

Translation Equivalents of Vague Language Items: A Study of General Extenders in a Parallel Corpus

Jūratė Ruzaitė

Abstract. The present paper aims to show how a contrastive analysis of English and Lithuanian based on a parallel corpus can be used in studying translation equivalents. The paper examines the category of general extenders, which are often treated as a vague language category. General extenders include such lexical items as *etc./ et cetera*, and so on, *or/ and something like that*. The analysis is based primarily on the Parallel Corpus of the Lithuanian Language. Where relevant, the issue of equivalence between English and Lithuanian is discussed by comparing the data to the results obtained from monolingual English and Lithuanian corpora. 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. The equivalents of vague language items are also interesting from the perspective of cross-cultural differences; the data have revealed the tendency to omit vague language categories in the target language, which might point to the possibility of some cross-cultural differences or the general tendency to underestimate vague language categories.

Key words: *vague language, general extenders, parallel corpus, translation equivalents, dictionaries.*

Introduction

The present paper examines general extenders, which are often treated as a *vague language* (VL) category. General extenders include such lexical items as *etc./ et cetera*, and so on, *or/ and something like that*. These items are primarily used for pragmatic purposes (such as information management and text structuring) and have very low informative content. Therefore, an analysis of ways of translating such items can provide some interesting implications as to how the communicative impact and informative content of the *source text* (ST) changes depending on the way a general extender is translated in the *target text* (TT).

There exist two predominant approaches to vagueness. First, vagueness is perceived as a pervasive language phenomenon; vagueness viewed from this approach is called “linguistic vagueness”. This approach is based on the philosophic tradition which has studied vagueness from the perspective of the truth-value logic since Aristotelian times. The truth-based approach to vagueness is based on a rigid approach to meaning (e.g. Williamson, 1994); therefore, linguistic vagueness is often studied in terms of indeterminacy of the referential boundary of words. The phenomenon of vagueness addressed by different philosophers is based on the ancient sorites paradoxes, e.g. *How many grains make a heap?* or *When is a man bald?* The questions raised in sorites paradoxes aim to determine strict boundaries between words, e.g. baldness and non-baldness.

The second approach to vagueness is called “pragmalinguistic vagueness” and focuses on some specific categories of vague language used as an important communicative strategy. This approach is adopted in the present study as well. In the present paper, vague language is understood as a set of specific categories that are “themselves vague and/

or which add vagueness to utterances” (Wang and Piao, 2007). According to Channell (1994), the main categories that can be ascribed to pragmalinguistic vagueness are as follows:

1. quantifiers (*many, a lot, a bit*);
2. approximators (*approximately, about, around*);
3. placeholders (*thingummy, thingummajig*);
4. vague references to categories (called *general extenders* in Overstreet (1995, 1999)) (*and something like that, etc., and so on*).

The categories listed above include adverbs, certain types of determiners or multi-word phrases that have low semantic load and are used primarily for pragmatic purposes.

The present analysis of general extenders is based on the Parallel Corpus of the Lithuanian Language (<http://donelaitis.vdu.lt>). Where relevant, the issue of equivalence between English and Lithuanian is discussed by comparing the data to the results obtained from monolingual Lithuanian corpus (Corpus of the Contemporary Lithuanian Language, <http://donelaitis.vdu.lt>). On the basis of corpus data, the paper aims to answer the following research questions:

- What are the Lithuanian equivalents of English general extenders in the parallel corpus?
- How do corpus equivalents differ from those in the bilingual dictionary?
- What do the data from comparable monolingual corpora show about the use of general extenders in English and Lithuanian?
- How can a cross-linguistic analysis based on a parallel corpus be used for practical applications in translation, lexicography and language teaching?

The hypothesis is that equivalents of general extenders differ in the parallel corpus and the bilingual dictionary, and that vague language items are often misrepresented in translated texts (possibly due to the translator's attitude to vague language items and/ or cross-cultural or cross-linguistic differences).

