Acta Baltico-Slavica

Acta Baltico-Slavica, 36
SOW, Warszawa 2012

DOI: 10.11649/abs.2012.007

Roman Roszko
Institute of Slavic Studies of the Polish Academy of Sciences
Warsaw

The Importance of Bilingual Corpora in Polish--Lithuanian Comparative Studies

1. Introduction

The phenomena studied in this paper were previously analysed in R. Roszko (2011), where the semantic category of hypotheticality was defined in accordance with the theoretical framework of comparative studies. The category was divided into six degrees of probability and, consequently, each degree was assigned relevant Polish and Lithuanian lexical expressions.

The present article focuses on the issue of the use of parallel corpora (in this case, a Polish-Lithuanian corpus) in comparative studies and the impact it bears on the results. In the 1990s, I began to study modal categories in Lithuanian with Danuta Roszko. Subsequently, our research interests grew to encompass Polish as well as the Puńsk dialect of Lithuanian. We published some of the results of our research in a few articles and three books: R. Roszko (1993, 2004) and D. Roszko (2006). In the end of 2000s, we began compiling parallel corpora: a Bulgarian-Polish-Lithuanian corpus (Dimitrova, Koseska-Toszewa, Roszko, & Roszko, 2011) and a Polish-Lithuanian corpus. Based on the corpora in question, we decided to verify the compiled data using manual extraction.

Previous research on hypotheticality in Polish and Lithuanian made use of a corpus consisting of 11 texts, which yields 22 items in total, because each item appeared

Publisher: Institute of Slavic Studies PAS & The Slavic Foundation
[Wydawca: Instytut Slawistyki PAN & Fundacja Slawistyczna]

in both languages. The corpus included 5 texts translated from Polish into Lithuanian, four from Lithuanian into Polish, and two texts translated from Russian into Polish and Lithuanian. Presently, thanks to the Experimental Polish-Lithuanian Parallel Corpus, the number of texts has increased greatly. The new texts include not only parallel translations from Polish and Lithuanian but also numerous works of contemporary world literature in many languages, e.g. German, English or Portuguese. These include very popular authors, such as Dan Brown, William Golding, John Gray, J. K. Rowling, Paulo Coelho, Richard Bach, Françoise Sagan, Vladimir Sorokin and many others.

Since the phenomenon analysed in this article pertains to modality, it is predominantly restricted to fiction literature. Modality is virtually non-existent in operation manuals and legal documents, and the legislation of the European Union includes a very limited set of examples:

[1] **Polish:** Dwa pierwsze programy *z pewnością* wniosły wartość dodaną do wymiany informacji między administracjami, zaś nowy program *na pewno* przyczyni się do rozwoju lokalnego i regionalnego poprzez ułatwianie wymiany pomysłów i doświadczeń w różnych dziedzinach, takich jak zatrudnienie, rybołówstwo, rolnictwo, zdrowie, ochrona konsumenta oraz wymiar sprawiedliwości i sprawy wewnętrzne.

Lithuanian: *Reikia pripažinti, kad* dvi pirmosios programos *tikrai* padėjo administravimo institucijoms keistis informacija, o naujoji programa *neabejotinai* prisidės prie vietos ir regionų vystymosi, nes bus sudarytos geresnės sąlygos keistis idėjomis ir patirtimi įvairiose srityse, pavyzdžiui, užimtumo, žuvininkystės, žemės ūkio, sveikatos, vartotojų apsaugos ir teisingumo ir vidaus reikalų srityse.

English: The two initial programmes have clearly provided added value to the exchange of information between administrations, and the new programme will definitely contribute to local and regional development by facilitating the exchange of ideas and experiences in various fields such as employment, fisheries, agriculture, health, consumer protection and justice and home affairs.

