

CORPORA FOR APPLIED PURPOSES: A CASE STUDY OF QUANTIFIERS IN ENGLISH AND LITHUANIAN

Jūratė Ruzaitė

Abstract. The present paper aims to show how a cross-linguistic analysis based on a parallel corpus can be used for numerous practical applications in such areas as translation, lexicography and language teaching. The paper examines non-numerical quantifiers (e.g. *a bit, a few, few*, etc.) in English and Lithuanian. The analysis is based on the Parallel Corpus of the Lithuanian Language and The Corpus of Spoken Lithuanian. Where relevant, the issue of equivalence between English and Lithuanian is discussed by comparing the data to results obtained from the British National Corpus. The study shows that a parallel corpus is especially useful when searching for equivalents in the target language since it clearly displays the differences in the inventory of a particular category in two languages. The establishment of equivalent categories in English and Lithuanian is especially important in terms of lexicography and translation.

Keywords: parallel corpus, quantifiers, translation, equivalents, language teaching, English, Lithuanian

1. Introduction

The present paper is a cross-linguistic analysis of non-numerical quantifiers based on a parallel corpus, the results of which, as will be argued further, can be used for numerous practical applications in such areas as translation, lexicography and language teaching. The paper examines non-numerical quantifiers (e.g. *a bit, a few, few*, etc.) in English and Lithuanian. The analysis is primarily based on the Parallel Corpus of the Lithuanian Language, which contains almost 70 000 parallel sentences translated from English into Lithuanian and 1614 sentences translated from Lithuanian into English. In addition, the data are supplemented with the results

obtained from the Corpus of Spoken Lithuanian¹, the Corpus of the Contemporary Lithuanian Language², and the British National Corpus.

The present study shows that a parallel corpus is especially useful when searching for equivalents in the target language since it clearly displays the differences in the inventory of a particular category in two languages. The establishment of equivalent categories in English and Lithuanian is especially important in terms of lexicography and translation. Equivalents for most lexical items in Lithuanian bilingual dictionaries are often presented on the basis of lexicographers' intuition; synonymous equivalents are often provided as interchangeable items without any usage specifications. However, as corpus data show, there exist important usage differences between apparently synonymous equivalents presented as interchangeable items. Hence, a parallel corpus provides a more extensive inventory of cross-linguistic correspondences than a bilingual dictionary. Therefore, it can be argued that corpus-based contrastive analysis may at least partly solve the problem of equivalence, which is often encountered when using dictionaries, especially bilingual ones. As Granger (2003: 19) so aptly points out, despite some drawbacks of parallel corpora, they are "an ideal resource for establishing equivalence between languages since they convey the same semantic content".

Thus the main aims of the study are as follows:

- to analyse the Lithuanian equivalents of English paucal quantifiers in a parallel corpus;
- to show how English and Lithuanian differ in the use of quantifiers;
- to suggest how a cross-linguistic analysis based on a parallel corpus could be used for practical applications in translation, lexicography and language teaching;
- to show how a parallel corpus can be supplemented with data from comparable monolingual corpora.

Thus the present paper focuses primarily on the issue of equivalence from the perspective of corpus linguistics.

2. Theoretical preliminaries

Paucal quantifiers in the present investigation are defined as non-numerical quantifying expressions that refer to small quantities. The quantifiers under investigation are: *a few*, *few*, *a bit*, *a little bit*, *little*, *a little*, and *several*. Paucal quantifiers are contrasted to multal quantifiers, which refer to large quantities, e.g. *a lot*, *many*, and *much*. The distinction of paucal and multal quantifiers is based primarily on the classification of Quirk et al. (1985), according to which such quantifiers as *many* and *much* are assertive multal pronouns, whereas *a few* and *a little* are assertive paucal pronouns.

