A-Z entries or in their front and back matters. Nevertheless, they were produced for different target groups. Hence, the analysis of these dictionaries might provide valuable data to show the differences between the dictionaries for native speakers and foreign learners. Another reason for the selection of these dictionaries was that their publication dates are close to each other: 2001 and 2005 respectively. Therefore, there would not be much update differences among them. Another reason was that they are almost at the same size. Because of these reasons, these dictionaries were selected for comparison. The front and the back matters, and the A-Z entries of these dictionaries were compared and the similarities and the differences were listed.

Then, a Turkish to Turkish dictionary that was designed for the native speakers of Turkish was selected. In the selection of this dictionary,

the charecteristics like size and publication date were taken into consideration. That is to say, in order to be able to make valid comparsions, its size, length and publication date should have been close to the ones for the Oxford dictionaries. Therefore, *Arkadaş Türkçe Sözlük (Arkadaş Turkish Dictionary)* which was prepared by Ali Püsküllüoğlu in 2005 was selected. The structure of this dictionary was analyzed and the findings were compared with the ones that were obtanied in the analysis of two English-to English dictionaries. In the final stage, the modifications that can be made in the front-back matters and the A-Z entries of the Arkadaş Türkçe Sözlük (ATS hereafter) were given as a list. However, it should be noted that the purpose of this study is not to find the weak points of ATS and suggest modifications for them. ATS is a dictionary prepared for native speakers, and in this respect, it might be rather sufficient. However, as it has been expressed before, foreign learners have special needs and learner's dictionaries should meet these needs. Hence, the purpose of this study is to suggest modifications for the possible learner's version of ATS.

In the second part, the entry structures of three headwords presented in the ATS were modified to be used in the suggested "learner's version" of this dictionary. The study also aimed to show the advantages of leaning on a corpus in dictionary building process. Therefore, the structures of only three headwords were modified as a sample. These headwords were: "*eğlenmek*" *(to have fun),* "*döşek*" *(mattress)* and "*kurnaz" (cunning),* a verb, a noun and an adjective respectively. These headwords were selected arbitrarily. That is to say, no other creteria were applied in the selection of them except for being in different word forms. There might be differences among the entries of the nouns, verbs or adjectives etc. Hence, selection of words in different forms (a noun, a verb and an adjective) was necessary to make the study more comphrehensive.

3. THE ANALYSIS

3.1. THE ANALYSIS OF CONCISE OXFORD DICTIONARY AND OXFORD ADVANCED LEARNER'S DICTIONARY

Table 1 below presents the information that can be found in both dicttionaries. The criteria for comparsion have been determined by the author himself, but he mainly made use of the analysis methods of Atkins & Rundell (2008).

Table 1. The information found in both COD and OALD

No	Information found in the dictionaries
1	Guide to the use of the dictionary
2	Basic information about labels, spelling and pronunciation
3	A short list of abbreviations used in the dictionary
4	Senses which are labelled by numbers
5	Frequent use of labels (domain, region, register, style, meaning type)
6	Multiword expressions which are given under the same headword
7	Comparative and superlative forms of adjectives and adverbs (e.g. *fast- er- est*)
8	The grammatical labels (parts of speech: noun, verb, adj, adv etc.)
9	Derived forms of the words (e.g. *slow- slowly*)
10	Inflected forms of the verbs (e.g. *go, went, gone, going*)
11	The plural forms of the irregular nouns are given (e.g. *man- men*)
12	In every page, examples showing the pronunciation of vowels and consonants in English
13	List of countries of the world
14	Commonly confused pairs of words (e.g. *affect-effect*)
15	Basic information about nouns, verbs, commas, semi-colons etc.
16	The list of irregular verbs
17	Abbreviations used in electronic text messaging (e.g. *SMS*).

Some of these 17 charecteristics can be seen as basics for all dictionaries. That is to say, they can be found in almost every dictionary (guide to the use of the dictionary, a list of abbreviations, basic information about labels, spelling, pronunciation etc.). The other similar points must be the ones which are regarded as necessary for both natives and foreigners who are going to use the dictionaries. In other words, these charecteristics must have been viewed as necessary for both of the target groups by the dictionary producers.