The above example comes from the Opinion of the Committee of the Regions of the European Union, no. 2009/C 200/10. In the parallel translations there are corresponding modal expressions in Lithuanian and Polish, e.g. Pol. *z pewnością* and Lith. *tikrai* or Pol. *na pewno* and Lith. *neabejotinai*. Both expressions are lexical. The Lithuanian translation includes also a syntactic construction operating together with the lexeme *tikrai* mentioned above: *reikia pripažinti*, *kad … tikrai*.

Having analysed the legislation of the European Union in detail, I noticed that the Lithuanian modus relativus is not used in legal documents of this sort. It is also worth mentioning that the Polish and the Lithuanian translations of the same text are

likely to differ semantically. There might be cases when the Polish translation contains modal expressions, whereas its Lithuanian counterpart does not. The examples below show diverging interpretations of two translations of the same texts on the basis of another category of modality, i.e. imperceptivity:

[2] **Polish:** Sąd ten ustalił, że A. Achughbabian posiada obywatelstwo armeńskie, że zastosowano wobec niego środek w postaci zatrzymania, a następnie środek detencyjny ze względu na nielegalny pobyt i że powołuje on się na to, *jakoby* art. L. 621–1 Ceseda *był niezgodny* z dyrektywą 2008/115, w świetle wykładni przedstawionej w ww. wyroku w sprawie El Dridi.

Lithuanian: Šis teismas konstatavo, kad A. Achughbabian yra Armėnijos pilietis, kuris buvo sulaikytas, vėliau suimtas dėl neteisėto gyvenimo šalyje ir kuris teigia, kad Ceseda L. 621–1 straipsnis neatitinka Direktyvos 2008/115, kaip ji aiškinama minėtame Sprendime El Dridi.

English: The latter [i.e. the court] took note that Mr Achughbabian was of Armenian nationality, that he had been placed in police custody and then in detention for an unlawful stay, and that he had argued that Article L. 621-1 of Ceseda is incompatible with Directive 2008/115, as interpreted in El Dridi.

[3] **Polish:** Powyższego wniosku nie podważa ani okoliczność przedstawiana przez rząd francuski, *jakoby* na podstawie okólników kierowanych do instytucji sądowych kary przewidziane przez uregulowanie krajowe rozpatrywane w postępowaniu przed sądem krajowym *były rzadko wymierzane* z wyjątkiem wypadków, w których nielegalnie przebywająca osoba dopuszcza się, poza wykroczeniem w postaci nielegalnego pobytu, innego wykroczenia, ani okoliczność, również podnoszona przez ten rząd, *jakoby* A. Achughbabian *nie został* skazany na taką karę.

Lithuanian: Pirma padarytos išvados nepaneigia nei Prancūzijos vyriausybės nurodyta aplinkybė, kad pagal teisėsaugos institucijoms išsiųstus aplinkraščius nagrinėjamuose nacionalinės teisės aktuose nustatytos bausmės, išskyrus atvejus, kai asmuo be neteisėto buvimo šalyje padarė dar ir kitą baudžiamąjį nusižengimą, retai skiriamos, nei tai, kad, kaip taip pat nurodė ši vyriausybė, A. Achughbabian minėtos bausmės nebuvo skirtos.

English: The above conclusion is not called into question either by the fact, invoked by the French Government, that, pursuant to circulars sent to the courts, the penalties laid down by the national legislation at issue in the main proceedings are rarely imposed outside cases where the person staying illegally has, in addition to the offence of staying illegally, also committed another offence, or by the fact, likewise invoked by that government, that Mr Achughbabian has not been sentenced to those penalties.

The Polish versions of [2] and [3] contain three instances of jakoby ('supposedly'; 'allegedly'), which connects the subordinate and the matrix clause, at the same time endowing the subordinate clause with semantic properties of unreality, uncertainty or improbability. This is absent in the Lithuanian version, because there is no lexical, morphological or syntactic expression that would indicate any properties of modal imperceptivity.

2. The semantic category of hypotheticality

The definition of hypotheticality given in R. Roszko (2011) conforms to the theoretical assumptions of hypotheticality laid out in Maldžieva, Koseska-Toszewa & Penčev (2003).

