



DOI: <https://doi.org/10.15688/jvolsu2.2023.4.8>

UDC 811.512.161'374  
LBC 81.63.12-42

Submitted: 26.12.2022  
Accepted: 03.04.2023

## SYSTEM-FUNCTIONAL STRATIFICATION OF THE AVERAGE TURKISH-RUSSIAN DICTIONARY

**Zamira K. Derbisheva**

Kyrgyz-Turkish Manas University, Bishkek, Kyrgyzstan

**Alexey A. Kretov**

Voronezh State University, Voronezh, Russia

**Abstract.** The study is aimed at stratifying Rosa Yusipova's Turkish-Russian Dictionary (2005) in accordance with the system and functional weights of its constituent words. The method is a parametric lexicon analysis (PLA), developed and tested by scientists of Voronezh State University. PLA involves identifying four particular parametric weights for each word. This is the FUNCTIONAL weight (F-weight is indirectly estimated by the length of the word), the PARADIGMATIC weight (P-weight is estimated by the number of synonyms), SYNTAGMATIC weight (C-weight is estimated by the number of phrases and utterances with this word) and EPIDIGMATIC weight (E-weight is estimated by the number of meanings of a word in the dictionary). For each of the 4 parameters, partial core, counting at least 1000 words, was allocated. The words presented in all 4 particular cores, entered a Small parametric core. Words presented in 3 particular cores entered an Average parametric core, words represented in 2 particular parametric cores – a Large parametric core and words presented in 1 particular parametric core entered the core of the Dictionary. Words that are not presented in any particular parametric core make up the Periphery of the Dictionary. The analysis revealed the words of all 4 cores of the dictionary: Small – 140 words, Middle – 630, Large – 3234, the core of the Dictionary – 6861 and the Periphery of the Dictionary counts 18236 words. The dominant (the most important word in the dictionary) was the word *iş* 'work, labor', and the vice-dominant – the word *üst* 'the top'.

**Key words:** Turkish-Russian dictionaries, parametric analysis, functional word weight, paradigmatic word weight, syntagmatic word weight, epidigmatic word weight, core vocabulary, periphery of vocabulary.

**Citation.** Derbisheva Z.K., Kretov A.A. System-Functional Stratification of the Average Turkish-Russian Dictionary. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2023, vol. 22, no. 4, pp. 101-115. DOI: <https://doi.org/10.15688/jvolsu2.2023.4.8>

УДК 811.512.161'374  
ББК 81.63.12-42

Дата поступления статьи: 26.12.2022  
Дата принятия статьи: 03.04.2023

## СИСТЕМНО-ФУНКЦИОНАЛЬНАЯ СТРАТИФИКАЦИЯ ЛЕКСИКИ СРЕДНЕГО ТУРЕЦКО-РУССКОГО СЛОВАРЯ

**Замира Касымбековна Дербисева**

Кыргызско-Турецкий университет «Манас», г. Бишкек, Кыргызстан

**Алексей Александрович Кретов**

Воронежский государственный университет, г. Воронеж, Россия

**Аннотация.** Цель данного исследования – системно-функциональная стратификация лексики турецкого языка в соответствии с системными и функциональным весами составляющих ее слов, установленными по данным «Турецко-русского словаря» Розы Юсиповой. Метод исследования – параметрический анализ лексики, разработанный и апробированный российскими учеными кафедры теоретической и прикладной

лингвистики Воронежского государственного университета. Метод предполагает определение четырех частных параметрических весов для каждого слова. Это *функциональный* вес (косвенно оценивается по длине слова, поскольку, как отмечал еще Дж.К. Ципф, средняя длина и средняя частота слов взаимозависимы: по мере убывания средней частоты слов их длина возрастает; следовательно, максимальный функциональный вес имеют самые короткие слова, а минимальный – самые длинные), *парадигматический* вес (косвенно оценивается по количеству синонимов у данного слова; при этом синонимами признаются слова, толкующие части которых хотя бы в одном из значений имеют не менее 50 % общих метаслов), *синтагматический* вес (косвенно оценивается по числу фразеосочетаний и речений с данным словом) и *эпидигматический* вес (оценивается по числу значений слова в словаре). По каждому из четырех параметров выделено частнопараметрическое ядро размером не менее 1 000 слов. Слова четырех ядер, имеющие вес по всем четырем параметрам, вошли в малое параметрическое ядро. Слова, имеющие вес по трем параметрам, отнесены к среднему параметрическому ядру; слова, представленные в двух частнопараметрических ядрах, – к большому параметрическому ядру; слова, вошедшие в одно частнопараметрическое ядро, – к ядру словаря. Слова, не вошедшие ни в одно частнопараметрическое ядро, составляют периферию словаря. В результате анализа выявлены слова всех 4 ядер словаря: Малый – 140 слов, Средний – 630, Большой – 3 234, Ядро словаря – 6 861 и Периферия словаря – 18 236 слов. Доминантой оказалось слово *iş* ‘работа, труд’, а вице-доминантой – слово *üst* ‘вершина’.

**Ключевые слова:** турецко-русские словари, параметрический анализ, функциональный вес слова, парадигматический вес слова, синтагматический вес слова, эпидигматический вес слова, ядро лексики, периферия лексики.

**Цитирование.** Дербишева З. К., Кретов А. А. Системно-функциональная стратификация лексики среднего турецко-русского словаря // Вестник Волгоградского государственного университета. Серия 2, Языковедение. – 2023. – Т. 22, № 4. – С. 101–115. – (На англ. яз.). – DOI: <https://doi.org/10.15688/jvolsu2.2023.4.8>

## Introduction

The lexicon of the language is measured in tens or even hundreds of thousands words and seems to be unusable for comparative lexicology. Therefore, one of the important tasks of modern lexicology is to create well-ordered descriptions of vocabulary, that enable distinguishing its representative cores of about 1000 words, providing such a comparison. And if there is no shortage of dictionaries containing information about the vocabulary of the world's languages, the necessary tools for theoretical mastering of this information – Parametric Analysis of Lexicon (hereinafter – PAL) – appeared relatively recently [Titov, 2002; 2004a]. One example of the use of this toolkit is the collective monograph [Kretov et al., 2016].

