

DANUTA ROSZKO

University of Warsaw, Warsaw, Poland
d.roszko@uw.edu.pl

THE USE OF THE LEXICAL EXPONENTS OF HYPOTHETICAL MODALITY IN POLISH AND LITHUANIAN

Abstract

In this article the author focuses on the issue of hypothetical modality^a in Polish and Lithuanian. A list of the basic exponents of hypothetical modality in both languages is presented. However, the focus is mainly placed on the lexical exponents. On the basis of one of the six groups, which describes a high degree of probability (H5), the differences between the use of the lexical exponents in both languages are examined. In the study, multilingual corpora resources, including *The Polish-Lithuanian parallel corpus Clarin-PL.*, are utilized.

Keywords: Polish; Lithuanian; hypothetical modality; parallel corpora

^aIn the academic literature, for the notion described herein, the term of epistemic modality is also used. Nevertheless, in this paper I will continue to use the term of hypotheticality, which I borrowed from the studies on modality, conducted in Polish-Bulgarian cooperation (Slavic Institute of Polish Academy of Sciences and Institute for Bulgarian Language of the Bulgarian Academy of Sciences).

1 Introduction

Studies on hypothetical modality have been conducted for many years. Recently, however, in the age of the obvious development of parallel corpora, work on the exponents of hypothetical modality specified herein has gained new meaning. The exponents identified earlier can be easily verified, whilst the list of all the linguistic means used for expressing this modality has been considerably extended. It is possible to conduct statistical research and an equiponderant analysis of equivalent expressions in both languages on the basis of multilingual corpora resources.

In this article I refer to my earlier works (i.a. Roszko, D., 2006, 2011, 2015) and those conducted with Roszko, R. (e.g. Roszko, D. & Roszko, R., 2012), in which I gradually extended and revised the list of hypothetical exponents in the Polish and Lithuanian languages.

The methodological basis for the research conducted on hypothetical modality was taken from theoretical contrastive studies utilizing the interlanguage, which I will discuss in detail below.

2 The semantic category of modality

I adopted the basis for the semantic category of modality from V. Maldžieva (2003), who, quoting (Koseska-Toszewa, Maldžieva, & Penčev, 1995), concludes that according to the assumptions made

for *Polish-Bulgarian Contrastive Grammar*, modality is a sentence category. The phrasality of the semantic category results (1) from the orientation of the analytical research conducted from the contents plane to the form plane and (2) from an indirect approach to semantics (Maldźieva, 2003, p. 11). The modality itself is inseparable from the notion of a mental state, with which the addresser expresses lack of knowledge of the content conveyed by himself / herself. This mental state, just like the state of the utterance, is connected to the addresser. Both states (utterance and mental) are concurrent.

V. Maldźieva assumes that the mental state always accompanies the utterance state. As a consequence of such an assumption, it is necessary to conclude that the mental state may either correspond to knowledge (indicative modality) or not (non-indicative modality), comp. (Koseska-Toszewa, Maldźieva, & Penčev, 1995; Maldźieva, 2003, p. 12).

V. Koseska-Toszewa does not agree that the mental state is an inseparable element of every utterance. The very introduction of a notion which does not distinguish modal utterances from non-modal ones is a misunderstanding stemming from the perspective of dichotomous division (commonly used in science). Thus, she suggests a modality description within the Petri net discrete theory, based on such elements as fork (splitting), branching and local state (Koseska-Toszewa, Korytkowska, & Roszko, R., 2007, p. 60; Dimitrova & Koseska, 2014, pp. 57–59).

My article assumes that all which is not true (value “1”) or false (value “0”) is modal, i.e. elements to which the value of truth or the value of falsity cannot be assigned. In order to demonstrate this, I will quote the following example: The sentence *Jan zdał wczoraj egzamin z języka angielskiego*. ‘Yesterday John passed his English exam’ may have the value of truth or falsity, because this sentence has two mutually exclusive interpretations. Either the addresser’s utterance is true or false. That is why I consider such sentences to be non-modal. The following sentence is an example of a modal sentence: *Chyba Jan zdał wczoraj egzamin z języka angielskiego*. ‘Yesterday John may have passed his English exam.’ This sentence cannot be assigned the value of truth or the value of falsity. That is why I conclude that this mental state is correct only for modal sentences, comp. (Koseska-Toszewa et al., 2007, p. 500), where two types of modality are distinguished: possibility (containing the functor of possibility) and obligatory (containing the functor of obligation). In this article we will focus on the modality of possibility, especially on the semantic category of hypothetical modality.