Hypotheticality is assumed to constitute one of the categories typical for natural languages. It is used to express the subjective stance of the speaker regarding the states and events described in an utterance.[1] The category of hypotheticality pertains to propositions and can be expressed on three different levels: lexical, morphological, and syntactic. One of the features typical for hypotheticality are the different levels of probability it may carry: from complete falsehood (0) to complete truth (1). In D. Roszko (2006), the scale also included the value of ½, a middle value indicating the use of morphological expressions of hypotheticality typical for standard Lithuanian and the Puńsk dialect. Maldžieva (Maldžieva et al., 2003) does not mention any morphological expressions of probability, which adequately reflects the structures of Bulgarian and Polish. Hence, in Maldžieva et al. (2003), there is no possibility of expressing a situation when the probability of a proposition P(x) equals that of its negation ~P(x).

3. Expressions of hypothetical modality in Polish and Lithuanian

The semantic category of hypotheticality can be expressed on different levels: lexical and syntactic (in both languages in question) as well as morphological (only in Lithuanian).[2] The presence of modus relativus in Lithuanian enables the hearer to accurately determine whether a given fragment of discourse contains hypotheticality. In contrast, Polish lacks morphological expressions of hypotheticality and the use of only lexical and syntactic expressions does not allow speakers of Polish to determine the hypothetical properties of a proposition with full confidence. What is more, people usually avoid using overlapping lexical expressions for stylistic reasons.

[1] If there is an expression of subjective stance of the speaker, it means that the semantic structure of a given proposition contains an operator of possibility.

[2] For a more detailed account, see D. Roszko (2006).

3.1. Lexical expressions of hypotheticality

The analysis presented in this section is limited to lexical expressions of hypotheticality in the two languages studied. Table 1 contains a list of expressions of hypotheticality typical for Lithuanian and Polish – the classification is broken down into six different groups displaying growing degrees of probability:

Table 1. Lexical expressions of hypotheticality in Polish and Lithuanian

Group	Polish	Lithuanian
I	może i	gal ir
II	a może i, może zresztą	o gal ir
III	a może	o gal
IV	chyba, może jednak, może rzeczywiście, może naprawdę, przypadkiem, a może faktycznie	gal, turbūt, nebent, gal tikrai, gal vistiek, o gal faktiškai
V	być może, może, pewnie, zapewne, tak myślę, moim zdaniem, prawdopodobnie, przypuszczalnie, ewentualnie, możliwe, widać, wydaje się, zdaje się, snadź, bodaj/bodajże (sporadycznie w znaczeniu hipotetyczności)	galbūt, įtikimai, eventualiai, man rodos, taip manau, pagal mane, pasirodo, atrodo, rodos, berods, man atrodo, galimas daiktas, manyčiau, tikrai, faktiškai, rasi, matyt
VI	najpewniej, najprawdopodobniej, moim zdaniem,* jak widać, bez wątpliwości, jak sądzę, jak przypuszczam, jak mi się zdaje, na pewno, niewątpliwie, widocznie, najwidoczniej, bez wątpienia, niechybnie, nieomylnie, na mur, mur beton, iście, doprawdy, murowanie, jako żywo, ani chybi (ni chybi)	tikriausia, tikriausiai, veikiausiai, greičiausiai, mano galva, be abejonės, kaip manau, kaip man atrodo, man panašiausia, iš tikro, iš tikrųjų, tikras dalykas, žinoma

* Some lexical expressions of hypotheticality can be characteristic of two adjoining groups, cf. D. Roszko, 2015.