Theoretical Background: Definition, Functions and Usage of General Extenders

Previous studies of general extenders are not numerous though linguists' interest in this VL category has considerably increased (e.g. Channell, 1985, 1990, 1994; Crystal and Davy, 1979; Dines, 1980; Aijmer, 1985; Dubois, 1993; Ward and Birne, 1993; Ediger, 1995; Winter and Norrby, 1999; Cheng and Warren, 2001; Overstreet, 1995, 1999; 2005; Overstreet and Yule, 1997, 2001, 2002). In the existing literature, there is no unified term used to refer to this VL category. General extenders have been studied under such labels as "set marking tags" in Dines (1980) and Winter and Norrby (1999), "utterance final tags" in Aijmer (1985), "discourse particle extensions" in Dubois (1992), "vague references to categories" in Channell (1994), "summarizing lexical items" in Crystal and Davy (1979), and "general extenders" in Overstreet (1995, 1999). The terms employed by different linguists highlight different functions of general extenders. For example, Channell (1994) uses the term vague references to categories, which puts a special emphasis on the categorisation function of general extenders. Overstreet (1999), meanwhile, uses the term "general extenders" because they are nonspecific and function to extend otherwise grammatically complete utterances.

In the present analysis the term "general extenders" is used since it is least restricted in terms of functions; general extenders are perceived as tags that start with the conjunction *or* or *and*, and follow an item that can be a noun phrase, verb phrase, embedded sentence or prepositional phrase (Channell, 1994). The conjunction is usually followed by the pronouns *something* or *anything*, or the noun *things*. In this study, the general extenders under investigation include:

- 1) *et cetera*,
- 2) *etc.*,
- 3) *and so on*.

These general extenders have been selected due to two main reasons:

- (1) they are among the most common general extenders in English, and
- (2) they are typical of written discourse (this is an important factor in this paper since it focuses on translated written texts).

General extenders, just like other vague language categories, are especially context-dependent. The interpretation of these categories depends largely on the hearer's framework of knowledge. For instance, if some item is followed by a general extender (*she studies marine **biology or something like that***), the hearer interprets the general extender by relying on his/ her knowledge of the category that the "extended" item belongs to. It is also important to note that

general extenders are often interpreted by taking into account their function in the communicative act.

General extenders, being a VL category, perform some of the functions that are characteristic of VL in general. The main functions of VL can be referred to as (a) epistemic (propositional) and (b) affective meta-functions (Drave, 2002, p.27). In other words, VL categories are used to modify the propositional content of an utterance, to indicate the precision, reliability or categoricity of the utterance. When a VL category performs the affective metafunction, it contributes to the achievement of certain interpersonal goals. As Drave rightly sums up,

"the major function of VL is to tailor conversational contributions to the perceived informational needs of the other participant(s) so as to maintain and enhance the ongoing relationship" (2002, pp.26-27).

Some specific functions that VL can perform include:

- 1) Filling lexical gaps (where a speaker cannot recall a word or where one does not exist in the language);
- 2) Filling knowledge gaps (memory lapse);
- 3) Emphasising (and deemphasising) certain information;
- 4) Deliberately withholding specific information;
- 5) Conveying tentativeness;
- 6) Conveying an evaluation of, or expectation about, a proposition;
- 7) Maintaining an atmosphere of friendliness, informality or deference (Drave, 2002, pp.26-27).

To generalize, VL can be used by the speaker to manipulate the informative content and to manage interpersonal relationships between the speaker and hearer. Since VL is closely related to the manipulation of the degree of precision and can be used as a face-saving strategy, it can be treated as a sub-category of hedging (cf. Hyland, 1998). It is thus important to note again that VL, including general extenders, is used primarily for pragmatic purposes, but it carries very little semantic load.

As Winter and Norrby observe, some early investigations of general extenders often

"refer to the optionality of the forms and suggest that [discourse] coherence would be maintained if the forms were omitted" (1999, p.2).

However, the omission of general extenders may have some important negative effects on, for example, the informative content of discourse as these units are a crucial cue for the listener when interpreting the preceding item as an illustrative example of a more general category. Therefore, it may be expected in the present study that omissions of general extenders in translated texts may result in mistranslations at least to some extent.

