INTRODUCTION

This article deals with translation problems caused by a rather specific notion of lexical gaps which are understood as lack of direct lexicalisation for a certain concept. This analysis is focused on the way such problematic cases are rendered in translational dictionaries and in language use, i.e. in corpora. However, the comparison of data from the two sources is not the only goal of the article. The other and, probably, a more important aim is to construct and to test a corpus-driven methodology finding translation equivalents for lexical gaps prior to translation. Certain recommendations as to the structure of an interactive translator’s tool (a lexical data base or a lexicon) for translators (and, hopefully, for a machine translation system) follow.

DEFINING A LEXICAL GAP

Lexical gaps are instances of lack of lexicalisation detected in a language while comparing two languages or in a target language during translation. Although the problem seems to be minor and clear, one gets rather the opposite impression after an excursion through the linguistic literature on lexical gaps. The problem is a question of lexicalisation, so that, it is usually explained from this linguistic perspective. The definition of lexical gaps can be regarded as the result of linguists’ practical needs as well as their understanding of the process of lexicalisation. Thus, lexical gaps are defined as cases 1) when a source-language word does not have a direct equivalent without going into details about the notion of a direct equivalent itself (Janssen 2004); 2) when a source-language word is rendered by a target-language phrase without distinguishing it from idioms and collocations (Arnold et al. 1994, Santos 1990); 3) when a concept is not encoded by a lexical item, i.e. by a word, a complex word, an idiom and a collocation (Bentivogli and Pianta 2000, 2003, 2004; Bentivogli et al. 2000). The main difference between these definitions can easily be noticed: the specificity of a lexical item. For the purposes of this paper the main preconditions for identification of lexical gaps need to be outlined.

LEXICALISATION VERSUS LACK OF LEXICALISATION

A lexical gap means the absence of lexicalisation of a certain concept. A concept is lexicalised when a language has
a lexical item – a single word, a complex word, an idiom or a collocation – to express it. The existence of a lexical gap will be noted only when a concept lacks lexicalisation and is expressed by a free word combination or any other transformation (e.g., omission, translation by a different part of speech, etc.). Thus, the multi-word expression tarybos narys (a councillor) is not a lexical gap, because it is a fixed expression in Lithuanian, while atleidimas del darbo vietu mažinimo (redundancy) is a lexical gap, because it is a free-word combination. However, another problem that does not allow a simple way to identify a lexical gap as an aspect of lexicalisation should not be overlooked: the distinction between a free-word combination and lexicalisation by multi-word units. A semanticist or a lexicographer classifying lexical items and the absence of lexicalisation has to make a decision regarding the boundary between a free-word combination and a fixed expression (an idiom or a collocation).

Vershueren (1985) offers an easy solution to this problem. First he distinguishes between simple (a single word) and complex (idioms and collocations) lexicalisation. Then he explains the difference between complex lexicalisation and a free word combination. According to Verschueren,

(...) if someone encoding his or her thoughts would be able to produce the group of words in question without having learned it as a unity, then we have to do with a simple group of words, a free construction; if previous learning can be expected to be required, then the combination of words is a lexical item even if someone decoding the message would have no trouble at all to interpret it without having come across it before (1985:33).

Although such an explanation is useful, unfortunately, it is applicable in linguistic research only when performed by a human being. Finding translational equivalents has also become a task undertaken by computers, i.e., software programmes designed for machine or machine-aided translation. This has generated problems of different nature – “training” or programming computers to find equivalents in a target language. This poses several questions. If our lexical data source is a bilingual dictionary, how can a machine tell the difference between a lexical item and a free word combination, which is usually used to explain the unlexicalised concept? What is to be done with the explanation, which is usually unusable in translation? If it is an example-based (i.e., extracting possible equivalents from previous translations) translation system, how can a machine be programmed to find equivalents for lexical gaps? It is my hope that this study will at least partially offer some answers to the latter question.

The essential feature of a lexical gap declared by the proponents of the semantic field theory (Lehrer 1974a, 1974b; Lyons 1977) – empty spaces in a structure – is also related to absence of lexicalisation. Such an approach is valuable from the methodological point of view, when creating cognitive structures or ontology. However, explaining that a lexical gap means an empty space in a lexical field is not helpful to creators of electronic lexical resources or machine translation programmes. Knowing reasons for this phenomenon and its prior identification would allow compensating for it so that a machine translation tool produces a good quality result.