The distinction between multal and paucal quantifiers is of special importance since these quantifiers differ in their communicative function. Some results obtained in a study of quantifiers in spoken academic discourse (Ruzaitė 2007) suggest that small quantities have a special importance in spoken interaction taking place in academic settings; it has been observed that paucal quantifiers are often used

¹ See <http://www.vdu.lt/LTcourses/> (see MOKSLAS 'education') (23.01.2009).

² See <http://donelaitis.vdu.lt> (23.01.2009).

for mitigation and help to avoid categorical statements. Since paucal quantifiers are often used to mitigate statements, they are frequently classified as hedges or downtoners. For instance, Dubois (1987: 531) calls *a little* and *a bit* as hedges since they function as “a means of diminishing precision”. Some quantifiers, e.g. *a bit*, *a little*, *little*, *few*, are sometimes referred to as downtoners, which are understood as a category that reduces the scalar intensity of verbs and adjectives (e.g. Hinkel 2003).

Quantifiers perform the functions of mitigating or downtoning since, as Powell (1985) observes, the meaning of quantifiers frequently encodes some evaluative content concerning the significance of a quantity. The evaluative function is an important and intended speaker’s message, which is lost if reformulated into a precise expression. In addition, as Powell points out, not all expressions have an evaluative dimension, for example, *a few*, *many*, *a large number of*, and *a great many* are evaluative, whereas *some* is neutral.

Some studies of quantifiers do not distinguish between multal and paucal quantifiers, but they still emphasise that speakers consciously employ quantifiers instead of numbers for certain communicative purposes (e.g. Channell 1990, Swales 1990, Altenberg 1990, Drave 2002). The preference for non-numerical quantifiers over precise numbers is explained by Moxey, Sanford (1993, 1997), which are based on a psycholinguistic approach and focus on how speakers perceive quantifiers. They observe that “passages seem to be easier to recall when they are quantified with natural language expressions, rather than with numbers” (Moxey, Sanford 1993: 211). According to Moxey and Sanford (1997: 211), “natural language quantifiers can convey far more than is conveyed by mere numerical denotation”.

It is important to note that, as Channell (1994) observes, some quantifiers have a metaphoric meaning. Metaphorical extensions of literal meanings are characteristic of such multal quantifiers as *a load of*, *oodles of*, *a bag of*, *bags of*, *a lot of* and *lots of*, and such paucal quantifiers as *a bit* and *a little bit* (see also Biber et al. 1999). When used metaphorically, these quantifiers specify not a true measure (e.g. a bit or a bag), but a large or small quantity. Such expressions are of special importance in translation and language learning since they may pose problems when finding an appropriate equivalent in different languages.

Though quantifiers, as has already been pointed out, are especially important for their communicative functions, which may differ across language communities, cross-linguistic studies of quantifiers are not numerous. However, some of the most recent research shows that a proper command of quantifying expressions is to be taken into account in language teaching. For example, Labrador de la Cruz’s (2003) cross-linguistic investigation of the most prototypical quantifiers in Spanish and English suggests that the functions of quantifiers in the two languages are very similar, but they differ significantly in their speech realisations.

Hinkel (2003), who studies the use of downtoners in the essays of native and non-native speakers, demonstrates that downtoners are used infrequently by both native and non-native speakers. In contrast, emphatics (e.g. *v + a lot*) and amplifiers (e.g. *very much*, *a lot + comparative adj*, *much + comparative adj*), are considerably more frequent in the essays of non-native speakers than in those of native speakers. The excessive use of emphatics and a significantly less frequent use of downtoners make the essays of non-native speakers sound colloquial, too assertive and categorical (Hinkel 2003).

On the basis of previous research of English quantifiers (research of Lithuanian quantifiers is limited to just a few grammar books) it can be hypothesised that English and Lithuanian differ mainly in the inventory of quantifying expressions, since morphologically the two languages are very different. Furthermore, it can be hypothesised that there exist some differences in the metaphoric expressions used for quantification in English and Lithuanian. It can also be assumed that quantifying expressions, being a category of hedging devices, are used to a different extent in the two languages. It is expected that the results of the present analysis will show how corpus data can supplement the information available in bilingual dictionaries and how corpora can be used as an aid in translation and language teaching.