3 The semantic category of hypothetical modality

As I have mentioned above, we distinguish two kinds of modality: one with a possibility functor (it is possible that) and one with an obligatory functor (it is necessary that). Hypothetical modality is one of the modal categories containing such a possibility functor. We can also give examples of other semantic categories containing such possibility functors, such as conclusivity or imperceptibility. According to the theoretical assumptions of contrastive studies, the semantic category of hypothetical modality is a sentence category. The contents of the semantic category of hypothetical modality present the addresser’s subjective attitude to opinions uttered by himself/herself, comp. (Maldźieva, 2003). Maldźieva distinguishes six degrees used to express probability. She assumes that the first degree (H1¹) has the lowest probability and degree six (H6) has the highest. The degrees from second to fifth are positioned between the extreme degrees (first and sixth), see Figure 1.

As can be observed in Diagram 1, the probability value rises concurrently with the increase in the degree numbering of the hypothetical H units. According to the assumptions of the theoretical contrastive studies utilizing the semantic interlanguage, used in the description of modal categories,

¹Markings H1, H2, H3, H4, H5, H6 are not used in V. Maldźieva works. I have introduced them into the description of hypotheticality in my works due to having worked on multilingual corpora, in which hypothetical meanings are one of the many elements of the new semantic annotation. Using the explicit abbreviations in the annotation is standard.

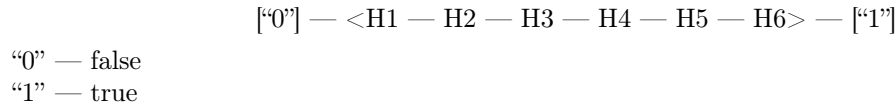


Figure 1: Hypothetical modality. Position of the respective probability degrees on the axis

V. Maldžieva rejects the extreme probability values, i.e. falsity (“0”) and truth (“1”). I agree with this claim, which is shown in Diagram 1 by isolating the range of the individual probability values $\langle H1 \dots H6 \rangle$ from the extreme values of falsity and truth. Moreover, V. Maldžieva eliminates intermediate values between falsity and truth, for which the arbitrary value „ $\frac{1}{2}$ ” can be accepted. As it can be seen in Diagram 1, I have not rejected this value. Accepting the “ $\frac{1}{2}$ ” value differentiates my approach from that of V. Maldžieva’s. The analysis of languages such as Lithuanian shows that the morphological exponents do have the „ $\frac{1}{2}$ ” value, which I describe in a monograph discussing the functions of perfectum in the Lithuanian language and the Lithuanian dialect of Puńsk (Roszko, D., 2006).

4 Hypothetical modality exponents in Polish and Lithuanian

On the basis of the research conducted in (Roszko, D., 2015) I conclude that in both Polish and Lithuanian the typical exponents of hypothetical modality are lexemes (e.g. Pol. *może* ‘maybe’, Lit. *gal* ‘maybe’). Some of the lexemes may cause the syntax to become subordinate/dependant in nature (e.g. Pol. *Wydaje mi się, że ...* ‘Seems to me that ...’, Lit. *Man atrodo, kad ...* ‘Seems to me that ...’), whereas other lexemes may function as interpolation (e.g. Pol. *widać* ‘as it appears’, lit. *matyt* ‘as it appears’). Furthermore, in the Lithuanian language we can observe morphological exponents, including ones based on *modus relativus*, described in the *Grammar of the Lithuanian language* (Ambrasas, 1997), as well as on forms of the Future Perfect Tense, subjunctive structures etc., whereas in the Polish language we can observe paramorphological structures (type of *musiał* (*musi*) + *bezokolicznik* ‘może’, *powinien* + *bezokolicznik* ‘może’, *mógł* + *bezokolicznik* ‘może’)

On the basis of the bilingual Polish-Lithuanian corpora (Experimental and Clarin-PL) it can be concluded that part of the Lithuanian lexical exponents may differ from the forms known in Polish, used unconventionally with this meaning, e.g.:

- forms of the imperative mood of the verb *sutikti* ‘to agree’: *sutik* (sg.) [literally: agree], *sutikite* (pl.) [literally: agree],
- or 2nd person future form of the verb *rasti* ‘find; come upon; find out; learn; meet; come across’ — *rasi* [literally: you will find] used as interjection.

It is worth noting that in the Lithuanian language the combination of the negative and positive forms is used more often than in Polish (e.g. *brolis ne broils* ‘maybe a brother’ [literally: a brother not a brother]) as a hypothetical exponent of group H4. A characteristic feature of the Polish language is the use of the improper intransitive verb (predicative) *widać*.