The main reason for lexical gaps, absence of lexicalisation, is not easily pinned down. However, a major group of lexical gaps can be explained by social and cultural differences of source and target language users. A lexical gap in a target language is identified when its users do
not know the concept encoded by a source language. For example, Lithuanian food names *didžukulis* or *skilandis* are untranslatable into English, because they represent specifically Lithuanian realia. Another example is English *hovercraft* and its explanation in Lithuanian *transporto priemonė su oro pagalve*. This absence of lexicalisation can be attributed to economic issues: such vehicles are not produced in Lithuania, so that the vocabulary in this field has to be borrowed, rendered by descriptive translation and, eventually, perhaps, adjusted to Lithuanian language norms. Another group of lexical gaps could be called pragmatic due to various differences in language paradigms. A subtype of this group is derivational gaps, originating from word formation differences. For example, *biculturalism* is translated by a free word combination *prisklausymas driem kultūroms* due to the “bi-”, which is absent in Lithuanian. Unfortunately, there are many instances, when no obvious explanation of the phenomenon is available. Such cases are deservedly called “lacunae” (Nida 1975, Armalytė and Pažūsis 1990), for example, *fisheries* – *žvejybos plotai* or *habitat* – *gyvenamoji aplinka*. Although Lithuanian speakers know these concepts, they are nevertheless unlexicalised.

IDENTIFYING LEXICAL GAPS

A lexical gap is identified in one language during a language contact: comparison, description or translation. Although Lehrer (1974b), Lyons (1977) and Kjelmer (2003) analyse the phenomenon within one language, most linguists tackling the problem hold the view that it is a question of language contact. A lexical gap found through language contrast can be called pure. Lehrer calls lexical gaps that really reflect a speaker’s need to express a concept (excluding illogical concepts like a *castrated woman*) functional gaps (1974b:104-106). Lexical gaps identified within the limits of one language (usually in the structure of a semantic field) are valuable for the linguistic interest; however, their practical value should derive from the needs of language users. It is important to note that a lexical gap appears only in one language. In other words, during translation from English to Lithuanian, we identify a lexical gap only in Lithuanian and vice versa. The same lexical gap in both languages is impossible, because, for its identification, one of a language pair has to be the point of departure or *tertium comparationis*.

A lexical gap is identified on the level of one meaning, not of a lexeme, which is usually polysemous. In translation we deal with one meaning, realising itself in a context specific only to a particular meaning. A translator is interested in individual meanings but not the semantic structure of a word. Therefore, lexical gaps are identified on the level of individual meanings.

In general we can say that a lexical gap is a concept which is not lexicalised by a lexical item (a single word, a complex word, an idiom and a collocation) in a language due to cultural, language/systemic or other reasons, it is identified while comparing or translating individual meanings of lexical items in two languages and is usually expressed by a free word combination in translation.

LEXICAL GAPS IN TRANSLATION

One of the reasons to study lexical gaps is that it is difficult to identify them in advance. Only during translation does one understand that the target-language
lacks a certain word, while a dictionary in such cases provides a mere explanation of the concept encoded by a source language. Unfortunately, such meaning explanations usually are not good in natural language use. For example, *The English-Lithuanian Dictionary* (Piesarskas 2000, further in the text – ELD) offers the following translation equivalents for the noun *deforestation*: *(iš)kirtimas/* *(su)deginimas* *(miško tam tikrame rajone)*. The thesaurus EUROVOC translates it as *mišky naikinimas*. Such dictionary translations pose several questions. First, if the two translations are compared, we get confused over the specificity of the meaning. Does the noun really mean either cutting or burning of woods or does it only imply the general destruction of forests? Second, the remark in the parenthesis puts forward a condition for the usage of the noun: it is the destruction of forests in a certain area or region. Third, the failure to provide one and single-word equivalent suggests the need for a multi-word, extended unit of translation that could cover a concept encapsulated by a single English noun.

This problem is not acute when human translators have to deal with it. They can select an appropriate strategy of compensation, make certain changes in a sentence and render the meaning of the source text properly. However, if a lexical gap is not known beforehand and is not recorded, a machine translation program cannot solve the problem by a mere transformation. Therefore, lexical gaps have to be identified and described for a machine translation program in advance. This is a twofold task. First, we have to construct a methodology that could select possible or acceptable equivalents for lexical gaps, i.e., for cases, when a target language actually lacks a lexical item. Second, the product of such selection should conform to qualitative target language norms. This implies that the selection should not be performed randomly and should be based on quantitative data.

Filling lexical gaps influences the quality of a translated text. Thus building a lexical data base or a lexicon which addresses this issue not only helps a translator to save time on looking up dictionaries, consulting other reference sources, and going over the already translated texts, but also approaches the broader notions of equivalence and non-equivalence. Although lexical gaps represent a very specific linguistic problem, they are just as important as regular correspondences and should be recorded in a database to be a kind of a desktop translation aid for a translator. The realisation of lexical gaps in translation, or, rather, methodology for finding translation equivalents can serve for a broader phenomenon of non-equivalence. If the proposed methodology is effective with lexical gaps, it is also applicable in finding regular or unproblematic translation equivalents.

**THE CORPORA**

The choice of data for this study was determined by two factors: the availability of representative corpora and the methodology, which is based on the analyses carried out by Tognini-Bonelli (1996, 2002), Tognini-Bonelli and Manca (2004). She makes use of both a translator’s or linguist’s competence and data from comparable corpora. The method will be outlined in the following section, but first, another important source of data, a parallel or translational corpus, needs to be discussed. Tognini-Bonelli, too, admits that exploring data from a parallel corpus would have given interesting insights to her study. Unfortunate-
ly, a parallel English-Italian corpus was unavailable when she carried out her research (2002).

