SPECIFIC PHRASEMES WITH ETHNONYMS AND THEIR STUDY BY CORPUS ANALYSIS

Iveta Dinžíková

Comenius University

E-mail: idinzikova@hotmail.sk

1. Introduction

The focus of our paper is phrasemes comprising an ethnonym in the source language (French) as well as the target language (Slovak). Firstly, we will classify them according to the type of equivalence of phrasemes (total equivalent, partial equivalent and phrasemes without equivalent), then all the researched phrasemes from the point of view of general and professional language as general and professional phrasemes. With the help of the quantitative methods of corpus linguistics such as *relative frequency* and the *logDice* association measure, we will try to determine the extent of their presentation and the extent of their specificity in four monolingual corpora, each of them comprising from approx. 130 million to approx. 10 billion words. Based on the corpus analysis, it is possible to propose which of the phrasemes could be ranked high when searching for the entries of their individual components or a phraseme as a whole, and thus contribute to supporting the creation of current French and Slovak lexicography as well as phraseography.

The aim of the article is a corpus analysis of the specific type of phrasemes, i.e. ethnonyms in French and Slovak. ***Research objectives*** are set to find the frequency of phrasemes in various types of texts and the extent of their specificity in each of the corpora.

2. Research methods

Due to the heterogeneity of data, the imbalance of sources and insufficient representativeness, we use hybrid methods and sources such as the analysis of static and dynamic data. The analysis of static data is based on the actual linguistic and lexicographic description of researched phenomena; e.g. studies, dictionaries. The analysis of dynamic data researches into the real functioning of language phenomena in the corpus, which means that we rely on corpus linguistics (Hunston, 2006, p. 244)[1], and we analyse the collected data in four monolingual text corpora. There are two corpora in French: I) the corpus *frTenTen12* comprising approx. 10 billion words – on the webpage https://the.sketchengine.co.uk/ and II) the corpus *Emolex* comprising approx. 137 million words – on the webpage http://emolex.u-grenoble3.fr/emoBase/), and two corpora in Slovak: I) the corpus *skTenTen11* comprising approx. 540 million words – on the webpage https://the.sketchengine.co.uk/ and II) the Slovak National Corpus named *prim-7.0-public-all* comprising approx. 1 billion words – on the page http://korpus.juls.savba.sk/. Subsequently, we have used the quantitative methods of corpus linguistics such as the *relative frequency* and the *logDice* association measure.

2.1. Comparison of corpora

Each of the four above-mentioned corpora has some **quantitative** (the size of corpus: *frTenTen12*, *skTenTen11*, *prim-7.0-public-all*; the diversity of language styles: *frTenTen12*, *skTenTen11*, *prim-7.0-public-all*) **and/or qualitative advantages** (orthographic cleaning up of texts: *prim-7.0-public-all*, *Emolex*; used functions: *frTenTen12*, *skTenTen11*, *prim-7.0-public-all*; precise stylistic determination of texts: *Emolex*, *prim-7.0-public-all*) compared to the other three corpora, and thanks to this heterogeneity of corpora, the analysis of collected phrasemes can be at a better level of excellence.

2.2. Quantitative methods of corpus linguistics

We have used two quantitative methods for the analysis of individual phrasemes: the *relative*

---

[1] Essentially, corpus linguistics means the following: looking at naturally occurring language; looking at relatively large amounts of such language; observing relative frequencies, either in raw form or mediated through statistical operations; observing patterns of association, either between a feature and a text type or between groups of words.

frequency and the *logDice* association measure. The *relative frequency* enables the mutual comparison of individual corpora. On the contrary, it is not possible to compare the values of the *logDice* association measure between several corpora but only inside one corpus.

2.3. Frequency[2]

Frequency is considered one of the fundamental indicators to assess the importance of a researched phenomenon and it points out the frequency of the occurrence of phenomena in the corpus. We differentiate between *absolute frequency* and *relative frequency*. The *absolute frequency* represents the number of the occurrence of phenomena in the corpus as a whole (*Tables No. 1 and No. 2*), which shows how many times a phraseme occurs in a corpus; e.g. the phraseme *roulette russe* in the corpus *frTenTen12* (occurred 4,499 times). The *relative frequency* represents the *absolute frequency* averaged over the corpus as a whole (in tokens), i.e. 1 million tokens, which means the *frequency of a phenomenon per 1 million occurrences* (*Tables No. 1 and No. 2*). It is possible to calculate according to the following formula:

$$REL = \frac{ABS}{N} \times 1000000.$$

Based on the comparison of *relative frequency* values we can find out, which of the researched phrasemes occur in the corpus more frequently and thus those that are statistically more significant compared to the other collected phrasemes.