The expressions presented in Table 1 have been compiled as a result of the research in the Experimental Polish-Lithuanian Corpus. When only traditional methods of research were used, i.e. non-corpus methods, the list was significantly smaller: the expressions found in such studies were those which are typical for both languages, characterised by the greatest frequency (e.g. Pol. *chyba* and Lith. *gal*) or those which display substantial formal similarity, e.g. Pol. *a może i* and Lith. *o gal ir*. Only by taking advantage of a large corpus were we able to discover other parallel combinations of expressions of hypotheticality, which had not been noticed previously due to their low frequency in examples extracted manually, e.g. Pol. *iście*, *nieomylnie*, *doprawdy*, *ani/ni chybi* and Lith. *rasi*, *manyčiau*, *mano galva*. A detailed analysis of the reasons for the discrepancies between the number of lexical expressions identified in both languages has led to a conclusion that the set of 11 parallel Polish and Lithuanian texts was inadequate to obtain satisfactory results. The fact that it consisted solely of mutual translations or translations of Russian texts only aggravated the problem. Subsequently, an experimental mini-corpus was compiled out of the 11 texts in

question and then analysed. The analysis using corpus techniques yielded results that did not diverge much from those obtained in the study conducted in a traditional manner. This may have been caused by the nature of the novels included or the characteristics of the language used in the period when they were written. The Lithuanian and Russian texts composed in the USSR period contain very specific language.

For objective reasons, the data obtained by means of traditional extraction were far from comprehensive, which becomes clear when one takes into consideration that 11 novels is quite a large sample of texts for a study of this kind. Hence, it was possible neither to accurately determine the frequency of the expressions found, nor to compile a comprehensive list of expressions of hypotheticality. That is why previous studies often reported fuzzy boundaries between some groups. It was argued that the expressions in Group 4 could be used to express probability associated with Group 3 and Group 5. An analysis of corpus data corroborated this claim, albeit the magnitude of the phenomenon was lower than previously expected. The statistical data have shown that the fuzzy boundaries occur in 9% of all examples at most and below 5% on average. I shall illustrate this below with the example of Pol. *chyba* and Lith. *turbūt*.

In 68% of cases when the Polish expression *chyba*[3] was used, the counterpart in the corresponding Lithuanian text was *turbūt*. The other pairs yielded the following statistics:

Pol. *chyba* – Lith. *nebent* – 23%,
Pol. *chyba* – Lith. *gal* – 5%,
Pol. *chyba* – Lith. *rasi* – 2%,
Pol. *chyba* – Lith. *matyt* – 1%.

The remaining one percent comprises other correspondences between Polish and Lithuanian, but its omission does not alter the general conclusions. Table 2 below presents the detailed results of the corpus study.

Table 2. Typical Lithuanian equivalents of the Polish *chyba* (Group 4)

Polish expression	Lithuanian equivalent	Percentage of instances	Group
chyba	turbūt	68%	IV
chyba	nebent	23%	IV
chyba	gal	5%	IV
	Total	**= 96%**	**IV**
chyba	rasi	< 2%	V
chyba	matyt	< 1%	V
	Total	**= < 3%**	**V**

[3] There were 1964 instances in the corpus of the Polish expression *chyba* in its hypothetical sense.

Fuzzy boundaries have been found in less than 3% of cases and pertained only to Groups 4 and 5. There were no cases of using expressions from Group 4 in contexts typical for Group 3 and conversely.

Taking the above results into consideration, it might be relevant to investigate the usual Polish equivalents for the Lithuania *turbūt*.[4] The analysis did not yield as great a number of instances as in the case of the Polish *chyba*. The pair that occurs most frequently in the corpus is Lith. *turbūt* and Pol. *chyba* (96% of instances); another 3% of instances consisted of the pair Lith. *turbūt* and Pol. *bodaj*, cf. Table 3 below:

Table 3. Typical Polish equivalents of the Lithuanian *turbūt* (Group 4)

Lithuanian expression	Polish equivalent	Percentage of instances	Group
turbūt	chyba	96%	IV
	Total	**= 96%**	**IV**
turbūt	bodaj	3%	V
	Total	**= 3%**	**V**

It was quite surprising that also in this case the fuzzy boundaries occur between Groups 4 and 5. The expressions in Group 4 are never used to convey probabilities typical for Group 3, which is corroborated by other corpus data not included in this study – the differences between the probabilities conveyed by the expressions from Groups 3 and 4 might be so great that it is impossible to use the expressions in these groups interchangeably. In contrast, the difference between the probabilities in Groups 4 and 5 is significantly smaller, which finds its reflection in the fact that the expressions in these groups can be used interchangeably, albeit not frequently so.

