UDC 821.163.2.09-84:39=112.2
Original scientific paper
Received on 15.08. 2011
Accepted for publication on 23.11. 2011

Hrisztalina Hrisztova-Gotthardt[1]
Zoltán Gotthardt[2]
[1]University of Pécs
[2]Pécs

Който търси, намира:[1] Searching for Bulgarian proverbs on the Web

The present study can be seen as the second stage of a larger research concerning the real proverb usage in contemporary written Bulgarian language. While the first survey aimed to determine the most frequently used proverbs in newspaper articles, the present study examines which Bulgarian proverbs occur on the Web and with what frequency. The results of the two separate studies showed similarities and helped to gain a better understanding of the current usage of proverbs in the modern Bulgarian language.

Key words: Bulgarian; proverbs; Web search; frequency of occurrence.

1. Introduction

Up to the present time, there have been no empirical studies conducted with the aim of evaluating the current paremiological situation in Bulgaria. There is no information on which Bulgarian proverbs are currently used in the written and spoken modern Bulgarian language, and with what frequency. Not even the latest proverb collections can answer this question. They are usually based on older compilations, often listing obsolete data, most probably unfamiliar to language users. For this reason, these collections are deserved to be classified under the so-called "data graveyards of proverbs" (see Baur and Chlosta 1996: 92). Even the latest collection of Bulgarian proverbs, according to its bibliography, con-

[1] He who seeks, will find.

tains primarily selected texts from old compilations published between the beginning and the late 1990s (see Stojkova 2007: 19).

In this regard, there is a tremendous need for conducting new sets of research and evaluating the current paremiological situation in Bulgaria. One way to obtain reliable empirical data on the real usage and frequency of occurrence of proverbs in the language is to work with written corpora (Ďurčo 2001; 2006; Čermák 2003).

Based on a written corpus, an empirical study in the field of Bulgarian paremiology was accomplished between 2008 and 2009. During that study, Hrisztova-Gotthardt gained results concerning the usage and frequency of occurrence of Bulgarian proverbs in newspaper articles (see Hrisztova-Gotthardt 2010). Articles published in the period between April 4, 2000 and March 3, 2008 in the daily newspaper *Standart* (*Стандарт*) were searched through by a special search algorithm in order to determine the frequency of occurrence of 2301 Bulgarian proverbs[2]. In the end a list of 225 proverbs was compiled, in which the texts were ranked by their occurrence in the corpus. In the findings, there were not only exact matches, but many qualitative and quantitative variants and numerous modifications (anti-proverbs) as well.

As Hrisztova-Gotthardt already mentioned, there is a need for involving various text types and language varieties in the research in order to achieve more reliable and complete results (Hrisztova-Gotthardt 2010: 64). For this reason, the present study was carried out.

2. Aim of the present work

The aim of this paper is to introduce the results of a further research concerning the real proverb usage in contemporary written Bulgarian language. The corpus for this study is the World Wide Web (Web). With the aid of two well-known search engines (Google and Bing) we are going to examine which Bulgarian proverbs used in the former study occur in the Web corpus and with what frequency.

[2] The 2301 proverbs were selected from two Bulgarian proverb collections (Grigorov and Katzarov 1986 and Vlahov 1996) which were considered as the latest ones at the time of conducting the study.

Our secondary purpose is to determine whether there is any relationship (similarity) between the frequency ranking of the proverbs from the Web and the frequency ranking of the proverbs from the newspaper *Standart*.

3. The corpus

Computer and corpus linguists consider the Web as an excellent source for phraseological research in addition to the annotated and structured text corpora. The Web contains an extraordinarily large amount of texts from different language varieties and text types (Ptashnyk, Hallsteinsdóttir and Bubenhofer 2010: 10-11).