2.4. *LogDice*[3] association measure

Currently, there are many association measures such as *logDice*, *MI-Score*, *MI3*, *log-likelihood*, *T-score* (all the association measures can be used for searching a collocation with the *Collocations* function in the corpora *frTenTen12*, *skTenTen12* and *prim-7.0-public-all,* and in all the three corpora), the *logDice* association measure is pre-set, and we decided to use it when analysing the collected phrasemes (*Tables No. 3 and No. 4*). This association measure shows values of statistical specificity for each phraseme. The highest their value, the highest their rank.

The *logDice* measure depends on the frequency of the words *x* and *y* and on the frequency of the bigram *xy*, i.e. N (the size of corpus in tokens) is not present for the calculation.

$$logDice(xy) = 14 + log_2 \frac{2f(xy)}{f(x) + f(y)}$$

The value of *logDice* can be from 14 to minus infinity. According to P. Rychlý (2006, p. 9), "comparing two scores, plus 1 point means twice as often collocation, plus 7 points means roughly 100 times frequent collocation."

Based on the comparison of *logDice* values inside one corpus we can find out which of the researched phrasemes have the highest *logDice* value and thus those that are statistically more significant compared to the other collected phrasemes.

3. Phrasemes in the French and Slovak linguistics

3.1. General phraseology

We can see diversity in French linguistics with respect to the name of the basic unit in phraseology as well as to its definition. More frequent are terms such as *locution figée*, *locution figurée*, *expression figurée* or *expression figée*, less frequent – the terms *unité phraséologique*, *phrasème* or *phraséologisme*. There also exist many definitions characterising the basic unit of phraseology, but we prefer the approach of L. Rosenbaum Franková (2010, p. 27), who based on the definitions of various francophone phraseologists such as G. Gross (1996), G. Gréciano (1993, 1997), I. González-Rey (2002), S. Mejri (2000) who compiled the defining characteristics of general phrasemes. There are fundamental and additional characteristics. The fundamental defining characteristics include: polylexicality, stability and semantic opacity, and additional defining characteristics: figurativeness, iconicity, reproducibility, anomaly, expressiveness and untranslatability.

In Slovak phraseology, two terms dominate to indicate the fundamental unit of phraseology, i.e. *phraseme* and *phraseologism* and its definition is relatively stable. According to J. Mlacek, P. Ďurčo et al (1995, p. 28), phraseme and phraseologism represent a "specific type of stable multi-word expressions typical of expressiveness and figurativeness, the parts of which have been fully or partially de-semantized." However, J. Mlacek (2007, p. 21) adds that "individual characteristics or the groups of the characteristics of phrasemes should be considered in relation to other principles that in phraseology and its functioning are mainly determined in new literature such as the principle of language economy, the principle of motive, the principle of syncretism, the principle of the secondariness of phraseological denomination, the principle of functional separation, the principle of transposition, the principle of analogy and idiomaticity, the principle of shaping

---

[2] https://www.sketchengine.co.uk/documentation/frequency-per-million/

[3] https://www.sketchengine.co.uk/documentation/statistics-used-in-sketch-engine/

and the principle of emancipation." At the same time, J. Mlacek (2010, p. 244) admits that the issue of the definition of phrasemes is still open and unsolved.

3.2. Professional phraseology

Professional phraseology researches into professional phrasemes. In French phraseology, the basic unit of professional phraseology is named *phrasème spécialisé* (specialised phraseme), *phrasème de spécialité* (professional phraseme), *phraséologisme spécialisé* (professional phraseologism), *terme phraséologique* (phraseological term), *phraséoterme* (phrase-term). The term *odborná frazéma* (professional phraseme) dominates Slovak phraseology.

Based on the definition of a professional phraseme by francophone phraseologists (G. Grecia-no, L. Gautier) and Slovak phraseologists (P. Ďurčo, J. Mlacek, Ľ. Mešková, M. Olejárová), the author L. Rosenbaum Franková (2010, p. 59) suggested the common characteristics of professional phrasemes, and this author characterised it as the "phraseological unit regularly occurring in professional texts with the typical characteristics of phraseme such as figurativeness or expressiveness, but at the same time taking the denominative function of the term". Thus, she included the presence of the term, the figurativeness or imagery of the meaning of at least one part of the phraseme among the fundamental defining characteristics of the professional phraseme. However, contrary to general phrasemes (J. Mlacek, P. Ďurčo et al, 1995, p. 95), professional phrasemes "neither have such a weakened reference part of the meaning nor have they an expressive pragmatic (specifically expressive and evaluating) component of the meaning". Some professional phrasemes might seem just terms, as if the boundary was not clear enough, but at least one phraseme's component of the professional phraseme should have a figurative or metaphoric meaning.