Corpus analysis yielded a number of previously unmentioned expressions of hypotheticality. The Lithuanian *rasi* is one of these expressions. It was probably derived from 2SG.FUT of Lith. *rasti* 'find'. The examples below illustrate the use of *rasti* and compare it with its Polish equivalents:

[4] **Lithuanian:** Kuris podraug bjaurisi savo geidimo objektu ir kraustosi iš galvos dėl jo, ir atiduotų dėl jo gyvybę, *rasi*, prilygdamas jausmais Romeo ir Džiuljetai…

Polish: Który jednocześnie brzydzi się obiektu swej pożądliwości i szaleje za nim i gotów jest narazić dla niego życie dorównując, *być może*, uczuciom Romea dla Julii…

English: A man who at one and the same time is ashamed of the object of his desire and cherishes it above everything else, a man who is ready to sacrifice

[4] There were 1355 instances in the corpus of the Lithuanian expression *turbūt* in its hypothetical sense.

his life for his love, since the feeling he has for it is perhaps as overwhelming as Romeo's feeling for Juliet.

[5] **Lithuanian:** Modelis, bet, *rasi*, natūralaus dydžio.

Polish: Model, ale *chyba* naturalnej wielkości.

English: A model, but lifesize.

The frequency of *rasi* in the Experimental Polish-Lithuanian Corpus is relatively low; the are only 362 instances of *rasi*. In contrast to other expressions of hypotheticality, *rasi* does not have one predominant equivalent in Polish. Usually, it corresponds to *być może* (25%), *może* (33%), *bodaj* (33%) or *chyba* (9%). Identically to *rasi*, the first three Polish equivalents belong to Group 5; only *chyba* is situated in Group 4.

4. Corpus data vs dictionary entries

It has often been claimed that corpus analysis might improve the quality of bilingual dictionaries. This section contains the results of a cohesion test which consisted in a comparison of the corpus data of a selected Lithuanian expression of hypotheticality with its entry in *PolLit'as dictionary* (Vaitkevičiūtė, 2003). I chose an electronic dictionary for the study, because it is much faster and much more convenient to look up an entry there than to browse through a few hundred pages of a conventional paper dictionary. The results of the analysis are presented in Table 4 below.

Table 4. The cohesion of corpus data and dictionary entries based on the Lithuanian *rasi*

Item	Dictionary entry		Corpus data		Group
	Lithuanian synonyms	**Polish equivalents**	**Lithuanian synonyms**	**Polish equivalents**	
1.	galbūt matyt		galbūt matyt	może (33%), bodaj (33%), być może (25%)	V
2.			įtikimai, eventualiai, man rodos, taip manau, pagal mane, pasirodo, atrodo, rodos, berods, man atrodo, galimas daiktas, manyčiau, tikrai, faktiškai		V
3.		chyba		chyba (9%)	IV
4.		pewnie			V
5.	turbūt				IV
6.	tikriausiai, be abejonės				VI

Five synonyms of the Lithuanian *rasi* were found in the analysis of dictionary entries. Only two of them correspond to these found in the corpus data, namely *galbūt* and *matyt*; both of them belong to Group 5. The remaining expressions, *turbūt*, *tikriausiai* and *be abejonės* are given by the author of the dictionary as synonyms to *rasi*. However, the corpus analysis has demonstrated that they in fact belong to Group 4 and Group 6, respectively. Therefore, the dictionary should have included only *galbūt* and *matyt* in this entry.

The accuracy of the dictionary entry amounted to 40%; 2 out of 5 expressions were coherent with those found in the corpus. In contrast, the list of Polish equivalents was coherent to a very small extent. Of the Polish equivalents of the Lithuanian *rasi* that the dictionary lists, only one belongs to Group 5, namely *pewnie*. The corpus data, however, did not support this equivalence, since there was no occurence of the mutual correspondence of *rasi* and *pewnie*. The other equivalent, *chyba* (Group 4), amounts to a mere 9% of all correspondences appearing in the corpus.

