The Real-life-based
School Dictionary for Turkish*

Bülent Özkan, *Turkish Language Teaching Department,*
Mersin University, Turkey (ozkanbulent@mersin.edu.tr)

Abstract: This report aims to describe the development of a new, original, and monolingual dictionary as an educational resource named the Real-life-based School Dictionary (RLBSD). There is a general lack of corpus-based school dictionaries in the Turkish context. The existing dictionaries are based on native speaker intuitions and are far from being empirical, indicating the need for the development of a corpus-based dictionary. In this respect, the RLBSD meets this critical need. A specialized corpus consisting of texts from a range of natural language environments such as coursebooks (from primary to high school) approved by the Ministry of Education (MoNE), children's literature books, and periodicals was compiled and forms the basis of the RLBSD. The dictionary has been designed with the use of information technologies and corpus linguistics methods and is accessible to users on all digital platforms (the Internet, smart board, tablet, smartphone) as an instructional tool. It is believed that the current study may serve as an exemplary work with its methods and outputs for Turkish lexicography studies.

Keywords: SCHOOL DICTIONARY, LEXICOLOGY, TURKISH EDUCATION, REAL-LIFE-BASED DICTIONARY, CORPUS LINGUISTICS

Opsomming: Die werklikheidsgebaseerde skoolwoordeboek vir Turks. In hierdie verslag word die ontwikkeling van 'n nuwe, oorspronklike, en eentalige woordeboek as opvoedkundige hulpbron, naamlik die Real-life-based School Dictionary (RLBSD), beskryf. Daar bestaan 'n algemene tekort aan korpusgebaseerde skoolwoordeboeke in die Turkse konteks. Die bestaande woordeboeke is gebaseer op die intuïsies van moedertaalsprekers en is nie naastenby empiries gefundeer nie, wat dui op die behoefte aan die ontwikkeling van 'n korpusgebaseerde woordeboek. Ten opsigte hiervan, voorsien die RLBSD in hierdie kritiese behoefte. 'n Gespesialiseerde korpus, bestaande uit 'n groot verskeidenheid tekste uit natuurlike taalomgewings soos handboeke (van primêre tot hoërskool) wat deur die Ministerie van Opvoedkunde (MoNE) goedgekeur is, kinderlektuur, en tydskrifte is saamgestel en vorm die basis van die RLBSD. Die woordeboek is ontwerp met gebruikmaking van inligtingstegnologie en korpuslinguistiese metodes en is

op alle digitale platforms (die Internet, interaktiewe witbord, tablet, slimfoon) vir gebruikers toe-ganklik as onderrighulpmiddel. Daar word aanvaar dat die huidige studie met sy metodes en resultate as voorbeeld kan dien vir Turkse leksikografiese studies.

Sleutelwoorde: SKOOLWOORDEBOEK, LEKSIKOLOGIE, TURKSE OPVOEDKUNDE, WERKLIKHEIDSGEBASEERDE WOORDEBOEK, KORPUSLINGUISTIEK

1.   Introduction

The terms *school dictionary*, *children's dictionary,* and *college dictionary* in lexico-graphic literature all refer to "a type of dictionary written for school-age chil-dren" (Hartmann and James 1998: 122) and have certain common characteristics. They contain a controlled vocabulary with a simple design often supported with visuals. The lemma list in this type of dictionary requires systematic selection.

In the development stage of a dictionary, certain issues such as *readability or usability* (readership), *dictionary format*, *content*, *lexical item presentation* (e.g., part of speech, derivational/inflectional affixes and dependent morphemes to be provided or not), the stages of compiling the dictionary and its publication should be considered (Bowern 2008). Similarly, in developing a school diction-ary, these issues should also be taken into account.

User research in lexicography has gained much importance in the devel-opment of dictionaries. Therefore, what users expect from a dictionary and what would be useful to include in it are among the common practices referred to in the literature. In this respect, the determination of the dictionary's target users is crucial (Jackson 2002). Various field studies such as surveys, observa-tions, or expert opinions are applied to determine the target users of a diction-ary (Atkins and Rundell 2008: 30). The target users, their age range, level of education, and intended purpose of the dictionary have profound effects on the whole structure of a dictionary.

There is a consensus in the literature on the necessity of the target popula-tion's needs in the development process of a dictionary. For this reason, needs-analysis using empirical methods should be performed to determine the needs of the target users (Atkins and Levin 1995: 85). In the current project, the needs analysis was conducted to determine the needs of the target dictionary users. Moreover, in the Turkish context, ineffective lexicographical traditions such as using index cards, finding example sentences manually, basing headwords, and example selection on intuitions are common practices in the development of a Turkish dictionary. However, in the development of a dictionary, the use of corpus is becoming the mainstream as it eliminates these ineffective lexico-graphical traditions. For this reason, RLBSD in the current project was devel-oped with the use of corpus.

Individual preferences and intuitive approaches of linguists and lexicographers necessarily yield different results (McEnery et al. 2006: 145). With an approach based on empirical results and language use research, modern lexicography develops dictionaries through language databases or corpora representing the language they are compiled for, yielding significant results for both lexicographers and users.

Along with the developments in corpus linguistics and corpus tools and their effect on lexicography today, there are few dictionary practices not based on corpus data. In the Turkish context, however, there has been a profound lack of corpus-based dictionaries. Dictionaries on Turkish should be developed based on current lexicographical practices including the development of dictionaries with corpus-based approaches. In this regard, the development of RLBSD as described in this report can be considered as a starting point to fill this important gap.