I would like to examine the lexical exponents in greater detail. That is why, without going into detail as far as non-lexical exponent structures are concerned, I will proceed to the description of lexemes, common in both languages, in the hypothetical modality function. In Polish and Lithuanian, the lexical exponents create a clear structure whose main feature is, in addition to expressing the very content of hypothetical modality, an indication of the degree of probability. The concept of distinguishing six groups (from H1 to H6) has already been presented above, and in which the group of lexemes described by H1 expresses the lowest degree of probability (close to the logical value of “0” / falsity), and each successive one respectively expresses a higher level of probability. The group marked H6 is assigned the highest level of probability (close to the logical value “1” / truth).

In Table 1² I present examples of the lexical exponents which express the contents of hypothetical modality with division into groups. For more on this see (Roszko, D., 2015).

Table 1: Selected lexical exponents of hypothetical modality in Polish and Lithuanian

Group	Polish	Lithuanian
H1	może i	gal ir
H2	a może i	o gal ir
H3	a może	o gal
H4	chyba	gal tikrai
H5	być może	galbūt
H6	najpewniej	tikriausia

The division into groups is not precise. It needs to be emphasized that verifying this very feature is possible due to analysis of the multilingual corporal resources. Individual lexical exponents may represent a small proportion of their use attributed to expressing the degree of probability that is proper for the neighbouring group. For instance, the exponent of group H4 may be occasionally used instead of the exponent of group H5. Then the same exponent of group H4 (appearing in the place reserved for the exponent of group H5) cannot replace the exponent of group H3. Therefore, co-sharing part of the exponents for the neighbouring groups is only possible unilaterally, i.e. exponent *a* assigned to group H4 can be used to express the degree of probability typical for groups H3–H4, whereas exponent *b* of the same group H4 can be used to express the probability degree typical for groups H4–H5 etc.

However, on the basis once more of the two Polish-Lithuanian corpora, it has to be mentioned that some exponents dissent from the general rule. These are: (1) the Lithuanian *matyt* ‘może’ whose use has been noted for three groups- H4-H5-H6 and (2) the Lithuanian *gal*, which seems to have an even broader scope. It needs to be emphasized that we get such a result only through a concurrent analysis of two paralleled language resources. If the analysis of these Lithuanian lexemes was conducted using a monolingual resource, then the Lith. form *gal* ‘może’ might be considered a neutral exponent of hypothetical modality even-handedly with the morphological exponent, i.e. *modus relativus* forms.

In Table 2 I present statistic data which describe individual groups of exponents of hypothetical modality on the basis of the *Polish-Lithuanian parallel corpus Clarin-PL* (Roszko, D. & Roszko, R., 2016) and the *Experimental Polish-Lithuanian corpus*. I have divided the data into two groups. The first group contains a number of unique lexemes in individual groups. The second includes the total number of noted appearances of these forms, also divided into groups. The total volume of both of the corpora, used for defining the particular degrees of probability, amounts to 8 million words.

Analysis of the data contained in Table 2 (columns 2 and 3) confirms low exponent representation in groups H1–H3. Then, an increase in the exponent number in group H4 can be observed. As can be clearly seen, the climax is reached in group H5, and then the number of unique exponents drops (noticeably for the Polish language) in group H6, although it remains high. Analysis of the data contained in Table 2 (columns 3 and 4) shows that in groups H1–H5 the frequency of the exponent use rises in tandem with the rise in probability degree. However, for group H6 we can observe a decrease in frequency compared to group H5 (for both languages). Nevertheless, the lexemes of groups H5 and H6 are the most frequently used lexical exponents by both groups of languages users.

²The issue of the lexical exponents of hypothetical modality in both languages was raised, amongst others, in Roszko, D. and Roszko, R. (2012) and later developed in Roszko, D. (2015), also with regard to the interrelatedness of the Lithuanian dialect of Puńsk, the Polish language and the Lithuanian language.

Table 2: Selected lexical exponents of hypothetical modality in Polish and Lithuanian

1.	2.	3.	4.	5.
Group	Number of exponents		Total no. of appearances in EKorpPL-LT	
	Polish	Lithuanian	Polish	Lithuanian
H1	1	1	49	76
H2	5	1	33	42
H3	2	2	266	224
H4	6	10	519	478
H5	49	31	4613	3017
H6	25	27	2002	1561
Total:	88	72	7482	5398

A lower total frequency of the Lithuanian exponents of hypothetical modality certainly draws attention. However, it is not a result of hypothetical contents being less frequently exposed in the Lithuanian language (compared to Polish). Instead, it is a result of the common use of morphological exponents in the Lithuanian language.

Similar conclusions can be drawn on the basis of the total number of unique lexemes in both languages. Namely, the Polish language with a total of 88 lexemes seems to be more diverse than the Lithuanian language with a total of 72 unique lexemes.

