

Phraseology in Multilingual Society

Edited by

Elena Arsenteva

CAMBRIDGE
SCHOLARS
P U B L I S H I N G

TABLE OF CONTENTS

Phraseology in Multilingual Society
Edited by Elena Arsenteva

This book first published 2014

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2014 by Elena Arsenteva and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-5584-7, ISBN (13): 978-1-4438-5584-6

List of Images	ix
List of Tables	x
Preface	xi
Part I: Phraseology at Kazan University	
Chapter One	2
Kazan School of Phraseology Elena Arsenteva and Rimma Safina	
Chapter Two	16
Baudouin de Courtenay and Phraseology Wolfgang Eismann	
Part II: Semantic Aspect of Phraseological Units	
Chapter Three	34
Basic Components of the Connotative Aspect in Phraseological Units: (As Seen by A. V. Kunin and his Disciples) Tatiana Fedulenkova	
Chapter Four	47
Fishing for Productive Models in the Troubled Waters of Figurative Language Antonio Pamies, Margaret Craig and Yara El Ghalayini	
Chapter Five	71
Phraseological Units with Onomatopoeic Component and Transparent Inner Form Rimma Salieva and Alsu Nurullova	
Chapter Six	83
The Reflection of Society in Professional Idioms Marina Solnyshkina and Farida Ismaeva	

Chapter Seven.....	92
Phraseological Creativity from a Linguoculturological Perspective Irina Zykova	

Part III: Comparative Study of Phraseological Units

Chapter Eight.....	106
Colour Symbolism of Biblical Idioms with Components Indicating White and Black in Bible Translations into English, German and Russian Louise Bayramova and Rose Muhametdinova	

Chapter Nine.....	122
English Phraseological Units with the Components <i>FIRE</i> and <i>WATER</i> and their Counterparts in Russian, Spanish and Tatar Natalia Konopleva and Albina Kayumova	

Chapter Ten.....	141
On Some Aspects of Borrowing of Phrases from English into German and Russian Lyubov Nefedova and Oleg Polyakov	

Chapter Eleven.....	156
Widespread Idioms in Europe and Beyond: The Lesser-Used European Languages and the “Lexicon of Common Figurative Units” Elisabeth Piirainen	

Part IV: Pragmatic Aspects of Phraseology

Chapter Twelve.....	174
To add Fresh Fuel to the Fire: Modifications of Widespread Idioms Pavel Dronov and Svetlana Bochaver	

Chapter Thirteen.....	184
Ways of Using Phraseological Units in TEFL Gulnara Gimaletdinova and Liliia Khalitova	

Chapter Fourteen.....	195
Formulaic Genres as Cultural Artefacts: Monolingual Speech Communities as Multicultural Societies Koenraad Kuiper	

Chapter Fifteen.....	210
Instantial Stylistic Use: A Cognitive View of Phraseological Units in Discourse Anita Naciscione	

Chapter Sixteen.....	228
Phraseological Units in Press: Cultural Peculiarities Aida Sadykova and Elena Smirnova	

Chapter Seventeen.....	239
Impact of Peculiarities of Context on Instantial Phraseological Units and their Transference into the Language of Translation Elena Semushina and Nina Soboleva	

Part V: Paremiology

Chapter Eighteen.....	250
Quantitative Appraisal in Russian Proverbs and Weather-Lore Tatyana Bochina and Marija Kul'kova	

Chapter Nineteen.....	269
On Proverb Translation Within a Literary Context: Preliminary Theoretical Considerations Pedro Martins	

Chapter Twenty.....	278
“What’s Sauce for the Goose is Sauce for the Gander”: The Proverbial Fight for Women’s Rights by Elizabeth Cady Stanton and Susan B. Anthony Wolfgang Mieder	

Chapter Twenty-One.....	297
Representation of Moral Values in Russian and Brazilian Proverbs Maria Luisa Ortiz Alvarez	

Part VI: Phraseography

Chapter Twenty-Two.....	316
Types of Phraseological Unit Definition Roza Ayupova and Radif Zamaletdinov	

Chapter Twenty-Three.....	329
Lexicographic Presentation of Phraseological Transforms Diana Davletbaeva and Iskander Yarmakeev	
Chapter Twenty-Four.....	339
On a New German-Russian Dictionary of Idioms: Aspects of Corpus-Based Bilingual Phraseography Dmitrij Dobrovol'skij	
Chapter Twenty-Five.....	350
Computational Corpus of Tatar Proverbs and Sayings: Electronic Database of Paremiias Kamil Galiullin, Elena Gorobets, Guzel Karimullina and Rezeda Karimullina	
Chapter Twenty-Six.....	363
Cross-Linguistic Equivalence of Idioms: Corpus vs. Dictionary Irina Parina	
Chapter Twenty-Seven.....	377
A Look at Carinthian-Slovenian Phraseology as Represented in the "Thesaurus of the Slovenian National Language in Carinthia" Heinrich Pfandl	