3. Methods and data

In this study quantifiers will be analyzed within the framework of corpus linguistics and contrastive linguistics. Corpora and especially parallel corpora can be said to have instigated the revival of contrastive linguistics in recent years since corpus linguistics offers a reliable methodology of collecting and systematizing data in more than one language (for some examples of such analysis see Hallebeek, Spaans 2000, Granger 2003, Frankenberg-Garcia 2006, McEnery, Xiao 2007). As Frankenberg-Garcia (2006: 142) rightly observes, “using the technology of corpus linguistics ... it is possible to analyse enormous quantities of translated text in unprecedented ways”. Corpus-based contrastive linguistics offers a chance to study both language-specific and cultural differences and similarities, as well as universal features. The results obtained in such a cross-linguistic analysis can be used for numerous practical applications in such areas as translation, lexicography and language teaching.

Computer corpora can serve as a reliable source of empirical data in a contrastive study of two or more languages, for several reasons. A corpus-based approach enables a linguist to make well-substantiated generalizations on the basis of naturally occurring stretches of language. The main advantages of such a study are the reliability of data (due to its abundance and carefully proportioned design of the corpus) and the naturalness of the language. Corpus data also bring to attention such cases that may otherwise go unobserved and can even be unsuspected in the language intuitively. A combination of different types of corpora, e.g. monolingual and multilingual corpora, can further increase the validity of the results. Since both monolingual and multilingual corpora have advantages and disadvantages, in cross-linguistic studies both types of corpora should be used concurrently to enhance the accuracy of the results (Granger 2003). The main advantages and disadvantages of the two types of corpora are well summed up by Granger (2003: 19):

Comparable corpora have the major advantage of representing original texts in the two or more languages under comparison, i.e. language spontaneously produced by native speakers of those languages. They are therefore in principle free from the influence of other languages, which is obviously not the case of translation corpora as the original source text is in a different language and will quite naturally exert some kind of influence on the target text. The main drawback of comparable corpora lies in the difficulty of establishing comparability of texts. Some types of text are culture-specific and simply have no

exact equivalent in other languages. Translation corpora are an ideal resource for establishing equivalence between languages since they convey the same semantic content. The main drawback of translation corpora, however, is that they often display traces of the source text and therefore cannot really be considered as reliable data as regards the target language, especially in frequency terms. In addition, it is not always possible to find translations of all texts.

These disadvantages can be compensated for by the combination of a parallel corpus with comparable monolingual corpora, as each can be used to supplement the other's data and double-check the findings obtained in any one of them.

Thus, to make the present cross-linguistic analysis more accurate, I will make use of several different corpora, both bilingual and monolingual. The primary source for collecting data will be The Parallel Corpus of the Lithuanian Language (PCLL).³ It contains almost 70 000 parallel sentences translated from English into Lithuanian and 1614 sentences translated from Lithuanian into English.

In addition to the parallel corpus, the data will be supplemented with the findings obtained from three monolingual corpora, one of which is the British National Corpus (BNC).⁴ The latter will be used to study quantifiers in English. The BNC contains 100 million words and is composed of samples of written and spoken language from a variety of spoken and written sources.

Lithuanian quantifiers will be further studied in The Corpus of the Contemporary Lithuanian Language (CCLL), which consists of 100 million words, thus being of a similar size as the BNC. The CCLL represents a wide range of contemporary written Lithuanian and contains mainly printed material from Lithuania's independence period (since 1990). The largest part of the corpus consists of general press (texts from regional and national newspapers), popular press, and special press (specialised newspapers and magazines). The rest of the corpus comprises fiction, memoirs, other literature (scientific and popular), and various official texts. Spoken language forms just a small part of this corpus; it consists of Parliament debates that constitute 2% of the corpus (or approximately 20 000 words). To represent spoken Lithuanian more fully, the Corpus of Spoken Lithuanian will be used to supplement the data obtained from the Corpus of the Contemporary Lithuanian Language. The Corpus of Spoken Lithuanian consists of 50 000 words.