Not by chance do I mention in the paragraph above that the Polish language seems to be more diverse than the Lithuanian language, because a detailed analysis of the corpora resources presents new facts, which make us look at both languages from a different perspective.

5 Analysis of the hypothetical modality exponents, included in group H5, in Polish and Lithuanian

The choice of group H5 for a detailed parallel Polish-Lithuanian analysis of lexical exponents in both languages is not accidental. Firstly, in group H5 we can observe a large number of lexical units: 49 Polish lexemes in group H5 and 39 lexemes in the remaining five groups (H1, H2, H3, H4 and H6). Secondly, there is a high frequency of lexemes in group H5 in Polish: 4613 lexeme uses in this group against a total of 2869 of all the lexemes included in the remaining groups. There is an analogous situation with the Lithuanian language: 3017 against 2381. Thirdly, there is a significant difference in the frequency of use between the Polish and Lithuanian lexemes of the very same group (H5): 4613 (Polish exponents) against 3017 (Lithuanian lexemes). Numerical data is contained in Table 3.

The group H5 exponents express a high degree of probability, much higher than value. On the basis of excerption of the two aforementioned corpora, the following exponents can be put into group H5 (after each exponent we see the number of uses in the two corpora):

Polish: *może* — 2261, *oczywiście* — 679, *być może* — 161, *możliwe* — 154, *z pewnością* — 147, *zapewne* — 126, *oczywista* — 126, *naturalnie* — 119, *zdaje się*³ — 105, *przypuszczam* — 71, *wyraźnie* — 69, *sądzę* — 56, *może nawet* — 49, *pewnie* — 39, *rzeczywiście* — 37, *kto wie* — 35, *widocznie* — 33,

³This Polish exponent, it seems, appears in contexts marked hypothetically and imperceptibly (in the academic literature the term *evidential/evidentiality* is also used). Sometimes an insufficient context does not allow us to unambiguously conclude whether this Polish exponent is an exponent of hypotheticality or imperceptibility. On the basis of paralleled Polish-Lithuanian resources the Polish seems to be able to be disambiguated in relation to unambiguous Lithuanian forms.

okazuje/okazało się — 29, *mam (takie) wrażenie* — 28, *uwazam(, że)* — 28, *widać* — 21, *może to (właśnie)* — 21, *moim zdaniem* — 20, *czy to aby nie* — 19, *(chyba) jasne, że* — 19, *należy sądzić* — 14, *zdawał (-o, -a, -i, -y) (mi) się* — 13, *rozumie się* — 13, *bodaj*⁴ — 12;

Lithuanian: *gal* — 1071, *žinoma* — 505, *atrodo* — 267, *tikrai* — 252, *galimas daiktas* — 169, *savaime aišku* — 128, *galbūt* — 109, *be abejo* — 79, *berods* — 77, *man atrodo* — 69, *man rodos* — 56, *kas žino* — 42, *rodos* — 42, *matyt* — 36, *man regis* — 35, *žinia* — 28, *mano nuomone* — 23, *savaime suprantama* — 19, *aiškiai matyti* — 14, *jaučiu* — 14, *paaiškėjo (-a)* — 14, *manding* — 12.

The exponents of group H5 below are relatively seldom registered: Polish: *czy co(?)*, *a zatem, istotnie, wygląda na, całkiem prawdopodobne, zrozumiałe, według moich przypuszczeń, według mnie, coś mi mówi, przeżywam, czuję, mieć/odnosić wrażenie, wydaje się, jakby, więc nie, czy to ... czy to, wyraźnie widać, tak myślę*;

Lithuanian: *taip manau, lyg, iš tiesų, rodėsi, visai aišku, aiškus daiktas, pasirodo, pagal mane*.

Table 3: Lexical exponents of hypotheticality in Polish and Lithuanian in a statistical approach. Group H5 vs remaining groups (on the basis of the *Polish-Lithuanian parallel corpus Clarin-PL* (Roszko, D. & Roszko, R., 2016) and *Experimental Polish-Lithuanian Corpus*)

1.	2.	3.	4.	5.
Group	No. of unique exponents		Total no. of appearances in EKorpPL-LT	
	Polish	Lithuanian	Polish	Lithuanian
H1, H2, H3, H4, H6	39	41	2869	2381
H5	49	31	4613	3017
Total:	88	72	7482	5398

5.1 Selected examples of the use of the lexical exponents of group H5 in Polish and Lithuanian

- PL** — *Może jutro* — powiedział chłopiec, odchodząc.

