COGNITIVE STUDIES | ÉTUDES COGNITIVES, 14: 193–202
SOW Publishing House, Warsaw 2014

DOI: 10.11649/cs.2014.016

MARK KIT[1,A] & **ELENA BERG**[2,B]

[1]Language Interface Inc., New York
[2]Ural State Law Academy, Language Interface Inc.
[A]mark.kit@langint.com ; [B]elenabkct@gmail.com

LEXICAL NEED AS A TWO-WAY REALITY COGNITION TOOL

Abstract

In this paper a concept of lexical need is introduced and its application in research of cognitive aspects of translation is discussed. Further discussion elaborates mechanisms of development of translator's lexical space in the course of translation. Authors discuss the importance and special nature of low-frequency lexical units and difficulties encountered when studying their usage and suggest that the lexical need concept help these studies. Lexical need analysis can be also used to learn specifics of translator's lexical space and then to take measures for selection of translators and improvement of their skills.

Keywords: lexicography, dictionary, translation, semantic representation, lexical unit, lexical need, cognition, linguistic competence.

Introduction

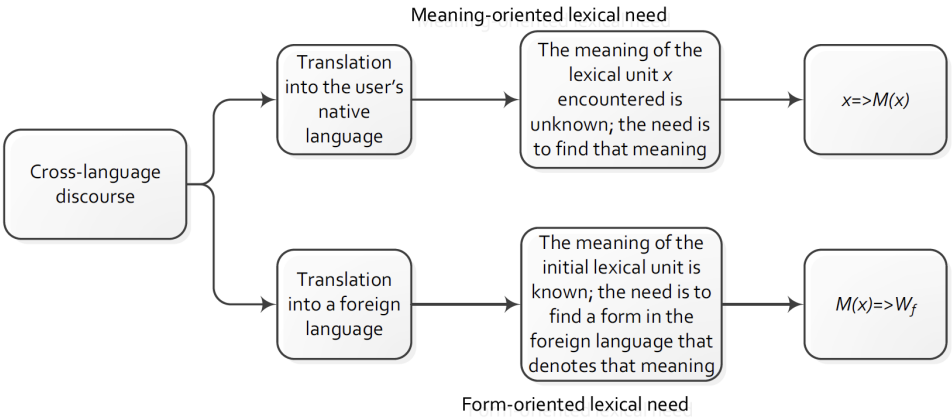
Making a dictionary query manifests a need experienced by a person when participating in discourse. Discourses are made possible by using words and combinations of words, which in this paper we call "lexical units".

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. This need would have not arisen if the user had 100% competence in the languages used in the discourse. This situation is unachievable because users cannot be completely competent even in their native languages — no living person is capable of knowing all words, expressions and professional terms of his native language. If the communication takes place across languages, the situation complexity is considerably greater.

In this paper we limit our discussion with cross-language communications only. The source of information can be either a speech or a text; in either case the original information needs to be translated, which can be done either by the recipient of the information or through a translator. In the latter case translation can go both

ways — from foreign to native languages or vice versa, so the lexical need can be of different orientation: it is either form-oriented or meaning-oriented. Classification of lexical needs based on the orientation is shown in Figure 1.

Figure 1 Different situations lead to different nature of lexical needs



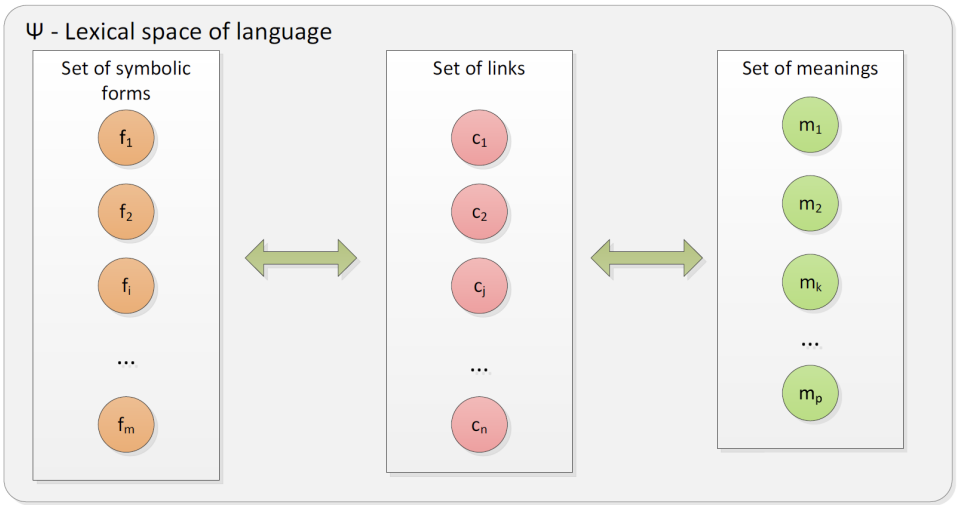
This diagram demonstrates that different communicative situations result in different types of lexical needs. When translating into the translator's native language the translator has to find out the meaning $M(x)$ of an unknown lexical unit x (emergence of a meaning-oriented lexical need), while translation into the native language calls for the necessity to find a foreign lexical unit W_f that corresponds to the known meaning $M(x)$ of the lexical unit x encountered in the native language text (form-oriented need). This is a manifestation of a lexical need that calls for a cognitive action whereupon a person enhances his linguistic competence and perception of reality.