Nevertheless, as Colson states, there are some objections to using the Web as corpus. For example, many topics are over-represented on the Web (sex, music, sport, etc). The Web has not been assembled by linguists and it contains all kinds of language errors (Colson 2010: 27). In addition, the Web is very dynamic by its nature, therefore its content changes constantly. There are no tools to search the entire Web; and search engines like Google cover only a fraction of it.

However, despite its disadvantages the Web still stands for a useful linguistic resource due to its huge size and the enormous diversity of text types. For this reason it is well suited for obtaining information on the usage of proverbs in written language.

4. Search tools

There are several possible methods to gather data from the Web. One method is to build a custom software tool to crawl (read) the Web by following the links, jumping from one page to another and storing the addresses and contents of the relevant pages. After ‘enough’ text has been collected, a manual step would have to be performed to filter the data further and determine the frequency of each proverb found. This solution is very purpose-oriented, but it takes practically an infinite amount of time and a lot of other resources like network bandwidth and computing power.

The other method is to utilize the help of search engines. The greatest advantage of employing search engines is that they have been crawling the Web for some years; therefore they have already indexed an enormous number of pages.

Moreover, it is very easy to use a search engine for looking up specific words or finding exact matches.

Some disadvantages should also be mentioned: although the search engines list results from a large amount of websites, it is still just a fraction of the whole Web. Furthermore, for the sake of a very short response time, most of the search engines do not provide an exact number of hits, they just attempt to give a quick estimation, which is based on the absolute frequency of the individual words. Considering the exact number of unique hits, Konecny pointed out that if the same content is found on different pages or in different formats (HTML/Text, MS Word doc, PDF) the search engines may count each one as an independent unique hit (Konecny 2010: 83). Although this problem does not affect all the search engines, any estimation to the number of hits should be taken with a grain of salt. However, apart from all its drawbacks, using a search engine appears to be the only practical solution for searching the Web for exact matches of proverbs.

At the beginning of this study, we have decided to use three search engines. The more search engines we utilize, the more objective the results will be. Our decision was primarily based on the index size (number of searchable unique web pages) of the different search engines. Unfortunately, search engine companies report less and less often the index sizes, therefore comparing different search engines based on this criteria is difficult. Maurice de Kunder elaborated a method for estimating the index size in the frame of his Master thesis at Tilburg University (Kunder 2007). Using corpus linguistic methods in his study, he elaborated on a statistical approach to approximate the number of indexed unique Web pages of four of the most popular search engines. This approximation is calculated and released to the public every day (see World Wide Web Size). On the basis of his study, we can conclude that at the time of composing this paper (March 2011) the two largest web search engines are Google (26 billion unique pages) and Bing (12 billion unique pages).

We also wished to utilize a search engine specialized in searching in our target language (Bulgarian) and so we included Search.bg as well. In the following section, we will discuss the relevant services and features of our search engine candidates.

4.1. *Google*

Google, as stated on The Official Google Blog, hit a milestone in 2008 with indexing 1 trillion unique URLs (also known as links) (see Google Blog). Fur-

thermore, it is said that they can successfully eliminate duplicate content found at different URLs.[3] Moreover, Google is able to search in more than 100 languages or dialects including Bulgarian. This search engine supports exact phrase search by surrounding a phrase with quotation marks which also disables word replacement with synonyms (see Google Web Search Help).

However, the estimated number of hits on the result pages seems to be rather high for any exact phrase search. Google fortunately offers a solution. It gives us a hint to "navigate to the last page to see how many results the search engine really delivered" (see Google Help Center). Let us note that visiting the last page would not always work because "Google does not serve more than 1000 results for any query" as it is displayed on the result page, if a customer wants to navigate beyond the limit of 1000. Since looking up proverbs yields to a relatively small number of hits, the 1000 result limit has no impact on our research.

4.2. *Bing*

Most of the statements for Google are valid for Bing as well. Bing supports exact phrase search the same way Google does (see Bing Advanced Search Options). By advanced search keywords, it also offers language-specific search including Bulgarian (see Bing Advanced Search Keywords). It eliminates duplicates as well. Navigating to the last result page reveals the exact number of matched web pages.