LIST OF IMAGES

Figure 11-1: Languages of the Project "Widespread Idioms in Europe and Beyond"	162
Figure 11-2: Equivalents of "to rise like the Phoenix from the ashes" in European languages	163
Figure 11-3: Equivalents of "to be caught between two fires" in European languages	167
Figure 15-1. Anti-Thatcher graffiti in West Belfast, Northern Ireland...	222

LIST OF TABLES

Table 7-1. Phraseological creativity of the 11 macro-metaphorical conceptual models (in accordance with the <i>quantitative</i> dimension)..	98
Table 9-1. Frequency of phraseological and non-phraseological counterparts of English PUs with the components FIRE and WATER in Spanish, Russian and Tatar	134
Table 9-2. Frequency of phraseological counterparts of English Pus with the component FIRE	135
Table 9-3. Frequency of phraseological counterparts of English Pus with the component WATER.....	135
Table 9-4. Frequency of non-phraseological counterparts of English PUs with the component FIRE	136
Table 9-5. Frequency of non-phraseological counterparts of English PUs with the component WATER.....	136
Table 15-1. Extended metaphor: metonymic links between the base metaphor and metaphorical sub-images.....	215
Table 25-1. Frequency list of letters	358
Table 25-2. Usage frequency of words containing definite quantity of letters	359
Table 25-3. Usage frequency of words with different quantity of syllables	360

PREFACE

In fact, the field of phraseology is remarkably little known outside of northern Europe.

Koenraad Kuiper

In comparison with other branches of linguistics with many centuries of development phraseology can be considered a young child though rather intelligent and shrewd. Its domain is constituted by picturesque and vivid elements termed phraseological units (PUs), which are characterized by a certain transference of meaning.

The term “phraseology” was introduced by a prominent Swiss scholar of French origin Charles Bally at the beginning of the twentieth century. The first to raise the question of phraseology as a linguistic subject was Professor Ye. D. Polivanov, a well-known Russian scientist (Polivanov 1931). Academician V. V. Vinogradov was the first to work out the classification of Russian phraseological units, which gave rise to extensive investigation of phraseology in other languages (Vinogradov 1974).

Since that time much has been done in the field of phraseology. The sole scholarly society for the furtherance of research in phraseology, the European Society of Phraseology “Europhras” was founded at the end of the twentieth century to coordinate the investigations of scientists from different countries and even continents. Each year international conferences are held which are devoted to the problems of phraseology. These are organized by members of “Europhras” in different European countries.

This book is based on the best presentations of the international conference “Phraseology in Multilingual Society” organized in partnership with “Europhras” and held at Kazan Federal University, Russia, in August 2013.

The main goal of the book is to give a rich understanding of phraseology and to present the work of scholars from different countries investigating phraseological units in many languages. That is why different aspects of phraseological research are presented in the book: semantic, pragmatic, comparative, etc.

The **first part** of the book is devoted to the description of the role of the Kazan School of Phraseology and the contribution of the well-known Polish-Russian scientist, the founder of the Kazan School of Linguists Jan

CHAPTER TWENTY-FIVE

COMPUTATIONAL CORPUS OF TATAR PROVERBS AND SAYINGS: ELECTRONIC DATABASE OF PAREMIAS

KAMIL GALIULLIN¹, ELENA GOROBETS²,
GUZEL KARIMULLINA³
AND REZEDA KARIMULLINA⁴

1. Introduction

Linguists and programmers of Kazan Federal University realized the project *Computational Corpus of Tatar* and have created the cumulative computer database of Tatar proverbs and sayings as a part of this corpus.

Tatar paremiography at the present stage has a great number of various collections of proverbs and sayings (mono- and polythematic, alphabetical, mono- and multilingual, explanatory and demonstrating, etc.). All this variety offers an opportunity to get acquainted with the most extensive, fundamental and informative references:

1) the three-volume collection of Tatar folk proverbs by Naki Isanbet (Isänbät 2010), its re-edition includes more than 38 thousand proverbs, excluding variants; the first edition was published in 1959-1967;

2) the twelve-volume collection of Tatar folklore (Tatar khalyk ijaty 1987), its volume "Proverbs and sayings" contains more than 17 thousand of these units.

According to the results of our investigation, nowadays it is incredibly necessary to provide an entire registration and description of materials

which are presented in overword units of different types (phraseological units, paremias, eponyms and so on); besides, it is obligatory to include all linguistic levels into this analysis, and cumulative electronic databases are able to solve these problems successfully and effectively.