Lexicography is, by its nature, a discipline that offers empirical and scientific ways for determining vocabulary also including the phases such as specifying lexical entries and providing definitions and examples for a dictionary. As in large-scale dictionaries, school dictionaries' compilation requires considering certain issues such as consistency, being free of discrepancies, and testability of lemma selection by the experts. School dictionaries should be constantly developed to be better by correcting the mistakes and making sure that there is harmony with interrelated, appropriate, and reasonably organized entries. If lemma selection in a dictionary is not based on scientific criteria, it cannot fully represent the vocabulary of the target language. Therefore, it is of great importance to choose lexical items to be included in a school dictionary from the lexical items of the target language vocabulary (Bozkurt 2017).

Research on Turkish school dictionaries indicates that the need to use a dictionary increases with age and school grade in middle schools due to the increasing intensity of the course contents and high vocabulary variation in the texts (Melanlıoğlu 2013). Akcan (2012) examined Turkish workbook exercises and found that students are directed to use dictionaries for 123, 174, 183, 186, 167, and 853 words in the first, second, third, fourth, fifth, and sixth grades, respectively. The same study emphasized that dictionaries should be used as a resource for word meaning and form and usage (Akcan 2012).

The studies on existing school dictionaries report that students, as dictionary users, mention problems such as lack of bold letters for word entries, examples for connotations and figurative meanings, and visual elements (Melanlıoğlu 2013). As is observed in the literature, school dictionaries are developed without initially-set criteria, a reasonable explanation for the visuals to be included that accompany words principles to be used for limiting the content of lexical entries (Çotuksöken 1999).

The current project set out to meet the user group's needs by identifying the vocabulary used in natural language environments that children encounter, evaluating the effectiveness of a dictionary in educational settings, and calculating fit indices based on expert opinion. The RLBSD will contribute to the literature as an original and contemporary work in terms of its processes for development, expansion/distribution, and evaluation, and being a resource based on empirical results. In this regard, an original and real-life-based school dictionary requires a *lexicographical* practice featuring *corpus linguistics* principles and methods and also a special purpose *corpus* containing texts from natural language environments such as coursebooks, children's literary books, and periodicals that children may encounter as an educational resource for the primary and elementary school students.

1.1   Aim and objectives

This report aims to describe the development of a new, original, and monolingual Real-life-based School Dictionary (RLBSD) as educational material. To this end, two objectives were sought to be achieved: the compilation of a special purpose *corpus* containing texts from natural language environments such as MoNE (Ministry of National Education)-approved and current coursebooks, children's literary books, and periodicals that children may encounter starting from elementary school to high school, and development of the Real-life-based School Dictionary based on the corpus in question.

This project's primary objective is to develop a real-life-based school dictionary as educational material by using information technologies and corpus linguistics methods. In the literature, there is a lack of a corpus-based school dictionary in Turkish lexicology studies. Moreover, existing dictionaries are based on native speaker intuitions making these dictionaries far from being empirical. In this respect, the RLBSD as an educational material meets a critical need that has been felt so far. The vocabulary that belongs to the natural language environments that school-age children encounter was determined through lexicographical practices.

1.2   Compilation of the corpus

The research corpus was developed using corpus linguistics principles and methods as a discipline that can yield real-life and empirical outcomes based on texts from natural use environments, potentially providing the researchers the most suitable and up-to-date opportunities for creating a new, original, and real-life-based school dictionary.

The MoNE-approved coursebooks and instructional materials were included in the specialized corpus to reveal their vocabulary. Other natural lan-

guage environments (e.g., children's literary books, periodicals, etc.) were demonstrated in a real-time and empirical way in the context of achieving the competencies stated in the Turkish Qualifications Framework (TQF). The RLBSD is a holistic and inclusive attempt that would directly contribute to students' language development as the most important life skill, going beyond the limitations of the curricula implemented by MoNE.

Researchers and educators across the country will be able to use the RLBSD in their educational environment with its dissemination/distribution through information technologies. It will be easily accessible through the web, tablet, smartphone, and smartboard applications, and thus take its place in education with an active function. Considering the study's objectives, the RLBSD may serve as a sustainable model as it provides lexicographers a flexible and updatable information platform. The development, implementation, evaluation, and cyclical improvement of the RLBSD constitute this study's main phases.

2.   Methodology

In this section, the methods and techniques, data collection tools, and their analyses are elaborated concerning the literature in line with the study's aim and objective.

2.1   Focus group interview and needs analysis

Data collection techniques used in qualitative research include observation, interview, focus group interview, and document analysis. In this study, a focus group interview (FGI) was employed to determine needs in the development of the RLBSD. FGI aims to obtain in-depth, detailed, and multidimensional qualitative data on the perspectives, experiences, interests, tendencies, thoughts, perceptions, feelings, attitudes, and habits of participants about a set topic (Bowling 2002; Gibbs 1997; Kitzinger 1994, 1995; Krueger 1994; Stewart and Shamdasani 1990).

In this sense, the first phase of developing the RLBSD was to determine users' needs. Focus group interviews were conducted with students, teachers, and lexicographers as the stakeholders of the dictionary to be developed, and details for the lexicographical structure of the RLBSD were determined (Çokluk 2011).

The questions that are planned to be asked in the focus group interview are presented in Figures 1 and 2.