LT — *Galbūt rytoj*, — sumurma atsitraukdamas.
- PL** *Ale nikt go nie słuchał, może tylko dżdżownica Zygmus, który zwinąwszy się w kłębek drzemał w pudełku zapalek.*

LT *Bet niekas jo nesiklausė, gal tik sliėkas Zigmūtis, kuris susirangęs į kamuoliuką ramiai snaudė degtukų dėžutėje.*
- PL** *Może to leży w naturze ludzkiej, że najmniej litości mamy dla tych, którzy przez pokorę, słabość lub obojętność najwięcej znieść gotowi.*

LT *Galimas daiktas, pati žmogaus prigimtis yra linkusi krauti viską tam, kas viską neša — ar iš tikro nusizėminimo, ar iš silpnumo, ar iš abejingumo.*
- PL** *Może to złodziej lub wspólnik złodziejski, który żyje jak żebrak, udając głupotę i niedołęstwo, aby tym bezpieczniej oddać się swemu rzemiosłu. Student wyprostował się przy tych słowach, lecz po chwili przyłożył znów oko do dziurki od klucza.*

LT — *Ar jis vagis, ar vogtų daiktų slėpikas, kuris, norėdamas išvengti įtarinėjimų, apsimeta kvailiu, iškaršėliu ir gyvena lyg elgeta?" — tarė sau Eugenijus, atsitiesdamas valandėlę. Studentas vėl prikišo akį prie rakto skylutės.*

⁴Another case of an ambiguous Polish exponent. Context and comparison to unambiguous Lithuanian exponents allows the Polish exponent to be disambiguated.

5. **PL** I wie pan co, *uwazam, że* tu jest zupełnie niezłe.

LT Ir žinote, *man atrodo, kad* šičia visai neblogai.

The examples above have been selected in order to present the larger diversity of exponents in the Lithuanian language. In examples 1–3, the Polish lexeme *może* has a corresponding lexeme in the Lithuanian language: in the first sentence it is *galbūt*, in the second — *gal*, and in the third it is *galimas daiktas*. However, in the Lithuanian sentence 4 there is no lexeme directly corresponding to the Polish *może*. Nonetheless, we can note a form of interrogative sentence, which in an equivalent way reflects the addresser’s uncertainty in relation to the contents conveyed. In sentence 5 we can see the Polish and Lithuanian lexical exponents which cause the syntax to become subordinate in nature: Pol. *uwazam, że* (I think that) ... and Lith. *man atrodo, kad* ...

5.2 Statistical analysis of the hypothetical modality lexical exponents in Polish and Lithuanian, included in group H5

Firstly, it is worth ranking the above-mentioned exponents in descending order of their frequency in the corpora. It is certainly surprising to find the Polish exponent *może* in the top position, with a disproportionately high number of appearances against other exponents, amounting to 2261. As a reminder, in this group 49 unique exponents have been noted, with a total of appearances amounting to 4613.

Table 4: Hypothetical modality lexical exponents in Polish and Lithuanian. (group H5) with a frequency above 10 appearances in the *Polish-Lithuanian parallel corpus Clarin-PL* (Roszko, D. & Roszko, R., 2016) and *Experimental Polish-Lithuanian corpus*

1.	2.	3.	4.
No.	Language	Exponent	No. of noted uses
1.	PL	może	2261
2.	LT	gal	1071
3.	PL	oczywiście	679
4.	LT	žinoma	505
5.	LT	atrodo	267
6.	LT	tikrai	252
7.	LT	galimas daiktas	169
8.	PL	być może	161
9.	PL	możliwe	154
10.	PL	z pewnością	147
11.	LT	savaime aišku	128
12.	PL	zapewne	126
13.	PL	oczywista	126
14.	PL	naturalnie	119
15.	LT	galbūt	109
16.	PL	zdaje się	105
17.	LT	be abejo	79
18.	LT	berods	77

1.	2.	3.	4.
19.	PL	przypuszczam	71
20.	LT	man atrodo	69
21.	PL	wyraźnie	69
22.	LT	man rodos	56
23.	PL	sądzę	56
24.	PL	może nawet	49
25.	LT	kas žino	42
26.	LT	rodos	42
27.	PL	pewnie	39
28.	PL	rzeczywiście	37
29.	LT	matyt	36
30.	LT	man regis	35
31.	PL	kto wie	35
32.	PL	widocznie	33
33.	PL	okazuje/okazało się	29
34.	LT	žinia	28
35.	PL	mam (takie) wrażenie	28
36.	PL	uważam(, że)	28
37.	LT	mano nuomone	23
38.	PL	widać	21
39.	PL	może to (właśnie)	21
40.	PL	moim zdaniem	20
41.	LT	savaime suprantama	19
42.	PL	czy to aby nie	19
43.	PL	(chyba) jasne, że	19
44.	LT	aiškiai matyti	14
45.	LT	jaučiu	14
46.	LT	paaikškėjo (-a)	14
47.	PL	należy sądzić	14
48.	PL	zdawał (-o, -a, -i, -y) (mi) się	13
49.	PL	rozumie się	13
50.	LT	manding	12
51.	PL	bodaj	12