Paremias in most cases present unique mini-texts and contain important material for different spheres of the humanities, especially for source study in linguistics and for didactical linguistics. Sometimes paremias are used and analyzed in linguistic studies; they are able to illustrate some examples in textbooks and manuals. Nevertheless our analysis shows that a lot of important materials of Tatar paremias are not involved in scientific processes.

Paremias are the important source of linguography – an interdisciplinary sphere of linguistics, theory and practice of creating language references and dictionaries (see *Kompyuternaya lingvografiya* 1995); the parts of linguography are lexicography, phraseography, morphemography, paremiography and so on). In linguographic activities paremias are often used as rather short and demonstrative examples of word usage; thus, the entries of a two-volume Tatar-Russian dictionary represent 1547 proverbs and 895 sayings (Tatarcha-ruscha süzlek 2007 (TRS)). But paremic material is not always used to the full extent; compare the above-mentioned Tatar-Russian dictionary (TRS) and an explanatory dictionary published in 2005 (Tatar 2005 (TTAS)) by the parameter of proverb quantity: proverbs with initial "A" – 118 (TRS)/47 in (TTAS); "Б" – 190/22; "М" – 71/13; "Т" – 391/217; "X" – 23/2; "Ю" – 37/2; "Я" – 16/6.

Didactical linguistics provides a possibility of applying paremias for solving various problems; particularly, proverbs and sayings are used as high-grade cognitive material helping to analyze and characterize different language phenomena and units with all their functional features; nevertheless, according to the results of special research, important paremic material is poorly included into the majority of training books. It can in no small way be explained by the difficulties of access to this information (we mean either quantitative or qualitative parameters).

The electronic corpus made it possible to work out the series of linguographic sources describing the language of Tatar paremias (Galiullin, Karimullina and Karimullina 2009). They include concordance, inverse and frequency indexes of words and word forms.

¹ Kazan Federal University, Russia. galiullin.kamil@mail.ru.

² Kazan Federal University, Russia. elena_gorobets@mail.ru.

³ Kazan Federal University, Russia. guzelika@list.ru.

⁴ Kazan Federal University, Russia. re_ka@mail.ru.

2. Main Part

2.1. Electronic Corpus of Tatar Paremias: Information Potential

The electronic corpus of Tatar proverbs and sayings is formed on the basis of the collections described above; in addition, it includes paremic materials presented in different dictionaries of Tatar: the defining dictionary (Tatar telenen anlatmaly süzlege 2005), the Tatar-Russian dictionary (Tatarcha-ruscha süzlek 2007), etc. The merits of traditional paremic collections are supplemented with plenty of advantages given by the opportunity to realize the project with the help of modern technologies.

The fund contains proverbs and sayings followed by a paremiographic description which consists of the following information zones:

- a) the exact address of a selected unit with the indication of its source and its concrete place in this source;
- b) the thematic group or subgroup of the selected paremia;
- c) the variants of the paremia (including choices connected with different optional paremic components);
- d) the concrete type of selected paremic unit (a proverb or a saying), etc.

This information essentially enlarges the options of a data search. A user can set the volume and the type of required material if he defines as a search array a) one source b) all sources, or c) some sources selected by any characteristic or by a group of them (for example, only training references or only bilingual dictionaries).

Precise addressing simplifies the work with corpus sources: a user receives the possibility to get quick access to them in order to choose the required supplementary data (for instance, to get acquainted with plenty of comments in Naki Isanbet's three-volume collection) (Isänbät 2010). The information about thematic groups of paremias extends the search by contents with the help of a selection possibility given to users if they make a query in thematic groups or subgroups. For instance, if a user chooses the 3.1.1.5.4 index processing the data from Naki Isanbet's collection, he can define the thematic group "Бәйрәмнәр" (Holidays) as a search array.

The subcorpus of words and word forms fixed in Tatar proverbs and sayings raises the information potential of an electronic database.

Traditional (paper) collections of paremias more often realize an alphabetical entry only by the initial word form or only by a keyword included into a paremic unit. An electronic collection allows the

organization of search by all words and word forms included in proverbs and sayings, and by the components of these words and word forms as well (ex. search by morpheme).

A concrete word or word form can be used as a search image. For instance, a user will get *Ятаган чыкса яу булыр* [6613 – here and further square brackets are used for the numbers of paremias in (Tatar khalyk ijaty 1987)] if he makes a query <ятаган>.