The table below demonstrates the information that are found only in *Concise Oxford Dictionary*:

Table 2. The information found only in COD

No	Information found in the dictionary
1	Weights and measures (e.g. *milligram, kilogram*)
2	Terms for groups of animals (e.g. *a pride of lions, a herd of elephants*)
3	Proofreading Marks (symbols for correction, deletion etc.)
4	Alphabets (e.g. *Arabic, Hebrew, Greek*)
5	Etymological information for some words

These five charecteristics can be obtained only in the Concise Oxford Dictionary. Etymological information for the words, and the alphabets that are used by different nations may not be very necessary for a non-native learner of English. Thus, it is plausible that they are not included in the learner's dictionary. The other three charecteristics (weights and measures, terms for groups of animals and proofreading marks), must have been regarded as not "vital" for the non-native learners of English either; so, they are not included in the learner's dictionary, either. In the table below, the information found only in *Oxford Advanced Learner's Dictionary are* demonstrated:

Table 3. The information found only in OALD

No	Information found in the dictionary
1	Indication of syllable divisons of headwords
2	IPA transcription of all headwords
3	Word stress which is shown with an apostrophe in the phonetic transcription
4	Countable and uncountable nouns are displayed by abbreviations C and U
5	5 Maps of English speaking countries
6	24 pages of colour illustrations
7	Cues for writing essays, CV's, letters, e-mails in English
8	Common suffixes and prefixes in English
9	Common personal and geographical names
10	The differences between American and British English
11	The most frequently used English proverbs and sayings
12	The place of English in the world as a lingua franca
13	The most basic 3000 words in English which are used in the definitions of the dictionary
14	Brief information about English grammar (the tenses, passives, conditionals, modal verbs, reported speech, phrasal verbs, relative clauses ..etc)
15	Culture-related numbers (their meanings of the numbers in American and British societies)
16	Detailed explanations for Key to Dictionary Entries (detailed information about how to use the dictionary);
17	Key to Verb Patterns (transitive- intransitive verbs, linking verbs, verbs used with clauses or phrases.. etc.

As it is obviously noticed, OALD provides far more information for its users compared to the other dictionary which was prepared for the native spakers of English. The 17 charecteristics that can be found in the learner's dictionary are not included in COD. The obvious reason

for this situation is that the non-native dictionary users need more information compared to the native users. Therefore, it is plausible that the dictionary that is prepared for them includes more information. For instance, it can be said that the non-native users of the English dictionaries need more information about the culture-related words, idioms or proverbs. They may need to know the most common personal and geographical names that are used in English speaking countries. Such information may not be vital for a native speaker of English, but they may be useful for non-natives. Thus, the learner's dictionary includes more information about the cultural points.

The non-native users of the dictionary, may need to get more information about the grammar, pronunciation and spelling of the target language as well. They may have just a little information about the grammatical structures in English. They may have difficulty in seperating the syllables, or finding out the stressed syllable in the words. Such information may not be necessary for a native language user, either. However, the non-native users of English may need such information.

3.2. THE ANALYSIS OF ARKADAŞ TÜRKÇE SÖZLÜK (ARKADAŞ TURKISH DICTIONARY)

In this section, the front-back matters and the A-Z entries of ATS were examined. The following front and back-matter information is presented in ATS:

3.2.1. THE FRONT MATTER

1- A preface; 2- The explanations about the dictionary; 3- A list of abbreviations used in the dictionary; 4- A list of symbols used in the dictionary.

3.2.2. BACK MATTER

There is not any information in the back matter of ATS.

In the analysis of OALD, it was seen that it displays detailed information about the grammar points, vocabulary and cultural issues in its front and back matter. Since foreign users are not as competent

as native users in the grammar forms or cultural issues, it is for sure that such detailed explanations are very beneficial for its target users. Compared to OALD, the information presented in the front and back pages of ATS is very limited. There is even no information at the back-matter. Therefore, when ATS is re-edited for the foreign users, the amount and the quality of the information presented in its front and back matter should be extended.

Yet, this does not mean that all of the information displayed in OALD should be presented in the new version of ATS. The difference between the languages should always be taken into account. That is to say, for an English dictionary, a list of irregular verbs, or a list that shows the differences between American and British English might be useful for its users; however, since there are not such irregularities in Turkish, there is no need to provide information on such points.