For example, when translating the Building Code Requirements for Structural Concrete a translator encountered the term *two-way slab system*. Finding out the concept behind this term enhanced the translator's competence both in English and in the modern concrete technology.

Discussion

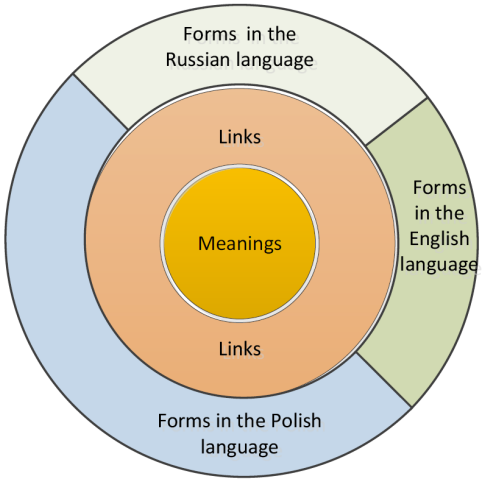
Lexical system of a language can be viewed as a space ψ composed of a set of forms (i.e. symbolic representations of lexical units, such as *street* or *heat exchanger*) and a set of meanings denoted by these forms. Each form corresponds at least to one meaning, but can denote several of them (due to homonymy or polysemy). Each meaning can be expressed with at least one form, but can be denoted by several forms (in case of synonymy). A form that is not connected to at least one meaning is meaningless and useless; a meaning that is not denoted by at least one form is useless because it cannot be expressed in a discourse. This model of a lexical system of language consists of a set of forms, a set of meanings and a set of links that connect elements of the first two sets, thus determining relations between them, as shown on Figure 2.

Figure 2 Model of language lexical space represented by sets of symbolic forms, meanings and links that establish relations between the first two sets



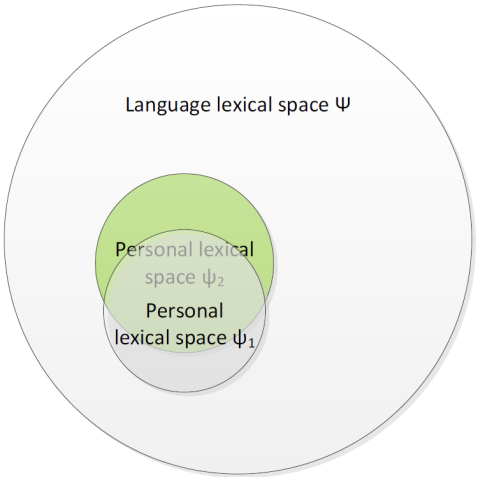
Any individual with linguistic competence in language L_i is in possession of his/her personal lexical space ψ_i, which is a limited subset of the entire lexical space ψ of that language L_i due to incompleteness of the individual's linguistic knowledge. This applies to competences in native and foreign languages. If the individual has competence in more than one language, his mind holds several subsets of forms that share the same set of meanings. A diagram on Figure 3 shows an example where the sets of forms that belong to English, Polish and Russian are connected to a single set of meanings in the mind of an individual who has competence in these languages.

Figure 3 Lexical space of individuals with competence in more than one language



It should be noted that incompleteness of the individual's lexical system may not obstruct communications because success of communications is driven not by the magnitude of linguistic competences of the communicating parties, but rather by a degree to which their lexical spaces overlap. This is shown on Figure 4, where ψ_1 and ψ_2 are lexical spaces of two individuals. Even if these spaces are very small but their overlap area is great, communication will be successful. This is true also in situations where communications take place between a text and a person, since any text is a reality created by humans.