Searching different forms of any word users can effectively apply a sign * – in a search query it means any sign, any quantity of signs and also the situation of sign absence. For example, the program will process the query <барнагыл*> and give the results *Барнагыл да барнагыл дин, Барнаулга бармагыл* [7776] and *Байның бар малы ятар, ярлының барнагылы йөгепер* [7271]. Consequently, the sign * with the invariant part of a word helps to provide more efficient an effective search. Let us analyze the difference: <ата-ана> query gives 12 results with the word form *ата-ана* (ex. *Ата-ана догасы утка-суга батырмас* [14395]), <ата-ана*> query gives these 12 results and 25 more: *ата-анага, ата-анадан, ата-аналы, ата-ананы, ата-ананыкы, ата-ананың, ата-анаңны, ата-анасы, ата-анасын, ата-анасына* (ex. *Чапкан атка дала кадерле, ата-анага бала кадерле* [14450]).

The advantages of this search help to find different derivatives as well: *чүлмәк* "a pot" and *чүлмәкче* "a potter", *кыю* "brave" and *кыюлык* "bravery" and so on.

The sign * placed into the beginning of a queried component allows for a list of words or word forms which include an askable component not only in preposition but also in postposition and in the middle of a word. In English and in Russian these types of query are actual either for compounds or for derivatives with different prefixes, but in Tatar these queries are aimed mainly at compounds (two- and multirooted derivatives) because of the practically full absence of derivatives with prefixes. For instance, <*яман*> query will give as a result either various forms of the word *яман* (*яманга, яманны* etc.) or derivatives of this base morpheme – it is necessary to mention that this base morpheme can take not only the first position (*яманлау, яманлык, яманчылык* и др.) but other positions as well (*яхшы-яман, яхшы-яманлык – Ерак юлга чыкмыйча, атыңның яхшы-яманлыгын белмәссең* [6758]).

Queries by sample permit the organization of the search of information about letters (and – through them – sometimes about sounds), syllables, morphemes, words – in one word, to realize different types of linguographic description: graphemographic (phonemographic), syllabographic, morphemographic, lexicographic; for example, the

subcorpus of lemmas gives the possibility of receiving as a result of <э*>, <*ш> queries precise information about the usage of these letters in correspondent positions in different words. Queries <*сыз> / <*сез> will select the words with these components from the database of proverbs and sayings.

The quality of search queries and the quality of their processing gets better if we include information about various alternations (ex. *б/н, з/к*, etc.: *китан – китабы, азык – азыгы*), different affixes dependent on the type of base morpheme (ex. *-дал-дәл-мал-мә*, etc.).

Queries by sample can be supplemented with queries by the quantity of letters, symbols, syllables, and by frequency as well due to the electronic format of a database. Supplementary queries can even be processed separately (see below, paragraph 2.2.).

The computational version of a reference fund enables the possibility for widening the set of requested characteristics; for instance, we plan to include the information about parts of speech and also about usage labels connected with lexical units.

Corpus contains a reference subcorpus with materials providing the semantization of words with limited usage (archaisms, historicisms, dialecticisms, professionalisms, etc.). To provide the desired information in our subcorpus we use the data of different sources: explanatory dictionaries, dialect dictionaries, encyclopedic dictionaries, dictionaries of foreign words, etc.; see some samples of data in subcorpus (they are placed after ▲ sign).

Ага-эне тату булса, ат күп булып; абсын тату булса, аш күп булып [14543] ▲ **Абсын** “Килендэшләр”. (Tatar telenen zur dialektologik süzlege 2009, 14)

Алэт барда куэт бар [6335] ▲ **Алэт** “Корал, эш коралы, инструмент, прибор”. (Timergalin 2007, 22)

Кулавызыз юлга кергән юлдан язар [6921] ▲ **Кулавыз** “Юл-үткел күрсәтүче, юламан, рәһбәр, гид”, *бор*. (Timergalin 2007, 265)

In perspective we plan to place some components of the cumulative computer database of Tatar paremias on the Internet: <http://www.klf.kpfu.ru> – the portal of Kazan Federal University, in the section “Kazan linguographic fund”.

Scientists have analyzed the development of lexicography and the information potential of language references and defined main tendencies. They show that the most efficient form of a dictionary is its Internet version. Here is a list of its advantages:

- 1) widely available; the development of computer technologies influences the development of Internet technologies, and it is a considerably new information revolution, the result of which is the global generalization of language references;
- 2) convenient exploitation;
- 3) multiple choice of entries, the possibility of a search by different parameters;
- 4) a dictionary can be supported in actual status, it can be constantly developed and improved; textual parts of a dictionary are corrected operatively, all desired supplements are included if necessary, etc.;
- 5) the restrictions on the volume of material in a dictionary are removed;
- 6) it is easy to connect with similar web references and to enlarge linguographic Internet databases with the help of other web sites.