Focus Group Interview Form (for Teacher and Experts)

Focus Group Interview: *Determination of layers / structures to be included in Real Life Based School Dictionary (RLBSD)*

Field Expertise:	**Year of Professional Experience:**

Questions

1. What kind of problems do you encounter while using the dictionary in the course?

2. Do you think the use of the dictionary is functional in your course? How?

3. What do you think dictionaries can contribute to the classroom environment? How?

In which situations do you think that the dictionary use is most useful in a course?

4. What is the contribution of existing dictionaries to the course environment? Are they any aspects in which dictionaries need improvement?

5. Do you think your students can make full use of dictionaries in courses? What are the drawbacks?

6. What are the situations in which students have difficulties in using the dictionary in the classroom environment?

7. What information should a school dictionary contain and what needs do you think a dictionary should meet?

8. Do you think that students can reach sufficient content in dictionaries related to a word they encounter in a course?

9. If it comes to conveying a concept / term / word to students with the logic of lexicography, what do you think that the dictionaries should include about that concept / term / word?

10. To what extent accessing easily via electronic environments to a concept / term / word encountered in the course can affect the success in teaching and learning?

Figure 1: Questions for the focus group interview

RLBSD EVALUATION QUESTIONNAIRE (Student, Teacher, Lexicography Experts)

This questionnaire was prepared in order to determine the opinions of students, teachers and lexicography experts about the use of Real Life Based School Dictionary (**RLBSD**) in classroom environments.

Participant Information

Student	O				
Teacher	O		Age :		
Lexicography Experts	O		Gender :	M: O	F: O

Items	(1)	(2)	(3)
1. The number of headwords in the dictionary is sufficient.	Yes	No	Partially
2. Screen designs of the dictionary are sufficient.	Yes	No	Partially
3. The dictionary meets the needs of lessons.	Yes	No	Partially
4. Explanations of the words in the dictionary are sufficient.	Yes	No	Partially
5. The pronunciation of the headwords in the dictionary is sufficient.	Yes	No	Partially
6. The sample sentences in the dictionary are enough to explain the words.	Yes	No	Partially
7. Collocations in the dictionary are sufficient.	Yes	No	Partially
8. Explanatory visuals at the beginning of the entries(headwords) in the dictionary are sufficient.	Yes	No	Partially
9. Additional texts included in the beginning of the entries are adequate to explain the entries	Yes	No	Partially
10. Word type of the head words in the dictionary are given correctly.	Yes	No	Partially
11. Entries in the dictionary include the vocabulary encountered in the lesson.	Yes	No	Partially
12. I could find other words I was looking for in the dictionary.	Yes	No	Partially
13. The sample sentences in the dictionary adequately explains the content.	Yes	No	Partially
14. The colors on the dictionary screens are compatible.	Yes	No	Partially
15. After querying on a word, query results can be reached easily.	Yes	No	Partially
16. The dictionary is easy to use.	Yes	No	Partially
17. The dictionary can be accessed whenever needed.	Yes	No	Partially
18. I experience difficulty using the dictionary.	Yes	No	Partially

There are some shortcomings in the dictionary. I can express these shortcomings as follows:..
....

I found the dictionary useful because ....
....
....

Figure 2: RLBSD evaluation questionnaire

2.2   Evaluation of the RLBSD by stakeholders

Another data collection used in line with the aim and objectives of the study was a questionnaire. A questionnaire is a technique and/or means for written semi-structured interviews to gather data. It is widely used in social sciences, especially for survey purposes, due to its relative ease of preparation and the possibility of reaching a large number of individuals in a short time. Questionnaires are used to identify/describe a current situation (Erkuş 2013: 161, Büyüköztürk et al. 2016: 124). The questionnaire developed for this study will be used to receive the opinions of lexicographers, teachers, and students on the content and usage of the RLBSD, and the feedback will improve the RLBSD (see Figure 2).

2.3   Evaluating the effectiveness of the RLBSD

One of the current study's objectives is to evaluate the effectiveness of a school dictionary that will be developed based on the methods and principles of corpus linguistics and lexicography. To this end, experimental and control groups were formed from the existing groups/classes by using a reading comprehension test to compare the two groups regarding the effects of the RLBSD. In this sense, the implementation is quasi-experimental. In the quasi-experimental method, existing classes/groups are used in cases where it is difficult or not possible to form experimental and control groups randomly (Robson 1998), so it is also known as the *non-equivalent control group method* (Karasar 2009). In quasi-experimental studies, a control group that is not influenced by the independent variable is used along with the experimental group subjected to this variable (Christensen 2004).

Designed by the researchers involved in the study, the RLBSD will be employed in the experimental group, whereas in the control group, a printed dictionary will be made available to students. The experimental and control groups will include middle school fifth and eighth-graders. In other words, a control group and an experimental group from fifth and eighth grade will be compared within each grade. The reason for choosing the fifth and eighth grades in the study is that they contain students who have just started middle school and those who are about to graduate, and thus, the effectiveness of the RLBSD will be evaluated at different levels. Firstly, peer groups (experimental and control groups) will be formed among the students who have the same level of reading comprehension by administering the comprehension skill test to fifth and eighth-graders, and then it will be revealed whether there is a statistically significant difference in-between after having the experimental group use the RLBSD and the control group a printed dictionary (see Table 1).