Parametric analysis of the Turkish vocabulary is presented in a number of papers [Bugaeв, 2006; Kretov et al., 2016; Semenova, 2018]. However, in these studies, the object of analysis were **Small** Turkish-Russian dictionaries measuring about 10,000 words: Turkish-Russian and Russian-Turkish Dictionary (Rybalchenko, 2001) was investigated by V.P. Bugaeв [Bugaeв, 2006] and by I.D. Semenova [Semenova, 2018], Brief Turkish-Russian Dictionary (Scherbinin, 1977) was investigated in the collective monograph [Kretov et al., 2016, p. 411].

In this regard, it seems appropriate to explore a larger Turkish-Russian dictionary and put the parametric analysis of Turkish vocabulary on a more complete and more modern basis. The purpose of this article is to study the connections that make up the lexicon system of the Turkish language and to stratify the vocabulary of the source dictionary according to the systemic weight of the components of its words.

## Data and methods

The object of the study is the “Turkish-Russian Dictionary” (Yusipova, 2005), rich by information and the most modern of the available Turkish-Russian dictionaries of this type. When counting one-word lemmas (without lemmas-phrases and reference articles), the volume of the dictionary has 25,097 words. The choice of bilingual dictionary is conditioned by the need for a single basis to compare the lexicons of Turkic languages, both among themselves and with the lexicons of any other languages of the world. For this purpose the Russian language, which performs the function of meta-language, has been accepted.

We proceed from the widespread notion of a field organization of the world's languages vocabulary the principle of “core in cores”, that

is according to the fractal principle. The core of the vocabulary of any language is the root words, followed by the derivative words sector, and finally, the periphery of the lexical-semantic system is formed by composite nominations (including phrases). Peripheral vocabulary can change as quickly as possible, but it does not affect the core of the lexical-semantic system, the selection of which is the purpose of our analysis. In comparative historical linguistics the change of the lexical core of language (“basic vocabulary”) is recognized as the most important event that can occur in language and with language: “Such cases are known and invariably qualified as a change of language. <...> ...If the basic vocabulary begins to be actively borrowed, the rest of the vocabulary of the language tends to be saturated with borrowings even more... as a result there is virtually nothing left of the original language – it can be stated that the people have switched to another language” (here and further English translation is ours. – Z. D., A. K.) [Burlak, Starostin, 2005, p. 14].

The subject of the study is the system-forming parameters of Turkish vocabulary. The method of research is Parametric Analysis of Lexicon (PAL), described, substantiated and tested in studies [Titov, 2002; 2004a; Voevudskaya, 2015; Kretov, 2011; 2017; Kretov, Cherechecha, 2020; Kretov et al., 2016; Kretov, Gasuns, Leonchenko, 2021; Merkulova, 2018; Semenova, 2018]. PAL is a method of analyzing vocabulary according to the data of foreign-Russian dictionaries. As part of the parametric analysis of vocabulary, the indicators of different dictionaries of the same language were repeatedly compared in order to assess the ratio of objectivity and subjectivity of their data. The result of the research is: “lexicographic sources, on average, by 2/3 reflect the realities of the language’s lexical system, and only 1/3 of the information they contain depends on the subjective factor” [Voevudskaya, 2015, p. 206].

Thus, all existing bilingual dictionaries representing a subjective image of objective reality today are the only source of information for the construction of lexical-semantic typology of languages. The dictionaries have already analyzed and represented the epidigmatics (polysemy) and (although sparingly) syntagmatics (and in implicit form – also paradigmatics) of the vocabulary of the corpus of texts that formed the basis of the dictionary

file. As a parametric analysis of foreign-Russian bilingual dictionaries, PAL accepts each of the dictionaries analyzed, including (Yusipova, 2005), and criticism of source dictionaries is carried out post factum – through comparison with the results of analysis of other dictionaries (see: [Titov, 2004b]). PAL assumes the definition of four private parametric scales for each word represented in the dictionary by its vocabulary form – lemma: **functional** weight (indirectly estimated by the length of the lemma: the shorter is the lemma, the greater is **F-weight**), **paradigmatic** weight (P-weight is indirectly estimated by the number of synonyms for a given lemma), **syntagmatic** weight (**S-weight** is indirectly estimated by the number of combinations with this lemma in a dictionary article, including illustrative examples) and **epidigmatic** weight (**E-weight** is indirectly estimated by the number of meanings allocated by lemma in the dictionary article). The addition of private weights of each lemma gives **integral** parametric weight (**I-weight**).

Each of the scales is calculated on the same formula:

$$Pr_i = \frac{\Sigma r - R_{1-i}}{\Sigma r}$$

where  $\Sigma r$  – the sum of lemmas of all ranks,  $R_{1-i}$  – the sum of lemmas from the first rank to the given, and  $Pr_i$  – the weight of the lemmas of the  $i$ -rank.  $Pr_i$  values fluctuate in the interval from 0 to 1.

The logic of the formula is simple: the fewer participants showed the same or better result, the higher the place (rank) of the participant. The weight of the lemmas of each rank depends on the number and weights of the lemmas of all other ranks. Thus, each of the words (lemmas) in the dictionary affects the weight of all the other words (lemmas) for each of the 4 parameters. This approach sharply narrows the freedom of research arbitrariness, increasing the scientific objectivity of the study.

## Results and discussion

This section is devoted to description of the analysis results for each parameter of the source dictionary, consisting in “weighing” each word within this particular parameter. *Syntagmatic* and *paradigmatic* connections are system-forming for the dictionary in synchrony, mutually defining each other: *syntagmatics* is represented by speech sequences, and *paradigmatics* is represented by

synonymous and other sets of words similar in any respect. *Epidigmatic* (derivational in a broad sense) connections characterize the dictionary as a developing and self-expanding object, which is associated with diachrony, and the *functional* parameter characterizes the dictionary as the most important part of a living, i.e. functioning language in the process of communication. Thus, a set of four system-forming parameters characterizes the dictionary as a developing, self-expanding and functioning system. Their totality provides the possibility of “weighing” words by their system-forming (integral) weight, i.e. by their place and importance in the lexico-semantic system. The novelty and scientific value of the results are presented at the end of each section.