Some of these advantages provided by the electronic form of a dictionary are simultaneously the attributes of other (non-web) language references, for instance, of CD-dictionaries and vocabularies, and some parameters (for example, the first and the sixth) describe only Internet dictionaries.

At the present time the network versions of basic Tatar dictionaries are placed on the Internet in parallel; our next step is to supplement the database with hyperlinks to corresponding Internet dictionaries. It will give an opportunity to create a reference complex connected with the resources of different sites and to fundamentally strengthen the information potential of the corpus.

The analysis shows the rapid “Internetization” of practical linguography and proves the perspectiveness and high level of self-descriptiveness in the sphere of Internet dictionaries.

2.2. Electronic Corpus of Tatar Paremias: Quantitative Data

The most important advantage of the electronic corpus is the ability to receive various quantitative data connected with different features of the corpus units (paremias, words, word forms, syllables, letters, etc.).

Quantitative information offers the field for qualificative characteristics of language (speech, text) facts, it is especially desired in comparative analysis (Bektaev 1978; Galiullin and Karimullina 2010; Piotrovskiy, Bektaev and Piotrovskaya 1977; Quantitative Linguistik

2005; Těšitelová 1992; Tuldava 1987, etc.); see the list of reference in (Quantitative Linguistik 2005).

Let us consider some quantitative materials which are the result of the processing and description of proverbs and sayings included in the twelve-volume collection of Tatar folklore (Tatar khalyk ijaty 1987). Also it would be useful to compare these results with data from other Tatar texts, particularly with prosaic (Rizvanova 2000) and poetic (Galiullin and Karimullina 2006; 2009; 2010; Galiullin, Karimullina and Mingazova 2006).

2.2.1. Proverbs and Sayings

The quantity of proverbs and sayings (N_P) is 17089. Their cumulative volume:

- in word usages, text words (N_{TW}) – 95405;
- in syllables (N_S) – 212714;
- in letters (N_L) – 494676.

The volume of paremia in text words – 2-29 words.

The average length of paremia in words (N_{TW}/N_P) – nearly 6 words (5.58).

The longest paremias (according to the quantity of words):

29 words – *Атның бер аягы ак булса – бер йөз сум, ике аягы ак булса – ике йөз сум, өч аягы ак булса – өч йөз сум, дүрт аягы ак булса – егерме биш сум* [2571]; 25 words – *Атадан ул туса – иге, ата юлын куса – иге, биленә садак буса – иге, битенә килер оятны үзе белеп юса – иге, шушыларны кылмаса, барыннан да югы – иге* [14357]; 24 words – *Кычыткан чыпчыгы: “Үз өем, үз илем!” – дип сайрый икән. Моньң өе кайда дип карап торсалар, кычыткан арасында кычыткан ефәгеннән үргән оясына кереп киткәнән күргәннәр* [4216].

The shortest paremias (according to the quantity of words):

2 words – *Атлы – көнле* [2557], *Ачу – аю* [12431]; *Аша, туным* [8950]; *Әжәт – назад* [8196]; *Кунак – бәби* [10094]; *Күз – дарья* [10414]; *Өчтә көч* [8012]; *Эш үтермәс* [6181].

Volume of paremia in syllables – 3-57 syllables.

Average length of paremia in syllables (N_S/N_P) – 12.45.

The longest paremias (according to the quantity of syllables):

57 syllables – see above “*Кычыткан ...*” [4216]; 56 syllables – see above “*Атадан ...*” [14357]; 53 syllables – *Кәжәнең сөте агып торган болак, эче тулы ылак, ите тамагыңа бап, тиресе аягыңа кап, мөгезе пычагыңа сап, эчәгесен аршинга тарт, скриткачыга сат* [3053].

The shortest paremias (according to the quantity of syllables):

3 syllables – *Күз – дарья* [10414], *Өчтә көч* [8012], *Тан та бак* [14047], *Чәч тә баш* [10366].

The volume of paremia in letters – 5133 letters.

The average length of paremia in letters (N_L/N_P) – 28.95.

The longest paremias (according to the quantity of letters):

133 letters – see above “*Кычыткан ...*” [4216]; 118 letters – *Аргамакның баласы аз утлар да күп юшар, атасы тарлауның дәмен татканга; азамат ирнең баласы аз сөйләр дә күп тыңлар, атасы йортка киңәш иткәнә* [14346]; 116 letters – see above “*Кәжәнең ...*” [3053].

The shortest paremias (according to the quantity of letters):

5 letters – *Ачу – аю* [12431], 7 букв – *Өчтә көч* [8012], 8 letters – *Аша, туным* [8950], *Күз – дарья* [10414], *Тан та бак* [14047], *Чәч тә баш* [10366].