4.2. *Search.bg*

This Bulgarian search engine supports exact phrase search the same way Google and Bing do (see Search.bg). Although it is possible to find phrases in other languages, its primary target language is Bulgarian. Search.bg states on its website that it has indexed 12 071 705 Bulgarian web pages (as of March 2011). This search engine displays the exact number of hits on the first result page so there is no need to navigate to the last page. Unfortunately, Search.bg has a serious drawback: after examining the returned results one can easily see that the retrieved pages are not unique – many of them are duplicates. Since the main focus of our research is to determine the exact frequency of the proverbs by recording unique hits, Search.bg was excluded from the study.

[3] Comparing the 1 trillion URLs to the 26 billion unique Web content one can conclude that the Web contains a lot of duplicates.

5. Search method

In the frame of the previous study we compiled a tokenized list of 2301 proverbs (e.g. we removed all punctuation marks and converted all the uppercase letters to lowercase). This same list was deployed in the present study as well. We developed a custom program (a PERL script) to read all the proverbs from our input list and write the results (number of hits) to a result (output) list.

Our program reads one proverb at a time from the list, constructs a query and sends it to the search engines. As the search engines return the last result page, our tool processes its content, extracts the number of hits (exact matches) and stores it in the result list.

5.1. *Google*

Constructing a query URL for Google is fairly simple. A valid exact phrase query has the following structure:

http://www.google.bg/search?q="exact phrase"&hl=bg&start=990

In this case, the so-called base URL is http://www.google.bg/search and the parameter 'q' refers the search term (namely, the proverb). The quotation marks around the phrase make Google perform an exact phrase search. We filter the language with the 'hl' parameter to Bulgarian, and immediately navigate to the last result page by setting the 'start' parameter to '990'. This is the maximum value of this parameter[4] (see section 'Search tools'/'Google'). The last result page contains the exact number of hits, and this is what we need to record for each proverb.

The task of the PERL script is to construct the URL for each proverb, send it to Google and finally analyze the returned HTML text data (the last result page). Google's reply contains the exact number of hits which resembles to one of the following items:[5]

Около 491000 резултата[6]<nobr>
Страница 6 от около 58 резултата[7]<nobr>

[4] Since the maximum exact matches for a proverb was 637, we did not reach the limit of 1000.

[5] Please note that this would be a subject to change any time as Google develops its services.

[6] 'Approximately 491000 results'.

Страница 2 от 11 резултата[8]<nobr>
1 резултат[9]<nobr>
При търсенето на - +"exact phrase" - не бяха открити съответстващи документи.[10]

The first sample shows an estimated number of hits, therefore it would not be displayed on the last page. Thus, it can safely be ignored. The next three items report exact matches and the last item indicates that no match was found. For all of these replies we constructed a PERL regular expression to extract and output the number of hits or output '0' in the last case.

5.2. *Bing*

The structure of a Bing query URL is similar to that of a Google query:

http://www.bing.com/search?q="exact phrase" language:bg&first=990

In this case, the base URL is http://www.bing.com/search and the parameter 'q' contains the search phrase, similarly to Google. The quotation marks force Bing to perform an exact phrase search. Unlike Google, Bing requires the language to be specified within the parameter 'q': the search term is followed by a space, the 'language' option, a colon and the language code 'bg'. Bing also lets us navigate immediately to the last result page by setting the parameter 'first' to '990'. Bing has the same limit of 1000 hits as Google does, however Bing does not give an error message for any higher value.

A reply from Bing contains a HTML fragment similar to one of the following:

1-10 of 7,200 results
461-470 of 470 results
<h1>No results found for +"exact phrase" language:bg.</h1>

The first item indicates an estimation of hits which does not occur on the last page so we will safely ignore it. The second item shows a sample from a last page ('470' exact matches). The last sample indicates that there was no match.