Table 1:   Research design

FORMING THE GROUPS	GROUPS	IMPLEMENTATION	EVALUATION
Reading Comprehension Achievement Test	Control 1 (fifth graders)	Text Analysis with Printed Dictionary	Evaluation of Effectiveness
	Experimental 1 (fifth graders)	Text Analysis with the RLBSD	
Reading Comprehension Achievement Test	Control 2 (eighth graders)	Text Analysis with Printed Dictionary	Evaluation of Effectiveness
	Experimental 2 (eighth graders)	Text Analysis with the RLBSD	

2.4   Principles and methods of corpus linguistics and lexicography in the development of the RLBSD

This phase of the study was designed in descriptive and relational models. Descriptive research helps determine a phenomenon as it is, whereas relational research refers to studies in which cause-effect relationships cannot be established, and change and supervision may not be possible either by nature or due to practical reasons (Büyüköztürk et al. 2016; Erkuş 2009; Karasar 2009). Data collection techniques, methods, techniques, and instruments developed for specific purposes are used in research studies (Erkuş 2009). In this project, the methods and techniques introduced by corpus linguistics that can yield empirical results were employed.

The principles and methods of corpus linguistics and lexicography, on which the current study is mainly based, are outlined as follows:

> Today, large-scale corpora that have been developed with their unique features and applications based on written and spoken data from natural language environments make major contributions to language learning and teaching as they contain real-time language data isolated from intuitive examples and secondary data sources (Özkan 2010).

A *corpus* (plural: corpora) can be defined as a body of texts organized based on set standards by selecting written and spoken texts to sample a language following linguistic criteria (McEnery et al. 2006; for other definitions, see Say 2003). The primary criteria for a corpus are its ability to *sample* (sampling) the language in which it is created and its *power* to represent that language. Besides, it should have *a certain limit* (finite size) or not (dynamic size), and be a standard reference for describing the target language in a machine-readable form (McEnery and Wilson 2001).

In the literature, existing corpora vary depending on their aims and foci. In other words, corpora are developed depending on research questions (For

types of corpora containing written and/or spoken texts and that can reflect linguistic diversity in a language, see McEnery et al. 2006).

A set of criteria is considered in the selection of texts to be included in a corpus. The most common ones are taking *literary/academic merit* as the basis, *random selection*, *currency*, *availability*, *demographic sampling*, being *empirical*, and selection from *a broad range of sources*, some of which have advantages and disadvantages over the others (Summers 1993). One or more of these criteria can be used when creating a corpus.

The most crucial issue that would be considered while building a corpus is corpus design. For example, the type of texts included, the number of texts, the selection of texts, sample selection among the pieces of texts, and the sample's length are the measure of whether the design is made consciously (Sampson and McCarthy 2005). Depending on the research question, it is necessary to use a *representative* corpus compiled from the characteristic and typical usage environments of the language to reveal relationalities in language studies fully. Shaping a corpus's quality and content is one of the main issues that make language studies *corpus-based*.

The *power of representation* is a significant defining characteristic of a corpus. This characteristic distinguishes a *corpus* from a *collection of texts* (that are randomly selected). This is a vital practice in corpus design. On the other hand, it is not possible to examine all of a language as a whole. For this reason, forming a sample seems inevitable, and one should be sure of the *representability* of the sample formed for that language while working on it. Furthermore, the balance (the distribution of text types), the quality of the text strata selected for each type, and the sample are among the main elements that constitute the *representative power* of a corpus (McEnery et al. 2006).

Today, *representativeness* that involves *text-type distribution/distributiveness* and *sample selection/sampling* is a prerequisite for corpus applications. On the other hand, *representativeness* comprises a sample with all the variables in a population. In studies on sample size, researchers agree that representativeness, which refers to how many texts a corpus should contain and how many words a text should contain, is the most important decision-making point. Accordingly, the variety of texts in a language and their linguistic distributions are among the issues that should be considered in corpus design (Sampson and McCarthy 2005).

Considering the arguments mentioned above and corpus characteristics, the corpus used for the Real-life-based School Dictionary can be regarded as *a specialized corpus*. The specialized corpus to be created contains two main sub-corpora. The first sub-corpus is the *Corpus of Turkish Children's Literature*, a specialized corpus that focuses on the literary works for Turkish children in the context of qualitative and quantitative research topics. *This corpus was created within the scope of another national project funded by TÜBİTAK.* With the project in question, Turkish Children's literature was used as the research population for various analyses such as *the readability and age-appropriateness of children's books,*

lexical variety and lexical domain patterns, morphological, lexical, and syntactic features, and internal and external structural features. This specialized corpus of "Turkish Children's Literature" was compiled according to corpus linguistics principles and methods and contains 8,639,522 million words. This corpus containing 1,089 texts in different types including texts from children's periodicals, will be evaluated in the proposed Real-life-based School Dictionary. As for the second sub-corpus, it includes MoNE-approved coursebooks. The research population was completed by digitizing the coursebooks based on corpus design steps.

2.4.1 Morphological, lexical, and syntactic structure of the corpus to be developed for the RLBSD

The texts that included in the corpus were sententially parsed, and metadata for every sentence such as sub-corpora and text types [*Texts of Turkish Children's literature* (novel, poem, story, memoir, periodicals, etc.), *MoNE-approved coursebooks* (by grades and courses), *Other Types* (other texts that children may encounter)], author, work title and year of publication will be tagged and added to the corpus.

At this stage, lemmatization was performed on the corpus, and the lemmas were deduplicated.