With respect to the selection of general and professional phrasemes, our paper relies on the approach of defining characteristics according to L. Rosenbaum Franková.

3.3. So-called mixed, general-professional phrasemes

Furthermore, there is another type of phrasemes that can be added to general and professional phrasemes, the so-called mixed phraseme. It is a phraseme with at least two meanings, i.e. the meaning of the general phraseme, but also possibly the meaning of the professional phraseme. For example, the phraseme *roulette russe* [*Russian roulette* literally also *Russian roulette*] within the meaning of the general phraseme means *a dangerous and risky solution,* and within the meaning of the professional phraseme represents a type of *suicidal game with a revolver and bullets*.

3.4. Phrasemes and ethnonym

We have analysed a specific type of phrasemes comprising an ethnonym, which is a noun or adjective identifying a nation, ethnicity, tribe or an inhabitant in some place. For example, the ethnonym *Polonais* [Pole] – *Poliak* [Pole] is a phraseme *soûl/saoul comme un Polonais* [*as drunk as a lord*, literally *drunk as a Pole*] – *opitý ako Dán* [literally *drunk as a Dane*].

For the purpose of the corpus analysis we have excerpted individual phrasemes comprising an ethnonym from statistical data, i.e. from various dictionaries and studies (we have excluded the professional phrasemes representing some kind of dish or an animal species such as *sauce tartare* [*Tartar sauce*, literally *Tartar sauce*] – *tatárska omáčka* [literally *Tartar sauce*] or *berger allemand* [*German Shepherd*, literally *German Shepherd*] – *nemecký ovčiak* [literally *German Shepherd*] because they are so numerous that they deserve a more detailed analysis in a separate paper.

4. Equivalence of phrasemes

Equivalence represents one of the fundamental issues for contrastive linguistics and translation. The equivalence of phrasemes can be considered from the qualitative or quantitative points of view (P. Ďurčo 2012, p. 91). We have decided to analyse its qualitative aspect and we distinguish three types of equivalents[4] M. Ţaran Andreici (2016), C. Teodorovici – R. Pisot Díaz (2008), J. Brumme (2008), P. Romero Ganuza – J. J. Baptista Rodríguez – D. García Padrón (2008): I. Total phraseological equivalence, II. Partial phraseological equivalence (where we distinguish partial equivalent phrasemes with the same ethnonym from partial equivalent phrasemes without ethnonym), III. Phrasemes without equivalent, but our attention is on the first two types.

I. Total equivalent phrasemes that comprise the same ethnonym in French and in Slovak; the compared phrasemes have the same component structure and the same motive (at a formal, lexical, semantic and figurative level). There are nineteen occurrences of this:

---

[4] All the collected phrasemes have been the subject of structural-semantic-typological analysis based on which we have classified them according to the level of equivalence into three types. We state the English equivalent to the phrasemes and the verbatim translation into English in square brackets to better point out lexical similarities as well as the differences between the source language and culture as well as the target language and culture.

1. *roulette russe* [*Russian roulette*, literally *Russian roulette*] – *ruská ruleta*

2. *filer à l'anglaise* [*to take French leave*, literally *to leave in an English way*] – *zmiznúť po anglicky Fin du meeting. Dès la fin de la Marseillaise, le président de l'Assemblée nationale, Bernard Accoyer file à l'anglaise. Enthousiaste sur l'idée de réduire le nombre de députés, «une très bonne idée", il bredouille sur la proportionnelle: «euh, c'est à voir avec le reste.»* Meaning "Bernard Accoyer is leaving without saying goodbye". frTenTen12: Web domain: liberation.fr, Crawl date: 2012-02-25, doc.url http://www.liberation.fr/politiques/01012390946-legislatives-sarkozy-propose-d-introduire-a-la-marge-la-proportionnelle

3. *comme une horloge suisse/ une montre suisse* [*like clockwork*, literally *like a Swiss watch*] – *ako švajčiarske hodinky*
   Son rituel avant chaque plongeon est lui aussi réglé comme une horloge suisse. Meaning "indeed accurately, evenly".
   Emolex: published => Le Monde; date => 13/08/2008