3.2.3. THE A-Z ENTRIES

The main characteristics of the A-Z entries of the dictionary are as follows:

3.2.3.1. THE HEADWORDS

- Letter by letter alphabetization is used.
- Secondary headwords are not used.
- Multiword expressions are not treated as headwords. They are presented under the first content word of the expression. If the first word is a private name, than the expression is displayed under its content word.
- Derived forms of the words are always presented as headwords (e.g. *cerrah (surgeon), cerrahi (surgical), cerrahlık (surgery)*).
- Senses are labelled with numbers.
- Homographs are listed as different headwords (e.g. *kara (black), kara (land)).*

The structure of the headwords in ATS looks similar to that of OALD. Yet, there is a vital difference between these dictionaries. The

headwords in OALD were selected from a corpus, yet, the ones in ATS do not lean on any corpus work. Thus, when a new dictionary is prepared in Turkish, it must lean on a corpus to be able to represent the vocabulary of Turkish in a more reliable sense.

3.2.3.2. THE SYLLABIFICATION, STRESS AND PRONUNCIATION

- The divison of syllables are not indicated.
- The phonetic transcription of the headwords are not presented.
- Word stress is not indicated.

As it is clear, ATS contains no information on syllabification, stress and pronunciation. Since its target users are native speakers of Turkish, there may not be a need for such information. However, for a foreigner, these points may be problematic. That is to say, when a foreigner wants to use this dictionary, it is likely that s/he will have some problems in seperating the syllables, finding out the stressed syllable, or pronuncing the words correctly. Therefore, when ATS is re-edited for foreign users, it should contain information about the syllabification, stress and pronunciation of all headwords.

3.2.3.3. THE GRAMMATICAL LABELS

Grammatical labels are presented just after the headword (e.g. a- for ad (noun); e for eylem(verb); s for sıfat (adjective), be for belirteç (adverb)). Intransitive verbs are marked by the nsz (nesnesiz (intransitive)) label.

3.2.3.4. OTHER LABELS (DOMAIN, REGION, REGISTER, STYLE, TIME, MEANING TYPE)

The dictionary possesses an extensive use of labels. The types and frequencies of the labels used in Arkadaş Türkçe Sözlük is similar to that of Oxford Advanced Learner's Dictionary. Both dictionaries use sufficient amount of labels. The only remarkable difference is that OALD contain limited amount of eytmological information while ATS presents information about the origins of the words extensively. Since OALD is a dictionary that is prepared for foreign users, the limited amount of etymological information is plausible. Thus, when the "learner's" version of ATS is prepared, there might be less (even no) information about the origins of the words, since such information

may not have much importance for the target users. The labels used in ATS are as follows:

a- Domain labels: The dictionary contains 36 domain labels Some of these labels are: (antr) anthropology, (gökb) astronomy, (tar) history, (hayv) zoology, (mat) mathematics.

b- Region Labels: The dictionary does not contain any region labels.

c- Register Labels: The following register labels are used in the dictionary: (arg) slang, (şaka) joke, (ka) vulgar, (ha) colloquial.

d- Meaning Type Labels: The meaning type labels that are used in the dictionary are: (mec) metaphorical, (hkr) insult, (al) mock.

e- Time Labels: Only Esk (old) time label is used in the dictionary.

f- Etymological Information: The dictionary provides etymological information for the words that were borrowed from foreign languages like: (Ar) Arabic, (Al) German, (Bul) Bulgarian, (Cer) Germenic, (Erm) Armanian, Far (Persian), (Fr) French, (İbr) Hebrew, (İng) English, (İsl) Slavic, (İsp) Spanish, (İt) Italian, (Lat) Latin, (Mac) Hungarian, Osm (Ottoman Turkish), (Rus) Russian, (Yun) Greek.

3.2.3.5. THE INFLECTED FORMS

Inflected forms of the verbs or nouns are not presented in the dictionary. This application is plausible since Turkish is not an inflected language. Similarly, comparative and superlative forms of the adjectives and adverbs are not presented in the dictionary. Since Turkish does not have irregular comparative and superlative forms, there is no need to include such information in the dictionary.

3.2.3.6. THE DEFINITIONS

The definitions are usually short and to the point, so they are easy for foreigners to understand. Yet, it would be useful to contain a list of most basic words in Turkish in the back matter. OALD contains a list of 3000 basic words of English in its back matter. Such an application would be useful for foreign users.

3.2.3.7. THE EXAMPLES

The dictionary does not provide examples for all headwords. There are examples for some of them, but the number of such cases is very limited. OALD provides examples for every headword and even for every sense. Compared with OALD, the examples presented in ATS are obviosly insufficient. Morever, the examples presented in ATS are not taken from a corpus, either. In its learner's version, it is for sure that more examples should be included and these examples should be taken from a corpus. Such examples are presented in the application part of this paper.