A parallel corpus, i.e. a collection of aligned source and target language texts (unidirectional or bidirectional, where translation has been performed in both directions), can be a solution to various translation questions arising from mismatches of various nature between two languages. A parallel corpus not only provides translation equivalents that are in actual language use (as opposed to the information from most bilingual dictionaries), but also offers data on language variation (when a source language unit can be translated by several target language units) due to conceptual, contextual or stylistic differences, thus offering more options for a linguist or a translator using the data. Moreover, a parallel corpus can show a strategy employed by a translator in case of non-equivalence and lexical gaps; whereas bilingual dictionaries offer a mere gloss of a meaning captured by a source language unit. For example, in ELD (2000) the verb *supersede* is translated as *ištumti, daryti ne(be)reikalingą; užimti (ko) vietą, or commute* is explained as *regulariai važinėti į darbą ir atgal*. These explanations only help to understand what concept is encoded by a foreign lexical unit. However, its use in language, especially for multi-word meaning explanations, in language would produce a cumbersome, unnatural structure. In such cases a parallel corpus is really useful in providing real-language use equivalents.

Despite the new opportunities arising from a parallel corpus, the issue of data quality must be addressed. On the one hand, a large number of translations is more useful than a few examples provided in a bilingual dictionary. Besides, a lexicographer is preoccupied with finding equivalents for isolated items and in the case of lack of equivalence, explaining the concept encoded by a source language item. The main purpose of a translator, on the other hand, is to render the function of a source language item into the target situation, culture and language. However, caution should be used due to the fact that a translated text is written in “a mediated language” (Tognini-Bonelli 2002:81). Although Teubert says that “a parallel corpus of a reasonable size contains more knowledge about translational equivalence than any bilingual dictionary”, he also declares that “translations, however good and near – perfect they may be (but rarely are), cannot but give a distorted picture of the language they represent … rather than representing the language they are written in, they give a mirror image of the source language” (1996:249; 247). Altenberg and Granger (2002) also have ambiguous ideas about the use of parallel corpora. They maintain that “although every translation is worth considering as a potential translation equivalent as it reflects the translator’s competence, translations are rarely literal renderings of the original. Translators transfer texts from one language (and culture) to another and the translation therefore tends to deviate in various ways from the original” (2002:17). Both Teubert and Altenberg and Granger believe that parallel corpora could be complementary to comparable corpora.

The English – Lithuanian parallel corpus became publicly available in the autumn of 2005 and is still in its initial stage in regard to its size. Although it is quite neatly programmed (offering several options for search and display of search results), it is too small to arrive at substantial generalizations: today it contains only 6768 aligned sentences. Although one can study translations of
frequent words, it is hard to perform a reliable research on rare words. Lexical gaps, unfortunately, usually represent rarer words. For this reason, the parallel corpus will not be the main source of data in this study. However, whenever a translation for a lexical item in question is available, it will be considered prior to data from bilingual dictionaries.

Although not parallel, comparable corpora are also greatly useful for linguistic research. Comparable corpora are used to collect translation equivalents, to explore usage patterns of terms and other lexical items, to analyse cultural issues reflected in language and for other purposes (Marcinkevičienė 2000:14). Data for the English lexical items were collected from the British National Corpus (further in the text – the BNC) of 100 million words. The search of possible counterparts in Lithuanian was performed in the Corpus of Contemporary Lithuanian, which is an electronic corpus of more than 100 million running words. The two corpora can be considered as comparable because they comprise various genres and are of similar size, while the analysis of such corpora can be considered valid.

THE METHODOLOGY

The methodology for establishing equivalents for lexical gaps is based on the essential postulate of corpus linguistics, the extended unit of meaning. Tognini-Bonelli metaphorically calls this notion “a new currency unit” (2002). The extended unit of meaning comprises collocation (the relation of a word with its collocates), colligation (the relation of a word with certain grammatical categories), semantic preference (the relation between a word and a set of semantically related words) and discourse prosody (the evaluative, emotive aspect of a word) (Sinclair 1996, Stubbs 2001, Tognini-Bonelli 2000). Tognini-Bonelli also calls such extended units of meaning “functionally complete units of meaning”, as they merge two dimensions, the contextual one and the functional one (2002:79). On the basis of the extended unit of meaning, she sets out a certain procedure of establishing translational equivalence. It involves three stages:

1) identifying and classifying the formal patterning in the context of a given word or expression against the evidence of an L1 corpus, followed by the matching of a specific meaning/function to each specific pattern;
2) identifying a prima facie translation equivalent for each meaning/function, ideally from a parallel corpus or, if such is not available, from a bilingual dictionary or using the analyst’s intuition;
3) deconstructing a function in L2, realised by the prima facie equivalent, into its formal realisations (collocational and colligational patterning), against the data of a comparable corpus in L2 (2002:81-83).