4. *douche écossaise* [1. *series of ups and downs*, 2. *hot and cold shower*, literally *Scottish shower*] – *1. studená sprcha* [*series of ups and downs*, literally *cold shower*], *2. škótska sprcha* [*hot and cold shower*, literally *Scottish shower*]

5. *froid sibérien* [*freezing cold*, literally *Siberian cold*] – *sibírska zima*

6. *grippe espagnole* [*Spanish flu*, literally *Spanish flu*] – *španielska chrípka*
   Dans un éditorial, Nature rappelle que les taux de mortalité de la grippe espagnole de 1918 ont été vingt-cinq fois supérieurs à ceux des autres pandémies mondiales. Meaning the "worldwide flu pandemic during 1918–1920."
   frTenTen12: Web domain: sciences-technologies.lefigaro.fr, Crawl date: 2012-03-05, doc.url: http://sciences-technologies.lefigaro.fr/homme-revue-nature.html

7. *guitare espagnole* [*Spanish guitar*, literally *Spanish guitar*] – *španielska gitara*
   Na dekách posedávali zaľúbené páry. Múrmi zrúcaniny sa ozývali tóny španielskej gitary sprevádzané spevom mladých. Meaning "a classical guitar, which is also named the Spanish guitar".
   prim-7.0-public-all: doc.bibl: SME. Daily. Bratislava: Petit Press 12.05.2009, doc.id: 2010-10-08

8. *bain turc* [*Turkish bath*, literally *Turkish bath*] – *turecký kúpeľ*

9. *massage thaïlandais* [*Thai massage*, literally *Thai massage*] – *thajská masáž*

10. *gazon anglais* [*well-kept lawn*, literally *English grass*] – *anglický trávnik*

11. *petit-déjeuner anglais* [*full breakfast*, literally *English breakfast*] – *anglické raňajky*

12. *chiffre arabe* [*Arabic numeral*, literally *Arabic numeral*] – *arabská číslica*

13. *chiffre romain* [*Roman numeral*, literally *Roman numeral*] – *rímska číslica*

14. *calendrier romain* [*Roman calendar*, literally *Roman calendar*] – *rímsky kalendár*

15. *droit romain* [*Roman law*, literally *Roman law*] – *rímske právo*

16. *football américain* [*American football*, literally *American football*] – *americký futbal*

17. *couteau suisse* [*Swiss Army knife*, literally *Swiss knife*] – *švajčiarsky nožík*

18. *médecine chinoise* [*Chinese medicine*, literally *Chinese medicine*] – *čínska medicína*

19. *signe chinois* [*Chinese sign*, literally *Chinese sign*] – *čínske znamenie*

II. Partial equivalent phrasemes that comprise a different ethnonym in French and Slovak; the compared phrasemes have a partially identical component structure and similar motive (identical in formal and/or lexical and/or semantic and/or figurative aspect). There are eight occurrences of this:

1. *soûl/ saoul comme un Polonais* [*as drunk as a lord*, literally *drunk like a Pole*] – *opitý ako Dán* [literally *drunk like a Dane*]

2. *boire comme un Polonais* [*to drink like a fish*, literally *to drink like a Pole*] – *piť ako Rus* [literally *to drink like a Russian*]

3. *c'est du chinois* [*it's all Greek to me*, literally *it is Chinese*] – *(to je) španielska dedina* [literally (*it is) a Spanish village*]
   *Au début, les racks, les motions, le chariot, avec ses 120 mètres de fil dans chaque bobine, tout ça c'*était du chinois pour moi. Meaning "something unknown".
   frTenTen12: Web domain: liberation.fr, Crawl date: 2012-02-21, doc.url http://www.liberation.fr/economie/0101403563-decoincer-la-tulle

4. *c'est de l'hébreu* [*it's all Greek to me*, literally *it is Hebrew*] – *(to je) španielska dedina* [literally (*it is) a Spanish village*]

5. *(d'un) calme olympien* [*as cool as a cucumber*, literally *(with) the Olympic calm*] – *(s) kľudom Angličana* [literally (*with) the calm of an Englishman*]
   Dans son bureau municipal de la place Sant Jaume, en plein quartier gothique, le leader anticorrida Jordi Portabella se montre d'un calme olympien: «Je ne suis pas inquiet...» Meaning "really calmly".
   Emolex: published => Le Figaro; Date => 25/07/2007

6. *assis en indien* [*sitting cross-legged*, literally *sitting in the Indian way*] – *turecký sed* [literally *Turkish sitting*]
7. *clé anglaise* [*monkey wrench*, literally *English wrench*] – *francúzsky kľúč* [literally *French wrench*]
8. *canne anglaise* [*crutch*, literally *English cane*] – *francúzska barla* [literally *French crutch*]

We have analysed the first two types in four corpora, i.e. both general and professional phrasemes which comprise the ethnonym in French as well as in Slovak. The first type comprises nineteen phrasemes, the second type – eight phrasemes, total – twenty-seven phrasemes. We have verified each of the phrasemes in all four corpora, but we only refer to such examples we consider to be more difficult for comprehension.