While analysing the numerical data contained in Table 4 it is necessary to remember that the total number of uses of any hypothetical modality exponent in group H5 in the Polish language amounts to 4613, whilst in the Lithuanian language it is substantially lower and amounts to only 3017 (it is noted that for 65% of the Polish exponents appearances an equivalent Lithuanian lexeme is used). Despite this, it can be observed that with the first seven exponents, 5 of them are Lithuanian forms (*gal*, *žinoma*, *atrodo*, *tikrai*, *galimas daiktas*) and only 2 are Polish (*może*,

oczywiście). In the Polish language the most dominant exponent is *może*, which appears 3.3 times more often than the second most dominant Polish exponent, and 14 times more often than the third most dominant — *być może*. As regards the Lithuanian language, these proportions are significantly lower, i.e. the most dominant exponent *gal* is noted 2.1 times more often than the second most dominant *žinoma* and 4 times more often than the third most dominant Lithuanian exponent *atrodo*.

The comparison of the total number of appearances of Polish and Lithuanian exponents, although very informative, does not fully reflect the facts. This is because the disproportion of the total number of uses of these lexemes in both languages is quite distinct. (comp. 4613 appearances of Polish lexemes to 3017 appearances of Lithuanian lexemes). That is why it has been concluded that presenting the percentage participation will better illustrate this Polish-Lithuanian disproportion in using the individual lexemes. The percentage values have been calculated on the basis of uses of a given lexeme against the total number of uses of all the lexemes (in group H5) in a given language. The results are presented in Table 5⁵.

Table 5: Percentage participation of the hypothetical modality lexical exponents, presented in table 4, in Polish and Lithuanian (group H5)

1.	2.	3.	4.	5.
No.	Language	Exponent	Percentage	No. of noted uses
1.	PL	<i>może</i>	49,01%	2261
2.	LT	<i>gal</i>	35,50%	1071
3.	LT	<i>žinoma</i>	16,74%	505
4.	PL	<i>oczywiście</i>	14,72%	679
5.	LT	<i>atrodo</i>	8,85%	267
6.	LT	<i>tikrai</i>	8,35%	252
7.	LT	<i>galimas daiktas</i>	5,60%	169
8.	LT	<i>savaime aišku</i>	4,24%	128
9.	LT	<i>galbūt</i>	3,61%	109
10.	PL	<i>być może</i>	3,49%	161
11.	PL	<i>możliwe</i>	3,34%	154
12.	PL	<i>z pewnością</i>	3,19%	147
13.	PL	<i>zapewne</i>	2,73%	126
14.	PL	<i>oczywista</i>	2,73%	126
15.	LT	<i>be abejo</i>	2,62%	79
16.	PL	<i>naturalnie</i>	2,58%	119
17.	LT	<i>berods</i>	2,55%	77
18.	LT	<i>man atrodo</i>	2,29%	69
19.	PL	<i>zdaje się</i>	2,28%	105
20.	LT	<i>man rodos</i>	1,86%	56

⁵The percentage participation of the Polish exponents is calculated in reference to a total of 4613, which is the total number of uses of all Polish exponents in group H5, whilst the percentage participation of the Lithuanian exponents is calculated in reference to a total of 3017, i.e. the total number of uses of all Lithuanian exponents in group H5.

1.	2.	3.	4.	5.
21.	PL	przypuszczam	1,54%	71
22.	PL	wyraźnie	1,50%	69
23.	LT	kas žino	1,39%	42
24.	LT	rodos	1,39%	42
25.	PL	sądzę	1,21%	56
26.	LT	matyt	1,19%	36
27.	LT	man regis	1,16%	35
28.	PL	może nawet	1,06%	49
29.	LT	žinia	0,93%	28
30.	PL	pewnie	0,85%	39
31.	PL	rzeczywiście	0,80%	37
32.	LT	mano nuomone	0,76%	23
33.	PL	kto wie	0,76%	35
34.	PL	widocznie	0,72%	33
35.	LT	savaime suprantama	0,63%	19
36.	PL	okazuje/okazało si	0,63%	29
37.	PL	mam (takie) wrażenie	0,61%	28
38.	PL	uważam(, że)	0,61%	28
39.	LT	aiškiai matyti	0,46%	14
40.	LT	jaučiu	0,46%	14
41.	LT	paaiškéjo (-a)	0,46%	14
42.	PL	widać	0,46%	21
43.	PL	może to (właśnie)	0,46%	21
44.	PL	moim zdaniem	0,43%	20
45.	PL	czy to aby nie	0,41%	19
46.	PL	(chyba) jasne, że	0,41%	19
47.	LT	manding	0,40%	12
48.	PL	należy sądzić	0,30%	14
49.	PL	zdawał (-o, -a, -i, -y) (mi) się	0,28%	13
50.	PL	rozumie się	0,28%	13
51.	PL	bodaj	0,26%	12