### *Function stratification of Turkish vocabulary*

The word usage is an unobservable factor. The frequency of the word in any text is equivalent to two independent patterns: objective – linguistic and subjective – textual. The author of the text has power only over subjective regularity. The objective one is imposed by the language: in any Russian text the most common word will be *i* ‘and’, in any English – *the*, in any Turkish – *bir* ‘one, some’ (Göz, 2003). This information does not give anything to highlight the lexical cores of the language. That is why in parametric analysis of vocabulary it is more expedient to evaluate the usage of words on such an objective observed parameter as the length of the lemma

(representing the word in the dictionary): over the length of the full-digit word (as opposed to its frequency) the author of the text is not in power.

Functional stratification of vocabulary raises the question: in which units to determine the length of the lemma. For the Turkish language this question can be removed: adopted in 1928 latinized Turkish alphabet quite accurately reflects the sound composition of Turkish speech. In the mass (many thousands!) study of the Turkish vocabulary, we have the right to put an equal sign between the length of the Turkish lemma in letters and its length in sounds. The prospect of using these parametric analysis of the vocabulary of the source dictionary in comparison with other Turkic languages, especially the Kyrgyz language, forces us to deviate from the form of lemmas in the dictionary-source to ensure that the result is comparable to the dictionaries of those Turkic languages in which the verb form is given in its purest form and marked with a hyphen. When calculating the length of the lemmas and calculating their functional weight (F-weight), *-mak/-mek* morphemes did not affect the length of the verb lemmas and their F-weight: the length of each lemma with these affixes was reduced by 3 letters of sound. For example, the length of the lemma *çıkarmak* ‘pull out, take out, extract smth.’ is not 8, but 5, the length of the lemma *tutmak* ‘hold’ – not 6, but 3, the length of the lemma *almak* (‘take’) is not 5, but 2 letters. The distribution of lemmas in the dictionary-source by length in letters based on the functional transformation supposition is presented in Table 1.

Table 1. Distribution of the source dictionary lemmas by length

Letters	Lemme	Cumul.	F-weight	Example	Meaning
1	5	5	0,99980	<i>o</i>	he, she, it
2	161	166	0,99339	<i>aç</i>	hungry
3	934	1100	0,95617	<i>ana</i>	mother
4	1838	2938	0,88293	<i>açık</i>	outdoor
5	4703	7641	0,69554	<i>abır</i>	honor, dignity
6	4258	11899	0,52588	<i>açınım</i>	development
7	4417	16316	0,34988	<i>adamlık</i>	humanity
8	3863	20179	0,19596	<i>adaletli</i>	just
9	2153	22332	0,11017	<i>akrabalık</i>	kinship
10	1361	23693	0,05594	<i>alışkanlık</i>	habit
11	794	24487	0,02431	<i>adaletlilik</i>	justice
12	317	24804	0,01167	<i>bağdaştırıcı</i>	adapter
13	171	24975	0,00486	<i>cesaretsizlik</i>	indecision, timidity
14	66	25041	0,00223	<i>dayanışmacılık</i>	solidarity
15	38	25079	0,00072	<i>değerlendirilme</i>	score
16	14	25093	0,00016	<i>rutubetlendirici</i>	humidifier
17	2	25095	0,00008	<i>toplumsallaştırma</i>	nationalization
18	1	25096	0,00004	<i>elektrokardiyogram</i>	electrocardiogram
21	1	25097	0,00000	<i>erkânharbiyeiumumiye</i>	General Staff

The distribution of lemmas by length in the source dictionary is presented in Figure 1.

Figure 1 indicates the heterogeneity of word distribution in the dictionary, as evidenced by the presence of two peaks: 5 (mode) and 7. Since the words of spoken speech are frequent and therefore short, the formation of the second peak of the distribution with a length of 7 indicates the prevalence of longer derived words characteristic of written speech.

The heterogeneity of the distribution of words by length also indicates the genetic heterogeneity of the vocabulary, the formation of the vocabulary of the standard language is largely due to derivative and borrowed words: “Borrowed words in the Turkish language are represented mainly by Arabic and Persian vocabulary, the number of which in the 17<sup>th</sup> and 19<sup>th</sup> centuries reached 80–90% in some works. <...> The oldest lexical borrowings from European languages are acquisitions from the Greek language... <...> Borrowings from Armenian, Albanian, Hungarian, Romanian, South Slavic and Russian languages played a role in the formation of the dictionary of modern standard Turkish” [Kononov, 1997, pp. 409-410].

The shortest (that means – the most important, having the biggest F-weight) Turkish content words are two-letter: *aç(mak)* ‘open’; *aç* ‘hungry’; *ad* ‘name’; *af* ‘forgiveness’; *ağ* ‘net’; *ağ(mak)* ‘rise up’; *ak(mak)* ‘flow, pour’; *ak* ‘white’; *al(mak)* ‘take’; *al* ‘scarlet’; *al* ‘cunning’; *an* ‘moment’; *an* ‘reason’; *an(mak)* ‘remember someone’; *ar* ‘shame, modesty’; *as(mak)* ‘hang’; *as* ‘ermine’; *aş(mak)* ‘overcome’; *aş* ‘food’; *at(mak)* ‘throw’; *at*