2.2.2. Word Forms

19118 word forms are singled out in paremias; we consider that text words with the similar graphic image are equal to one word form.

Different word forms have diverse frequency in paremia corpus (from 1 to 1230). Here is the list of paremias with a frequency equal to 100 and more word usages.

да (1230), *белән* (1197), *бер* (1100), *дә* (893), *юк* (755), *булса* (658), *кеше* (632), *бар* (576), *булыр* (501), *яхшы* (453), *булмас* (418), *ат* (397), *эт* (384), *була* (373), *күп* (373), *үз* (370), *ике* (352), *булмый* (346), *түгел* (321), *яман* (315), *сүз* (297), *дип* (296), *хатын* (258), *ир* (256), *булсын* (254), *көн* (244), *ди* (235), *үзе* (230), *су* (227), *кыз* (218), *эш* (207), *бала* (202), *ни* (187), *кара* (183), *бай* (179), *мал* (175), *баш* (165), *алтын* (158), *аш* (155), *таш* (155), *ил* (153), *бүре* (145), *кешенең* (142), *жирдә* (141), *өчен* (140), *кебек* (139), *күрә* (138), *та* (137), *тел* (137), *акыл* (130), *егет* (130), *кирәк* (130), *башы* (129), *ярлы* (128), *аз* (126), *булсаң* (125), *акча* (122), *кунак* (122), *кешегә* (118), *ут* (118), *карга* (115), *мең* (115), *агач* (115), *килсә* (113), *жир* (112), *теле* (111), *юл* (110), *булын* (109), *күз* (109), *һәр* (109), *бал* (107), *бул* (106), *дус* (106), *ул* (105), *үзәң* (105), *артык* (104), *өй* (103), *ач* (102), *зур* (102), *ит* (102), *кош* (101), *жиргә* (100).

The average frequency of a word form in proverbs and sayings (N_{TC}/N_{CF}) is 4.99. This index is higher than in poetry (4.05), prose (4.17/4.26) and newspaper texts (4.05/4.42) (Rizvanova 2000, 38). It could be in no small way explained by the fact that proverbs and sayings contain the higher percentage of words in common use.

More than half of word forms (10397) in analyzed paremias have an index of frequency equal to 1.

2.2.3. Letters

Analyzed proverbs and sayings contain 494676 letters. The frequency of letters is in the range from 2 (*u*) to 60624 (*a*). This base does not contain letters *ě* and *u*.

The frequency list of letters is in Table 25-1.

Table 25-1. Frequency list of letters

1. Range of a letter in # 2.

2. Letter: **a** – vowel; **b** – letters Ь, Ъ; **c** – consonant.

3. Absolute frequency.

4. Relative frequency.

#1	#2			#3	#4	#1	#2			#3	#4
	a	b	c				A	B	c		
1	A			60624	12.26	21		Ч	8190	1.66	
2			К	34545	6.98	22		З	8165	1.65	
3	Ы			33806	6.83	23	Я		6488	1.31	
4			Н	33723	6.82	24	Ө		6005	1.21	
5	Е			32603	6.59	25	О		5432	1.10	
6	Ә			27994	5.66	26		П	4634	0.94	
7			Р	27962	5.65	27		Ж	2763	0.56	
8			Л	27368	5.53	28		Х	2508	0.51	
9			Т	24856	5.02	29	Ә ⁵	Ә	2122	0.43	
10			Б	17754	3.59	30	Ю		1964	0.40	
11			С	17635	3.56	31		В	1564	0.32	
12			Г	15126	3.06	32		Ь	522	0.11	
13			М	15094	3.05	33		Һ	442	0.09	
14	И			14051	2.84	34		Ф	419	0.08	
15	У			13352	2.70	35		Ъ	54	0.01	
16			Д	10865	2.20	36		Ж	47	0.01	
17			Ш	10232	2.07	37		Ц	2	0.00	
18			Ң	8822	1.78	38	Ё		0	0	
19	Й			8669	1.75	39		Щ	0	0	
20	Ү			8274	1.67				494676	100 %	

⁵ The letter ә in most cases (2121) indicates a correspondent vowel, and only in the word *тәэсирсез* does it reflect a specific consonant vowel (paremia *Тәэсирсез сүз-яңгырсыз күкрәү* [16102]).

Paremias contain words of various length (the quantity of letters) – from 1 to 18. Table 25-2 reflects the usage frequency of words containing definite quantity of letters.