Taking the result of the corpus study into consideration, a dictionary entry for *rasi* could look as follows:

rasi (Pm:H)

może, *bodaj*, *być może* (↑); *chyba* (↓)

Where **Pm** denotes the possibility of modality, **H** stands for hypotheticality, and ↑ and ↓ mean high and low frequency equivalents, respectively.

5. The impact of source language on Polish-Lithuanian correspondences

Another experiment involved analysing corpus data for parallel texts. Each text was analysed separately, depending on whether the source text was in Lithuanian (a), Polish (b) or a different language (c). The results for (b) and (c) were quite similar, the analysis of corpus in (a), however, yielded quite different conclusions. The number of equivalent pairs of Polish and Lithuanian expressions in corpus (a) was lower than in the two remaining corpora. Hence, variation among the Polish equivalents in Lithuanian texts and their Polish translations was relatively small. It is also interesting that the form of Lithuanian multi-word expressions to a large extent influenced the choice of their Polish equivalents, e.g. Lith. *o gal faktiškai* – Pol. *a może faktycznie*, Lith. *be abejonės* – Pol. *bez wątpliwości*. In some cases, the original hypothetical meaning was replaced with a different degree of probability, e.g. Lith. *gal vistiek* (Group 4) and Pol. *może zresztą* (Group 2).

In the light of the above, it was expected that an analogical phenomenon would occur in (b), i.e. the form of the Polish expression would influence the form of its Lithuanian equivalent. The corpus data falsified this prediction.

As regards the morphological expressions of hypotheticality typical for Lithuanian, if modus relativus appeared in the Lithuanian sentence, it did not influence the Polish translation, cf.:

[6a] **Lithuanian:** Atsiprašau! *Būsiu pamiršęs* prisegti maketą prie ankstesnio savo laiško. Siunčiu dabar.

Polish: Przepraszam, lecz zapomniałem podpiąć makietę do wcześniejszego listu. Posyłam teraz.

English: I'm sorry. I forgot to attach the model to the previous letter. I'm sending it now.

[6b] **Expected Polish translation:** Przepraszam! *Najwyraźniej* zapomniałem podpiąć makietę do wcześniejszego listu. Posyłam teraz.

Expected English translation: I'm sorry. *Apparently*, I forgot to attach the model to the previous letter. I'm sending it now.

The original Lithuanian sentence with its Polish translation found in the parallel corpus are presented in [6a]; [6b] provides the expected translation of this sentence, including the marker of hypotheticality. A combination of lexical and morphological expressions of hypotheticality did not cause any difficulty to translators and they were able to convey the hypothetical features of the source text:

[7] **Lithuanian:** Sauliau, *greičiausiai* prieš dvi ar tris dienas *būsiu pamiršusi* išgerti tabletę.

Polish: Saulius, *najprawdopodobniej* przed dwoma czy trzema dniami zapomniałam o tabletce.

English: Saulius, *I think* I forgot to take the pill two or three days ago.

6. Conclusions

Corpus methods appear to have many advantages over traditional methods of research, where the extraction was performed manually – traditional analyses were very laborious and limited in scope. The study of corpus data showed a greater number of expressions of hypotheticality in both languages studied. Moreover, the boundaries between the different groups of expressions lost some of their fuzziness and became clearer. The corpus data also lent support to the claim that some expressions from one group can be used to convey probabilities associated with adjoining groups. Nonetheless, this phenomenon is not be as great in scope as it had been predicted in the results of traditional studies.