5. Research results

5.1. The results related to the classification of phrasemes into general, professional and mixed

We classify all the collected phrasemes in both types according to the type of general and professional language into general and professional phrasemes. The first type consists of nineteen phrasemes: thereof two general phrasemes (*filer à l'anglaise, comme une horloge/ une montre suisse*), two so-called mixed phrasemes (*russe roulette, douche écossaise*) and the remaining fifteen phrasemes are professional (*froid sibérien, grippe espagnole, guitare espagnole, bain turc, massage thaïlandais, gazon anglais, petit-déjeuner anglais, chiffre arabe, chiffre romain, calendrier romain, droit romain, football américain, couteau suisse, médecine chinoise, signe chinois*). The second type consists of eight phrasemes: thereof five general phrasemes (*saoul/ soûl comme un Polonais, boire comme un Polonais, c'est du chinois, c'est de l'hébreu, (d'un) calme olympien*) and three professional phrasemes (*assis en indien, clé anglaise, canne anglaise*).

Based on the analysis of the results from the first and the second type of phrasemes, we state with respect to French and Slovak culture the following: **I.** The first type is twice larger compared to the second type; **II.** After summarising all the phrasemes in both types, there have been twice more professional phrasemes (eighteen phrasemes) than general (seven phrasemes); **III.** We found out two so-called mixed phrasemes (the phraseme *roulette russe – ruská ruleta* and the phraseme *douche écossaise – 1. studená sprcha, 2. škótska sprcha*). As mentioned under Point *3.3. So-called general-professional phrasemes,* the phraseme *roulette russe – ruská ruleta* (*Russian roulette*) within the meaning of general phraseme is a

dangerous and risky solution and within the meaning of the professional phraseme – a *type of suicidal game with revolver and bullets*. The phraseme *douche* écossaise within the meaning of the general phraseme means a *nasty surprise*, the Slovak equivalent being *studená sprcha* (literally *cold shower*) and within the meaning of the professional phraseme representing *a type of shower where cold and hot water can be alternated*, and the Slovak equivalent is škótska sprcha (literally *Scottish shower*); **IV.** In the first type, there are seven times more (fifteen phrasemes) professional phrasemes compared to general (two phrasemes) or mixed (two phrasemes); **V.** In the second type almost twice more general phrasemes (five phrasemes) than professional (three phrasemes).

5.2. Results linked with relative frequency

According to the *absolute frequency* of phrasemes in the corpus and based on the size of every corpus (in tokens), we have calculated the *relative frequency* of phraseme, *i.e. the frequency per 1 million* (*Tables No. 1 and No.2*). The first nine or ten phrasemes with the highest frequency from both types are interesting for our assessment (these measures of specificity are highlighted in grey in *Tables No. 1 and No. 2*).

On the basis of the analysis of the results from the first and the second type of phrasemes we conclude the following for French and Slovak culture: **I.** If phrasemes occur minimum in two corpora at the same time, we evaluate them as statistically significant with respect to their frequency, specifically ten phrasemes: *roulette russe, grippe espagnole, massage thaïlandais, gazon anglais, chiffre romain, droit romain, football américain, couteau suisse, médecine chinoise, (d'un) calme olympien*; **II.** The vast majority of phrasemes with the highest frequency are professional phrasemes.

5.3. Results linked with the *logDice* association measure

We have calculated the *logDice* association measure (*Tables No. 3 and No. 4*) according to the *absolute frequency* of the individual components of phrasemes *x* and *y* and the *absolute frequency* of bigram *xy*. The first nine phrasemes of each corpus with the highest level of specificity from both types are interesting for the evaluation (the measures of specificity are highlighted in grey in *Tables No. 3 and No. 4*).

On the basis of the analysis of the results from the first and the second type of phrasemes we conclude the following for French and Slovak culture: **I.** If phrasemes occur minimum in two corpora at the same time, we evaluate them as statistically significant with

respect to the measure of their specificity, in particular thirteen phrasemes: *roulette russe grippe espagnole, massage thaïlandais, filer à l'anglaise, comme une horloge/ une montre suisse, bain turc, football américain, couteau suisse, médecine chinoise, (d'un)* *calme olympien, c'est du chinois, c'est de l'hébreu, assis en indien*; **II.** Approximately one half of the phrasemes with the highest *logDice* are general phrasemes, the other half are professional phrasemes, and there is also one mixed phrasemes *roulette russe*.