Table 25-2. Usage frequency of words containing definite quantity of letters

1. Quantity of letters

2. Quantity of text words with the indicated quantity of letters

3. Relative frequency

#1	#2	#3	#1	#2	#3	#1	#2	#3
1	42 ⁶	0.044	8	7269	7.619	15	9 ⁷	0.009
2	7637	8.005	9	2748	2.880	16	5 ³	0.005
3	12337	12.931	10	1433	1.502	17	1 ³	0.001
4	15449	16.193	11	573	0.601	18	2 ³	0.002
5	22276	23.349	12	164	0.172		95405	100 %
6	15311	16.048	13	56	0.059			
7	10068	10.553	14	25	0.026			

The average length of a text word in proverbs and sayings (N_L/N_{TW}) is 5.19 letters. This index is closer to the statistic data of poetic texts – 5.30 (Galiullin and Karimullina 2010, 38)/5.34 (Galiullin and Karimullina 2006, 60)/5.47 (Rizvanova 2000, 49) than to the data of prosaic (5.64) and newspaper (6.23) (Rizvanova 2000, 49) texts.

2.2.4. Syllables

Examined proverbs and sayings contain 212714 syllables.

Paremias include words with different syllabic length – from 1 to 8. Table 6-3 shows the usage frequency of words with different quantities of syllables.

⁶ Single-letter words: *a*, *ә* (8), *u* (7), *ю*, *я* (25).

⁷ Multiple-letter words: 18 letters – *тирләмәдем-нешмәдем* (2); 17 letters – *эрләмәгән-сукмаган*; 16 letters – *әссәләмәгаләйкәм*, *баккакау-баккакау*, *заманындагыларны*, *йөрмәгән-күрмәгән*, *йөрмәгән-тормаган*; 15 letters – *баһадирландырыр*, *бишекле-билтерле*, *булыр-булмасныкы*, *губернаторыннан*, *йортсыз-семьясыз*, *йортыңны-әжиреңне* (2), *сәләмәтлегеннән*, *ышанычсызлыккың*.

Table 25-3. Usage frequency of words with different quantity of syllables

- # 1. Quantity of syllables
 # 2. Quantity of text words with the indicated quantity of syllables
 # 3. Relative frequency

#1	#2	#3	#4	#5	#6
1	16777	17.585	6	69	0.072
2	46760	49.012	7	6 ⁸	0.006
3	25717	26.955	8	1 ⁴	0.001
4	5573	5.842		95405	100 %
5	502	0.526			

The average length of a text word in proverbs and sayings (N_{CT}/N_{TC}) is 2.23 syllables. This index (like the letter index above) is closer to statistic data of poetic texts – 2.17 (Galiullin and Karimullina 2010, 38)/2.24 (Galiullin and Karimullina 2006, 62)/2.27 (Rizvanova 2000, 54) than to the data of prosaic (2.41) and newspaper (2.81) (Rizvanova 2000, 54) texts. The words are comparatively shorter in paremias than the average. It is determined by the fact that lexical units of everyday language predominate here and they are characterized by high frequency – these units are usually rather short.

Other quantitative data characterizing the linguistic features of paremias will be presented in our special statistic appendix to the computational corpus of Tatar proverbs and sayings.

3. Conclusions

The electronic interface of the database containing Tatar paremias allows the realization of different types of linguographic description: paremiographic, lexicographic, morphemographic, and, besides, graphemographic (it presents the information connected with the usage of definite letters according to their position in a word) and syllabographic (it presents the information connected with the usage of definite syllables in a word). The e-database widens the performance capacity of search and scientific processes connected with the materials of Tatar proverbs and sayings.

⁸ Polysyllabic words: 8 syllables–*баккакау-баккакау*; 7 syllables–*эссаламегалэйкем, заманындагыларны, тирләмәдем-пешмәдем* (2), *эрләмәгән-сукмаган, эченә-бавырына*.

Our research shows the value of quantitative data, because it allows the defining of the typicality and occurrence of any feature in the paremic array, to describe the correlation of significant characteristics with other features or to fix the absence of this correlation.

All this information is very important to the comparative analysis of results which received two or more subsets if they reflect inclusions, crossings, and exceptions. These conditions can possibly reveal general and distinctive features of the subsets.

According to the results of our analysis, the cumulative electronic corpus of Tatar will play a considerable role in the optimization of access to the rich and important data of proverbs and sayings. It will be a contribution to a multifold description of paremias. The corpus will help put the received materials into linguistic and didactic use, to process them and to get the desired data in order to use it in different spheres.