4. Results

4.1. Frequency of quantifiers in the parallel corpus

The English-Lithuanian part of the parallel corpus has provided 943 occurrences of the seven quantifying expressions under investigation. The raw frequency of each quantifier is presented in Table 1.

Table 1 shows that *a few* is the most frequent quantifier (309 occurrences), as opposed to the least frequent *a little bit*, which occurs only 6 times. The main types of equivalents of these quantifiers will be discussed in greater detail in the following section, which will compare the equivalents available in a bilingual dictionary and those that can be obtained from a parallel corpus.

³ See <http://donelaitis.vdu.lt> (23.01.2009).

⁴ Available via the University of Zurich (English Seminar Corpus Server), see <http://escorp.unizh.ch/> (23.01.2009).

Table 1. Frequency of quantifiers in the parallel corpus

Quantifier	Frequency
<i>a few</i>	309
<i>a little</i>	172
<i>several</i>	167
<i>little</i>	142
<i>few</i>	113
<i>a bit</i>	34
<i>a little bit</i>	6
Total	943

4.2. Main types of equivalents of paucal quantifiers in Lithuanian

One of the major observations that can be made on the basis of the data obtained from the parallel corpus is that the parallel corpus provides a considerably larger number of equivalents than a bilingual dictionary. The equivalents of *a bit* presented in Table 2 clearly demonstrate that there is a greater variety of equivalents in the parallel corpus (PC) than in the bilingual dictionary (BD).

Table 2. Equivalents of *a bit* (frequency in brackets)

Bilingual dictionary (BD)	Parallel corpus (PC)
<u><i>truputi</i></u> <i>nedaug</i> <i>gana</i>	<u><i>truputi</i></u> (8) <i>šiek tiek</i> <i>kiek</i> <i>lyg</i>
	grammatical equivalent (affix) (4)
	emphatic equivalent (e.g. <i>kur kas, kaip</i>) (4)
'not a bit': <i>visai ne, nė kiek</i>	'not a bit': <i>nieko panašaus</i>
	no equivalent (8)

As can be seen in Table 2, almost no equivalents coincide in the PC and BD; the only equivalent that coincides (*truputi*) is underlined in the table above. In addition, there are cases (8 occurrences) where there is no equivalent in the translated texts.

The comparison of the information available in the parallel corpus and the bilingual dictionary (Piesarskas 2005) has also demonstrated that some quantifying expressions (e.g. *a little bit*) are not represented in any way in the BD. *A little bit* is not defined either in a separate entry or in the entry of *bit* in the BD. In the parallel corpus, in contrast, the latter quantifier is left untranslated only once, but in other cases it is rendered by using a variety of equivalents such as *šiek tiek*, *truputis*, *ne taip*, and the emphatic equivalent *labai*.

As the corpus data have revealed, in Lithuanian quantifiers are often realized only grammatically or both grammatically and lexically. There are numerous cases in the PC where equivalents of quantifiers are not lexicalized but are represented only grammatically. In such instances affixes (esp. suffixes) are added to nouns (*gurkšnelis*), verbs (*prunkštelėjo*, *šnekteldavo*) and adjectives (*keistoku*) that are modified by quantifiers in English, as can be seen in examples (1)–(4):

- (1) The savages sniggered **a bit** and one gestured at Ralph with his spear.
Laukiniai prunkštel **jo**, ir vienas parodė į Ralfą ietimi.
- (2) Perhaps **a little** water would restore her.
Gal gurkšnelis vandens ją atgaivins?
- (3) Langdon's friends had always viewed him as **a bit** of an enigma – a man caught between centuries.
Lengdono draugai jį visuomet laikė **truput** keistoku – jiems jis buvo žmogus, pasimetęs tarp kelių šimtmečių.
- (4) He usually stopped to talk with Mr. Charrington for **a few** minutes on his way upstairs.
Prieš lipdamas viršun, Vinstonas sustojęs visada šnekteldavo su juo **kelet** minučių.