Following this procedure Tognini-Bonelli analyses the functional equivalence of expressions containing English word case and the Italian caso. For expressions in the case of, in case of and in case and their Italian counterparts nel caso di, in caso di and se per caso she constructs and compares functionally complete units of meaning, which are found to be similar (2002). In another study, Tognini-Bonelli employs a similar approach for defining functional equivalence between an English adjective real and its prima facie equivalent reale (1996). The procedure leads to the conclusion that the Italian counterpart reale only partially corresponds to the English real. Therefore, the same procedure is carried out with another possible equivalent, vero, arriving at specific recommendations for a translator’s choice (206-214).
The methodology proposed by Tognini-Bonelli can be useful in case of lexical gaps, i.e. when a concept in a target language lacks a *prima facie* equivalent, but needs to be expressed. The notion of an extended unit of meaning or a functionally complete unit of meaning becomes handy. It is obvious that collocational and colligational patternings in English and Lithuanian can differ. On the other hand, these are not the notions we look for to establish translational equivalence. Rather, they are the end products to be used in translation. If a target language lacks a direct equivalent for a source language item, it is worth exploring the latter’s wider context. To put it in terms of corpus linguistics, to establish the relationship of equivalence, the semantic preference of a source language unit needs to be investigated and mapped on the one in the target language. Needless to say, the comparison of discourse prosody in these semantic areas of source and target languages is also a decisive factor in determining equivalence.

Unfortunately, here the first methodological problem that lies in the nature of the phenomenon being investigated appears. The absence of lexicalisation means the absence of an equivalent so that no direct equivalents can be derived from a bilingual dictionary. In such cases, a dictionary provides explanations of the meaning, but not a possible equivalent. Two remaining sources of information for *tertium comparationis* are the linguist’s subjective intuition or, preferably, a parallel corpus. In the case of this study, the parallel corpus will be consulted first and, if produced, the translations will be considered as *prima facie* equivalents. Otherwise, any available bilingual dictionary data will be tested against the comparable Lithuanian corpus in the third stage.

**THE DATA USED IN THE ANALYSIS**

The English words that will be analysed here represent derivational lexical gaps in Lithuanian because they lack direct equivalents. I adopted the term “derivational gaps” from Lehrer (1974b) and Kjellmer (2003) who identify derivational gaps within one language. According to them, derivational gaps are words produced from partially productive stems and suffixes, which are understandable, but not acceptable in a language. For example, although we understand the meaning of *ungood*, *mistelephone*, *converse*, *fryable* or *bakable*, they are unacceptable in the English language (Lehrer 1974b:96-97). The derivational gaps in this study, differently from the just mentioned ones, are gaps in Lithuanian, i.e. in the target language, identified within the translational framework of English and Lithuanian. I assert that these lexical gaps result due to derivational processes in the source language. In other words, the concepts encoded by the English words are not lexicalised in Lithuanian by one word due to different derivational processes, prefixation or suffixation. For this article translation equivalents for two words – an adjective *overwhelming* and a noun *overcrowding* – will be searched. The two words have a common prefix “over-”, which adds the element of excess to their meaning and which seems to cause certain translational problems in Lithuanian. The two nouns are also attractive to study for methodological reasons. *Overwhelming* appeared in the English – Lithuanian parallel corpus, while *overcrowding* did not. Therefore, it was interesting to see how and if different data sources could influence the methodology structure proposed by Tognini-Bonelli.
THE RESULTS OF THE ANALYSIS

OVERWHELMING MAJORITY, OVERWHELMING DESIRE AND OTHERS

The reason for selecting overwhelming was its "empty" nature. Sinclair (1992) distinguishes "selective" (broadly speaking, adding a separate or additional meaning to its collocational partners) and "focusing" (repeating a part of meaning of a noun) adjectives. For example, scientific in scientific analysis adds little to the overall meaning of the expression (Stubbs 2001: 32-34). Overwhelming, although translated by single words in Lithuanian, seems to be at least partially focusing and dependent on the nouns it combines with. The New International Webster's Comprehensive Dictionary of the English Language (1996) defines overwhelming as crushing by reason or force, weight or numbers; irresistible. ELD plainly provides the following information for the adjective:

1. triuškinantis
2. didžiulis, nesuskaičiuojamas; overwhelming majority didžioji dauguma
3. neįveikiamas, nenugalimas

It is obvious that a dictionary user will be puzzled over the choice between similar meanings enumerated as different senses (especially the first and the third one), while a translator will not be able to rely on such data. Can we combine overwhelming with enemy (because we can say neiveikiamas priešas in Lithuanian) or overwhelming house (didžiulis namas)? In such fuzzy cases, it would seem logical to provide longer stretches of translation, showing the possible combinations of the word in question. To render its meaning precisely, its contextual partners should be examined in texts. For the sake of space and time, only the first most frequent 200 surrounding words or contextual partners within the span 5:5 of overwhelming were analysed. They are provided in Table 1.