It has been observed in some studies that general extenders, being pragmatically a highly important category, vary across cultures, registers, age groups and other socially pre-determined units. However, not many studies have been carried out in this area; Winter and Norrby (1999), Cheng and Warren (2001), Drave (2002), Overstreet (2005) are among the few studies that focus on the issue. For instance, Winter and Norrby (1999) examine the discourse meanings

of general extenders, or set marking tags (SMTs) as they call them, in the talk of Australian English (AE) and Swedish speaking adolescents. They notice that general extenders are a salient feature of adolescents, but differences between the two languages, Australian English and Swedish, are minor. Overstreet (2005) presents a comparative analysis of a group of general in recorded English and German conversation among adult familiars. Her study shows that general extenders are used in both languages for similar purposes.

Thus cross-cultural differences in the use of general extenders have already been studied at least to some extent; however, studies of translation of VL categories are extremely limited. The only available studies on translation of vague language items focus on ways of rendering quantifiers from English into Chinese (Wang and Piao, 2007) and from English into Lithuanian (Ruzaitė, 2009). All comparative studies of general extenders are based on the data obtained from monolingual comparable corpora and from databases of transcripts of intercultural communication.

Data (1)

The corpus that was used to obtain results for the present investigation is the Parallel Corpus of the Lithuanian Language (henceforth, PCLL; available at <http://donelaitis.vdu.lt>). It contains 70,000 parallel sentences translated from English into Lithuanian and 1,614 sentences translated from Lithuanian into English. For the purposes of the present study, only the English-Lithuanian part of the corpus was used. Where relevant, the data from the PCLL are compared to the data from two monolingual corpora, namely, the British National Corpus (BNC; <http://corpus.byu.edu/bnc>) and the Corpus of Contemporary Lithuanian Language (henceforth, CCLL; available at <http://donelaitis.vdu.lt>). The size of the two monolingual corpora is around 100 million words, which makes them comparable in terms of size.

The general extenders under investigation differ considerably in their frequency in the PCLL. As can be seen in Table 1, the frequency of the general extender *and so on* is 28 occurrences. The abbreviated form *etc.* is the most frequent one; it occurs 164 times in the corpus. The full form *etcetera*, meanwhile, is extremely rare and occurs only once in the corpus. Due to its low frequency, the general extender *etcetera* will not be discussed further.

Table 1. Items investigated in the PCLL

Items under investigation	Number of examples
<i>etc.</i>	164
<i>et cetera</i>	1
<i>and so on</i>	28
TOTAL:	193

As Table 1 demonstrates, in total there are almost 200 instances of the general extenders under analysis; these will be further discussed in the next section by focusing on the translation equivalents used in the TT. The equivalents in the corpus will be compared to those that are provided in the bilingual dictionary (Piesarskas, 2005).

Results

The analysis of the information in the bilingual dictionary (BD) and the PCLL has revealed several major tendencies

that can be important in such areas as translation and lexicography. The tendencies observed in the data are as follows:

- 1) greater variety of equivalents in the PCLL than in the BD;
- 2) differences in the equivalents provided in the PCLL and BD;
- 3) different frequency of (a) seemingly interchangeable *general extenders* and (b) seemingly interchangeable *equivalents* of general extenders;
- 4) numerous cases of omission in the translated texts.

These tendencies will be discussed in greater detail below by referring to corpus examples illustrating different translation patterns observed in the corpus.

Discussion of the Equivalents in the Parallel Corpus and Bilingual Dictionary

The corpus data provide some important evidence that similar or apparently interchangeable forms are of considerably different frequency. First of all, even a small corpus can indicate which synonymous item in a set of related terms (in our case, a pair of general extenders) is more salient in the source language. As Tables 1 and 2 indicate, the apparently synonymous general extenders are of drastically different frequencies, the abbreviated form *etc.* being the most frequent one. The BD, however, provides a longer entry for its full form *et cetera*, which occurs in the corpus only once. The dictionary information suggests that the more basic form is the unabbreviated form since the entry for it is longer and more equivalents are provided (see Table 2). The PCLL data are supported by a much larger monolingual British National Corpus, which comprises 100 million words. As the data in this corpus shows, the frequency of the abbreviated form *etc.* amounts to as many as 3016 occurrences, whereas the full form occurs only in 509 instances.