In examples (3) and (4) English quantifiers are rendered into Lithuanian by providing a lexical equivalent (*truputį* in (3) and *keletą* in (4)) and by adding suffixes to the quantified notions. Those suffixes reinforce the small quantity expressed by the quantifier since they convey the idea of smallness: a limited amount of a quality (in the adjective *keistokas*) and a short duration of an action (in the verb *šnekteldavo*).

Though paucal quantifiers typically function as mitigators, they are sometimes translated into Lithuanian by providing emphatic equivalents (EN 'paucity' LT 'multitude'). It is difficult to find an explanation for such a tendency without knowing the real translators' intentions but such results may suggest some cross-cultural/cross-linguistic differences. As far as cross-linguistic differences are concerned, some words typically are not mitigated in LT as their most typical collocational patterns suggest; see examples (5)–(7):

- (5) The Vatican, it seemed, took their archives **a bit** more seriously than most.
Vatikanas, atrodo, į savo archyvų apsaugą žiūrėjo **kur kas** rimčiau nei dauguma kitų įstaigų.
- (6) **a few** emotional storms
daugel emocinių audrų
- (7) There are times when, for your sake, I have been a **little** uneasy at his marked preference, and have wished to put you on your guard...
Aš **labai** rūpinaus, matydama, kad jis kreipia į jus ypatingą dėmesį, ir, mylėdama jus, rengiausi su jumis pasikalbėti.

The Lithuanian equivalents in (5)–(7) *kur kas*, *daugelį* and *labai* express multitude or intensity but not paucity; therefore, they function as intensifiers.

A set of equivalents that can be of special interest in cross-cultural investigations are metaphoric equivalents. As the data have revealed, there exist cases when a non-metaphoric English quantifier is translated by using a metaphoric expression in Lithuanian, as in example (8):

- (8) ...but restricted his generosity to **those few** supporting the PRI (Partido Revolucionario Institucional), the government party.
 ...bet buvo dosnus tik **tai saujelei** žmonių, kurie rėmė vyriausybę partiją
 PRI (Institucinė revoliucinė partija).

The English quantifier *few* is translated by using a metaphoric equivalent *saujelė*, which literally means ‘a handful’, but metaphorically refers to a small quantity. Such metaphoric quantifiers exist in English as well (e.g. *a bit*, *a load of*, *oodles of*, *a bag of*, *bags of*, *a lot of* and *lots of*; for a discussion of such quantifiers, see also Channell (1994), Biber et al. (1999)).

One of the most unexpected findings was that in some instances the target text (TT) provides numerical equivalents for non-numerical English quantifiers. A more specific numerical reference is used instead of a non-numerical quantifier in examples (9)–(10):

- (9) A few
 du trys (‘two three’)
- (10) his spectacles needed wiping every few minutes
 akinius reikėjo šluostyti kas penkios minutės

Such a use of equivalents is revealing since it shows how the speakers of a language perceive and interpret quantifiers. A non-specific quantifier is a fuzzy concept whose interpretation may vary in different contexts and language communities. To study how quantifiers are perceived by English language speakers, Channell (1994) devised a questionnaire that respondents had to fill out. As the present research shows, a large database of translated texts could provide a reliable source for analysing the perception and interpretation of quantifiers, which could supplement, or be even more informative than, a questionnaire.

Interestingly, in a number of cases quantifiers are not translated into Lithuanian. This is especially typical of *a little*; in 48 cases (or 28% of the total number) *a little* is not translated into Lithuanian, as in example (11):

- (11) “I don’t mean to be baffled by **a little stiffness** on your part; I’m prepared to go to considerable lengths.”
 “Aš nesutrikšiu nuo šito jūsų **manieringumo**; taip lengvai neišsisuk-site.”