2.4.1.1 An example of a sentence and sentence tagging

< Tales was a great scientist who calculated the height of the Egyptian pyramids and "drew a right-angled triangle into a circle". >

      <Corpus of Turkish Children's Literature> **First sub-corpus**
          <Essay> **text type**
              <mavisel yener> **author name**
                  <Pyramids> **work title**
                      <2009> **year of publication**

2.4.1.2 An example of morphological analysis[1]

Mısır [???] | *pramitlerinin* [piramit(pyramid)+Noun+A3pl+P3sg+Gen] | *yüksekliklerini* [yüksek(height)+Adj^DB+Noun+Ness+A3pl+P3sg+Acc] | *hesaplayan* [hesapla(calculate)+Verb+Pos^DB+Adj+PresPart] | *ve* [ve(and)+Conj] *bir* [bir(a)+Det] | *dairenin* [daire(circle)+Noun+A3sg+P2sg+Gen] | *içine* [iç(in/into)+Noun+A3sg+P3sg+Dat] | *dik* [dik(right-angled)+Adj] | *üçgen* [üçgen(triangle)+Noun] | *çizen* [çiz(draw)+Verb+Pos^DB+Adj+PresPart] | *büyük* [büyük(great)+Adj] | *bir* [bir(a)+Det] *bilim* [bilim(science)+Noun] | *adamıydı* [adam(man)+Noun+A3sg+P3sg+Nom^DB+Verb+Zero+Past+A3sg] | *Tales* [???]

2.4.3   Data processing in the RLBSD

Data processing was performed via an online platform. The data processing platform created is a previously experimented platform by which other studies based on the Turkish language have also been conducted. Various lexicogra-phy applications supported by TÜBİTAK and university funds were carried out on the said platform (http://turkcederlem.mersin.edu.tr/) (see Figure 3).
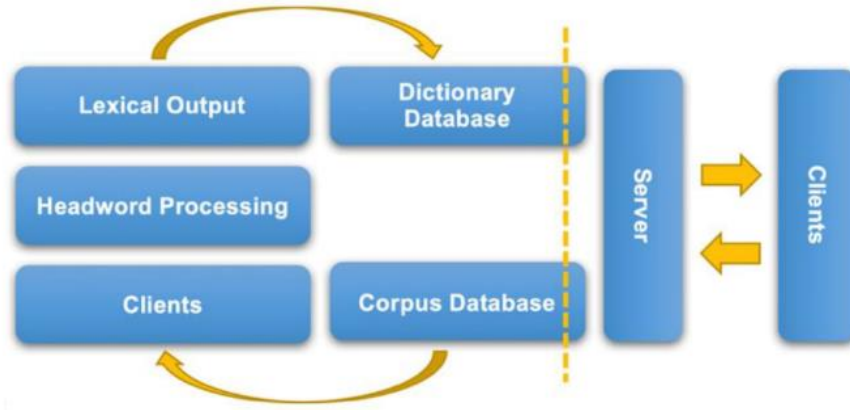

Figure 3:   Data processing platform (Özkan 2013)

In the 'data-processing platform' shown above, two separate databases on the server provide an uninterrupted service for users to process and view data. Simply, users make lexical queries on the corpus database and then perform data processing on the query results, followed by saving entries on the diction-ary database. The saved and processed data are published instantly on the server with query screens.

2.4.3.1 Data processing in the RLBSD over the corpus

The RLBSD is subjected to the processing of tagging and reporting over a sys-tem represented in Figure 4.
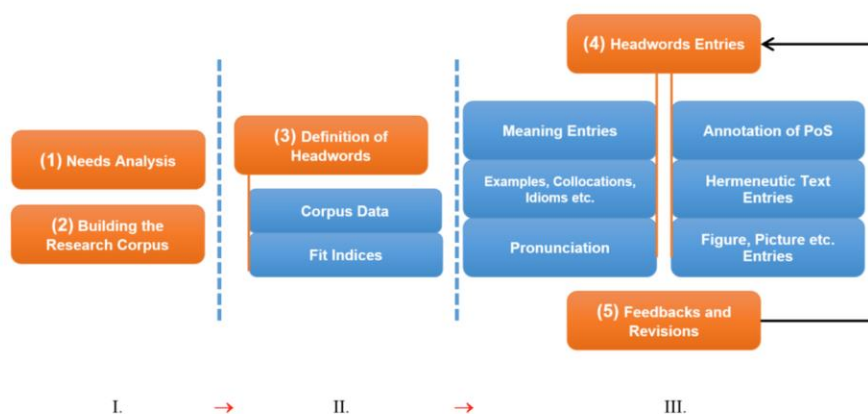
Figure 4:   Data processing stages

In brief, the proposed project includes five main steps in three phases. *The first phase involves* (1) determining the needs and (2) creating the corpus, *the second phase* (3) determining the lexical entries, and *the third phase* (4) structuring the lexical entries which include meaning entries, word type tags, selection of examples, identifying collocates and related idioms, pronunciation, reader-friendly text, figures and visuals, and (5) feedback and corrections.

These steps along with the methods and techniques to be implemented in the projects are explained in the following:

(1)   Determining the needs

In brief, determining the needs of the target population is one of the most decisive elements when developing a dictionary. From this perspective, the needs of the target population and/or stakeholders must be determined precisely. Focus group interviews with subject-matter teachers and linguistics/lexicography experts will be employed in this step (see the methodology section above).

(2)   Creating the corpus

The corpus that would be used for the Real-life-based School Dictionary is a specialized corpus. The specialized corpus created contains two main sub-corpora. The first sub-corpus is the specialized Corpus of Turkish Children's Literature (see http://turkcederlem.mersin.edu.tr/cocuk/ and Özkan 2014). As for the second sub-corpus, it includes MoNE-approved coursebooks and the periodicals that children may encounter.