Table 2. General extenders in the Parallel and Monolingual Corpora and the Bilingual Dictionary

General extender	Freq. in PC (raw)	Freq. in BNC (raw)	Dictionary info
<i>and so on</i>	28	13	2 equivalents (<i>ir panašiai</i> , <i>ir taip toliau</i>); interchangeable with <i>and so forth</i>
<i>et cetera</i>	1	509	longer entry, 2 equivalents (<i>ir taip toliau</i> , <i>ir kita</i>)
<i>etc.</i>	164	3 016	shorter entry, one equivalent (<i>ir taip toliau</i>)

The frequency of the general extender *and so on* is relatively low: 28 occurrences in the PCLL and 13 occurrences in the BNC.

Tables 3 and 4 demonstrate that there is greater variation of the equivalents of general extenders in the PCLL than in the BD. As Table 3 shows, in the PCLL *and so on* is translated in 4 ways: *ir taip toliau* (“and so on”), *ir taip be galo* (“and so endlessly”), and two abbreviated forms *ir t. t.* and *ir pan.* In the BD, meanwhile, there are only two equivalents presented: *ir taip toliau* (“and so on”) and *ir panašiai* (“and the like”). The equivalents coinciding in the

two sources are presented in bold in Table 3; these are *ir taip toliau* and *ir panašiai/ir pan.* In contrast to the corpus data, the information in the BD indicates that the unabbreviated forms are treated as more basic equivalents of *and so on*.

Table 3. Equivalents of *and so on* in the BD and PC

BD	PC	Freq.
ir taip toliau (sense 2)	ir taip toliau	13
	ir t. t.	8
		21
ir panašiai (sense 1)	ir pan.	3
	ir taip be galo	2

The difference in the variation of the equivalents of *etc.* is even greater. The BD provides two equivalents as two different senses of the general extender *etc.*, namely, *ir taip toliau* (“and so on”) and *ir kita* (“and other”). In the translated texts, there are 9 different equivalents of *etc.* The equivalents that coincide in both sources are *ir taip toliau* and *ir kita*; the latter equivalent appears in a different grammatical form in the corpus but has the same meaning “and other”. The abbreviated forms of *ir taip toliau* ((*ir*) *t. t.*) are not presented in the dictionary though they are highly frequent equivalents in translated texts. The different forms of *ir/ ar pan.* (“and/ or the like”) are not provided in the BD though it is the second most frequent equivalent in the PCLL.

Table 4. Equivalents of *etc.* in the BD and PCLL

BD	PCLL	Freq.
ir taip toliau (sense 1)	ir t. t.	91
	ir taip toliau	3
	t. t.	1
		95
–	ir pan.	30
	ar pan.	2
	arba pan.	1
		33
ir kita (sense 2)	ir kt.	5
	kt.	1
	ir kitų	1
		7
	Cases of omission	21

The data of the present investigation also suggests that not only is the variation of possible equivalents greater in the PCLL, but also the salience of these equivalents is different in the BD and PCLL. The frequency information in Tables 3 and 4 shows that the equivalents presented in the BD as the most basic equivalents are of relatively low frequency if compared to other equivalents. Table 3 shows that *ir panašiai* appears as the first sense in the BD. However, its abbreviated form *ir pan.* is of very low frequency in the PCLL (only 3 occurrences). In the PC, the most frequent equivalent of *and so on* is *ir taip toliau* (its abbreviated and unabbreviated forms constitute 21 instances). However, in the BD this equivalent appears in the second position.

The equivalents of *etc.* show a similar tendency; the dictionary equivalents are among the least frequent corpus equivalents. The unabbreviated form *ir taip toliau* appears as the first sense of *etc.* in the dictionary; its frequency in the corpus, however, is only 3 instances. In contrast, the abbreviated form (*ir t. t.*), which is not provided in the

dictionary, is the most frequent equivalent in translated texts; its frequency amounts to 91 occurrences.