Figure 4 Success of communications depends on the degree of overlapping lexical spaces ψ_1 and ψ_2 of communicating parties



Personal lexical spaces are rarely large. As a rule, everyday communications require only a small vocabulary that covers a significant portion of the ordinary communication needs. Beyond that small lexical space the frequency of occurrence of lexical units (LU) drops rapidly with the increase of rank of that LU in the frequency table this is true, in particular, for professional communications). This dependency is described by the Zipf law (Zipf, 1949) and can be expressed as $f = Ck^{-B}$, where B≈1 and C is a constant. This law was later enhanced by Mandelbrot who suggested another formula that fits the experimental results better, $C = C(K + V)^{-B}$, where C, V and B are constants (Mandelbrot, 1966). When V=0 and B≈1, this expression is equals Zipf law.

The exponential reduction of frequency (a Zipfian curve) results in a negligible probability of occurrence of words beyond the first 10,000 — about 1 per 100,000 words (Montemurro, 2001). It puts constraints on studies of distribution of low-frequency words since the statistics of their occurrences cannot be derived even from large corpora. This problem is mostly irrelevant to regular corpora studies conducted within the range of common vocabulary. An average educated English speaker's vocabulary includes no more than 20,000 words. Research in a group of university graduates older than 22 demonstrated that their vocabulary sizes

are between 13,200 and 20,700 words with the average of 17,200 words (Goulden, Nation & Read, 1990).

In contrast, lexical needs for low-frequency LU that arise in communications involving professional subjects are of a scientific and practical interest. For example, professional translators who graduated from foreign languages universities with excellent marks, often fail to translate even relatively simple technical texts. This phenomenon deserves attention that may yield results leading to changes in educational programs and in approaches to qualification requirements.

Since studies of this lexical need cannot be done through conventional statistical analysis, other methods shall be explored. Large corpora obscure low-frequency words and do not reflect their nature and significance. These properties of low-frequency words emerge only in special literature.

For example, the word *debridement* is in the tail of the frequency distribution curve derived from large corpora, such as the British National Corpus or the Corpus of Contemporary American English. Probability of distribution of this word is negligible: 3 per 100 million in spoken language, 1 per 100 million in fiction and 5 per 100 million in magazines, but in academic papers its frequency is 134 times greater than in fiction (1.34 per million) and it can be expected that in special literature on surgery it is even greater (Davies, 2013). Low-frequency LU studies can make use of the concept of lexical need through analysis of dictionary queries. With this approach dictionary is viewed not as a mere information retrieval system, but rather a cognitive tool.

It is relatively easy to screen out dictionary users who request commonly used words and expressions. The remaining users are mostly professional translators who use dictionaries mainly in search for professional terminology and rare words or expressions. This makes dictionary queries a good source of information on how and in what context low-frequency lexical units are used.

Lexical needs can be also used to learn about lexical space of a dictionary user and make inferences on his cognitive capacity.

Translator can develop his lexical competence in two ways: through contextual analysis or using dictionaries. In the former case a "guess" takes place where the meaning of an unknown word or expression is derived from the reality of the context and a native language form is linked to that meaning. For example, when seeing a road sign "Vancouver, BC" a translator can guess that the acronym "BC" cannot mean the usual "before Christ" and it is somehow related to the Canadian city of Vancouver. Further analysis tells the translator that Vancouver is located in the Canadian province of British Columbia, thus "BC". From that point on the meaning of the acronym "BC" is linked in the translator's mind to both "before Christ" and "British Columbia". This mechanism, however does not work when translating into a foreign language, where the uncertainty can be resolved only by search in dictionaries. It should be noted that contextual cognition of language involves analytical reasoning mechanisms employed to make inferences by means of context analysis.

The other way to meet lexical needs is to turn to dictionaries for help. In this case the translator usually faces a problem of choosing the right LU from the list of words and expressions offered by the dictionary. For example, in response to query

debris the dictionary LexSite 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 action. Detection of patterns in lexical needs of dictionary users is essential for creation of "smart" dictionaries that respond with the translations most relevant to the tasks performed by the users (Kit, M. & Kit, D., 2012).

Studies of lexical needs of users can also help solve certain organizational and professional problems, including determination of linguistic competence of a user and deficiencies in his knowledge, the nature of the text being translated and dictionary shortfalls.