Functional words, as well as verb forms have, had, is, was, etc., were disregarded as they have little significance for the

<table>
<thead>
<tr>
<th>NOUNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. People, population, world</td>
</tr>
<tr>
<td>2. Majority, number, numbers, proportion, mass</td>
</tr>
<tr>
<td>3. Support, response, interest, approval</td>
</tr>
<tr>
<td>4. Impression, impact, influence, emphasis, importance</td>
</tr>
<tr>
<td>5. Power, victory, force, success, strength, predominance, advantage, forces, war</td>
</tr>
<tr>
<td>6. Evidence, vote, opposition, government, election, party, case</td>
</tr>
<tr>
<td>7. Desire, sense, feeling, need, urge, demand</td>
</tr>
<tr>
<td>8. Pain, fear, relief, loss, joy, compassion, temptation, confidence, love, passion, stress, experience, emotion</td>
</tr>
<tr>
<td>9. Time, fact, odds, seats, nature, presence, consensus</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VERBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Felt, gave, found, face, state, feel, won, given, received, became, become, becomes, say, showed, seems, said, see, show, made</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ADJECTIVES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical, military, public, early, popular, remember</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OTHER FUNCTIONAL WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Against, in spite of, despite, in favour of</td>
</tr>
</tbody>
</table>
meaning of the adjective, because they are not linked by a strong syntactic relation. Verbs, adjectives or other functional words which are provided in the table do not seem to have much influence on the meaning, either. Needless to say, the most important lexical and syntactic partner for an adjective is a noun and, in this case, contextual partners that are nouns are abundant. They could be classified into several semantic groups, thus revealing a certain semantic preference by the adjective. Especially interesting are the surrounding nouns related to quantity or amount (group Nr. 2), power (group Nr. 5), wish (group Nr. 7) or certain feelings or emotions (group Nr. 8). Its contextual partners partially support the idea of overwhelming as a focusing adjective. Thus, majority still means majority, even though it is overwhelming. Power, force, victory, desire do not acquire much from their collocate, either.

The adjective forms two major colli-gational patterns:
1. Noun + be/become + overwhelming
2. overwhelming + noun

It is also interesting that the second pattern is often preceded by functional words almost, so, such, quite, very and too, highlighting the intensity of the combination. In summary, it can be said that the adjective shows a strong tendency to combine with abstract nouns and that its function is to emphasize the intensity of its immediate neighbours.

Following the methodology outlined by Tognini-Bonelli, we now need to establish prima facie equivalents for overwhelming. The English-Lithuanian parallel corpus provides only three translations as presented in Figure 1.

Three translations are not enough to draw sound conclusions, but the implication is still clear: the collocation overwhelming temptation is rendered by a transformation. The noun-adjective phrase is transformed into an adverb+verb phrase. On the other hand, the colligational patterning “be+overwhelming” is translated by an equivalent offered by a translational dictionary. However, didžiulis, as well as other suggested equivalents – nesuskaitėjomas, nėjveikiamas, nenugalimas – can be used as translational equivalents for other English adjectives. Although the parallel corpus is unidirectional – English – Lithuanian – and the method of back-translation is not applicable, it is still possible to check for how many adjectives, for instance, didžiulis is used as an equivalent. The following adjectives were translated by didžiulis:

---

**Originalo tekstas**

<s>The work was overwhelming, all the more so because the processes that it involved could not be called by their true names.</s>

<s>He had an almost overwhelming temptation to shout a string of filthy words at the top of his voice.</s>

<p><s>For a moment he had an overwhelming hallucination of her presence.</s></p>

<s>Akimirką ji užvaldę nenugalimas požūtis, kad ji yra šalia.</s>

---

**Vertimas**

<s>Darbas buvo didžiulis, juoja kad procesai, reikalingi jam atlikti, nevadintini tikrais vardais.</s>

<s>Labai norėjosi visu balsu rėkinti nepadorius žodžius.</s>

---

Figure 1. Overwhelming in the English-Lithuanian Parallel Corpus
enormous, great, large, vast, huge, tremendous, mighty, big. Overwhelming is related to this group of words, but it has an additional meaning component of excess or a crushing intensity. This implies, that overwhelming lacks precise equivalents in Lithuanian and dictionary data should take a different shape, perhaps by presenting functionally complete, extended units of translation, in both source and target languages. It is understandable that this is an improbable task to demand from a printed dictionary, but electronic lexical resources have this ability.

The next step in Tognini-Bonelli’s methodology of tracking matching functionally complete units of meaning in two languages is the construction of a function performed by prima facie equivalents in an L2 corpus. This is complicated in the case of overwhelming, which seems to lack an exact equivalent in Lithuanian. On the other hand, as a focusing adjective, it forms strong collocations. In the BNC corpus, the strongest collocation is overwhelming majority (167 times). Other collocations are: overwhelming desire (25), overwhelming support (19), overwhelming evidence (19), overwhelming number (12), overwhelming victory (11), overwhelming response (11), overwhelming feeling (11), overwhelming importance (10) and others.