Importantly, in the case of *etc.*, as the corpus data show, there are numerous cases of omission (21 occurrences where *etc.* is not translated at all). Such instances deserve special attention since the communicative effect of utterances without the general extender is different from that in the original. Such instances will be discussed separately in the section below.

Cases of Omission

It has been observed in the corpus data that the general extender *etc.* is not translated in the TT in as many as 21 cases (or 13% of the total number of its uses); *and so on* is translated in all instances. Omission is a typical strategy of translation when the equivalent for some item in the source text is missing in the target language. However, in the case of *etc.*, there is no lexical gap in Lithuanian, and it has a very evident and straightforward equivalent. Therefore, it can be hypothesised that such omissions occur due to the underestimation of the importance of vague language categories.

The importance of such categories becomes evident if we compare the original proposition with *etc.* and the proposition in the TT without a general extender. Changes in the communicative function of the proposition where the general extender is omitted can be observed in examples (1) and (2):

- (1a) *It is employers' money, not yours, that maintains on-line sites, newspaper advertising, etc.*
- (1b) *Ne jūs, o darbdaviai moka pinigų įvairioms interneto svetainėms ir laikraščiams.*
- (2a) *With products bought for emotional reasons, there sometimes surfaces a “rational regret” which means that the buyer wants to find rational reasons (reading tests, asking opinions and confirmation from authorities etc.) for the buying decision after he/ she has made it.*
- (2b) *Įsigyjant prekę emociniu pagrindu kartais susiduriame su taip vadinamu „protingu apgailestavimu“. Tai reiškia, kad pirkėjas ieško racionalių priežasčių, kodėl jis įsigijo šį produktą ar paslaugą (skaito testus, teiraujasi nuomonių ar patvirtinimų iš atitinkamų institucijų).*

In the ST in example (1), the general extender *etc.* indicates that “online sites” and “newspaper advertising” do not constitute a complete list, whereas in the TT the list “internetų svetainės ir laikraščiams” (“online sites and newspapers”) suggests that the list of two items joined with the conjunction *and* is complete, and there are no other relevant items. The proposition in (1b) is more categorical than the proposition in the ST in (1a), which conveys tentativeness and implies possible variation of the listed items. Similar changes in the communicative function can be observed in example (2), where the list of “rational reasons” is made complete and categorical in the TT by removing the general extender. The difference between the ST and TT in these and similar cases, however, is not only in the communicative function of the propositions. The ST and TT versions of the same proposition differ here also content-wise, since the referential scope of the list in the ST is wider than in the TT. In examples (1) and (2), the more specific items following a more basic category are illustrative examples of the category, which

here are a crucial cue for the listener when interpreting the preceding or implied category (“rational reasons” in (2) and the implied category “all the areas where employers’ money is spent” in (1)).

Some omissions in the data obtained from the PCLL do not modify the content or communicative function of the ST to a large extent since the impact of general extenders is compensated by some other lexical categories that indicate the incompleteness of the enumerated items in the proposition. Such compensated omissions are illustrated in example (3), where the categories used to introduce an incomplete list are presented in *italics*:

- (3a) In turn, the Structural Funds should generally focus on the development of infrastructure linked to measures to stimulate economic growth (*such as* tourism development, improvements to increase the attractiveness of industrial sites, *etc.*).
- (3b) Savo ruožtu struktūriniai fondai turėtų daugiausia orientuotis į infrastruktūros plėtojimą, susijusį su ekonomikos augimą skatinančiomis priemonėmis (*pavyzdžiui*, turizmo plėtra, patobulinimai, skirti pramoninių vietovių patrauklumui didinti).

The introductory item *such as* and its Lithuanian equivalent *pavyzdžiui* show that some information in the proposition is withheld, and the provided list is open-ended. Therefore, it can be argued that this lexical item suffices in such an instance since it already indicates that the listed items are not the only examples of a larger category.