[7] 'Page 6 of approximately 58 results'.
[8] 'Page 2 of 11 results'.
[9] '1 result'.
[10] 'Searching for "exact phrase" did not return any results'.

We constructed another PERL regular expression to extract the number of hits or output '0' in the last case when there are no results.

6. Results

6.1. *Appointing proverb lemmas*

In the course of the first study conducted between 2008 and 2009 (Hrisztova-Gotthardt 2010), the author analyzed the search list of 2301 proverbs (see Introduction) and noticed that it consisted not only of unique proverbs. There were qualitative and quantitative variants of some proverbs listed among the texts. Therefore, it was reasonable to 'unite' the different variants under one lemma (also called main form or main variant). The lemmas were not selected on the basis of a subjective decision. After conducting the search in the *Standart* corpus, the proverb variant with the most hits was appointed as a lemma. For instance, *Една птичка пролет не прави* ('One bird does not make spring') and *Една лястовица пролет не прави* ('One swallow does not make spring') are certainly two variants of the same proverb. In the *Standart* corpus the first form provided 9 hits, while the second provided only 2. Based on this result, the first form was appointed as a lemma.

The final search result was prepared in tabular format. For a particular proverb the appointed lemma was displayed together with its (bracketed) variants in the first column. In the second column, the number of hits for each form was displayed. It was followed by the total number of hits for both the lemma and all of its variants. Finally, the proverbs were ranked according to the total of their hits.

In the course of the present study, we have followed the same approach. In the majority of cases, the Web search showed similar results to those of the *Standart* search, regarding the numbers of hits for a particular proverb and its variants. In the case of the above mentioned Bulgarian proverb, the results are the following: Google provided 455 hits for the lemma and 103 hits for its variant, while Bing found 317 and 60 matches, respectively. As the numbers indicate, the web search led to a much higher number of results and therefore to a more accurate ranking. Based on this new ranking, in some cases, another lemma was appointed. Nevertheless, this change affected only a small number of proverbs with extremely low number of hits in the *Standart* corpus. For instance, the proverbs *Бог помага ала в кошара не вкарва* (God helps but he does not drive (the sheep) into the sheepfold) and *Господ дава ама в кошара не вкарва* (The Lord gives but he does not drive (the sheep) into the sheepfold) provided 9 and 5

hits respectively in the *Standart* corpus. Therefore the first form was assigned as a lemma. However, for the same proverbs, the Google search engine provided 51 and 68, while Bing gave 22 and 29 hits. Both of these search results indicate that the second form is most likely the more widespread one, which makes us conclude that actually the second proverb should be appointed as a lemma.

6.2. *Top 10 lists of Google and Bing*

Due to space limitations, we are unable to show the full result lists provided by both search engines. In fact, Google found 1495 different proverb forms on 52 914 pages. In contrast, Bing detected only 1000 unique proverb forms on 31 508 pages.

The tables below show the ten proverbs with the most hits provided by Google and Bing, respectively:

Table 1. Top 10 proverbs found by Google.

Proverb (Variants)	**Translation**	**Individual hits**	**Total hits**	**% (of total hits)**
Който търси, намира. (Търсете и ще намерите.)	He who seeks, will find. (Seek and you shall find.)	439+268	707	1,34%
Глас народен - глас божи. (Глас народен - глас божий.)	The voice of the people [is] the voice of God. (The voice of the people [is] the voice of God almighty.)	426+150	576	1,09%
Една птичка пролет не прави. (Една лястовица пролет не прави.)	One bird does not make spring. (One swallow does not make spring.)	455+103	558	1,05%
Бързай бавно.	Hurry slowly.	478	478	0,90%
Старост - нерадост.	Old age brings no joy.	465	465	0,88%
Апетитът идва с яденето.	Appetite comes with eating.	461	461	0,87%
Всяко зло за добро.	No evil without good.	456	456	0,86%
Всяко чудо за три дни.	Every miracle (lasts) for three days.	456	456	0,86%
Клин клин избива.	A wedge drives out a wedge.	451	451	0,85%
От всяко дърво свирка не става.	You can't make a pipe from each tree.	449	449	0,85%

Table 2. Top 10 proverbs found by Bing.