If prima facie equivalents are not obvious, the function matching the one of overwhelming could be found by looking for similar collocations in Lithuanian. Thus, we looked at the first 200 immediate contextual partners (in case of an adjective, the most important collocational partner is a noun, the first to the right) of one of its translations in the dictionary – didžiulis (in the corpus realised by various cases of feminine and masculine both in singular and plural). Didžiulis appears in the following collocations:

didžiulė įtaka (171), didžiulis įspūdis (64), didžiulė atsakomybė (66), didžiulė žala (31), didžiulišia akmenys (28), didžiulė klaida (27), didžiulė įtampa (26), didžiulės galimybės (24), didžiulis darbas (21), didžiulis akmuo (21), didžiulis buryš (17), didžiulės automobiliai (16), didžiulės energijos (15), didžiulė jėga (15), didžiulė dauguma (15), didžiulė aistra (9), didžiulės bažnyčios (6), didžiuliaiakiniai (6), didžiulės akys (5).

The immediate contextual partners could be classified into two groups: 1) abstract nouns corresponding to those combining with overwhelming and 2) concrete nouns referring to physical size. Although some collocations with didžiulis match those of overwhelming (influence-įtaka, impression-įspūdis, tension-įtampa, force-jėga, majority-dauguma, desire-aistra), didžiulis lacks the meaning component of excess and this is not compensated by any further contextual partners in its concordance. Thus, didžiulis only partially corresponds to the English overwhelming.

The combinability of didžiulis could also be tested against corpus evidence for the most frequent collocates of overwhelming. Majority is the top collocational partner for overwhelming: in the BNC the adjective and the noun appear together 167 times. The most frequent collocations with its Lithuanian counterpart, dauguma, related to amount or intensity are the following:

valdančioji dauguma (275), absoliuti dauguma (126), didžioji dauguma (49), didesnė dauguma (11), didelė dauguma (8), dominuojanti dauguma (1)

Although the first collocation, valdančioji dauguma, is strongly linked to a political context, it is also related to a certain power or force, which are implied by the discourse prosody of overwhelming. We can see that dauguma forms stronger collocations with other adjectives as well. However, none reflect the crushing nature of overwhelming.

The equivalent of the next most often collocate of overwhelming, desire (25), aist-
ra, showed a great variation of collocates related to intensity:
beprotiška (24), kunkuliuoja (20), begalinė (14), didelė (13), nežabota (12), didžioji (10), didžiulė (9), nenumaldoma (8), didžiausia (7), didesnė (7), aukščiausia (6), nenugalima (6), stipri (6), kunkuliųjančios (5), nesutramdoma (4), stirpiausia (4), galinga (4), žmogų draskanti (3), nesibaigianti (3), nepaprasta (2), nesuvaldoma (2), didi (1), neapsakoma (1), pragaistina (1)

The collocates of aistra, contrary to the ones of dauguma, reflect the meaning of overwhelming. The huge variation shows that using corpus evidence solve transatlational questions and produce new effective lexicographic data – desktop tool for translators. However, it is also evident that the unit of translation, matching between target and source language, should expand from single words to multi-word expressions.

POPULATION, SCHOOL AND PRISON OVERCROWDING

The noun overcrowding is interesting for two reasons. First, it seems to constitute a derivational gap in Lithuanian due to the combination of the prefix “over-” and “population” and the absence of a similar combination or amalgamation of the two concepts into one word in Lithuanian. Second, the noun denotes a concept that, although known to the Lithuanian speakers, is not psychologically salient, i.e. it is not vitally significant in the Lithuanian society. By contrast, underpopulation or its danger might be a more important issue in the present day context.

The noun overcrowding appeared in the BNC 289 times. Although it does not form striking collocational or colligational patterns, the findings are still worth mentioning. Together with semantic preference and discourse prosody, the noun seems to perform a clear function in the English language. It turns out that most contextual content (as opposed to functional) partners presented in Table 2 of the noun have a heavy negative connotation.

It can be seen from the table that the discourse prosody related to overcrowding is rather negative. Contextual noun-partners could be divided into three groups. The first one helps to differentiate between two senses of the noun: 1) overcrowding in prisons, schools or other public institutions and 2) population overcrowding. It must be said that the latter meaning is not strongly expressed and only the contextual partners urban or areas suggest a more general meaning. The second group of nouns discloses the problematic nature of the phenomenon encoded by the noun. Overcrowding causes poverty, pollution, disturbances, unemployment