‘horse’; *av* ‘hunting’; *ay(mak)* ‘regain consciousness’; *ay* ‘moon’; *az(mak)* ‘become violent’; *az* ‘insufficient’; *de(mek)* ‘talk, say’; *eğ(mek)* ‘tilt’; *ek* ‘supplement’; *ek(mek)* ‘sow’; *el* ‘hand(s)’; *el* ‘stranger’; *em(mek)* ‘suck’; *em* ‘medicinal remedy’; *en* ‘width’ *en* ‘brand for cattle’; *er* ‘man’; *er(mek)* ‘reach sth’; *es(mek)* ‘blow (about the wind)’; *eş* ‘couple, partner’; *eş(mek)* ‘rake the ground’; *eş(mek)* ‘gallop’; *et(mek)* ‘do’; *et* ‘meat’; *ev* ‘house’; *ev(mek)* ‘hurry’; *ez(mek)* ‘crush, mash’; *iç* ‘the inside, the inside (of something)’; *iç(mek)* ‘drink’; *iğ* ‘spindle’; *ih(mak)* ‘kneel (about a camel)’; *il(mek)* ‘weakly tie’; *il* ‘(administrative unit in Turkey) *il* ‘vilayet, province’; *il(mek)* ‘weak knot’; *im* ‘sign, signal’; *in(mek)* ‘go down’; *in* ‘den, hole’; *ip* ‘rope’; *is* ‘soot’; *iş* ‘work, labor’; *it* ‘dog’; *it(mek)* ‘push’; *iv(mek)* ‘hurry’; *iz* ‘trace’; *öç* ‘revenge’; *od* ‘fire’; *öd* ‘bile’; *öd* ‘smell of the burning the scarlet tree’; *ok* ‘arrow’; *ol(mak)* ‘to be, to happen’; *öl(mek)* ‘to die’; *öl* ‘soil moisture’; *om* ‘thickened/rounded end of the bone’; *on(mak)* ‘improve, correct’; *ön* ‘place (in front of something)’; *öp(mek)* ‘kiss ‘whom’; *ör(mek)* ‘knit’; *ot* ‘grass’; *öt(mek)* ‘sing; chirp’; *ov(mak)* ‘knead; rub’; *öv(mek)* ‘praise’; *oy* ‘opinion’; *oy(mak)* ‘make a’ recess/deepening’; *öz* ‘the essence (of a person)’; *öz* ‘native (about relatives)’; *öz* ‘river, stream’; *sl(mak)* ‘smash, break’; *su* ‘water’; *ti* ‘bugle signal’; *uç(mak)* ‘fly’; *uç* ‘the point, the pointed end (of a knife, etc.)’; *um(mak)* ‘hope, hope for someone’; *un(mak)* ‘organize’; *un* ‘flour’; *ün* ‘voice, sound’; *ün* ‘fame’; *ur* ‘neoplasm, tumor’; *us* ‘mind’; *üs* ‘base’; *üş(mek)* ‘to gather in a crowd’;

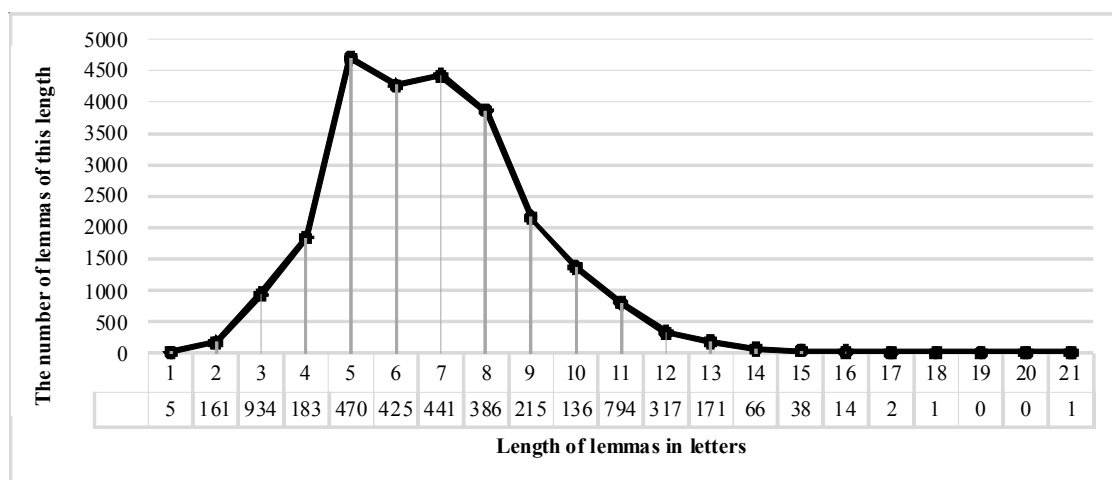


Fig. 1. Distribution of lemmata by length in the dictionary-source

*ut* ‘shame’; *ut* ‘ud’; *ut(mak)* ‘to win’; *üt(mek)* ‘to scorch, burn with flame’; *üt(mek)* ‘to win in the game’; *uy(mak)* ‘to match’; *uz* ‘good; beautiful’; *üz(mek)* ‘to upset’; *ye(mek)* ‘to eat’; *yu(mak)* ‘wash’. It follows from the table that the functional core of vocabulary in the dictionary source consists of words that are no longer than 3 letters long. After excluding the vocabulary groups described above from this set, the size of the F-core was 983 words.

### *Syntagmatic stratification of Turkish vocabulary*

Usually even explanatory dictionary stingily reflects the word compatibility. However, the degree of completeness-wealth of the representation of the syntagmatics of dictionaries does not affect the objectivity of their data: the most syntagmatically rich words remain so, no matter how many of their phraseological combinations (hereinafter PhC) are taken into account: 100 or 10. The less syntagmatically important words have no phraseological combinations. Another thing is that those words that, with a maximum of 10 PhC, had 0 phrases, in a dictionary with a maximum of 100 can have from 1 to 9 phrases. The scale and details of the syntagmatic curve change depending on the completeness of the data, but the form of the curve (in its objective part) remains the same: this is the idea of parametric “weighting” of words, and this is the objectivity of the data obtained at such weighting. The limitation of syntagmatic information in bilingual dictionaries makes the syntagmatic “weighing” of words take into account all the vocabulary evidence of compatibility presented in the dictionary: both stable phrases with the word, and the compatibility of the word in illustrative speeches.

We consider this method of syntagmatic “weighing” of words to be objective, since each lemma has theoretically equal chances to be represented in the dictionary by a phrase combination or illustrative speech. The more phraseological combinations with this lemma are presented in a dictionary article, the more its syntagmatic weight (S-weight) is. It is unlikely that the non-distinguishing of composite nominations and phrase combinations leads to errors in calculating the syntagmatic weight of a word: after all,

composite nominations are also syntagmas, so the participation of a word in composite nominations should be taken into account when studying its syntagmatic activity. On the contrary, ignoring this circumstance can lead to a distorted view of the syntagmatic activity of a word expressed by its S-weight. See Table 2 for the distribution of Turkish vocabulary about C-weight.