A similar tendency to omit VL categories in translation has been observed in some previous research on translation of quantifiers. As the study of equivalents of quantifiers in the PCLL (Ruzaitė, 2009) demonstrates, quantifiers, which are another VL category, are also frequently omitted in translations. Interestingly, the quantifiers that have no equivalent in the TT most commonly include two synonymous quantifiers *a little* and *a bit* (in the PC 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). Thus the quantifiers that are most commonly omitted in the TT are those that are used primarily not for their informative content, but that are mainly important as hedges or mitigators. Therefore, it may be argued that quantifiers and general extenders, which perform some important pragmatic functions, are omitted since their communicative function is underestimated by translators. This suggests that translators still give too much prominence to informative content and thus neglect lexical items mainly necessary for other purposes such as deliberately withholding specific information and conveying tentativeness or an evaluation of a proposition.

Another possible explanation for the frequent omissions of general extenders in the TT is cross-linguistic differences; such an assumption was a part of the initial hypothesis in this paper. It was initially assumed that general extenders are often omitted in translation since they are generally infrequent in Lithuanian. The monolingual Lithuanian corpus (CCLL), however, shows that general extenders are of extensive use in Lithuanian. The frequency of different general extenders in Lithuanian is presented in Table 5.

Table 5. Frequency of Lithuanian general extenders in the Corpus of Contemporary Lithuanian (ranked by frequency)

General extender	Raw freq.	Relative freq. (per mln)
<i>ir kiti</i> (“and other(s)”)	12 943	129.43
<i>ir panašiai</i> (“and something like that”)	1 541	15.41
<i>ir taip toliau</i> (“et cetera”)	838	8.38
<i>ar panašiai</i> (“or something like that”)	337	3.37
<i>arba panašiai</i> (“or something like that”)	17	0.17
Total:	15 676	156.76

The general extenders in Table 5 appear in the CCLL as statistically stable units, and are of highly frequent usage. The most frequent general extender is *ir kiti* (“and other”). Its high frequency is especially important here since it demonstrates that general extenders are used to a great extent in Lithuanian. This strongly suggests that omissions in translations most probably occur not due to cross-linguistic differences but because of the general attitude to vague language items, which are often treated as unnecessary linguistic items that optional and can be easily excluded from a proposition.

Conclusions

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. First of all, the equivalents in bilingual dictionaries and those obtained from the parallel corpus coincide only to a limited extent. Some equivalents that appear in the parallel corpus are not provided in the bilingual dictionary (the corpus equivalents of *ir t. t.*, *t. t.*, *ir/ ar/arba pan.*, *ir kt.*, *kt.*, and *ir kitų* do not appear in the dictionary entry for *etcetera*). In addition, cases of omission are of special interest since such instances strongly suggest that vague language items are often treated by translators as items of low informative content and thus are frequently disregarded in the target text.

The data have revealed that in a number of cases the general extender *etc.* is omitted in the TT. As argued above, such omissions, if not compensated by other linguistic means, can change the communicative effect and informative content of the translated proposition. Therefore, such omissions of vague language items should be avoided. On the other hand, examples with omissions can be used in language teaching when designing tasks for English language learners to highlight the communicative effect of vague language items.

Equivalents for most lexical items in Lithuanian bilingual dictionaries are often presented on the basis of the lexicographers’ intuitions; synonymous equivalents are often provided as interchangeable items without any usage specifications. However, as corpus data show, in actual translations there exist important usage differences between apparently synonymous equivalents presented as interchangeable items. A parallel corpus thus can provide a more extensive inventory of cross-linguistic correspondences than a bilingual dictionary. Therefore, dictionary equivalents should be necessarily supplemented with corpus data. It is important, however, to note that the parallel corpus can be used as an *aid* for different language users, not a *substitute* for a bilingual dictionary. Such a corpus should be employed as a

translator assistant, teacher assistant and learner assistant but it cannot completely be a substitute for other reference sources.

The present study has highlighted some important aspects of the usage of general extenders on the basis of a parallel corpus. A closer study of general extenders in monolingual corpora in the two languages could reveal a more detailed picture, especially with regard to collocational patterns and usage differences between different modes of speaking or discourse types.