List of References

- Bektaev, K. 1978. *Statistiko-informatsionnaya tipologiya tyurkskogo teksta*. Alma-Ata: Nauka.
- Galiullin, K. 2009. "Internet-lingvografiya: russkiye tekstoopisyvayuschiye slovary." In *Problemy istorii, filologii, kul'tury*. 2(24): 635-639.
- Galiullin, K., and R. Karimullina. 2006. "O slovarnom opisaniy proizvedeniy Musy Jalilya: «Moabitskiye tetradi»." In *Uchenye zapiski Kazanskogo universiteta*. 148(3): 55-64.
- Galiullin, K., and R. Karimullina. 2009. *Gabdulla Tukay: shigriyat tele: süzlek*. T. 1-2. Kazan: Mägarif. [Internet version: <http://www.klf.kpfu.ru/tukay>. Accessed 30.07.2013.].
- Galiullin, K., and R. Karimullina. 2010. *Tukay shigerläre tele: lingvostatistik beleshmälek*. Kazan: Izdatel'stvo Kazanskogo universiteta.
- Galiullin, K., and D. Mart'yanov. 2006. *Slovar' yazyka russkikh poslovits i pogovorok kontsa XVII – pervoy poloviny XVIII veka: Sbornik Petrovskoy galerei*. Kazan: Kazan universitet. [Internet version: http://www.klf.kpfu.ru/parem17_18. Accessed 30.07.2013.].
- Galiullin, K., Karimullina, A., and R. Karimullina. 2009. "Tatarskaya folklornaya ligvografiya: slovarnoye opisaniye yazyka poslovits i pogovorok." In *Aktual'nye problemy sovremennoy folkloristiki*, 135-136. Kazan: Alma-Lit.

- Galiullin, K., Karimullina, R., and L. Mingazova. 2006. *Musa Jalil. «Moabit дәftärläre tele: süzlek»*. Kazan: Mägarif. [Internet version: <http://www.klf.kpfu.ru/jalil>. Accessed 30.07.2013.]
- Isänbät, N. 2010. *Tatar khalyk mäkalläre: mäkallär jyelmasy*. Kazan: Tatarstan kitap nashriyaty.
- Kompyuternaya lingvografiya* 1995. Kazan: Izdatel'stvo Kazanskogo universiteta.
- Piotrovskiy, R., Bektaev, K., and A. Piotrovskaya. 1977. *Matematicheskaya lingvistika*. Moskva: Vysshaya shkola.
- Quantitative Linguistik: Ein internationales Handbuch* 2005. Berlin: New York: Walter de Gruyter.
- Rizvanova, L. 2000. *Kvantitativnaya kharakteristika tatarskogo slova*. Kazan: Tarikh.
- Tatar khalyk ijaty: Mäkallär häm äytemnär* 1987. Kazan: Tatarstan kitap nashriyaty.
- Tatar telenen anlatmaly süzlege* 2005. Kazan: Matbugat yorty.
- Tatar telenen zur dialektologik süzlege* 2009. Kazan: Tatarstan kitap nashriyaty.
- Tatarcha-ruscha süzlek* 2007. T.1-2. Kazan: Mägarif.
- Těšitelová, M. 1992. *Quantitative Linguistics*. Praha: Academia.
- Timergalin, A. 2007. *Milliyat süzlege*. Kazan: Mägarif.
- Tuldava, Yu. 1987. *Problemy i metody kvantitativno-sistemnogo issledovaniya leksiki*. Tallin: Valgus.

CHAPTER TWENTY-SIX

CROSS-LINGUISTIC EQUIVALENCE OF IDIOMS: CORPUS VS. DICTIONARY

IRINA PARINA¹

1. Introduction

The present investigation deals with the cross-linguistic equivalence of phrasemes and, specifically, idioms. This issue is important for translation, foreign language teaching and phraseography, and has come into focus once again with the advance of corpus-based dictionaries. Corpora are generally seen as a source of evidence about lexical items of any kind, which is objective and thus more reliable than the intuition of a lexicographer (cf. Lubensky and McShane 2007, 919; Moon 1998, 44). Examples of phraseme usage taken from corpora are considered to be advantageous for a phraseological dictionary, especially when they are provided with a translation into the target language (see for example Dobrovol'skij 2013, also this volume, on a New German-Russian Dictionary of Idioms). However, it is sometimes the case that an idiom in an authentic context cannot be translated with its traditionally accepted equivalent.

Equivalence in translation is claimed to be different from equivalence in the language system, because the former means the relationship between an idiom and its translation variant in a particular text while the latter considers the relationship between L1 and L2 idioms on the systemic level (Dobrovol'skij 2013; Korhonen 2007, 575). But there seems to be no strict separation between these two aspects, as the features of a L2-idiom which make it impossible to use it as an equivalent for a L1-idiom in a particular context can sometimes turn out to be relevant from the phraseographical point of view.

¹ Linguistics University of Nizhny Novgorod, Russia. parinai@yandex.ru.