<table>
<thead>
<tr>
<th>Part of speech</th>
<th>Contextual partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUN</td>
<td>Prison, prisons, school, urban, areas, trains, jail</td>
</tr>
<tr>
<td></td>
<td>Problem, problems, poverty, slum, congestion, clearance, squalor, pollution, unemployment, understaffing, disturbances, restrictions, difficulties, crime, overfeeding, homelessness</td>
</tr>
<tr>
<td>VERB</td>
<td>Housing, facilities, factors, population, staff, home, years, sanitation, families, space, health, effect, city, Britain, tenancies, immigrants, conditions</td>
</tr>
<tr>
<td></td>
<td>Lack, ease, caused, reduce, result, improving, designed, answer, became, report, spread, tackle, reducing, avoid</td>
</tr>
<tr>
<td>ADJECTIVE</td>
<td>Poor, bad, local, severe, physical, acute, worst, chronic, serious, greater, main, insanitary, undesirable, sanitary</td>
</tr>
</tbody>
</table>
and other problems. The third group of nouns could be regarded as neutral. However, these words easily aquire connotations from their neighbours in the texts. Thus, *immigrants* become negatively shaded when they are encountered in contexts about poverty and unemployment; likewise, *conditions* can be excellent or poor depending on the surrounding words. The same could be said about *facilities, sanitation* and other nouns in this group. The negative discourse prosody is also strengthened by adjectives loaded with exceptionally negative connotations. The surrounding verb-partners show that *overcrowding* spreads and causes problems which should be tackled, reduced or somehow stopped.

The noun *overcrowding* does not form numerous collocational patterns. However, the few that have been noticed help clarify the function performed by the noun in texts. Hopefully, tracking the same function in the target language could help find translational equivalents. Thus, collocations with the noun *overcrowding* are provided below:

- prison overcrowding, severe overcrowding in, relieve overcrowding, reducing overcrowding, due to overcrowding, because of overcrowding, ease overcrowding, urban overcrowding, avoid overcrowding, chronic overcrowding, acute overcrowding, overcrowding problem, relieving overcrowding, serious overcrowding

Although the colligational patterns with the noun are not significant, still they contribute to the functionally complete unit of meaning. Thus, the noun appears in the following structures:

1. Verb (from the semantic field of ending) + overcrowding
2. Noun (from the semantic field of problem) + overcrowding
3. Overcrowding in + noun (location)
4. Noun (problems related to overcrowding) and overcrowding
5. Overcrowding and noun (problems related to overcrowding)

Such colligational constructions on a more abstract level (showing semantic preference and discourse prosody) constitute the extended unit of meaning or in Tognini-Boneli’s terms – a functionally complete unit of meaning.

Having constructed the functionally complete unit of meaning for the noun *overcrowding*, looking for prima facie equivalents in Lithuanian is the next step in the analysis. Unfortunately, the *overcrowding* query in the English-Lithuanian parallel corpus produces zero results. Thus, the lexicographic data had to be used instead. *ELD* provides the following translations of the noun:

1. *(kalėjimo ir pan.)* perpildymas;
   overcrowding in classrooms *perpildytos klasės*
2. gyventojų perteklius *(mieste)*

The ES Thesaurus EUROVOC suggests only one translation: *gyventojų perteklius*.

Such translations produce several difficulties. First, *perpildymas* alone does not represent the meaning encapsulated by *overcrowding*. The Lithuanian counterpart alone means the excessive result of filling a container or a place with something. Thus it can mean a very crowded classroom or a prison, as suggested in the parentheses of the explanation in the dictionary, as well as a too-full glass of beer. This suggests that providing a dictionary user with one word with uncertain hints in parentheses is misleading. Moreover, in the expression *overcrowding in classrooms*, the noun *overcrowding* is translated by an adjective. The second sense is defined as related to overcrowding in towns. However, the expression is too general and could be used for overcrowding in rural areas, too.

When a word lacks a single-word translation, it is rather difficult to decide on the way a functionally complete unit of meaning in a target language should be constructed. Thus, concordances for the
single words *perpildymas* and *gyvenotoju*, the words appearing in the lexicographic data, were generated.

The concordance for *perpildymas* contained 17 lines, out of which 3 had the expressions *kalėjimų perpildymas* (prison overcrowding), 1 – *klasių perpildymas* (classroom overcrowding) and 1 – *gyvenotojų perpildymas* (population overcrowding). The remaining lines refer to overfilled containers or an occupied market. The numbers are too small to draw any sound conclusions. However, we can see that overcrowding in prisons or schools is a less important topic in Lithuanian discourse. For this reason, the concepts are not lexicalised by one Lithuanian word and to avoid ambiguity in translation, the noun overcrowding must not be rendered by a single word.

The analysis of the concordance for the noun *gyvenotojų* offers several expressions that could be used for the English noun *overcrowding* in its second sense, *population overcrowding*:

- Gyventojų perteklius – 38 lines
- Gyventojų skaičius – 26 lines
- Gyventojų tankumas – 18
- Gyventojų tankumas – 3

Since the concordance with *gyvenotojų* comprises 514 lines, the small numbers next to the expressions above are really telling. Indeed the problem of overpopulation is not acute for Lithuania. Therefore, lexicalisations of the concept are really poor and not numerous. It is also apparent that *gyvenotojų skaičius*, *gyvenotojų tankis* or *gyvenotojų tankumas* are more general than *gyvenotojų perteklius*. Their contextual partners were rather neutral: *auga, didėja, mažėja, siekia*. *Gyventojų tankis* in combination with its contextual partners – *nedidelis, mažas, didelis, išaugęs*, *per didelis* – especially the last one, suggest a functional relationship to the English *overcrowding*. However, the closest counterpart for the English *overcrowding* turned out to be *gyvenotojų perteklius*. Once again, I looked at the first 200 contextual partners were examined to see the semantic preference of the expression *gyvenotojų perteklius*. The contextual partners are provided in Table 3.