Tables 1 and 2. *Frequency of phrasemes*

X – without occurrence

1st type					
Phrasemes	Absoute frequency		Relative frequency		
roulette russe	4 499	68	0,4	0,59	
ruská ruleta	162	329		0,25	0,26
filer à l'anglaise	875	33	0,08	0,24	
zmiznúť po anglicky	3	5		0	0
comme une horloge/ une montre suisse	310	3	0,03	0,02	
ako švajčiarske hodinky	106	117		0,15	0,09
douche écossaise	595	24	0,05	0,18	
škótska sprcha	4	9		0,01	0,01
froid sibérien	914	5	0,1	0,04	
sibírska zima	16	85		0,02	0,07
grippe espagnole	2 429	0	0,21	0	
španielska chrípka	186	279		0,28	0,22
guitare espagnole	290	4	0,03	0,03	
španielska gitara	35	41		0,05	0,03
bain turc	1 735	14	0,15	0,1	
turecké kúpele	390	160		0,59	0,13
massage thaïlandais	1 693	0	0,15	0	
thajská masáž	381	320		0,58	0,26
gazon anglais	308	15	0,03	0,11	
anglický trávnik	127	198		0,19	0,16
petit-déjeuner anglais	385	0	0,03	0	
anglické raňajky	76	97		0,12	0,08
chiffre arabe	2 049	3	0,18	0,02	
arabská číslica	155	144		0,24	0,12
chiffre romain	3 123	8	0,27	0,24	
rímska číslica	203	360		0,31	0,29
calendrier romain	660	4	0,06	0,03	
rímsky kalendár	41	124		0,06	0,1
droit romain	5 473	2	0,48	0,01	
rímske právo	312	529		0,48	0,42
football américain	7 600	77	0,66	0,56	
americký futbal	712	1 866		1,09	1,49
couteau suisse	5 436	55	0,47	0,4	
švajčiarsky nožík	68	74		0,06	0,1
médecine chinoise	7 228	42	0,63	0,31	
čínska medicína	918	470		1,4	0,38
signe chinois	1 093	0	0,1	0	
čínske znamenie	24	15		0,04	0,01

2nd type					
Phrasemes	Absoute frequency		Relative frequency		
soûl/ saoul comme un Polonais	83	1	0,01	0,01	
opitý/ spitý ako Dán	2	4		0	0
boire comme un Polonais	19	0	0	0	
piť ako Rus	4	3		0	0
c'est du chinois	1 224	0	0,11	0	
to je španielska dedina	349	326		0,53	0,26
c'est de l'hébreu	408	0	0,04	0	
to je španielska dedina	349	326		0,53	0,26
(d'un) calme olympien	2 351	15	0,21	0,11	
(s) kľudom Angličana	93	32		0,14	0,03
assis en indien	123	0	0,01	0	
turecký sed	131	196		0,2	0,16
clé anglaise	566	12	0,05	0,09	
francúzsky kľúč	28	182		0,04	0,15
canne anglaise	524	21	0,05	0,15	
francúzska barla	112	178		0,17	0,14

Tables 3 and 4. *logDice association measure*

X – without occurrence

1st type				
	Test logDice			
Phrasemes	Corpus frTenTen12-all	Corpus Emolex-all	Corpus skTenTen11-all	Corpus prim-7.0-public-all
roulette russe	7,845	8,801		
ruská ruleta			6,542	5,669
filer à l'anglaise	6,966	7,357		
zmiznúť po anglicky			1,809	1,266
comme une horloge suisse/une montre suisse	3,686	4,527		
ako švajčiarske hodinky			8,151	7,176
douche écossaise	5,762	8,767		
škótska sprcha			3,154	3,574
froid sibérien	4,825	6,079		
sibírska zima			3,23	4,8
grippe espagnole	6,992	0		
španielska chrípka			7,505	6,554
guitare espagnole	3,604	4,403		
španielska gitara			5,045	3,962
bain turc	6,195	5,869		
turecký kúpeľ			8,478	6,196
massage thaïlandais	7,537	0		
thajská masáž			9,444	9,552
gazon anglais	3,059	6,797		
anglický trávnik			6,019	5,668
petit-déjeuner anglais	3,227	0		
anglické raňajky			5,504	4,899
chiffre arabe	5,501	3,038		
arabská číslica			4,894	4,085
chiffre romain	6,329	4,874		
rímska číslica			5,211	5,329
calendrier romain	4,933	4,414		
rímsky kalendár			6	5,817
droit romain	4,95	3,116		
rímske právo			5,296	5,478
football américain	6,913	7,756		
americký futbal			7,147	6,778
couteau suisse	7,484	8,011		
švajčiarsky nožík			6,528	5,055
médecine chinoise	7,828	7,954		
čínska medicína			9,303	7,312
signe chinois	4,39	0		
čínske znamenie			3,895	2,055

