# Online Bilingual Dictionary as a Learning Tool:
# Today and Tomorrow

## Mark Kit[1], Elena Berg[2]

## Abstract

*Online dictionaries are among the most essential components in the language learning process. While retaining the best characteristics of the traditional dictionaries, such as completeness and correctness, they possess new properties – fast response and reversibility, which greatly helps users who seek lexical help.*

*Unlike traditional dictionaries, online ones emerge and change fast. Besides, their operation depends on traffic and system tune-up. This makes their characteristics volatile and they are usually unknown. The paper discusses properties of online dictionaries and states that their specifications should be studied and tested on a regular basis. When developing and delivering their courses the teachers have to identify what dictionary characteristics are essential for their specific goals and select dictionaries that suit those goals best.*

*The paper also suggests that further innovations in the online lexicography shall be sought, such as application of cognitive faculty of dictionaries and addition of adoptive properties to the dictionaries. Analysis of the interaction between students and dictionaries can reveal information on students' knowledge and problem areas in their learning. This ability is viewed as cognitive features of online dictionaries and discussed in this paper as such. Examples show how inferences can be made based on user's queries sent to the dictionary.*

*The exploitation of cognitive faculty opens opportunities for dictionary optimization. In response to user's queries dictionaries produce large amount of data, leaving the job of finding the required translation in that data shown to the user. This takes considerable effort and time. Through analysis of user's queries, the dictionary can make inferences regarding translations most appropriate to the context the user works with and put these translations on the top of the results list.*

*Data obtained through experiments and analysis support the theoretical provisions discussed in the paper.*

## 1. Introduction

One of the most important activity entertained by individuals who learn languages is retrieval and analysis of lexical entries in dictionaries. In foreign language studies, the most in-demand dictionaries are bilingual ones. The use of these dictionaries form the basis for studies of lexical system of the language.

In the recent decades dictionaries as sources of information underwent considerable changes. Traditional printed dictionaries have been displaced by emerging electronic and online dictionaries that offer fast retrieval of lexical data are flexible when it comes to updating and appending data.

Rapid evolution of multilingual dictionaries, however, was not accompanied by studies aimed at *determination of their characteristics and scope of application.* As new lexical resources emerge, disappear, and change, selection of dictionaries that are best for each particular course is not as clear as it used to be. For this reason *identification and description of characteristics of dictionaries* becomes essential. Besides, online dictionaries open new opportunities for human-dictionary interaction and *development of new functions* that help optimize the foreign language learning process and professional translation work.

[1] University of Washington; Language Interface Inc., USA
[2] Ural State Law University, Russia; Language Interface Inc., USA

## 2. Discussion
### 2.1 Characteristics of Dictionaries
Some characteristics of online dictionaries and printed ones are the same. These characteristics include **up-to-date compliance**: the dictionary should be consistent with the contemporary state of the language. Next comes objectivity – interpretation of lexical entries shall be unbiased and free from any ideological or political views. Another characteristic is the **lexical coverage**, which is related both to the overall dictionary (how well the dictionary covers the relevant collection of lexical entries) and individual – how well it covers the data pertaining to individual lexical entries. Another important general characteristic is **correctness** of the interpretations and description of lexical unit given in the dictionary against linguistic norm.

Online dictionary **response time** is the time elapsed from user's query submission to displaying translations to the user. Most online dictionaries are fast, others lose popularity. However, apparent fast response can be deceiving since dictionary operation can slow down under heavy traffic. This problem often become acute when a dictionary gains popularity and the traffic exceeds its capacity. It is advisable, therefore, to check dictionary response time every so often at different times of the day. Another specific feature of online dictionaries is **reversibility** – the ability of the dictionary to find translations in both directions within the given language pare. The degree of reversibility depends on specifics of the language pair and tuning of the search engine. Perfect reversibility cannot be possibly achieved. For example, search engines experience difficulties in cases where one word of the source language is translated with two or more words of the target language.

This problem can be due to specifics of the lexical database architecture such as asymmetric many-to-many relationship. All online dictionaries possess the properties listed above, but they are manifested differently. Knowing specifics of dictionaries teachers can make educated recommendations on using one or another dictionary to their students. At this time, the authors conduct ongoing development of methodologies to be employed for identification and comparison of characteristics of dictionaries; however, this research is beyond the scope of this paper.

### 2.2 Exploiting Cognitive Features of Dictionaries
Dictionaries have another capacity although unexplored – their **cognitive** faculty, which in this context is understood as the opportunity to obtain information on the user and his text-related activities based on the user's queries sent to the dictionary. To do so it is important to recognize that the fact of turning to a dictionary means that the student has certain lexical need. The concept of lexical need is understood here as a necessity to obtain information about a lexical unit in order to go on with an adequate participation in the discourse [1]. This need would have not arisen if the user had perfect competence in the foreign language. Analyzing this need and exploiting the knowledge acquired helps improve and speed up the learning process.

Analysis of dictionary queries can reveal different types of deficiencies in student's linguistic competence, including limited vocabulary and knowledge of grammatical or cultural aspects needed for translation. Sometimes a student has to search a dictionary not due to the lack of lexical competence, but due to poor knowledge of the foreign language grammar or cultural realities. For example, queries consisting of more than one word may demonstrate the student's inability to break the text into semantic units.

By analyzing dictionary queries, it can be determined in what subjects the students feels more confident, what aspects of the grammar and vocabulary need to be improved. This analysis can also identify the need for re-grouping the students. It can also help choosing the best teaching methods and optimize the existing methodology employed for each particular groups or individual students.