As example (8) demonstrates, the quantified noun (underlined in the examples) is not preceded by a quantifier in the Lithuanian version. All the instances where English quantifiers are not represented in the target text (TT) are presented in Table 3.

Table 3. Quantifiers omitted in the target text

	<i>a little</i>	<i>a bit</i>	<i>few</i>	<i>a few</i>
Frequency	48	8	13	33
Percentage	28.1%	23.5%	11.5%	10.7%

As Table 3 demonstrates, the quantifiers that have no equivalent in the TT most frequently include two synonymous quantifiers *a little* and *a bit* (they are not translated into Lithuanian in 28.1% and 23.5% of the cases, respectively). *Few* and *a few* are omitted to a considerably lesser extent (11.5% and 10.7% of the cases, respectively).

There are several possible explanations for such a tendency. First, it is easily noticeable (see Table 3) that the quantifiers that are the most commonly omitted in the TT are those that are used primarily not for their informative content, but that are primarily important as hedges or mitigators. Therefore, it may be argued that quantifiers functioning as hedges are omitted since their communicative function is underestimated by the translator. This suggests that translators still give too much prominence to the informative content and thus may neglect the lexical items mainly necessary for other purposes such as persuasion.

A second explanation is based on the structural differences between the two languages under investigation. Lithuanian, being a synthetic language, can express quantification by the genitive case (*partitive use*) and thus a quantifier is of less importance in Lithuanian than in English. In English a paucal quantifier is absolutely necessary if the speaker needs to refer to just a part of something, whereas in Lithuanian the partitive use of the genitive case immediately suggests a part of the quantified notion.

Finally, since the translation equivalent is a unit typically larger than one word, the notion of quantification is often encoded in a longer phrase without a quantifier in the TT. This is especially common when set phrases, fixed expressions and idioms are translated, as in the following instances:

- (12) In the past/last **few** years – pastaruuju metu/pastaraisiais metais/per pastaruosius metus
- (13) In the next **few** days – artimiausiu metu
- (14) **a few** moments ago – neseniai
- (15) for **a few** moments/minutes – valand **l**
- (16) To name but **a few** – sunku viską ir išvardinti
- (17) you're **a few** crumpets short of a proper tea – iš tavęs jam menka nauda
- (18) A person of **few** words – nešnekus

The time references in (12)–(14) above contain no quantifiers in Lithuanian but they still refer to the shortness of certain time periods just as their English counterparts. In (15), the diminutive form is used to refer to a short period, not a quantifier. Examples (16)–(18) show how quantifiers are omitted in fixed expressions and idioms. Thus, the omission of quantifiers in the TT can sometimes, albeit not always, be explained by the dependence of the equivalent on a collocational pattern.

5. Concluding remarks

The present analysis is just a brief investigation of a limited set of quantifying lexemes in two languages and could profitably be expanded in many different respects. For instance, a closer study of monolingual corpora in the two languages could reveal a more detailed picture including more information on collocational patterns and usage differences between different modes of speaking or discourse types. However, even an investigation of such a limited scope enables some important generalisations with regard to some possible applications of such corpus results.

First of all, a contrastive study of English and Lithuanian can have some applications in lexicography, especially with regard to both monolingual Lithuanian dictionaries and bilingual English-Lithuanian / Lithuanian-English dictionaries. The use of corpora could contribute to the development of a new generation of grammars and dictionaries in Lithuanian, which so far have been highly conservative. The data have clearly demonstrated that dictionary equivalents should be necessarily supplemented with corpus data since very few of them coincide in both sources.