Some contextual partners for *gyvenotojų perteklius*, like the ones for *overcrowding*, also have a negative discourse prosody: *spaudinam, bado, inflacijā, nedarbas, neramumai, skursta, neproduktysi* and others. However, it must be stressed that the majority of the words within the semantic scope of the expression *gyvenotojų perteklius* are rather neutral. Most of them create the impression that the speaker of Lithuanian refers to *gyvenotojų perteklius* from a distance, as if the issue is relevant only to the extent of discussing

<table>
<thead>
<tr>
<th>Part of speech</th>
<th>Contextual partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUN</td>
<td>Kaimo, pasaulio, Anglijos, šalyse, Kinija, Kinijai</td>
</tr>
<tr>
<td></td>
<td>Spaudinam, bado, badas, baimės, ekspansijos, inflacijā, karas, katastrofa, korupcija, maras, nedarbas, neramumai</td>
</tr>
<tr>
<td></td>
<td>Armija, atlyginimams, ciklo, dėsnio, dėsnis, dirbančiųjų, doktrina, epochoje, darbo, paklausos, pasiūlos, perskirstymo, turto, administracija, argumentas, ateities, žemvaldžių, dalis, darvinizmo, demografinės, demografiniai, ekonomikos, faktas, kapitalizmo, kultūra, migracija</td>
</tr>
<tr>
<td>VERB</td>
<td>Apriboja, egzistuojā, skursta, atvesti, egzistavo, gabentis, gribės, grindžia, gyvename, įžengti, kritikuoti, nekentėjo, nemoka, nusikrąstytis, nusėpti</td>
</tr>
<tr>
<td>ADJECTIVE</td>
<td>Teisinga, didžiausias, ekonominio, valstybinio, agrarini, žemės, didelis, didesnis, intensyvi, kaimyninėei, klimatinės, laisva, neproduktivy, netinkamas, nunykusi, nuolatinis,</td>
</tr>
</tbody>
</table>
about economic figures somewhere further from Lithuania, in other countries (šalyse) or in the world (pasaulyje) in general.

Perhaps because of its more neutral nature, the expression does not appear in any specific colligational patterns that could be matched to the ones for overcrowding. Having considered its frequency, semantic preference and discourse prosody, the conclusion can be drawn that gyventojų perteklius can be considered a translational equivalent of overcrowding (i.e., population overcrowding), though with a more neutral connotation. It is unfortunate that the English-Lithuanian parallel corpus does not provide any instances of translations of overcrowding. It is possible that some techniques could be employed to compensate the lack of negative connotations, or even other expressions, gyventojų tankis, gyventojų tankumas or gyventojų skaičius, would be used.

CONCLUSION

The method proposed by Tognini-Bonelli for matching equivalents in source and target languages, which is based on the analysis of corpus data, is applicable to lexical gaps, when a language actually lacks a direct equivalent. However, it must be emphasized that it has to be customized to the specificity of the problem.

In case of the adjective overwhelming, which is rendered by a number of synonymous translational equivalents in the bilingual dictionary, the procedure is more effective when matching shifts from a single-word to multi-word expressions.

When a prima facie equivalent is already a multi-word expression (as in the situation with the noun overcrowding and its translational equivalents), the method also has to be modified. The suitability or unsuitability of the proposed equivalents is judged against the corpus evidence for all members of a multi-word expression. A comparison of semantic preference and discourse prosody is very important in deciding which expression best suits a source-language lexical item.

Tognini-Bonelli’s methodology could be improved if a parallel corpus of considerable size was available. This would lead directly to techniques employed in previous translations and thus reveal the realisation of lexical gaps in language use. Unfortunately, in this study the parallel corpus so that possible equivalents had to be sought from comparable corpora.

The results of the analysis carry strong implications about the shape of present day lexicographic data. It is clear that printed dictionaries cannot move from the single-word to the multi-word paradigm; however, electronic dictionaries have the potential to help translators to solve many translation problems. First, they should lead a translator to a concordance (preferably, a parallel concordance) of a word in question. Alternatively, the end user could get preprocessed corpus material. Such preprocessed electronic databases could contain semantic preference and marks about discourse prosody both of a source and target language words and, most importantly, the most frequent collocations that a word forms. Matching these elements would be a sound reason to establish an equivalence between a source and target language items.
REFERENCES


SOURCES

Dabartinės lietuvių kalbos tekstynas. http://donelaitis.vdu.lt/
EUROVOC. www.lrs.lt
vdu.lt/main.php?id=4&nr=7
The British National Corpus. http://view.byu.edu/

Gauta 2006 01 17
Parengta 2006 03 01
Jurgita CVILIKAITE

LEKSINIŲ ERTMIŲ PROBLEMA IR JOS SPRENDIMAI

Santrauka