The concept has been tested through studies of queries made by users of online dictionary LexSite. The first release of the dictionary was published in the Internet (`http://www.lexsite-dictionary.com`) in 2009 and provided lexical support in the English-Russian language pair, so the vast majority of data was acquired in those languages and our study is limited to that. In this paper we refer to a dataset containing about 400,000 queries made by 13,133 users in the English-Russian language pair.

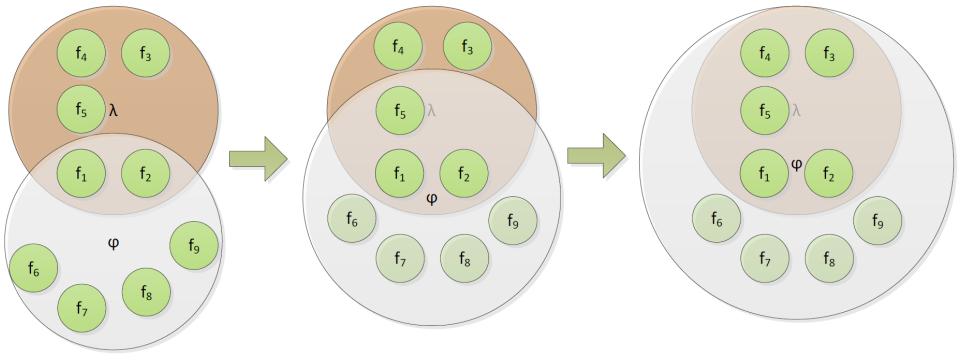
Analysis of dictionary queries can reveal different types of deficiencies in translator's linguistic competence, including limited vocabulary and knowledge of grammatical or cultural aspects needed for translation. Sometimes user has to search in a dictionary only due to poor knowledge of the foreign language grammar or cultural realities. Queries consisting of more than one word may demonstrate the user's ability to break the text into semantic units. For example, a query *furnace butt-welded pipe* manifests that the user failed to separate the semantic unit [*butt-welded pipe*] from the preceding unit [*... furnace*].

By analyzing dictionary queries it can be determined in what subjects the translators feels more confident, what aspects of the grammar, and vocabulary need to be improved. This analysis can also identify the translation team composition. Besides, translation and foreign languages can be taught using reinforced learning methods where students being trained or tested receive short texts selected based on the students' dictionary queries.

Not only meeting of lexical needs enables communications in a specific discourse, but also it develops the user's knowledge system. In cross-language communication the lexical space (LS) of the translator interacts with the lexical space of the text he works with. Ideally, these two spaces would be equal or the translator's space includes all elements of the text space; in that case no lexical need would arise. If this is not so, the translator's LS gradually absorbs the LS of the text in the course of communication with the text. This process is shown on Figure 5. In this example translator's lexical space ϕ consists of 6 elements, two of which also belong to the text's lexical space λ. When encountering an unknown lexical unit f_5, the translator finds its meaning in the dictionary or through contextual analysis and adopts it in his lexical space. If translation is perfect, upon completion of the discourse the translator's LS has completely absorbed all element of the discourse LS.

However, translator can misunderstand meanings of lexical units. In this case the translator's LS does not include all meanings of the text's LS even though it does include all forms of the text's LS. In our practice a translator took *New*

Figure 5 Expansion of lexical space in the course of translation



Orleans, LA for a list of cities New Orleans and Los Angeles while it actually meant *New Orleans, Louisiana*. The lexical space of the translator did not acquire the acronym *LA* (Louisiana) since he used the meaning that had already been available in his lexical space. It is worth noting that he did not experience a lexical need and because of that failed to search in dictionaries.

Comparing translations made by a particular translator with reference translations and analyzing queries sent to the dictionary in the course of the translation one can make inferences regarding the translator's learning capacity and his analytical capabilities. It is even simpler to evaluate that person's competence in certain areas of knowledge. Tables 1 (p. 200) and 2 (p. 201) show sequences of queries made by a user in a period of about 2 hours. The list of queries makes it clear that the user dealt with a text related to electric power generation. Queries such as "*combined cycle power plant*" prompt that the user did not know one of the basic concepts of the modern power generation. Poor knowledge of chemical terminology is clearly seen from queries "*chloride*" and "*chloride stress corrosion cracking*". Poor competence in that area manifests itself in such queries as "*gas turbine generator*" and "*steam turbine generator*". Such conclusions are very useful when making selection from a pool of translators for work in certain subject areas and for determination of fields of knowledge where the translator's competence should be improved.