2nd type				
	Test logDice			
Phrasemes	Corpus frTenTen12-all	Corpus Emolex-all	Corpus skTenTen11-all	Corpus prim-7.0-public-all
soûl/ saoul comme un Polonais	5,337	4,719		
opitý/ spitý ako Dán			9,573	2,753
boire comme un Polonais	-0,33	0		
piť ako Rus			1,991	0,955
c'est du chinois	-2,549	0		
to je španielska dedina			7,224	5,863
c'est de l'hébreu	-4,134	0		
to je španielska dedina			7,224	5,863
(d'un) calme olympien	6,851	7,486		
(s) kľudom Angličana			7,407	5,75
assis en indien	2,77	0		
turecký sed			8,068	7,358
clé anglaise	3,199	5,258		
francúzsky kľúč			3,76	4,959
canne anglaise	3,954	8,614		
francúzska barla			6,255	5,289

6. Conclusions

Thanks to the quantitative methods of corpus linguistics, *the relative frequency and the logDice* association measure, we have presented the measures of the frequency and specificity of individual phrasemes, and sixteen phrasemes show the highest values: *roulette russe – ruská ruleta*, *grippe espagnole* – španielska chrípka, *massage thaïlandais – thajská masáž*, *filer à l'anglaise – odísť po anglicky, comme une horloge/ une montre suisse – ako* švajčiarske *hodinky, bain turc – turecký kúpeľ, football américain – americký futbal, couteau suisse – švajčiarsky nožík, médecine chinoise –* čínska *medicína, (d'un) calme olympien – (s) kľudom* Angličana, *c'est du chinois – (to je) španielska dedina,* c'est de l'hébreu – to je španielska dedina, assis en indien – turecký sed, gazon anglais – anglický trávnik, chiffre romain – rímska číslica, droit romain – rímske právo. Based on the corpus analysis we can propose that it should be useful to place these sixteen above-mentioned phrasemes at the forefront when compiling French-Slovak dictionaries (i.e. to place them in such a way that they will draw the attention of a dictionary user then looking up a phraseme as a whole or its individual components). We see usefulness of the obtained results in supporting the creation of the actual lexicography and phraseology of general and professional French and Slovak dictionaries as well as in bilingual French-Slovak and Slovak-French dictionaries where the results of this research can be

helpful for updating and compiling entries in these dictionaries.

Furthermore, the phrasemes of the second type, which are relatively more frequent such as *c'est du chinois*, *c'est de l'hébreu*, *(d'un) calme olympien*, *assis en indien*, *clé anglaise*, *canne anglaise,* can be used from the didactic point of view as a contrastive factor because they are easily interchangeable with phrasèmes in the target language.