Proverb (Variants)	Translation	Individual hits	Total hits	% (of total hits)
Който търси, намира. (Търсете и ще намерите.)	He who seeks, will find. (Seek and you shall find.)	483+173	656	2,08%
Всяко зло за добро.	No evil without good.	637	637	2,02%
По-добре късно, отколкото никога.	Better late than never.	594	594	1,89%
Нищо ново под слънцето. (Няма нищо ново под слънцето.)	Nothing new under the sun. (There is nothing new under the sun.)	318+212	530	1,68%
Съединението прави силата.	Union makes strength.	484	484	1,54%
Целта оправдава средствата.	End justifies the means.	458	458	1,45%
Любовта е сляпа.	Love is blind.	428	428	1,36%
Всяко чудо за три дни.	Every miracle lasts for three days.	407	407	1,29%
Една птичка пролет не прави. (Една лястовица пролет не прави.)	One bird does not make spring. (One swallow does not make spring.)	317+60	377	1,20%
Кръвта вода не става.	Blood is thicker than water.	369	369	1,17%

6.3. Top 10 of Standart vs. Top 10 of 'Web'

We have unified the Google list with the Bing list to a common 'Web list' by simply selecting the number of hits from the search engine that found more results for a particular proverb. Summing up or averaging the two results from the two search engines would have caused duplicates because of possible common hits. We will now compare the results from *Standart* to the above constructed 'Web list':

Table 3. Exact proverb matches found in *Standart.*

Proverb (Variants)	Translation	Individual hits	Total hits	% (of total hits)
Нищо ново под слънцето. (Няма нищо ново под слънцето.)	Nothing new under the sun. (There is nothing new under the sun.)	1020+6	1026	51,10%

Proverb (Variants)	Translation	Individual hits	Total hits	% (of total hits)
По-добре късно, отколкото никога.	Better late than never.	53	53	2,64%
Всяко зло за добро.	No evil without good.	48	48	2,39%
Съединението прави силата.	Union makes strength.	41	41	2,04%
Всяко чудо за три дни.	Every miracle [lasts] for three days.	28	28	1,39%
Апетитът идва с яденето.	Appetite comes with eating.	24	24	1,20%
Кръвта вода не става.	Blood is thicker than water.	23	23	1,15%
Глас народен - глас божи. *(Глас народен - глас божий.)*	The voice of the people [is] the voice of God. (The voice of the people [is] the voice of God almighty.)	20+1	21	1,05%
Целта оправдава средствата.	End justifies the means.	16	16	0,80%
Клин клин избива.	A wedge drives out a wedge.	16	16	0,80%

Table 4. Exact proverb matches found on the Web.

Proverb (Variants)	Translation	Individual hits	Total hits	% (of total hits)	Source
Който търси, намира. *(Търсете и ще намерите.)*	He who seeks, will find. (Seek and you shall find.)	439+268	707	1,30%	Google
Всяко зло за добро.	No evil without good.	637	637	1,18%	Bing
По-добре късно, отколкото никога.	Better late than never.	594	594	1,10%	Bing
Глас народен - глас божи. *(Глас народен - глас божий.)*	The voice of the people [is] the voice of God. (The voice of the people [is] the voice of God almighty.)	426+150	576	1,06%	Google
Една птичка пролет не прави. (Една лястовица пролет не прави.)	One bird does not make spring. (One swallow does not make spring.)	455+103	558	1,03%	Google
Нищо ново под слънцето. *(Няма нищо ново под слънцето.)*	Nothing new under the sun. (There is nothing new under the sun.)	318+212	530	0,98%	Bing