The analysis of the usage of each exponent in comparison to the usage of all the group H5 exponents in a given language changes the order of the most frequent lexemes. The Polish exponent *może* still remains in the highest position with its 49.01% participation in the Polish H5 group. The second highest position goes to the Lithuanian *gal* with its 35.50% participation in the Lithuanian H5 group. The third most frequent exponent is the Lithuanian *žinoma* with its participation of 16.74%, which replaces the Polish *oczywiście*, noted in this position in Table 4. The Polish *oczywiście* occupies position 4 with its participation of 14.72%. Positions 5, 6 and 7, like in Table 4, are taken by the Lithuanian exponents *atrodo*, *tikrai*, *galimas daiktas*, which respectively have participations of 8.85%, 8.35% and 5.60%. A significant fact is the drop of the Polish *być*

może to position 10 (the participation of this exponent is 3.49%). It has been overtaken by two successive Lithuanian exponents, *savaime aišku* and *galbūt*, with respective participations of 5.60% and 4.24%. In summary, within the top ten most frequently used exponents we can find 3 Polish forms and 7 Lithuanian forms.

On the basis of the total number of unique exponents noted in both languages, we can define the average participation of the exponent in its group. As regards the Polish language, the average participation is 2.04%, and for Lithuanian it is 3.23%. When comparing these values against the participation of the individual exponents, we can conclude that in the Polish language the considerable domination of one exponent can be observed: (*może* — 49.1%). This allows us to draw the conclusion that practically every other exponent of hypothetical modality used for expressing the fifth degree of probability (H5) is the Polish *może*. When we take into account the respective participations of the remaining Polish lexemes and at the same time look at the facts registered in the Lithuanian context,⁶ it is necessary to conclude that the diversity of Polish exponents is low. Of course, the total number of Polish lexemes is considerably higher than their Lithuanian counterparts, but their distribution is apparently uneven.

6 Conclusions

Using Polish and Lithuanian as an example, it can be concluded that hypothetical modality may be expressed by lexical (including simple and complex expressions) and morphological means. In the Polish language the dominant exponents are lexemes. In the Lithuanian language there is a balance between the use of lexical and morphological exponents. This is why, in a numerical approach to the use of lexical exponents in both languages, a more frequent hypothetical marking of a sentence can be observed as a result of using the lexeme in the Polish context (comp. group H5 described herein: 4617 uses of Polish lexemes corresponding to 3017 uses of the Lithuanian lexical counterparts). Moreover, a larger number of unique lexemes in the Polish language is observed (for group H5 it amounts to 49 unique Polish lexemes and 31 Lithuanian counterparts).

However, a detailed analysis of the corpora resources reveals a disproportion between the total number of the lexemes and their frequency. It appears that in the Polish language only one lexeme (Pol. *może*) is noted in every second hypothetical sentence of group H5. Almost all the other lexemes are used much less frequently. In the Lithuanian language we do not observe an explicit domination of one of the exponents. Most of the Lithuanian hypothetical exponents have a higher frequency, which is followed — compared to the Polish language — by a lexeme diversity in the hypothetical function.

References

- Ambrazas, V. (Ed.). (1997). *Lithuanian grammar*. Vilnius: Baltos lankos.
- Dimitrova, L. & Koseska, V. (2014). *Semantic properties of selected universal language categories in digital bilingual resources*. Sofia: Demetra Ltd Publishers.
- Koseska-Toszewa, V., Korytkowska, M., & Roszko, R. (2007). *Polsko-bułgarska gramatyka konfrontatywna*. Warszawa: Wydawnictwo Akademickie „Dialog”.
- Koseska-Toszewa, V., Korytkowska, M., & Roszko, R. (2009). Contrastive studies and semantic interlanguage. *Cognitive Studies / Études cognitives*, (9), 33–52.
- Koseska-Toszewa, V., Maldžieva, V., & Penčev, J. (1995). *Gramatyka konfrontatywna bułgarsko-polska* (V. Koseska-Toszewa & J. Penčev, Eds.) (Vol. 6-1, *Modalność. Teoretyczne problemy opisu*). Warszawa: Slawistyczny Ośrodek Wydawniczy.
- Maldžieva, V. (2003). *Gramatyka konfrontatywna bułgarsko-polska* (V. Koseska-Toszewa & J. Penčev, Eds.) (Vol. 6-3, *Modalność: Hipotetyczność, irrealność, optytywność i imperatywność, warunkowość*). Warszawa: Slawistyczny Ośrodek Wydawniczy.