Another example of sequential queries is given in Table 2 (p. 201). In this example the user called the dictionary quite often within 35 minutes. It can be assumed that either the user is a speaker of Russian and has poor knowledge of common English words (such as *practice*, *essentially*, *deposit* and *place*) or he is a speaker of English and was looking for Russian equivalents of lexical units the meanings of which he knows. Further analysis can help narrowing our hypotheses so that the likelihood of inferences we make is fairly good.

Conclusions

Lexical need arises any time the lexical space of the translator is inadequate to understand text under translation. To meet the need translator has to expand his lexical space either through contextual analysis or with a dictionary. Either option is a cognitive action that is worth being studied.

Table 1 Sequence of queries made by a user

Query	**Assumed subject**	**Time**
momentary excess current	electrical engineering	7:33 am
excess current	electrical engineering	7:34 am
associated exciter	electrical engineering	7:58 am
combined cycle power plant	electrical engineering	8:04 am
combined cycle power plant	electrical engineering	8:06 am
contractor	business	9:11 am
contractor	business	9:12 am
sinopec engineering	business, petrochemical	9:13 am
Sinopec	business, petrochemical	9:14 am
chloride stress corrosion cracking	chemistry, material science	9:14 am
chloride stress corrosion cracking	chemistry, material science	9:15 am
chloride stress	chemistry, material science	9:16 am
chloride	chemistry	9:17 am
stress	material science	9:17 am
chloride	chemistry	9:19 am
corrosion under insulation	material science	9:21 am
insulation	technology, heat engineering	9:22 am
insulation	technology, heat engineering	9:22 am
gas turbine generator	heat engineering, electrical engineering	9:24 am
steam turbine generator	heat engineering, electrical engineering	9:25 am
heat recovery steam generator	heat engineering, electrical engineering	9:44 am

Studies of usage of low-frequency lexical units cannot be accomplished using conventionally used large corpora. Yet, these units are of interest for researchers since they serve important purpose in professional communications and top-level fiction. These studies can employ the concept of lexical need and its manifestation in dictionary queries.

Finally, by studying lexical needs of individual translators through dictionary queries one can learn specifics of translator's lexical spaces, which would help develop knowledge improvement programs and select best individuals for work in translation teams.

Table 2 Sequence of queries made by another user

Query	**Assumed subject**	**Time**
treatment atmosphere	material science	6:59 am
heat treatment atmosphere	material science	6:59 am
specify by	general	7:01 am
specify	general	7:01 am
service condition	general	7:02 am
significant surface	general	7:05 am
significant	general	7:05 am
blind	general, but can be a term	7:12 am
adhesion	material science	7:14 am
adhesion test	material science	7:14 am
practice	general	7:15 am
essentially	general	7:18 am
essentially free	general	7:18 am
pore	general	7:19 am
spot	general	7:25 am
deposit	general, but can be a term	7:30 am
deposit process	general, but can be a term	7:31 am
placed	material science	7:33 am
place	general	7:33 am
process	general	7:34 am
specimens	general, material science	7:34 am

References

Davies, M. (2013). Mark Davies / Brigham Young University. *Corpus of Contemporary American English*. Retrieved from `http://www.americancorpus.org/`.

Goulden, R., Nation, P. & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, *11*(4), 341–363. `doi: 10.1093/applin/11.4.341`.

Kit, M., & Kit, D. (2012). On Development of "Smart" Dictionaries. *Cognitive Studies | Ètudes Cognitives*, *12*, 115–127.

Mandelbrot, B. (1966). Information theory and psycholinguistics: a theory of words frequencies. In P. Lazafeld, N. Henry (Eds.), *Readings in Mathematical Social Science*. Cambridge, MA: MIT Press.

Montenurro, Marcelo A. (2001). Beyond the Zipf-Mandelbrot law in quantitative linguistics. *Facultad de Matematica, Astronomia y Fisica*. Universidad Nacional de

Cordoba, Ciudad Universitaria, 5000 Cordoba, Argentina. Retrieved from `http://arxiv.org/pdf/cond-mat/0104066.pdf`.

Zipf, George K. (1949). George K. *Human Behavior and the Principle of least Effort.* Cambridge, MA: Addison-Wesley.

Publisher: Institute of Slavic Studies PAS & University of Silesia in Katowice