Mutual translations of Polish and Lithuanian texts as well as Polish and Lithuanian translations from other languages were also analysed. The results did not show any significant

interference of the source texts with their translations – it is likely that the main reason for such a result was the quality of the translators' work. Although some expressions in both languages were similar in form, this did not bear influence on the choice of the equivalent. Subsequently, mutual translations of Lithuanian and Polish texts were analysed separately. The results showed that Polish translators are very careful when choosing the equivalents for some expressions in Lithuanian texts. In a number of translations the meanings arising from the Lithuanian modus relativus (e.g. hypotheticality) were omitted altogether. Modus relativus is usually used in Lithuanian to convey modal imperceptivity, hypotheticality, admirativity, conclusivity, etc. The inaccuracies in some texts might have been caused by an indirect translation, i.e. a translation from Lithuanian to Polish via Russian. In order to verify this hypothesis, a trilingual Polish-Lithuanian-Russian corpus is required. Such a corpus would enable scholars to conduct systematic research in this area.

Translated by Jarosław Józefowski

References

Dmitrova, L., Koseska-Toszewa, V., Roszko, D., & Roszko, R. (2010). Application of multilingual corpus in contrastive studies (on the example of the Bulgarian-Polish-Lithuanian Parallel Corpus). *Cognitive Studies | Études cognitives*, *10*, 217–239.

Dmitrova, L., Koseska-Toszewa, V., Roszko, D., & Roszko, R. (2011). Bulgarian-Polish-Lithuanian corpus – recent progress and application. In D. Majchráková & R. Garabík. (Eds.), *Natural language processing, multilinguality: Sixth international conference, Modra, Slovakia, 20–21 October 2011: Proceedings* (pp. 30–43). Bratislava.

Lietuviu kalbos žodynas. Retrieved from http://http://lkz.lt/

Maldžieva, V., Koseska-Toszewa, V., & Penčev, J. (2003). *Modalność: hipotetyczność, irrealność, optatywność i imperatywność, warunkowość*. Warszawa: Slawistyczny Ośrodek Wydawniczy.

Roszko, D. (2006). *Funkcjonalne odpowiedniki litewskiego perfectum w litewskiej gwarze puńskiej i w języku polskim*. Warszawa: Slawistyczny Ośrodek Wydawniczy.

Roszko, D. (2015). *Zagadnienia kwantyfikacyjne i modalne w litewskiej gwarze puńskiej (na tle literackich języków polskiego i litewskiego).* Warszawa: Slawistyczny Ośrodek Wydawniczy.

Roszko, R. (1993). *Wykładniki modalności imperceptywnej w języku polskim i litewskim*. Warszawa: Slawistyczny Ośrodek Wydawniczy.

Roszko, R. (2004). *Semantyczna kategoria określoności/nieokreśloności w języku litewskim: w zestawieniu z językiem polskim*. Warszawa: Slawistyczny Ośrodek Wydawniczy.

Roszko, R. (2011). Leksykalne wykładniki hipotetyczności w językach polskim i litewskim. *Acta Baltico-Slavica*, *35*, 81–90.

Vaitkevičiūtė, V. (2003). *Didysis lenkų-lietuvių kalbų žodynas*. Marijampolė: Myros Martišienės vertėjų biuras.

The Importance of Bilingual Corpora in Polish--Lithuanian Comparative Studies

Summary

In his article, the author compares and contrasts the results of his own research on the hypothetical modality in Polish and Lithuanian: a) carried out together with Danuta Roszko, using the traditional method (without use of bilingual corpora in the 1990s); b) with use of parallel Polish-Lithuanian corpora resources.

As for the contrast of the two methods, special attention has been drawn to the lexical exponents singled out.

The use of the corpora resources resulted in the fact that the number of exponents of hipothetical modality singled out in the two languages has slightly risen. Moreover, the borders between the corresponding groups of exponents have become more distinct and obvious. The study also confirmed a possibility of using the corresponding groups of exponents to express the meanings of the adjacent groups. The conclusion has been drawn that this phenomenon is as obvious as it was earlier expected (in studies without the use of bilingual corpora).