Data obtained from parallel and comparable monolingual corpora are also important in translation studies. A corpus-based contrastive analysis may at least partly solve the problem of equivalence, which is often encountered when using dictionaries, especially bilingual ones. As Granger (2003: 25) rightly points out, bilingual corpora are “an extremely valuable pedagogical resource in translation teaching”. She suggests that corpus-based classroom activities for translator trainees could “involve comparable and parallel corpora of general or specialised language” (Granger 2003: 25).

Finally, corpus results should be taken into consideration when teaching English or Lithuanian as a foreign language. The appropriate usage of such expressions as quantifiers should be treated as a special strategy of successful communication, the use of which contributes to the linguistic fluency and thus should be acquired by foreign speakers in order to sound natural and polite. Interesting cross-cultural and cross-linguistic differences (e.g. metaphoric uses) that have been highlighted by a parallel corpus can be useful in language teaching and translation. The results of the present analysis show that, although some important differences with regard to mitigation can be observed in English and Lithuanian, in general these two languages use quantifiers for similar purposes and in similar patterns. It should be emphasised that when teaching, both differences and similarities between the native and the target language should be taken into account. Making learners aware of cross-linguistic similarities could enhance the learning process: if students can draw parallels with their own native language(s), the feeling of familiarity can help them adopt certain language properties more easily.

References

- Altenberg, Brigitte 1990. Some functions of the booster. – Jan Svartvik (Ed.). *The London-Lund Corpus of Spoken English: Description and Research*. Lund: Lund University Press, 193–210.
- Biber, Douglas; Johansson, Stig; Leech, Geoffrey; Conrad, Susan; Finegan, Edward 1999. *Longman Grammar of Spoken and Written English*. London, New York: Longman.
- Channell, Joanna 1994. *Vague Language*. Oxford: Oxford University Press.

- Drave, Neil 2002. Vaguely speaking: A corpus approach to vague language. – Peter Collins, Pam Peters, Adam Smith (Eds.). *New Frontiers of Corpus Research*. Amsterdam: Rodopi, 25–40.
- Dubois, Betty Lou 1987. “Something on the order of around forty to forty-four”: Imprecise numerical expressions in biomedical slide talks. – *Language in Society*, 16, 527–541. doi:10.1017/S0047404500000361
- Frankenberg-Garcia, Ana 2006. Using a parallel corpus in translation practice and research. – *Actas da Contrapor 2006, 1ª Conferência de Tradução Portuguesa*, Faculdade de Ciências e Tecnologia. Universidade Nova de Lisboa, 11-12 de Setembro de 2006. Lisboa, 143–148. www.linguateca.pt/documentos/Frankenberg-GarciaContrapor2006.pdf (10.05.2008).
- Granger, Sylviane 2003. The corpus approach: A common way forward for Contrastive Linguistics and Translation Studies. – Granger Sylviane, Jacques Lerot, Stephanie Petch-Tyson (Eds.). *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam, Atlanta: Rodopi, 17–29.
- Hallebeek, Jos; Vaggroep Spaans 2000. English parallel corpora and applications. – *Cuadernos de Filología Inglesa*, 9 (1), 111–123.
- Hinkel, Eli 2003. Adverbial markers and tone in L1 and L2 students’ writing. – *Journal of Pragmatics*, 35, 1049–1068. doi:10.1016/S0378-2166(02)00133-9
- Labrador de la Cruz, Belén 2003. Typicality in the use of English and Spanish indefinite quantifiers. – *Babel*, 49/1, 41–58.
- McEnery, Anthony; Xiao, Zhonghua R. 2007. Parallel and comparable corpora: What is happening? – Margaret Rogers, Gunilla Anderman (Eds.). *Incorporating Corpora: The Linguist and the Translator*. Clevedon: Multilingual Matters, 18–31.
- Moxey, M. Linda; Sanford, Anthony J. 1993. *Communicating Quantities: A Psychological Perspective*. Hove: Lawrence Erlbaum Associates.
- Piesarskas, Bronislovas 2005. *The English-Lithuanian Dictionary*. Vilnius: Alma littera.
- Powell, Mava Jo 1985. Purposive vagueness: An evaluative dimension of vague quantifying expressions. – *Journal of Linguistics*, 21, 31–50. doi:10.1017/S002222670001001X
- Quirk, Randolph; Greenbaum, Sidney; Leech, Geoffrey; Svartvik, Jan 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Ruzaitė, Jūratė 2007. *Vague Language in Educational Settings: Quantifiers and Approximators in British and American English*. *Europäische Hochschulschriften. Angelsächsische Sprache und Literatur* 433. Frankfurt am Main: Peter Lang.
- Swales, John M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Jūratė Ruzaitė (Department of English Philology, Vytautas Magnus University, Lithuania). Her research interests include discourse analysis (esp. spoken academic discourse), discourse and ideology, corpus linguistics, and vague language.
j.ruzaitė@hmf.vdu.lt