References

1. Brumme, J., 2008, 'La traducción de la frasse hecha. El caso del castellano de Cataluña'. In *A Multilingual Focus on Contrastive Phraseology an Techniques for Translation.* Hamburg: Verlag Dr. Kovač, p. 75-92.
2. Díaz, R. P., Teodorovici, C., 2008, 'Una mirada a los frásicos del español y del rumano'. In *A Multilingual Focus on contrastive Phraseology and Techniques for Translation.* Hamburg: Verlag Dr. Kovač, p. 147-168.
3. Ďurčo, P., 2012, 'Typológia ekvivalentnosti vo frazeológii – na príklade ruských slovenských a nemeckých parémií'. In *Slovo v slovníku*, Bratislava: SAV, p. 91-102.
4. Gautier, L., 2004, 'Terme, phraséoterme, phrasème : question de délimitation en langue spécialisée'. In *Le Continuum en linguistique*, Jatlaoui, Héidi et al. (Ed.), Sousse: Faculté des Lettres et Sciences Humaines de Sousse, 2004, p. 153-172.
5. González-Rey, I. et al., 2002, *La phraséologie du français*. Toulouse: Presses Universitaires du Mirail, p. 268.
6. Gréciano, G., 1997, 'Collocations rythmologiques'. In *Meta,* XLII, I, p. 33-44.
7. Gréciano, G., 1993, 'Vers une modélisation phraséologique: Acquis et projets d' EUROPHRAS'. In *Terminologies nouvelles, No 10*, p. 16-22.
8. Gross, G., 1996. *Les expressions figées en français*. Paris: Editions Ophrys, p. 161.
9. Hunston, S., 2006, 'Corpus linguistics'. In *K. Brown (Eds., Vol. 3*, p. 234-248.
10. Mejri, S., 2000, 'Figement et dénomination'. In *Meta, XLV*, 4, p. 610-621.
11. Mešková, Ľ., 2015, 'La phraséologie française du point de vue de la phraséologie slave'. In *Certas palavra o vento nao leva*. Fortaleza: Parole, p. 209-223.
12. Mešková, Ľ.,Olejárová, M., 2004, 'Les phrasèmes spécialisés du domaine de la finance'. In *Phraseologismen als Gegenstand sprach- und wissenschaflicher Forschung.* Schneider Verlag Hohengehren, p. 325-335.
13. Mlacek, J., 2010, 'Pokiaľ siaha frazeológia'. In *Baláková – Walter (Eds.)*, p. 239-251.
14. Mlacek, J., 2007, Štúdie a state o frazeológii. Ružomberok: Katolícka univerzita, 375 p.
15. Mlacek, J., Ďurčo, P. et al.., 1995, *Frazeologická terminológia*. Bratislava: Stimul, 160 p.
16. Romero Ganuza, P., Batista Rodríguez, J. J. & García Padrón, D., 2008, 'Aproximación contrastiva a las UF del campo semántico-conceptual de la 'alegría' en español y alemán'. In *A Multilingual Focus on Contrastive Phraseology and techniques for Translation.* Hamburg: Verlag Dr. Kovač, p. 169-185.
17. Rosenbaum Franková, L., 2010, *Dynamické procesy vo frazeológii odborných textov z oblasti ekonómie*. Bratislava: Filozofická fakulta Univerzity Komenského, 216 p.
18. Rýchly, P., 2008, 'A Lexicographer-Friendly Association Score'. In *RASLAN 2008* (2nd *ed.)* Brno: Masarykova Univerzita, p. 6-9.
19. Ţaran Andreici, M., 2016, 'Translation of phrasemes. Equivalence et nonequivalence'. In *Professional Communication and Translation Studies*, 9/ 2016. Editura Politehnica, p. 153-166.

Summary

SPECIFIC PHRASEMES WITH ETHNONYMS AND THEIR STUDY BY CORPUS ANALYSIS

I. Dinžíková

This article studies the phrasemes comprising an ethnonym in the source language (French) as well as the target language (Slovak). This approach is contrastive and the phrasemes have been classified according to the type of equivalence (total equivalent, partial equivalent and phrasemes without equivalent). The aim of the research was to analyse 27 phrasemes with the help of the corpus linguistics method (*relative frequency and logDice* association measure), and four monolingual corpora (the corpora *frTenTen12*, *skTenTen11*, *Emolex*, *prim-7.0-public-all*) with approx. 130 million up to approx. 10 billion words in each of them, so it is a fairly wide range of language materials.

Firstly, we focus on the current state of French and Slovak phraseology. We present the distribution of phrasemes into three types: general, professional and so-called mixed (of which the last type represents our own proposition). Then, by translating the source language-culture into the target language-culture, we demonstrate the three basic types of phrasemes equivalence but our attention is on the first two types. Afterwards, we present quantitative methods of corpus linguistics (four

monolingual corpora, *relative frequency* and *the logDice* association measure). Then, we analyse 27 specific phrasemes. This is qualitative analysis (their distribution into three types of equivalence as well as their repair in general, professional and mixed phrasemes), but also quantitative analysis (analysis based on *relative frequency* and also on *logDice* association measure). In the end, we demonstrate and evaluate the results of our research.

The research objectives are set to find out the frequency of phrasemes in various types of texts and the level of their specificity within the framework of each of the corpora, based on which it is possible to propose which of the phrasemes should be placed at the forefront for looking up an entry and its individual components or a phraseme as a whole, and thus contribute to supporting the creation of current French/Slovak lexicography and phraseography.

Moreover, from the point of view of teaching foreign languages, we can use the second type of equivalence as a contrasting factor between French and Slovak language-cultures because they can either easily interfere with other phrasemes of the target language-culture or be not well understood in the target language-culture.