(3)   Determining the lexical entries

When the history of lexicography is examined, it can be observed that the language samples that linguists and lexicographers used in their attempts are mostly based on native-speaker intuitions (i.e., armchair lexicography), and these works have yielded results that are not right, and/or that could not fully reflect linguistic realities. Moreover, individual preferences and intuitive approaches of linguists and lexicographers necessarily mean differing results for every language expert (McEnery et al. 2006). In current lexicographical traditions, practices such as collecting examples, using individual knowledge, and forming index cards are methods that are accepted as non-functional in developing a dictionary (Atkins and Levin 1995). On the other hand, school dictionaries are important resources in first language education, and studies on how and from which sources these dictionaries should be developed have been widely addressed in the literature (Bergenholtz and Gouws 2012; Malkiel 1967).

In a lexicographical study, all steps such as *defining linguistic components, determining the frequencies of occurrence for these components, forming concordances,* and *extracting collocational structures* provide quality and valid data for the researchers in the selection of dictionary entries. In this respect, dictionary entries in the RLBSD are structured in the light of scientific and experimental methods.

While determining lexical entries, there are two ways that are followed. The first of these is to obtain lists of lexical entries from the lexemes that occur in the corpus. This is one of the standard data processing steps on the corpus platform. Regarding the vocabulary retrieved from the MoNE-approved coursebooks that can be lexical entries in the RLBSD, structures are determined as lexical items employing simple fit index calculations after asking for the opinions of three subject matter teachers for each coursebook, in addition to frequency and distribution criteria. Furthermore, since the corpus will be lemmatized the words that would be retrieved through the *simple fit index* calculations and those obtained from the corpus will be structured as lexical entries.

Thus, the lexical entries will be those: **a.** that are obtained from the corpus, and **b.** that are retrieved through simple fit index calculations based on the opinions of field experts. This type of process means that the selection of lexical entries will be done with an empirical approach.

While identifying the subject areas (i.e., science, social studies, or mathematics in the context of courses), the metadata about the texts included in the corpus such as *sub-corpus, text type, author name, work title,* and *date of publication* can be automatically obtained through the corpus formed digitally. This and similar metadata will form the basis for further research as standard corpus output that can provide researchers the data that can be used in the detailed analyses of the corpus.

(4)   Structuring the lexical entries

When creating a dictionary, a set of lexical entry configurations is used in accordance with its purpose. These are *writing* (spelling), *pronunciation, inflections, word forms (*part of speech), *meaning(s), definitions, examples, usage, other derivations* (run-ons), and *root information* (etymology) (Jackson 2002; Hanks 2003). Other configurations (picture, sound file, etc.) can be added to these by using information technologies. In the structuring phase, what a dictionary entry includes is decided based on the lexicographer's aim. The data obtained from corpora form the basis for this decision.

In other respects, as mentioned before, although school dictionaries have a similar structural feature with general dictionaries (entry, meaning, part of speech, example, etc.), it has to contain certain configurations specific to the target audience, unlike general dictionaries. The basic arrangements to be included in the RLBSD and their sources can be summarized as follows:

a.   Meaning/definition arrangement

Experts of linguistics/lexicography arranged the meanings to be provided for lexical entries in a context-sensitive way. At this stage, the target population was considered, and the meaning arrangements were shaped by making use of the meaning patterns existing in general dictionaries. The corpus's most significant contribution to this study is that all lexical contexts can be seen, inspected, and included while assigning meanings to lexical entries. This is of significance in making a simple, clear presentation of meaning in line with the target audience's needs by making comparisons in connotations. The development of specialized dictionaries with a corpus-based or corpus-driven approach is also important for words to be presented in a semantic harmony (Yu and Cai 2009).

As stated in the literature (Günay 2007), the existing definitions in lexicography (e.g. *definitions about meaning and extension, descriptive meaning, operational definition, explicative definition,* and *conventional definition*) are formatted depending on the characteristics of lexical entries in the stage of meaning and definition. When providing definition and meaning, for example the lexeme "**ak**" can be defined as "beyaz" (white) in general dictionaries, whereas its definition can be "Kar veya süt renginde olan." (the color of snow or milk) in a school dictionary. Such an approach can present a meaning/definition directly to users, considering the dictionary's target audience.

b.   Pronunciation

Another arrangement to be included in the RLBSD in accordance with its purpose is the pronunciation of the lexical entries. Audio recordings were added to the entries of the RLBSD.

c.   Determining the part of speech and conceptual field

The part of speech was tagged by sampling all uses of words in a context-sensitive way. This is a step that can be implemented easily through corpus linguistics applications.

d.   Selection of examples

In corpus queries, all contextual patterns about the lexical entry being searched can be viewed and listed for selection. In this sense, linguists/lexicographers can choose the right and functional examples in accordance with the purpose of the dictionary and have the opportunity to configure the entries.

e.   Determining the collocational and idiomatic expressions

Revealing collocational and idiomatic expressions out of a corpus by using n-grams means finding out the main structures in the formation of meaning patterns. In corpora, the retrieval of multi-word combinations [... n-3, n-2, n-1 | | n+1, n+2, n+3 ... (bigram, trigram...)] is one of the standard processing steps, the tagging of the listed constructions simplifies the work of linguists/lexicographers as they identify collocational and idiomatic expressions.

f.   Explanatory texts

Another characteristic of the RLBSD is enriching the entries with supportive/explanatory texts that explain them. At the end of some entries, additional information about the headword can be found. This information includes extra information about a place, a person, a thing, a process. With this explanatory information, the meaning of the headword becomes clearer for the dictionary users.

g.   Figures, images, etc.