3.3. THE LIST OF SUGGESTED MODIFICATIONS FOR THE LEARNER'S VERSION OF THE ARKADAŞ TÜRKÇE SÖZLÜK

After the comparison of ATS with COD and OALD, the following changes can be made on the dictionary to prepare learner's version of it. All of the modifications suggested below might be beneficial for L2 learners of Turkish who are assumed to have limited information about Turkish culture and the charecteristics of Turkish language.

The points that are suggested to be removed from the dictionary:

1- There should be less (even no) etymological information. Such information might not be necessary for L2 learners.

The points that are suggested to be added to the new version:

3.3.1. TO THE FRONT-MATTER

1- Detailed information about how to use the dictionary.

2- Key to Verb Patterns in Turkish (Transitive-intransitive verbs, linking verbs, verbs used with clauses or phrases. etc.).

3.3.2. TO THE BACK- MATTER

1- Colour illustrated pages for vocabulary teaching.

2- Commonly confused pairs of words in Turkish.

3- The most basic 3000 words in Turkish which are used in the definitions of the dictionary.

4- Brief information about Turkish grammar (the tenses, passives, conditionals, modal verbs, reported speech, phrasal verbs, relative clauses etc.).

5- Common suffixes and prefixes in Turkish (with examples).

6- Cues for writing essays, CV's, letters, e-mails in Turkish

7- Common personal and geographical names in Turkish.

8- The most frequently used Turkish proverbs and sayings.

9- Culture-related numbers.

10- The place of Turkish in the world.

3.3.3. TO THE A-Z ENTRIES

1- In every page, examples that show the pronunciation of vowels and consonants in Turkish.

2- Indication of syllable divisions of headwords.

3- IPA transcription of all headwords.

4- Word stress which is shown with an apostrophe in the phonetic transcription.

5- Countable and uncountable nouns which are displayed by abbreviations after noun headwords.

6- One or more example sentence(s) or phrase(s) taken from a corpus for all senses of the headwords.

All of the modifications suggested above might be rather beneficial for the foreigners who have got limited knowledge on Turkish culture and Turkish grammar. For instance, providing information about the commonly used suffixes and prefixes in Turkish; showing syllable divisions of headwords, their IPA transcriptions and word stress might be beneficial for them to learn the chracteristics of Turkisk language. In a similar vein, providing information about culter-related numbers, the place of Turkish in the world, common personal and geographical names in Turkish, most ferquently used proverbs and sayings in Turkish might help them to get information about Turkish culture.

4. THE APPLICATION

In this part, the entry structures of three headwords presented in the Turkish dictionary were modified to be used in the suggested "learner's version" of this dictionary. To prepare the corpus of the study, the words *eğlenmek (to recret)*, *döşek (mattress)* and *kurnaz (cunning)* were searched by the Google Search Engine and 1000 results for each word (in total 3000 results) were analyzed. Since the verb *eğlenmek (to have fun)* has many inflected forms like *eğlendi (had fun)*, *eğlenecek (will have fun), eğleniyor (is having fun)* or *eğlendirdi. (made someone have fun).* etc, during the Google search, a different procedure was followed for this word. Five different forms of this word, *eğlendi (had fun), eğlenmek (to have fun), eğleniyor (is having fun), eğlenecek (will have fun)* and *eğlenir (has fun)* were written on the search engine and 200 results for each inflected forms were taken into account. In the search results, the full sentences that contained the target words were selected and added to the corpus. The ones which did not contain the use of the target words in full sentence were disregarded. For instance, short phrases like *döşek imalatı (mattress manufacturing), kurnaz kedi (cunning cat), eğlenmek kelimesinin anlamı (the meaning of the term "have fun")* etc are some of such ignored usages. The results in which the target words were used as private names were also excluded (*Ahmet Kurnaz, Ali Kurnaz, etc).* It was also noticed that the same sentences were repeated in many different websites, only one of such results were included to the corpus and the rest were ignored. Thus, from 3000 results, 433 sentences were selected and the corpus of the study was formed by these 433 sentences: 200 for *eğlenmek (to have fun)*, 159 for *döşek* (mattress) and 74 for *kurnaz (cunning).*

4.1. THE APPLICATION FOR THE HEADWORD "EĞLENMEK" (TO HAVE FUN)

For the headword *eğlenmek (to have fun)*, 200 sentences were selected from the Google search results. In the analysis of the 200 sentences, the following senses of this word have been determined:

1- Hoşça vakit geçirmek (to have fun, to recreate oneself): in 184 sentences.