The data of Table 2 is clearly presented on Figure 2.

As you can see from Table 2 and Figure 2, the compatibility in the source dictionary is worked out unevenly: there is a compact syntagmatic core of 1000–2000 words and an extensive periphery. 18.573 one-word lemmas out of 25.097 (which is 74% of the source dictionary) have no information about compatibility.

In the source dictionary, the syntagmatic *dominant* with 168 PhC is the noun *el* I ‘hand(s)’, and the syntagmatic *vice-dominant* is the noun *iç* ‘interior’ with 126 PhC. Next in descending order of the number of PhC are words, among which nouns predominate: *su* ‘water’ 113; *yer* ‘earth’ 105; *iş* ‘work, labor’ 97; *ağız* I ‘mouth, jaw’ 91; *ayak* ‘leg(s)’ 82; *dil* ‘tongue’ 82; *yüz* II ‘face’ 77; *baş* ‘head’ 74; *üst* ‘upper part’ 71; *can* ‘soul’ 67; *Allah* ‘Allah, God’ 63; *kan* ‘blood’ 63; *söz* ‘word, speech’; 59; *kafa* ‘head’ 58; *akıl* ‘mind’; 56; *etmek* ‘do’ 52; *gönül* ‘soul, heart’ 50, etc. The syntagmatic core (S-core) of the source dictionary includes 1,213 words with at least three phraseological combinations. Syntagmatic *nucleus, dominant* and *vice-dominant* in the dictionary-source revealed for the first time.

### *Paradigmatic stratification of Turkish vocabulary*

Paradigmatic stratification of vocabulary involves the identification of synonymous series from the smallest (2 words) to the largest (8 words). In order to implement the paradigmatic stratification of the Turkish vocabulary, a database containing a separate record of the interpretation (or Russian equivalent) of each individual meaning of each word was created. This is based on the assumption that a polysemous word can enter the synonymous series by any of its meanings, and the maximum number of synonymous series that includes the word is theoretically limited only by the number of its meanings.

Table 2. Distribution of Turkish vocabulary by S-weight

PhC	Words	Cumul.	S-weight	PhC	Words	Cumul.	S-weight
168	1	1	0,99996	29	2	45	0,99821
126	1	2	0,99992	28	1	46	0,99817
113	1	3	0,99988	27	2	48	0,99809
105	1	4	0,99984	26	5	53	0,99789
97	1	5	0,99980	25	5	58	0,99769
91	1	6	0,99976	24	2	60	0,99761
82	2	8	0,99968	23	7	67	0,99733
77	1	9	0,99964	22	5	72	0,99713
74	1	10	0,99960	21	4	76	0,99697
71	1	11	0,99956	20	9	85	0,99661
67	1	12	0,99952	19	13	98	0,99610
63	2	14	0,99944	18	4	102	0,99594
59	2	16	0,99936	17	10	112	0,99554
58	1	17	0,99932	16	10	122	0,99514
56	1	18	0,99928	15	14	136	0,99458
54	1	19	0,99924	14	20	156	0,99378
52	1	20	0,99920	13	18	174	0,99307
50	1	21	0,99916	12	17	191	0,99239
49	1	22	0,99912	11	33	224	0,99107
47	1	23	0,99908	10	41	265	0,98944
44	1	24	0,99904	9	38	303	0,98793
42	2	26	0,99896	8	55	358	0,98574
40	2	28	0,99888	7	83	441	0,98243
39	1	29	0,99884	6	109	550	0,97809
38	1	30	0,99880	5	130	680	0,97291
35	2	32	0,99872	4	198	878	0,96502
34	4	36	0,99857	3	335	1213	0,95167
33	1	37	0,99853	2	665	1878	0,92517
32	1	38	0,99849	1	4646	6524	0,74005
31	2	40	0,99841	0	18573	25097	0,00000
30	3	43	0,99829				

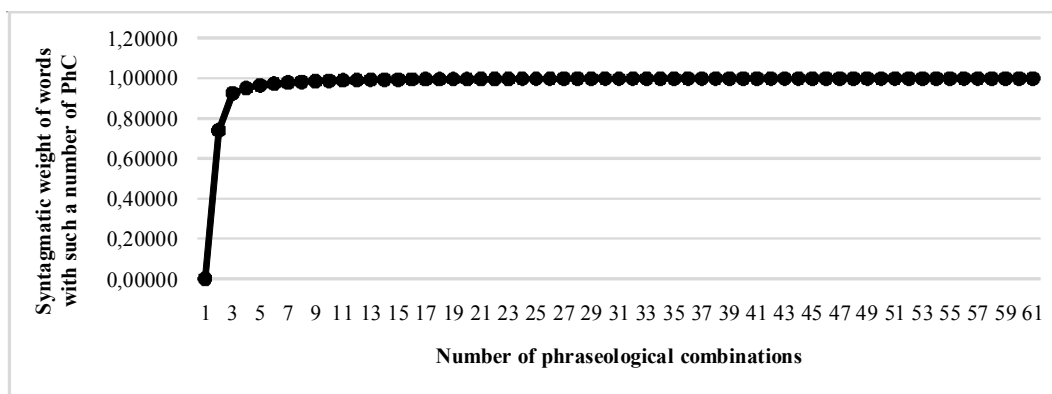


Fig. 2. S-weight of words in the dictionary-source depending on the number of PhC