In accordance with the purpose and the target audience, visuals containing, explaining, and representing the lexical entries were used. Accordingly, the list of lexical entries determined by lexicography experts through simple fit indices was visualized with images and figures.

h.   Vocabulary teaching, vocabulary games, tests for teaching spelling

Particularly concerning vocabulary teaching, certain digital applications (e.g. matching, comparing, completing), that are widely used in language education will take their place in the corpus-based entry configurations.

Consequently, a sample lexical entry can be represented as shown in Table 2.

Table 2:   Sample of lexical query of "üçgen"
(see also http://okulsozlugum.com/arama/ and
http://turkcederlem.mersin.edu.tr/okulsozluk/)

(5)   Feedback and corrections

Apart from corpora, another way an individual develops a general dictionary that can be used as a data source is through user suggestions. User suggestions may include the addition of missing words or word meanings to the dictionary etc. Although dictionary developers may have a vast and comprehensive data-base, they will always be outnumbered by those who use the dictionary. This means that it is likely for a dictionary user to not find a word they are looking up in a dictionary. No matter how many and diverse the texts that dictionary makers based their dictionary on, users always come across more texts. In this regard, the developer of a general dictionary who is open to user contributions and suggestions should be able to benefit from user feedbacks (Bozkurt 2017) meaning that user practices and preferences can also be taken into considera-tion to develop a more useful dictionary.

This project has an interactive and multi-stakeholder construct. Merely having the opinions of lexicographers would not be sufficient for developing a school dictionary. User practices and preferences influence the stages of dic-tionary development with user feedback and corrections. Necessary arrange-

ments and corrections will be made by receiving feedback from subject matter teachers, lexicographers, and students.

It is also common practice to log user queries [access logs] in modern corpus-based and IT-supported dictionary applications. Today, advanced and statistical software tools can provide the researchers data on which lexemes users search for as well as their user profiles (e.g. age, region, gender, etc.). The most well-known of these are open-access applications such as Google Analytics and Yandex Analytics. In this respect, relevant information technologies can be used to receive feedback and make corrections.

3.   Conclusion

In the current project, RLBSD differs from the school dictionaries that have been introduced so far in terms of both its creation method and the possibilities of reaching the end-user and/or accessing the dictionary. First of all, the focus group interview for dictionary creation is a new practice in itself, at least for Turkish lexicography. Besides, the use of "a special field corpus" that includes vocabulary elements that school-age children encounter can be considered the most appropriate approach to the content organization of a school dictionary. In this regard, it can be concluded that the use of corpus linguistics principles and methods and lexicology principles together in the creation of RLBSD can be considered a breakthrough in the Turkish lexicography tradition. Although creating a dictionary may seem like a lexicographer's job, a dictionary cannot be considered independent of user evaluation. From this perspective, the stages of evaluating the dictionary's effectiveness after the creation of RLBSD in educational environments are considered a brand-new approach that corresponds to an application that has not been experienced before in Turkish lexicography.

As is known, the corpora provide important contributions to the lexicology field with the experimental results they provide. When considered in this regard, corpus linguistics applications can be accepted as an empirical criterion for dictionary creation. From this perspective, RLBSD draws attention as a source that contains real-time data in terms of item selection and item configurations. With this feature, RLBSD can be placed on an axis outside the traditional lexicography understanding.

It can be said that the use of and access to RLBSD developed with the aforementioned experimental methods may become a standard for corpus-based dictionary creation due to the use of information technologies used in its e-spread and distribution. In this sense, RLBSD with its feature of ease of access through the Internet, tablet, smartphone, and smartboard applications will make it an effective resource as course material in educational environments. Also, RLBSD can be considered a sustainable model as it offers a flexible and updatable platform. In addition, since the effectiveness of RLBSD is to be evaluated, how it affects learning outcomes will also be tested.

Endnote

1. Every lexical item in the corpus to be created will be analysed using the morphological analysis tool (see http://www.hlst.sabanciuniv.edu.tr/TL/) developed by Kemal Oflazer.

References

Akcan İ.P. 2012. Türkçe Öğretiminde Sözlük Kullanımı: Türkçe Çalışma Kitaplarında Sözlüğe Gönderen Alıştırmalar. Aksan, M. and Y. Aksan (Eds.). 2012. *Türkçe Öğretiminde Güncel Çalışmalar*: 93-102. Ankar: Şafak Yayınları.

Atkins, B.T.S. and B. Levin. 1995. Building on a Corpus: A Linguistic and Lexicographical Look at Some Near-synonyms. *International Journal of Lexicography* 8(2): 85-114.

Atkins, B.TS. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

Bergenholtz, H. and R.H. Gouws. 2012. What is Lexicography? *Lexikos* 22: 31-42.

Bowern, C. 2008. *Linguistic Fieldwork: A Practical Guide*. New York: Palgrave Macmillian.

Bowling, A. 2002. *Research Methods in Health: Investigating Health and Health Services*. 2nd edition. Buckingham/Philadelphia: Open University Press.

Bozkurt, F. 2017. *Sözlükselleşme: Genel Sözlükler için Sözlük Birim Seçimi.* Istanbul: Kesit Yayınları.