KORPUSTE RAKENDUSI: HULGASÕNADEST INGLISE JA LEEDU KEELES

Jüraté Ruzaitė

Vytautas Magnuse Ülikool

Artikli eesmärk on juhtida tähelepanu paralleelkorpustel põhineva kõrvutava keeleanalüüsi rakendusvõimalustele tõlkimisel, sõnaraamatutöös ja keeleõpetuses. On uuritud ingliskeelseid hulgasõnu (nt *a bit*, *a few*, *few* jt) ja nende leedu vasteid, kasutades leedu keele paralleelkorpuse ja suulise leedu keele korpuse ning vajadusel ka briti inglise keele korpuse British National Corpus andmeid.

Selgub, et paralleelkorpus pakub tunduvalt rohkem vasteid kui kakskeelne sõnaraamat. Leedu keeles antakse kvantori tähendus sageli edasi kas ainult grammatiliselt või siis grammatiliste ja leksikaalsete vahendite koostoimel. Leksikaliseerimata, s.o puhtgrammatilisi kvantorivastendusi, kus inglise nimi-, tegu- või omadussõna on piiritletud hulgasõnaga, vastavale leedu sõnale aga lisandub liide (sagedamini järelliide), leidub paralleelkorpuses hulganisti.

Kuigi vähest hulka, kogust või määra tähistavatel kvantoritel on tavaliselt vähendav funktsioon, tõlgitakse neid vahel hoopis tunderõhulise leedu vastega. Tõlkija kavatsusi tundmata on nähtust küll raske seletada, kuid tegemist võib olla ka keelelis-kultuuriliste erinevustega. Nagu nähtub tüüpilistest kollokatsioonidest, mõne leedu sõna puhul vähendusvormi lihtsalt ei kasutata.

Sageli jäetakse inglise hulgasõnad üldse leedu keelde tõlkimata. Võimalikke seletusi on rohkem kui üks. Esiteks torkab silma, et kõige sagedamini loobutakse tõlkimast kvantoreid, mis toimivad pigem leevendaja või vähendajana kui kannavad muud informatsiooni. Niisiis võib väita, et tõlkijad alahindavad säärase kvantore kommunikatiivset rolli. Paistab, et tõlkijad kipuvad ikka veel üle tähtsustama teksti informatiivsust selle muude võimalike aspektide (nt sisenduslikkuse) ees. Teiseks võib nähtust seletada ka vaatlusaluste keelte struktuurilise erinevusega. Sünteesiva keelena võib leedu keel kvantori asemel (partitiivselt) genitiivi kasutada, mistõttu hulgasõnadel pole sama tähtsust mis inglise keeles. Kolmandaks, kvantifitseerida saab ka mitmesõnalise fraasiga (sellised ongi analüüsitud vastete hulgas enamuses), kus hulgasõna kui selline puudub. Selliselt tõlgitakse tavaliselt käibefraase, püsiväljendeid ja idioome.

Võtmesõnad: paralleelkorpus, hulgasõnad, tõlkimine, vasted, keeleõpe, inglise keel, leedu keel