2- Dalga geçmek (to make fun of): in 9 sentences.
3- Bulunmak (to exist): in 4 sentences.
4- Oyalanmak, vakit geçirmek (to play around): in 3 sentences.

As it is clear, the first sense of it, *hoşça vakit geçirmek (to have fun)*, is far more frequently used than its other senses, so it is clear that its this sense should come first in the dictionary. The place of other senses should be arranged in accordance with their frequency as well.

Having a corpus has another advantage. The example sentences for these senses could be selected from the corpus; thus, the real-life usages of them could be presented in the dictionary. For the senses mentioned above, following examples were selected from the corpus:

1- Hoşça vakit geçirmek (to have fun, to recreate oneself): Çanakkaleli romanlar gönüllerince eğlendiler. (Gypsies in Çanakkale had really great fun.)

2- Dalga geçmek (to make fun of): Eğlenir bizimle Hacı efendi keyifliylen (The hadji makes fun of us when he is cheery.)

3- Bulunmak (to exist): Arısız kovanda bal mı eğlenir? (Does honey exist in the hive without bees?)

4- Oyalanmak, vakit geçirmek (to play around): Bu sıcakta eğlenecek şey arayan millet anketörlerle dalgasını geçmiş! (The people who wanted to play around in this hot weather made fun of the pollsters!)

It should be made clear that the corpus of this mini- study is very limited and it may not reflect the real usage of this word. There might be more senses of it if it were to be analyzed in a larger corpus. That is to say that the application carried out here is just a sample. To prepare a full dictionary, it is for sure that a corpus that contains millions of words is necessary. Yet, the aim of this study is to show the objectivity of corpus usage in lexicography.

After the application of other suggested modification on the headword, the headword *eğlenmek (to have fun)* appears in our mini-dictionary as:

Eğ-len-mek (e:lenmek') gçz e. **1.** Hoşça vakit geçirmek, *Çanakkaleli Romanlar Gönüllerince* Eğlendiler. **2.** Dalga geçmek: *Eğlenir bizimle Hacı efendi keyifliylen.* **3**. Bulunmak. *Arısız kovanda bal mı* eğlenir? **4.** Oyalanmak, vakit geçirmek. *Bu sıcakta* eğlenecek *şey arayan millet anketörlerle dalgasını geçmiş!*

Eğ-len-mek (e:lenmek') int v. **1.** To have fun, to recreate oneself. *Gypsies in Çanakkale had really great fun.* **2.** To make fun of. *The hadji makes fun of us when he is cheery.* **3.** To exist. Does honey exist in the hive without bees? **4.** To play around. *The people who wanted to play around in this hot weather made fun of the pollsters!*

4.2. THE APPLICATION FOR THE HEADWORD "DÖŞEK" (MATTRESS)

159 sentences have been found for the headword *döşek (mattress)* in the Google search. After the analysis of these sentences, the following senses of it has been determined:

1- *Yatak (mattress, bed)*: 148 sentences, 86 of which were used in conventinal context.

2-Gemi bölümü (part of a ship): in 6 sentences.

3- Bina Tabanı (base of a building): in 5 sentences.

In ATS, the first sense of *döşek (mattress)*, has been simply defined as *yatak (bed).* Yet, when the corpus was analyzed, it was observed that defining it as simply *yatak (bed)* would be misleading for the foreigners. 86 of the 148 sentences found in the websites contain it in conventional contexts (like the conventional lifestyles of village life in Turkey, or in the ones which give information about Islam). A Turkish native speaker may realize that *döşek (mattress, bed)* has got a more conventional usage than *yatak (bed)*; but a foreigner may be misled if it is presented as synoynmous for *yatak (bed)*. Thus, there should be some information in the definition which makes its this usage clear.

Besides, ATS did not include the *bina tabanı (base of a building)* sense of this word. Yet in the five sentences of the corpus, this word was used in this sense. On the other hand, in ATS there is also another sense of this word: *çimlendirmek üzere tohum ekilen, fidan dikilen yer (the place where one seeds and dibbles to germinate).* In our mini-corpus, this sense of the word was not detected. It is for sure that in a larger corpus, it will also be possible to observe the use of that sense of *döşek (mattress)* as well (and may be other senses if there are any).