Proverb (Variants)	Translation	Individual hits	Total hits	% (of total hits)	Source
Съединението прави силата.	Union makes strength.	484	484	0,89%	Bing
Бързай бавно.	Hurry slowly.	478	478	0,88%	Google
Старост - нерадост.	Old age [brings] no joy.	465	465	0,86%	Google
Апетитът идва с яденето.	Appetite comes with eating.	461	461	0,85%	Google

7. Discussion

The search for proverbs in newspaper articles and on the Web showed similar results. The proverbs in the Top 10 list of the *Standart* corpus are included in the Top 10 of either Google or Bing. Nevertheless, the ranking of proverbs from the *Standart* corpus turned out to be rather biased. A single proverb, namely *Нищо ново под слънцето* ('There is nothing new under the sun'), has reached 1026 hits, which is more than half of all the matches. This occurred due to the fact that the proverb was used as a title of a newspaper column and therefore appeared in almost every issue.

In contrast, the unified result of the search engines (the 'Web list') ranks the proverbs very smoothly: after the Top 6, the total number of hits of each proverb is just slightly less than that of the previous in the list. There is no big 'jump' or 'gap' between any two proverb forms. This suggests that there is no borderline between the 'more frequent' proverbs and the 'less frequent' ones (with the exception of the Top 6). Still, the top 65 proverbs cover more than half (50.15%) of all Web hits and can be nominated as candidates for the list of the most frequently used proverbs in the modern Bulgarian written language.

8. Outlook

The present study can be seen as an important step towards establishing a list of the most frequently used proverbs in the modern written Bulgarian language. However, in order to draw final conclusions about the current paremiological situation in Bulgaria, there is a need for further research. There should be more language varieties and text types included in the investigation.

In this context, we are planning to carry out a new study to determine the frequency of proverbs from the Bulgarian National corpus. At the moment, the

Corpus consists of about 320 000 000 words and includes about 10 000 texts contributed from publishers and authors, obtained from the digitalization of printed texts and extracted from the Internet. The materials represent huge number of different text types and reflect the state of the Bulgarian language (mainly in its written form) from the middle of 20th century to the present. Due to its characteristics, the Bulgarian National corpus seems to be another ideal candidate for observing and estimating the real proverb usage in the language.

References

Baur, Rupprecht S., Christoph Chlosta (1996). Sprichwörter: ein Problem für Fremdsprachenlehrer wie -lerner? *Deutsch als Fremdsprache. Zeitschrift zur Theorie und Praxis des Deutschunterrichts für Ausländer* 33.2: 91–102.

Bing Advanced Search Keywords = http://onlinehelp.microsoft.com/en-us/bing/ff808421.aspx.

Bing Advanced Search Options = http://onlinehelp.microsoft.com/en-us/ bing/ff808438.aspx.

Bubenhofer, Noah, Stefaniya Ptashnyk (2010). Korpora, Datenbanken und das Web: State of the Art computergestützter Forschung in der Phraseologie und Lexikographie. Ptashnyk, Stefaniya, Erla Hallsteinsdóttir, Noah Bubenhofer, Hrsg. *Korpora, Web und Datenbanken.* Baltmannsweiler: Schneider Verlag Hohengehren, 7–18.

Colson, Jean-Pierre (2010). The Contribution of web-based corpus linguistics to a global theory of phraseology. Ptashnyk, Stefaniya, Erla Hallsteinsdóttir, Noah Bubenhofer, Hrsg. *Korpora, Web und Datenbanken.* Baltmannsweiler: Schneider Verlag Hohengehren, 23–36.

Čermák, František (2003). Paremiological minimum of Czech: The corpus evidence. Burger, Harald, Annelies Häcki Buhofer, Gertrud Gréciano, Hrsg. *Flut von Texten – Vielfalt der Kulturen*. Baltmannsweiler: Schneider Verlag Hohengehren, 15–31.