Büyüköztürk, Ş. et al. 2016. *Bilimsel Araştırma Yöntemleri.* 22nd edition. Ankara: PegemA Yayıncılık.

Christensen, L.B. 2004. *Experimental Methodology.* 9th edition. Boston: Pearson Education.

Çokluk, Ö. 2011. Nitel Bir Görüşme Yöntemi: Odak Grup Görüşmesi. *Kuramsal Eğitimbilim* 4(1): 95-107.

Çotuksöken, Y. 1999. Okul Sözlükleri Üzerine. *Kebikeç* 7–8: 61-65.

Erkuş, A. 2009. *Davranış Bilimleri için Bilimsel Araştırma Süreci.* Ankara: Seçkin Yayıncılık.

Erkuş, A. 2013. *Davranış Bilimleri için Bilimsel Araştırma Süreci.* 4th edition. Ankara: Seçkin Yayıncılık.

Gibbs, A. 1997. Focus Groups. *Social Research Update* 19(8): 1-8.

Günay, D. 2007. *Sözcükbilime Giriş.* Istanbul: Multilingual.

Hanks, P. 2003. Lexicography [Chapter 3]. Mitkov, Ruslan (Ed.). 2003. *The Oxford Handbook of Computational Linguistics*: 48-69. Oxford: Oxford University Press.

Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London/New York: Routledge.

http://turkcederlem.mersin.edu.tr/. *Date of access:* 17-12-2019

http://turkcederlem.mersin.edu.tr/cocuk/. *Date of access:* 17-12-2019

http://turkcederlem.mersin.edu.tr/okulsozluk/. *Date of access:* 5-10-2020

http://okulsozlugum.com/arama/. *Date of access:* 16-5-2021

http://www.hlst.sabanciuniv.edu.tr/TL/. *Date of access:* 9-02-2009

Jackson, H. 2002. *Lexicography: An Introduction.* London/New York: Routledge.

Karasar, N. 2009. *Bilimsel Araştırma Yöntemi.* 20th edition. Ankara: Nobel Yayın-Dağıtım.

Kitzinger, J. 1994. The Methodology of Focus Groups: The Importance of Interaction Between Research Participants. *Sociology of Health and Illness* 16(1): 103-121.

Kitzinger, J. 1995. Qualitative Research: Introducing Focus Groups. *British Medical Journal* 311: 299-302.

Krueger, R.A. 1994. *Focus Groups: A Practical Guide for Applied Research.* 2nd edition. Thousand Oaks, CA: Sage.

Malkiel, Y. 1967. A Typological Classification of Dictionaries on the Basis of Distinctive Features. Householder, Fred W. and Sol Saporta (Eds.). 1967. *Problems in Lexicography*: 3-24. Bloomington: Indiana University.

McEnery, T. and A. Wilson. 2001. *Corpus Linguistics: An Introduction.* 2nd edition. Edinburgh: Edinburgh University Press.

McEnery, T. et al. 2006. *Corpus-Based Language Studies: An Advanced Resource Book.* London: Routledge.

Melanlıoğlu, D. 2013. Ortaokul Öğrencilerinin Sözlük Kullanma Alışkanlıkları: Nitel Bir Araştırma. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi* 2(2): 266-284.

Özkan, B. 2010. Türkçenin Öğretiminde Sıfatların Eşdizim Sözlüğü: Yöntem ve Uygulama. *e-Uluslararası Eğitim Araştırmaları Dergisi* 1(2): 51-65.

Özkan, B. 2013. Yöntem ve Uygulama Açısından 'Türkiye Türkçesi Söz Varlığının Derlem Tabanlı Sözlüğü'. *bilig. Türk Dünyası Sosyal Bilimler Dergisi* 66(3): 149-178.

Özkan, B. 2014. Bir Özel Alan Derlemi Olarak 'Türk Çocuk Yazını Derlemi' ve Türk Çocuk Yazını Üzerine İleri Uygulamalar Projesi. Özkan, B. (Ed.). 2014. *Çocuk Yazını Araştırmaları-I:* 1-40. Adana: Karahan Kitabevi.

Özkan, B. et al. 2019. The Method of the Real Life Based School Dictionary. Gürlek, M. et al. 2019. *ASIALEX 2019. Proceedings of the 13th International Conference of the Asian Association for Lexicography, Istanbul*: 94-105. Istanbul: Asos.

Robson, C. 1998. *Real World Research: A Resource for Social Scientists and Practitioners.* Blackwell.

Sampson, Geoffrey and Diana McCarthy. 2005. *Corpus Linguistics: Readings in a Widening Discipline.* London/New York: Continuum.

Say, B. 2003. *Türkçe için Biçimbirimsel ve Sözdizimsel Olarak İşaretlenmiş Ağaç Yapılı Bir Derlem Oluşturma.* Proje No: 199E026. Proje Raporu-Ek-b: Türkçe Ağaç Yapılı Derlem Kılavuzu. Ankara.

Stewart, D.W. and P.N. Shamdasani. 1990. *Focus Groups: Theory and Practice.* Newbury Park, CA: SAGE.

Summers, D. 1993. Longman/Lancaster English Language Corpus — Criteria and Design. *International Journal of Lexicography* 6(3), 181-208.

Yu, P.-F. and J.-C. Cai. 2009. Semantic Prosody: A New Perspective on Lexicography. *US-China Foreign Language* 7(1): 20-25.