In total, there were 37,909 entries in the database. (Phrases were not included in the synonymous series: *only* one-word lemmas were taken into account.) The database rows were sorted by similarity of the right (Russian, i.e. metalanguage) parts. When definitions coincided, synonyms appeared in adjacent lines. Words which definitions coincide by 100% (at least in one of the meanings) were considered operationally potential synonyms. Meanwhile, not dictionary entries were compared, but definitions of lemmas presented in dictionary entries – lexico-semantic variants (hereinafter LSV). The number of meta words (Russian content words) in the definition is taken as 100%. Paradigmatic “weighing” of vocabulary, involving human participation, is the most time-consuming and least automated part of parametric analysis. However, when we adopt the strictest possible understanding of synonymy requiring 100% convergence of definitions, we risk losing some of the synonyms represented in the dictionary. For example, *fetiş* ‘amulet’; *maskot* ‘amulet’, and *muska* ‘amulet, talisman’; *tılsım* ‘amulet, talisman’. Formally, we get two two-member synonym series. If we reduce the threshold from 100% coincidence of definitions to 50%, then this will allow us to obtain a 4-member synonymic series combining all these words. Similarly, at 100% threshold we get 3 binomial synonymous series: 1) *mecalsız* ‘powerless, infirm’; *takatsız* ‘powerless, infirm’; 2) *dingin* ‘powerless, infirm; weak’; *kudretsiz* ‘powerless, infirm; weak’ and 3) *güçsüz* ‘powerless, weak, infirm’; *kuvvetsiz* ‘powerless, weak, infirm’. When the threshold is lowered to 50% and the restriction on the order of meta words is removed, all 6 words turn out to be synonyms. This approach may seem rough, but in most cases it gives a completely acceptable result, which can be considered as materials for a dictionary of synonyms. The massive and frontal nature of the dictionary survey inevitably leads to the approximation of semantic analysis. But the task of PAL is not to compile an impeccable computer dictionary of Turkish synonyms, but to get the paradigmatic weight of Turkish words, to “weigh” Turkish words according to the paradigmatic parameter.

Turkish languages, to which Turkish belongs, have their own specificity and, although the

scientific validity of PAL has been repeatedly proven and tested, the application of PAL to each new type of language requires “adjustments to linguistic reality”, which we did when calculating F-weight, shortening the verbs with the affix *-mak/-mek*. When researching the paradigmatic parameter of the vocabulary of any language (including Turkish) the following restriction is imposed on the concept of synonyms: **words with different roots are recognized as synonyms**. As a result of the solutions described above, “synonyms” with the same root are excluded from the series characterized by an operationally understood identity of semantics. Of all the word families in the formally identified synonymic series, one (as a rule, the shortest and least marked) word remains. A marked word is considered to have any restrictive or stylistic markings. The derived word, by the presence of an additional affix (and its inherent meaning), is marked in relation to the producing one. For example, from the synonymic series with the meaning ‘healthy’: *esen, iyi, pürsıhhat, sağ, sağlam, sağlıklı, salim, sıhhatli*, the lemmas *sağlam, sağlıklı* are excluded and one lemma *sağ* is left. The number of the synonyms changes, there is: not 8, but 6. Similarly to the synonymous series discussed above, the dimension of the synonymous series with the meaning ‘critical’ also changes: *eleştirel, eleştirici, eleştirmeci, eleştirmeli, kritik, tenkidî, tenkitçi*. The variants *eleştirici, eleştirmeci, eleştirmeli* are excluded, the shortest variant *eleştirel* is left. The *tenkitçi* variant is excluded from the *tenkidî-tenkitçi* pair. As a result, the number of synonyms meaning ‘critical’ is reduced from 7 to 3: *eleştirel, kritik, tenkidî*.

We take the dictionary source for granted by examining all the words presented in it. If there are markers in the dictionary that indicate the archaic and outdated nature of words, we can take them into account, if there are no such markers, we analyze what the dictionary gives. The application of the principles and approaches described above allowed us to obtain the results presented in Table 3.

The distribution of words by P-weight is clearly presented on Figure 3.

Paradigmatic dominant vocabulary is marked by a 8-member synonymous series in the source dictionary 4 (P-weight 0.99988): ‘strength, power’: *çelim, erk, güç, kudret, kuvvet, mecal,*



Table 3. Stratification of Turkish vocabulary by the size of synonymous series

Words	Series	Cumul.	P-weight
8	1	1	0,99997
7	4	5	0,99986
6	5	10	0,99971
5	17	27	0,99922
4	79	106	0,99694
3	339	445	0,98714
2	1916	2361	0,93177
1	32241	34602	0,00000

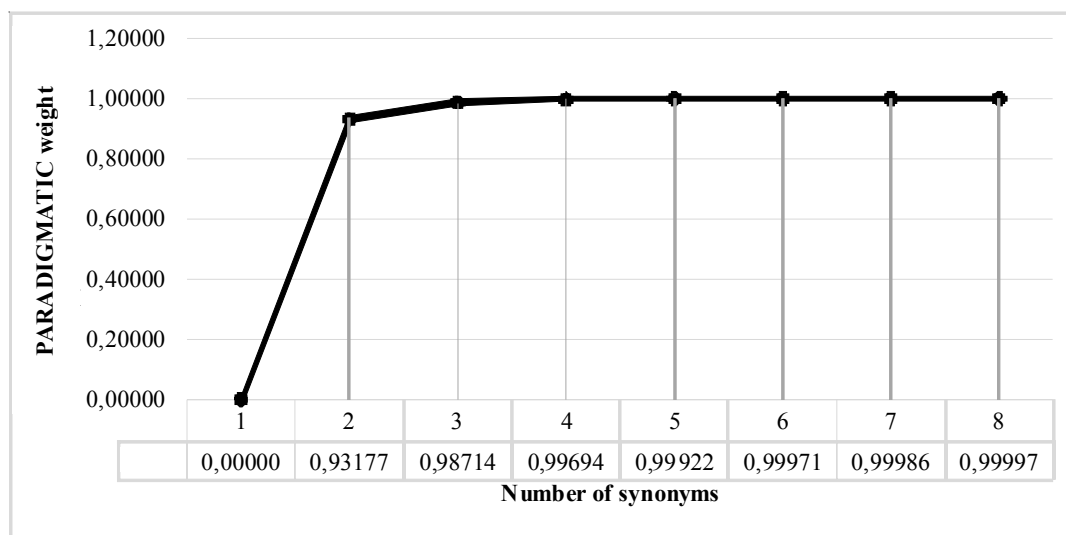


Fig. 3. Distribution of Turkish vocabulary by P-weight

*pehlivanlýk, zor*. The paradigmatic vice-dominants of Turkish vocabulary are represented by 7-member synonymous series with the meanings: ‘chest’ *bađır, döş, göğüs, koyun, meme, sadır, sine*; ‘sad’: *hüzünlü, içli, kederli, mađmum, mahzun, pürmelâl, üzgün* and ‘carefree’: *ferah, gairesiz, gamsız, geniş, kedersiz, meraksız, üzüntüsüz*.