The separate analysis of corpora resources with the division into the material being a) mutual Polish-Lithuanian translations (i.e. from Polish into Lithuanian and vice versa) and b) translations into Polish and Lithuanian from third languages (here: from German, English or Russian) yields that the target language does not significantly influence the number and diversity of the lexical exponents applied in the two languages. This fact proves a high competence of the translators. The formal resemblance of some of the Polish and Lithuanian exponents does not have a significant influence on which form is chosen in the target language.

In the translations from Polish into Lithuanian, part of the lexical exponents are conveyed with morphological exponents (which Polish lacks). The hypothetical modality understated in Polish is sometimes clarified in translations into Lithuanian with the help of morphological forms. In some translations from Lithuanian into Polish, the total omission of meanings (also the hypothetical) can be noticed, which results from applying the Lithuanian modus relativus forms. In several cases where some Lithuanian-Polish divergences in translations from Lithuanian into Polish have been noticed, a preliminary comparison of a Lithuanian original material and its translation into Russian can suggest that despite the alleged direction of translation (from Lithuanian into Polish), it can indeed be a translation from Russian into Polish. However, proving this hypothesis requires the establishing of a trilingual Polish-Lithuanian-Russian corpora for the selected material, to allow systematic and consistent studies in this direction.

The author gives statistical data for the Polish-Lithuanian lexical exponents of hypothetical modality to distinguish between mutual (Polish-Lithuanian) translations and those from third languages.

Keywords: Lithuanian; Polish; contrastive studies; hypothetical modality; parallel Polish-Lithuanian corpus.

О результатах использования ресурсов двуязычного корпуса на примере польско-литовского сопоставительного исследования

Резюме

В статье автор сопоставляет результаты двух научных исследований по гипотетической модальности – в польском и литовском языках: (а) традиционных исследований и (б) современных, в которых используются цифровые ресурсы (здесь экспериментальный польско-литовский параллельный корпус).

Описание гипотетической категории модальности основывается на методе теоретического сопоставления естественных языков с использованием так называемого языка-посредника (tertium comparationis). Выделяется 6 степеней вероятности (здесь возможности) и соответственно – 6 параллельных групп средств выражения гипотетичности в обоих языках.

Использование параллельного корпуса в исследовании гипотетичности приводит к новым фактам. Количество показателей гипотетичной модальности оказывается несколько выше, чем это было установлено в ходе традиционного исследования (вручную). Следующее, цифровые ресурсы подтверждают предложение об использовании показателей данной группы вероятности/возможности для выражения значений соседних групп, хотя во время проведения традиционных исследований ожидалось бóльшее число использования средств одной группы для выражения соседних степеней вероятности.

Проведенный отдельно анализ ресурсов корпуса, материал которых выбран по исходному языку оригинала: (корпус А) литовского языка, (корпус Б) польского языка, (корпус В) другого языка (напр. английского, португальского, немецкого, русского) показал, что только в одном случае установленных польско-литовских эквивалентных групп показателей гипотетичности заметно меньшее количество и разновидность тех же групп. Речь идет о корпусе А, в котором исходным для перевода является литовский язык.

Установлено также, что в переводе с литовского на польский язык литовские формы модус релятивус (modus relativus) обычно не переводятся. В таких случаях польский перевод теряет исконную модальную характеристику, разве что в оригинале формам modus relativus сопутствуют другие лексические или синтаксические показатели модальности. В некоторых случаях отсутствие семантического соответствия между литовским оригиналом и польским текстом допускает предпосылку непосредственного перевода с русского (а не литовского) на польский язык. Чтобы это доказать, нужен трехъязычный литовско-польско-русский корпус (ограниченный избранными исконно литовскими произведениями и их переводами на польский и русский языки).

В статье корпусные данные сопоставляются с литовско-польским словарем. Оказывается, что предлагаемые автором словаря литовско-польские соответствия лишь в небольшой степени подтверждаются цифровыми ресурсами польско-литовского корпуса.

Ключевые слова: польский язык; литовский язык; сопоставление языков; гипотетrическая категория модальности; двуязычный польско-литовский корпус.