References

- Aijmer, K., 1985. What Happens at the End of Our Utterances? The Use of Utterance-Final Tags Introduced by "And" and "Or". *Papers from the 8th Scandinavian Conference of Linguistics*, pp.366-389.
- Channell, J., 1985. Vagueness as a Conversational Strategy. *Nottingham Linguistic Circular*, 14, pp.3-24.
- Channell, J., 1990. Precise and Vague Quantities in Academic Writing. In: W. Nash, ed. *The Writing Scholar: Studies in the Language and Conventions of Academic Discourse*. Beverly Hills: Sage, pp.95-117.
- Channell, J., 1994. *Vague Language*. Oxford: Oxford University Press.
- Cheng, W., Warren, M., 2001. The Use of Vague Language in Intercultural Conversations in Hong Kong. *English World-Wide*, 22 (1), pp.81-104.
- Crystal, D., Davy, D., 1979. *Advanced Conversational English*. New York: Longman.
- Dines, E., 1980. Variation in Discourse – and Stuff Like that. *Language in Society*, 1, pp.13-31.
- Drave, N., 2002. Vaguely Speaking: a Corpus Approach to Vague Language in Intercultural Conversations. In: *Language and Computers, New Frontiers of Corpus Research. Papers from the Twenty First International Conference on English Language Research on Computerized Corpora Sydney 2000*. Peters, P., Collins, P. and Smith, A. (eds.). Amsterdam and New York: Rodopi, pp. 25-40.
- Dubois, S., 1993. Extension Particles, etc. *Language Variation and Change*, 4, pp.179-203.
- Ediger, A. M., 1995. *An Analysis of Set-Marking Tags in the English Language*. Unpublished PhD thesis, the University of California.
- Hyland, K., 1998. *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins Publishing Company.
- Overstreet, M., Yule, G., 1997. On Being Inexplicit and Stuff in Contemporary American English. *Journal of English Linguistics*, 25(3), pp.250-258.
- Overstreet, M., Yule, G., 2001. Formulaic Disclaimers. *Journal of Pragmatics*, 33, pp.45-60.
- Overstreet, M., Yule G., 2002. The Metapragmatics of and Everything. *Journal of Pragmatics*, 34, pp.785-794.
- Overstreet, M., 1995. *The Form and Function of General Extenders in English Interactive Discourse*. Unpublished PhD thesis, the University of Hawaii.
- Overstreet, M., 1999. *Whales, Candlelight, and Stuff like that. General Extenders in English Discourse*. Oxford: Oxford University Press.
- Overstreet, M., 2005. And stuff und so: Investigating Pragmatic Expressions in English and German. *Journal of Pragmatics*, 37, pp.1845-1864.
- Piesarskas, B., 2005. *The English-Lithuanian Dictionary*. Vilnius: Alma littera.
- Ruzaitė, J., 2009. Corpora for Applied Purposes: Translation of Vague Quantifiers from English into Lithuanian. In: *Eesti Rakenduslingvistika Ühingu aastaraamat [Estonian Papers in Applied Linguistics]*.
- Wang, A.Y., Piao, S., 2007. Translating Vagueness? A Study on Translations of Vague Quantifiers in an English-Chinese Parallel Corpus. *The Proceedings of the Corpus Linguistics Conference 2007*. Birmingham: UK. Available at: http://ucrel.lancs.ac.uk/publications/cl2007/paper/278_Paper.pdf [Accessed 5 December 2009]
- Ward, G., Birmer, B., 1993. The Semantics and Pragmatics of and Everything. *Journal of Pragmatics*, 19, pp.205-214.
- Williamson, T., 1994. *Vagueness*. London: Routledge.
- Winter, J., Norrby, C., 1999. *Proceedings of the 1999 Conference of the Australian Linguistic Society*. Available at: <http://www.linguistics.uwa.edu.au/about/research/als99> [Accessed 5 December 2009]

Jūratė Ruzaitė

Neapibrėžtos kalbos vertimas: bendrųjų plėtinių vertimo atitikmenys lygiagrečiajame tekste