The words *ferah* and *geniş* in this synonymous series may seem foreign. If we take into account only their first meanings, this is indeed the case: they form their own synonymous series with the meaning ‘wide, spacious, roomy’, characterizing rooms, but not people. However, if we pay attention to their figurative meanings: “*ferah* 2) figurative meanings ‘carefree, careless’; *geniş* 2) figurative meanings ‘carefree, careless’”, we will have to change our mind. This means that ‘latitude’ is transferred from physical space to the breadth of the human soul.

The paradigmatic vice-dominants of Turkish vocabulary are represented by 6-member synonymous series with the meanings: ‘healthy’; ‘in love’; ‘hashish’; ‘coquette’; ‘memory’. The

5-member synonymous series is represented by the meanings: ‘lightning’; ‘neutral’; ‘taste’; ‘pride’; ‘like-minded’; ‘earth’; ‘lowness, meanness’; ‘ordinary, mediocre’; ‘organ’; ‘offer’; ‘commitment’; ‘permission, approval’; ‘pimp’; ‘holiness’; ‘word’; ‘falcon’; ‘toilet, restroom’. Paradigmatic dominants in the source dictionary form the meanings of ‘strength, power’, ‘careless, carefree’, ‘chest’ and ‘sad, sorrowful’. Paradigmatic vice-dominants are 6-member synonymy series. The paradigmatic core (P-core) of the dictionary-source vocabulary consists of words that are included in all 2,360 selected synonym series.

Can you name the most important meaning in Turkish? The dictionary (Yusipova, 2005), treated with PAL, says: it is ‘strength, power’. This information is also received for the first time.

#### *Epidigmatic stratification of Turkish dictionary*

The epidigmatic depth of the source dictionary, measured by the maximum number

of meanings, is 33 meanings, which is a lot for a dictionary of such size. The distribution of words by the number of meanings is presented in Table 4.

The most polysemous words which have from 13 to 33 meanings are verbs; polysemous nouns occur in the range of 2–12 meanings. Consequently, according to the dictionary-source, the superpolysemy is a characteristic of Turkish verbs, and the superphraseology (see syntagmatic

stratification above) is a characteristic of Turkish nouns.

The data from Table 4 is clearly presented on Figure 4.

Figure 4 indicates that epidigmatics (polysemy) in the dictionary-source is worked out more evenly than (cf. Fig. 3) syntagmatics (compatibility).

The most polysemantic word in the dictionary-source (33 meanings) – **E-dominanta** – is the verb

Table 4. Distribution of words by the number of meanings in the dictionary (Yusipova, 2005)

Meaning	Words	Cumul.	E-weight	Example	Meaning
33	1	1	0,9999	<i>çıkmaq</i>	go
29	1	2	0,9999	<i>çekmek</i>	pull, drag
26	1	3	0,9998	<i>gelmek</i>	to come
23	1	4	0,9998	<i>çıkarmak</i>	pull out, take out
21	1	5	0,9998	<i>tutmak</i>	hold on, hold
19	1	6	0,9997	<i>almak</i>	take
18	1	7	0,9997	<i>yapmak</i>	do; perform
16	2	9	0,9996	<i>atmak; düşmek</i>	throw, fall
15	1	10	0,9996	<i>olmak</i>	be, happen
14	2	12	0,9995	<i>vurmak; kaldırmak</i>	beat, hit; raise
13	3	15	0,9994	<i>açmak; geçmek</i>	open, move on
12	7	22	0,9991	<i>kol</i>	hand
11	6	28	0,9988	<i>taban</i>	sole, foot
10	14	42	0,9983	<i>iç</i>	inside, inside
9	25	67	0,9973	<i>yüz</i>	face
8	33	100	0,9960	<i>baba</i>	father, dad
7	52	152	0,9939	<i>taş</i>	stone
6	91	243	0,9903	<i>boğaz</i>	throat, throat
5	245	488	0,9805	<i>ot</i>	grass
4	536	1024	0,9592	<i>öz</i>	native (relatives)
3	1449	2473	0,9014	<i>et</i>	meat
2	5613	8086	0,6778	<i>ay</i>	moon
1	17011	25097	0,0000	<i>aş</i>	food

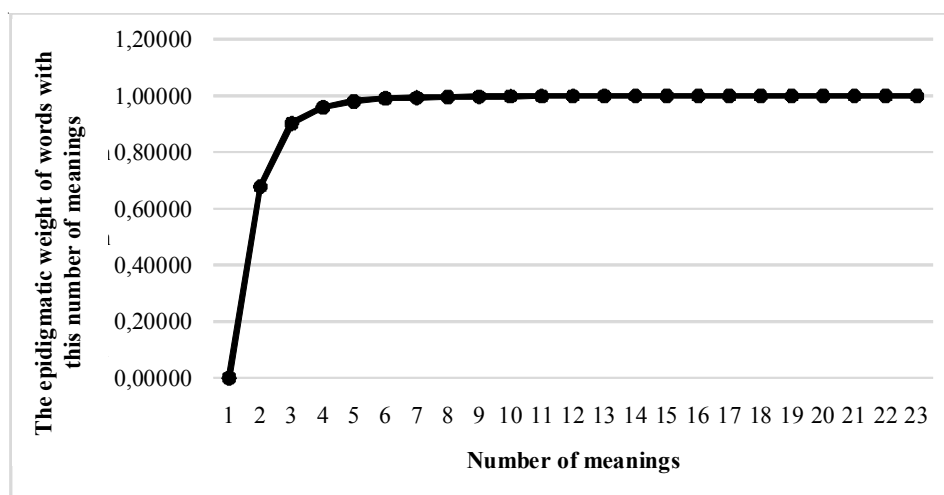


Fig. 4. Dependence of E-weight words on the number of meanings in the dictionary source

**çıkmaq** 1) ‘go’. The second most polysemantic word in the dictionary-source (29 meanings) – **E-vice-dominanta** – is the verb: **çekmek** 1) ‘pull, drag.’ We take 2473 words with 4 or more meanings as the epidigmatic core of the source dictionary. All this information about Turkish vocabulary was received for the first time.