After the analysis of the corpus, the senses of this headword, its definitions and example sentences for them were determined as follows:

1- Yatak (mattress, bed) (Geleneksel bağlamlarda, "yatak" kelimesi yerine "döşek" kelimesi daha çok tercih edilir. "In conventional contexts, the term "mattress" is preferred more than the term "bed"). *Salman usta döşekleri serdi, yastıkları koydu (Master Salman laid the mattresses and put the pillows).*

2- Gemi gövdesinde, su basıncı, çarpma, karaya oturma vb. durumlarda darbeleri karşılayabilecek yapı gereci (the building material in the body of the ship that can resist water pressure, slamming and grounding). *Döşekler düzgün yayılı su basıncının etkisi altındadırlar. (This part of the ship is subject to uniformly distributed water pressure).*

3- Binaların temeli altına konulan betonarme kısım (The ferro-concrete section which is put under the base of a building). Binanın güney cephesine yerleştirilen döşek ek ısı kazancı için özel olarak tasarlanmıştır (The section placed south side of the building has been specially designed for heat gain).

After the application of other suggested modification on the headword, it appears in our mini-dictionary as:

Dö-şek (dØ∫ek') syln i. **1.** Yatak (Geleneksel bağlamlarda, "yatak" kelimesi yerine "döşek" kelimesi daha çok tercih edilir). Salman usta döşekleri serdi, yastıkları koydu. **2.** Gemi gövdesinde, su basıncı, çarpma, karaya oturma vb. durumlarda darbeleri karşılayabilecek yapı gereci. Döşekler düzgün yayılı su basıncının etkisi altındadırlar. **3**. Binaların temeli altına konulan betonarme kısım. Binanın güney cephesine yerleştirilen döşek ek ısı kazancı için özel olarak tasarlanmıştır.

Dö-şek (dØ∫ek') count n. **1.** Mattress, bed (In conventional contexts, the term "mattress" is preferred more than the term "bed"). Master Salman laid the mattresses and put the pillows. **2.** The building material in the body of the ship that can resist water pressure, slamming and grounding. This part of the ship is subject to uniformly distributed water pressure. **3.** The ferro-concrete section which is put

under the base of a building. The section placed south side of the building has been specially designed for heat gain.

4.3. THE APPLICATION FOR THE HEADWORD "KURNAZ" (CUNNING)

In the 1000 entries that Google search engine provides for the headword *kurnaz* (cunning), 74 sentences were observed to contain the target word. Since the word *kurnaz (cunning)*, is also used as a surname in Turkey, majority of the search results displayed private names like (*Murat Kurnaz,* or *Ahmet Kurnaz etc.)*, and all these usages were ignored. Thus only 74 sentences that contain this word have been included in the corpus.

In the analysis of the corpus, it was noticed that in all of the sentences, *kurnaz* (cunning) was used in the same sense. So, it is possible to say that it has got only one sense. The definition presented in ATS was not changed for this word. Yet, an example sentence from the corpus was added. After the application of the other suggested modification on the headword, "*kurnaz*" appears in our mini-dictionary as:

Kur-naz (kʊrnʌz') s**.** Başkalarını kandırmasını ve ufak tefek oyunlarla amacına ulaşmayı beceren, açıkgöz (kimse). *Gerçek paraları sahteleriyle değiştiren kurnaz avukat yakalandı.*

Kur-naz (kʊrnʌz') adj**.** A vigilant (person) who succeeds in deceiving others and reaching at his/her goals with small tricks. *The cunning lawyer who changed the real money with the fake ones has been caught.*

5. CONCLUSION

This study aimed to draw attention to two of the problems that Turkish lexicography faces today. There is a need for a dictionary which has been prepared for the learners who acquire Turkish as a foreign or second language. These learners have different needs compared to the native speakers of Turkish. Therefore, it is necessary to to build a dictionary that appeals to their needs. In Western World, such dictionaries have already been produced: The Oxford Advanced Learner's Dictionary, Cambridge International Dictionary of English for English, Dictionnaire Du Français Contemporain for French,