Ďurčo, Peter (2001). Bekanntheit, Häufigkeit und lexikographische Erfassung von Sprichwörtern. Häcki Buhofer, Annelies, Harald Burger, Laurent Gautier, Hrsg. *Phraseologiae Amor. Aspekte europäischer Phraseologie. Festschrift für Gertrud Gréciano zum 60. Geburtstag.* Baltmannsweiler: Schneider Verlag Hohengehren, 99–106.

Ďurčo, Peter (2006). Methoden der Sprichwortanalysen oder auf dem Weg zum Sprichwörter-Optimum. Burger, Harald, Annelies Häcki Buhofer, eds. *Phraseology in Motion I.* Baltmannsweiler: Schneider Verlag Hohengehren, 3–20.

[Grigorov, Milko, Kostadin Katzarov] Григоров, Милко, Костадин Кацаров (1986). *Български пословици и поговорки.* Трето преработено издание. София: Издателство Наука и изкуство.

Google Blog = http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html.

Google Help Center = http://www.google.com/help/indexsize.html.

Google Web Search Help = http://www.google.com/support/websearch/bin/ answer.py?hl=en &answer=136861.

Hrisztova-Gotthardt, Hrisztalina (2010). Methoden und Ergebnisse einer korpusbasierten Untersuchung zur Vorkommenshäufigkeit bulgarischer Sprichwörter in zeitgenössischen Zeitungstexten. Ptashnyk, Stefaniya, Erla Hallsteinsdóttir, Noah Bubenhofer, Hrsg.

Korpora, Web und Datenbanken. Baltmannsweiler: Schneider Verlag Hohengehren, 55–66.

Konecny, Christine (2010). Lexikalische Kollokationen und der Beitrag der Internet-Suchmaschine *Google* zu ihrer Erschließung und Beschreibung. Ptashnyk, Stefaniya, Erla Hallsteinsdóttir, Noah Bubenhofer, Hrsg. *Korpora, Web und Datenbanken.* Baltmannsweiler: Schneider Verlag Hohengehren, 77–94.

Kunder, Maurice De (2007). *Geschatte grootte van het geïndexeerde World Wide Web.* http://www.dekunder.nl/Media/Scriptie%20Maurice%20de%20Kunder%20-%20Grotte%20geindexeerde%20web.pdf.

Ptashnyk, Stefaniya, Erla Hallsteinsdóttir, Noah Bubenhofer, Hrsg. (2010). *Korpora, Web und Datenbanken.* Baltmannsweiler: Schneider Verlag Hohengehren.

Search.bg = http://www.search.bg/bg/help.html.

[Stojkova, Stefana] Стойкова, Стефана (2007). *Български пословици и поговорки.* София: Издателство Колибри.

[Vlahov, Sergej] Влахов, Сергей (1996). *Съпоставителен речник на пословици*. София: Издателство Ето.

World Wide Web Size = http://www.worldwidewebsize.com.

Authors' address:

Hrisztalina Hrisztova-Gotthardt
University of Pécs
Foreign Language Centre
H-7633 Pécs
Szántó Kovács János utca 1/B
xpucuhu@gmail.com

Zoltán Gotthardt
Data Warehouse Expert
H-7632 Pécs
Olga utca 1.X.32.
gotthardt.zoltan@gmail.com

Който търси, намира

U POTRAZI ZA BUGARSKIM POSLOVICAMA NA INTERNETU

Ovaj se prilog može smatrati drugom fazom šireg istraživanja koje se bavi stvarnom porabom poslovica u suvremenom bugarskom jeziku. Dok je u prethodnoj fazi cilj bio ustanoviti najčešće poslovice u jeziku novina, ova studija ima zadaću utvrditi koje se bugarske poslovice pojavljuju na internetu i koliko često. Rezultati dva odvojena istraživanja pokazuju sličnosti i doprinose boljem razumijevanju porabe poslovica u suvremenom bugarskom jeziku.

Ključne riječi: bugarski; poslovice; web; frekvencija uporabe.