⁶The Author has in mind a number of Lithuanian morphological forms which correspond to the Polish lexical exponents.

- Roszko, D. (2006). *Funkcjonalne odpowiedniki litewskiego perfectum w litewskiej gwarze puńskiej i w języku polskim*. Warszawa: Slawistyczny Ośrodek Wydawniczy, pp. 177.
- Roszko, D. (2011). Leksykalne wykładniki hipotetyczności w litewskiej gwarze puńskiej z odniesieniami do języków polskiego i litewskiego, *Acta Baltico-Slavica*, 35, 159–170. DOI: <http://dx.doi.org/10.11649/abs.2011.012>
- Roszko, D. (2015). *Zagadnienia kwantyfikacyjne i modalne w litewskiej gwarze puńskiej (Na tle literackich języków polskiego i litewskiego)* [Quantification and modal issues in the Lithuanian local dialect of Puńsk (with reference to Polish and Lithuanian literary languages)]. Warszawa: Slawistyczny Ośrodek Wydawniczy. 342
- Roszko, D., & Roszko, R. (2012). Значения гипотетичности в литовском, польском языках и в литовском говоре окрестностей Пунска в Польше. *Baltistica*, 47(1). DOI: <http://dx.doi.org/10.15388/baltistica.47.1.2133>
- Roszko, D., & Roszko, R. (2016). *Polish-Lithuanian Parallel Corpus*, CLARIN-PL digital repository. <http://hdl.handle.net/11321/309>.

References (Transliteration)

- Ambrazas, V. (Ed.). (1997). *Lithuanian grammar*. Vilnius: Baltos lankos.
- Dimitrova, L., & Koseska, V. (2014). *Semantic properties of selected universal language categories in digital bilingual resources*. Sofia: Demetra Ltd Publishers.
- Koseska-Toszewa, V., Korytkowska, M., & Roszko, R. (2007). *Polsko-bułgarska gramatyka konfrontatywna*. Warszawa: Wydawnictwo Akademickie „Dialog”.
- Koseska-Toszewa, V., Korytkowska, M., & Roszko, R. (2009). Contrastive studies and semantic interlanguage. *Cognitive Studies / Études cognitives*, 9, 33–52.
- Koseska-Toszewa, V., Maldžieva, V., & Penčev, J. (1995). *Gramatyka konfrontatywna bułgarsko-polska* (V. Koseska-Toszewa & J. Penčev, Eds.) (Vol. 6-1, *Modalność. Teoretyczne problemy opisu*). Warszawa: Slawistyczny Ośrodek Wydawniczy.
- Maldžieva, V. (2003). *Gramatyka konfrontatywna bułgarsko-polska* (V. Koseska-Toszewa & J. Penčev, Eds.) (Vol. 6-3, *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. (2011). Leksykalne wykładniki hipotetyczności w litewskiej gwarze puńskiej z odniesieniami do języków polskiego i litewskiego, *Acta Baltico-Slavica*, 35, 159–170. DOI: <http://dx.doi.org/10.11649/abs.2011.012>
- Roszko, D. (2015). *Zagadnienia kwantyfikacyjne i modalne w litewskiej gwarze puńskiej (Na tle literackich języków polskiego i litewskiego)* [Quantification and modal issues in the Lithuanian local dialect of Puńsk (with reference to Polish and Lithuanian literary languages)]. Warszawa: Slawistyczny Ośrodek Wydawniczy.
- Roszko, D., & Roszko, R. (2012). Znacheniiia gipotetichnosti v litovskom, pol'skom iazykakh i v litovskom govore okrestnostei Punska v Pol'she. *Baltistica*, 47(1). DOI: <http://dx.doi.org/10.15388/baltistica.47.1.2133>
- Roszko, D., & Roszko, R. (2016). *Polish-Lithuanian Parallel Corpus*, CLARIN-PL digital repository. <http://hdl.handle.net/11321/309>

Acknowledgment

This work was financed as a part of the investment in the CLARIN-PL research infrastructure (www.clarin-pl.eu) funded by the Polish Ministry of Science and Higher Education.

The author declares that she has no competing interests.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (<http://creativecommons.org/licenses/by/3.0/pl/>), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.

© The Author 2016

Publisher: Institute of Slavic Studies, Polish Academy of Sciences, University of Silesia & The Slavic Foundation