Diccionario Para La Ensenza De La Lengua Espanola for Spanish etc. (Hartmann, 2001, p. 76). Hence, it is time to produce a monolingual Turkish dictionary for the foreign learners of this language. Secondly, a corpus based approach should be developed in the preparation of the dictionaries in Turkey. It is certain that the information that were collected from the web cannot be sufficient to produce a full dictionary. Nevertheless, the mini corpus that was used in this study aimed to show some of the advantages of having a corpus-based approach in dictionary building process. Finding the senses of the headwords, determining the order of these senses in accordance with their frequency of occurance and selecting real-life examples for the senses are some of such advantages observed in the present study. Hence, the dictionaries should lean on corpora. Fortunately, there are promising developments on this issue. "ODTÜ Sözlü Türkçe Derlemi Projesi" (Spoken Turkish Corpus Project by METU), "ODTÜ-Sabancı Türkçe Ağaç Yapılı Derlem Projesi" (The Project of METU-Sabacı Turkish Tree Bank), Türkçe Ulusal Derlemi (Turkish National Corpus), TS Corpus Project are some of the projects that have been carried out to produce extensive corpora for Turkish (Sarıgül, 2011; Aksan & Aksan, 2009; Say et al., 2002). Hopefully, in the near future, the dictionaries that are produced in Turkish will lean on such corpora.

REFERENCES

Aksan, Y. & Aksan M. (2009). Building a national corpus of Turkish: Design and implementation. *Working Papers in corpus-based linguistics and language education, 3* (pp. 299-310). Tokyo: Tokyo University of Foreign Studies.

Atkins, S. and Rundell M. (2008). *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.

Çal, A. (2015). Türkiye'de farklı dönemlere ait kelime sıklığı çalışmaları üzerine bir değerlendirme. *International Periodical for the Languages, Literature and History of Turkish or Turkic, 10 (8),* 715-730.

Çötüksöken, Y. (1996). Türk sözlükçülüğünde sorunlar. *Kuram Dergisi, 11,* 85-90.

Göz, İ. (2003). *Yazılı Türkçenin kelime sıklığı sözlüğü*. Ankara: TDK Yayınları.

Hartmann, R. R. K. (2001). *Teaching and researching lexicography*. Harlow: Pearson Education Limited.

Jackson, H. (2002). *Lexicography: An introduction*. London: Routledge, Taylor & Francis Group.

Kocaman, A. (1996). Sözlükçülükte ikinci boyut üzerine düşünceler. *Kuram Dergisi, 11,* 81-84.

Kocaman, A. (1998). Dilbilim, sözlük, sözlükçülük. *İnsanbilimleri İçin Kaynak Araştırmaları Dergisi, 6,* 111-113.

Meijs, W. (1996). Linguistic corpora and lexicography. *Cambridge Journals, Annual Review of Applied linguistics,* 99-114.

Ooi, V. B.Y. (1998). *Computer corpus lexicography.* Edinburg: Edinburg University Press.

Ölker, G. (2011). *Yazılı Türkçenin kelime sıklığı sözlüğü (1945-1950 arası).* Konya: Kömen Yayınları.

Özkan, B. (2013). Yöntem ve uygulama açısından Türkiye Türkçesi söz varlığının derlem tabanlı sözlüğü. *Bilig, 66*, 149-178.

Özkan, B., Tahiroğlu, B. T. & Özkan, E. A. (2014). *Türkçe üzerine derlem dilbilim uygulamaları.* Adana: Karahan Kitabevi.

Sarıgül, K. (2011). Dilbilim, sözlükbilim, bilgisayar. *Dil ve Edebiyat Araştırmaları Dergisi, 4*, 41-53.

Say, B., Zeyrek, D., Oflazer, K. & Özge, U. (2002). Development of a corpus and a treebank for present-day written Turkish. In the proceedings of the eleventh international conference of Turkish linguistics.

Tietze, A. & Kurtböke, P. (1996). Türk sözlükçülüğünün sorunları 1. *Kuram Dergisi, 11*. 65-69.

Vitayapirak J. & Ratiroch P. (2006). Computational approach for processing of control engineer text: applications for corpus lexicography (pp. 1-6). Cybernetics and Intelligent Systems, IEEE Conference: 7-9 June 2006.

DICTIONARIES ANALYZED IN THE STUDY

Arkadaş Türkçe Sözlük. (2005). A. Püsküllüoğlu (ed.). Ankara: Arkadaş Yayınevi.

Concise Oxford Dictionary. (2001). J. Pearsall (ed.). Oxford: Oxford University Press.

Oxford Advanced Learner's Dictionary. (2005). S. Wehmeier et al. (eds.). Oxford: Oxford University Press.