Table 1. Examples of cognitive inferences

| Identified pattern | Inference |
|---|---|
| Large portion of the user's queries are high frequency words | Limited vocabulary, improvements in lexical competence needed |
| Most user's queries contain low frequency words | Good lexical competence in the commonly used words, however the user might excessively rely upon guesses or have overconfidence in his knowledge of lexicon |
| User's queries contain many spelling errors | Writing skills improvement needed, more reading in the target language recommended |
| Queries consist mostly of phrases (more than one word) that are not multiword terms | The student's vocabulary and knowledge of grammar are poor and he/she seeks to obtain a text translation rather than a dictionary term. The student does not understand the difference between dictionaries and machine translators. If the student also breaks the text into wrong segments, his/her knowledge is very limited. |
| Queries contain many words in non-basic form (dogs, went, flown) | The student's knowledge of grammar needs improvement |

## 2.3 Optimization of Dictionaries

The versatile nature of cognitive faculty opens opportunities for improvement of the dictionary using state-of-the-art information technologies can introduce new features that have potential for optimization of foreign language studies and translation.

One of the most pressing problem the users of dictionaries face is excessive data produced by dictionaries in response to user queries. This list of translations can be very big and confusing. For example, in response to query "debris" the dictionary LexSite [2] offered a list consisting of 30 translations of the word and 62 translations of expressions that include "debris". Making the right choice from these options is a time-consuming intellectual effort. Detection of patterns in lexical needs of dictionary users is essential for creation of "smart" dictionaries that respond with the students most relevant to the tasks performed by the users [3].

Even seasoned translators or advanced level students experience difficulties finding the most appropriate translation in this abundance of data. The need to analyze large amount of suggested translations in a search for a single right one results in a considerable increase of time needed for the search.

Experiments run by the authors revealed the dependency of the term search time from the number of results produced by dictionaries in response to the user's query. Over 68% of the queries result in large lists of translations that take long time to look through.

Time is also spent for returning to the text being translated. Duration of this component is a function of user's ability to shift attention from one task to the other, his taken away from translation work. Our studies showed that out of the four phases of the dictionary use event (entering the query, waiting for response, search in results obtained, returning to translation) the last two contribute up to 94% of the total event duration. Reduction of time needed for search in translations produced by the dictionary will also reduce the duration of the last phase.

Some dictionaries limit the scope of data returned in response to translation request. For instance, Google Translate [4] responds to the query *barrel* with only 12 results sorted by their usage frequency.

This solution seems to reduce the user's efforts, however at the risk of completely fail to find the required translation. This way is similar to that chosen by authors of concise printed dictionaries that help users quickly find high-frequency lexical units.

Building a lexical resource that would combine information capacity of a large dictionary with the advantages of search in concise dictionary would offer a qualitatively new way of lexical search.

## 2.4 Introducing Interactive Dictionaries

Solutions for this problem can be found from studies of the cognitive properties of dictionaries, resulting in introduction of a new property – adaptability, i.e. a capability of a dictionary to change its behavior in order to better meet the lexical needs of the user.

Adaptability if formed in the course of user-dictionary interaction when inferences are made to detect potential patterns in the lexical need of the user. Initially, the knowledge of the user's lexical need is very limited and, therefore, the translations should be sorted by their frequencies in corpora. This knowledge can then be updated based on the collection of previous sessions – the user cycle. Moreover, for a new user it can be initially assumed that this user belongs to the most probable category of users and deals with the most probable category of texts, as determined based on the "corpus" of users' calls. Even such a simple analysis of the priors makes a great deal of difference.

Here is an example of a sequence of queries made by a user of the LexSite dictionary:

*Behind, pancreas, epigastrium, chief, undergo, underdog, extremely, carries, carrie, intracellular, formation, formation, occur, remains, remains, slightly, heredity, evolution, requirs, require, enzyme, suspect.*

From this example some inferences can be readily made. First, apparently the user does not know English very well; otherwise he would not search for the word "behind". Misspelled word "require" also supports this hypothesis, as well as other common words (extent, suspect, occur). Second, the user is dealing with a text related to medicine or health care ("pancreas", "epigastrium"). At this point the dictionary can start autonomously configuring its filters.

Smart dictionaries can greatly improve translators' performance. Developers of such dictionaries should understand that the user deals with the meaning of the texts while the linguistic content is only a means of transferring that meaning. By analyzing the prior history of queries made by all its users, the dictionary can configure its initial filters and prioritization algorithms for further interaction with new users (priors), while the analysis of the ongoing and future sessions allows it to tailor the results to specific users. Future work is aimed at further investigation of these ideas and their practical implementation.

## 3. Conclusion

Online dictionaries are one of the core components of the language learning process. While retaining the best characteristics of the traditional dictionaries, such as completeness, актуальность, объективность, and correctness, they offer new properties – fast response and reversibility, which greatly helps users seeking lexical help. When developing and teaching their courses the teachers should realize what characteristics are essential for their specific goals. To do so they need the information on characteristics of the available dictionaries. We also suggest that further innovations in the online lexicography shall be sought, such as application of cognitive abilities of dictionaries and addition of adoptive properties to the dictionaries. The authors keep working on research and development in these areas.

## References

[1] Kit, M., Berg, E. "Lexical need as a two-way reality cognition tool", Cognitive Studies, Warsaw, Institute of Slavic Studies of the Polish Academy of Sciences, 2014, p.193-202

[2] LexSite Dictionary: http://www.langint.com, 2016

[3] Kit, M., Kit, D. "On Development of "Smart" Dictionaries", Cognitive Studies, Warsaw, SOW Publishing House, 2012, p. 115-127

[4] https://translate.google.com/