Santrauka

Šiuo straipsniu siekiama parodyti, kaip lygiagrečiojo teksto duomenys, papildyti kitų tekstų duomenimis, gali būti naudojami lyginamajai dviejų kalbų analizei. Straipsnyje tiriama anglų ir lietuvių kalba remiantis keliais tekstais: lygiagrečiuoju tekstu ir dviem vienkalybiais anglų ir lietuvių kalbų tekstais. Pagrindinis duomenų šaltinis – Lygiagretusis tekstas (internetinė prieiga: www.donelaitis.vdu.lt), kuris yra sudarytas iš čekų ir anglų kalbų tekstų bei jų lietuviškųjų vertimų. Tyrimui naudoti tik tekstų vertimai iš anglų į lietuvių kalbą; šią tekstų dalį sudaro 70 813 lygiagretūs sakiniai (2 023 005 žodžiai: anglų – 1 129 768, lietuvių – 893 237). Tyrimo objektas – netikslios kalbos vertimo atitikmenys lygiagrečiajame tekste ir dvikalybiame anglų-lietuvių kalbų žodyne.

Neapibrėžta kalba šiame darbe suprantama kaip tam tikros leksinės kategorijos, kurių semantika yra sunkiai apibrėžiama, bet ypač svarbios jų pragmatinės ypatybės; todėl šios kategorijos dažnai įvardijamos kaip pragmatolingvistinės. Jos apima (a) kiekybinius žymiklius (*daug, mažai, šiek tiek, nedaug*), (b) apytikslūs skaičius (*maždaug, apytiksliai, apie + N*), (c) metalingvistinius įterpinius (kaip ten jis, koks ten jo vardas) ir (d) bendruosius plėtinius (ir taip toliau, ir panašiai, ir kiti). Šio straipsnio pagrindinis tikslas yra ištirti, kaip rašytinėje anglų kalboje dažniausiai vartojami bendrieji plėtiniai (*et cetera, etc. ir and so on*) yra verčiami į lietuvių kalbą bei rasti atitikmenys palyginti su dvikalybiame žodyne pateikiamais atitikmenimis. Lygiagrečiojo teksto duomenys lyginami ir su vienkalybiais tekstais: Britų nacionaliniu tekstu (British National Corpus (BNC)) ir Dabartinės lietuvių kalbos tekstu.

Kaip rodo tyrimo rezultatai, žodyne pateikiami ir realiaame tekste pavartoti vertimo atitikmenys sutampa tik iš dalies. Tie atitikmenys, kurie sutampa abiejuose šaltiniuose, skiriasi vartosenos dažnumu, kas neatspindi žodyne pateikiamoje informacijoje. Tyrimas parodė, jog dažnai bendrieji plėtiniai iš viso nėra verčiami, dėl to pakinta vertimo komunikacinė funkcija, taip pat iškraipoma originalaus teksto informacija. Dabartinės lietuvių kalbos teksto duomenys rodo, kad bendrieji plėtiniai lietuvių kalboje yra plačiai vartojami, tad tokie netikslūs vertimai negali būti paaiškinami tarpkalbiniais ar tarpkultūriniais skirtumais. Tokius vertimo netikslumus ko gero lemia vertėjų požiūris į pragmatolingvistinių kategorijų svarbą tekste. Tyrimas aiškiai parodė, kad lygiagretusis tekstas gali būti labai svarbi pagalbinių priemonė žodžių reikšmėms, vertėjams ir kalbos mokymo specialistams.

Straipsnis įteiktas 2010 01
Parengtas spaudai 2010 03

Apie autorę

Jūratė Ruzaitė, doc., dr., Vytauto Didžiojo universiteto Anglų filologijos katedra.

Mokslinės veiklos sritys: diskurso analizė, tekstų lingvistika, vaikų kalbos sutrikimai.

Darbovietės adresas: Vytauto Didžiojo universitetas, Anglų filologijos katedra, K. Donelaičio g. 52, 44248 Kaunas

El. paštas: j.ruzaitė@hmf.vdu.lt.

DOI: 10.5755/j01.sal.1.16.43359