### *Parametric stratification of Turkish lexicon*

Now we have come to the culmination of our research, in which the particular parametric stratification of Turkish words in (Yusipova, 2005) develops into a monolith of the systemic stratification of Turkish vocabulary. Moreover, we get an idea of the system stratification of the Turkish dictionary and the place of each word in this stratification.

The integral parametric weight of words was calculated as follows. For each of the particular parameters, sets of words (about 1000 words) with the maximum particular weight were taken: F-core was 983 words (2–3 letters long), C-core was 1213 words (the number of PhC from 3 to 168), the E-core consisted of 1024 words (meanings from 4 to 33), the P-core was 2,360 synonym series (2–7 synonyms). The addition of particular weights for the words included in these sets gave a picture presented in Table 5.

Words with an I-weight, rounded to 4, make up the Small parametric core. Words with an E-weight, rounded to 3 or more, make up the Middle parametric core. Words with an I-weight, rounded to 2 or more, make up the Large parametric core of the dictionary. Finally, words with an E-weight, rounded to 1 or more, make up the parametric core of the dictionary-source. Words that do not get into the core according to any of the parameters, make up the system periphery of the dictionary.

The Small parametric core of the source dictionary contains the following 140 words (after the meaning the integral weight of the word is given): **iş** ‘work, labor’ 3,95731; **üst** ‘the upper part, the top’ 3,94826; **ek** ‘supplement, app, addition’ 3,94325; **el II** ‘stranger’ 3,93964; **er I** ‘man’ 3,93560; **can** ‘soul’ 3,93319; **top** ‘ball’ 3,93175; **dil** ‘language’ 3,91997; **zor** ‘difficulty’ 3,91953; **iç** ‘inside’ 3,91794; **ruh** ‘soul, spirit’ 3,91690; **iyi** ‘good’ 3,91115; **almak** ‘take’ 3,91053; **düz I** ‘smooth, even, flat’ 3,90814; **ana** ‘mother’ 3,89777; **ak** ‘white’ 3,89579; **dip** ‘bottom’ 3,89431; **kaş** ‘eyebrow’ 3,88531; **ağız I** ‘mouth’ 3,88178; **baş** ‘head(also figurative meaning)’ 3,88119; **yüz II** ‘face’ 3,87944; **alt** ‘bottom’ 3,87801; **kol** ‘arm’ 3,87769; **bakmak** ‘look’ 3,87753; **ip** ‘rope’ 3,87499; **sıra** ‘row’ 3,87492; **tek I** ‘the only one’ 3,87482; **yapmak** ‘do, make perform’ 3,87458; **gelmek** ‘to come, to arrive from somewhere’ 3,87343; **dem I** ‘breath, sigh’ 3,87196; **tutmak** ‘hold’ 3,87171; **yan** ‘side’ 3,87136; **durmak** ‘stand, be / remain motionless’ 3,86880; **kalmak** ‘stay’ 3,86402; **kör** ‘blind’ 3,86132; **arka** ‘back’ 3,86131; **diş** ‘external / exterior side, external / exterior appearance, appearance’ 3,86064; **tam** ‘full, whole’ 3,85897; **kuru** ‘dry’ 3,85895; **baba** ‘father, dad’ 3,85860; **adam** ‘person’ 3,85812; **mal** ‘property, state’ 3,85681; **kök** ‘root, rhizome’ 3,85522; **çekmek** ‘pull, drag’ 3,84741; **vurmak** ‘beat, hit’ 3,84701; **yanmak** ‘burn, light up’ 3,84482; **dava** ‘lawsuit’ 3,84243; **gün** ‘day’ 3,84016; **hak I** ‘rights’ 3,83594; **asıl** ‘base, basis’ 3,83334; **boy II** ‘height’ 3,83274; **kırmak** ‘smash, break’ 3,83247; **kötü** ‘bad’ 3,83242; **ham** ‘unripe (about fruit)’ 3,82805; **usta** ‘master, craftsman, expert in his field’ 3,82502; **kalp I** ‘heart’ 3,82425; **ateş** ‘fire’ 3,82302; **dam I** ‘roof’ 3,81976; **ayrı** ‘separate, detached’ 3,81513; **dost** ‘friend’ 3,81513; **pis** ‘dirty, stained’ 3,81458; **sırt** ‘spin’ 3,81361; **ayak** ‘leg, paw (animal) foot (insect)’

**Table 5. Stratification of the vocabulary-source (Yusipova, 2005) by rounded integral weight**

Sets				IntRound	R.R. Jusipova Dictionary		
Dictionary	Large	Middle	Small	Weight	Words	Cumul.	AccNum
Core	Core	Core	Core	4	140	140	0,56%
			Periphery	3	490	630	2,51%
		Periphery	Periphery	2	2604	3234	12,89%
			Periphery	1	3627	6861	27,34%
Periphery			0	18236	25097	100,0%	

*Note.* *IntRound* – Integral, total parametric weight of words, rounded to whole; *AccNum* – accumulated number of words = lemmas in the database.

3,80493; **orta** ‘middle’ 3,80413; **hava** ‘air’ 3,80189; **sıkı** ‘tight, narrow’ 3,79863; **âlem** ‘world’ 3,79860; **iğne** ‘needle’ 3,79230; **kese I** ‘bag’ 3,79071; **oyun** ‘game’ 3,77923; **askı** ‘hanger, hook (for hanging clothes)’ 3,77313; **kara II** ‘black’ 3,76731; **ağır** ‘heavy’ 3,76660; **yapı** ‘building, construction’ 3,76301; **sert** ‘hard, solid’ 3,76270; **ocak I** ‘hearth, furnace, oven, stove’ 3,75823; **yürümek** ‘go, move, walk’ 3,75481; **boya** ‘paint’ 3,75086; **kıyı** ‘coast’ 3,74652; **sulu** ‘juicy’ 3,74134; **küme** ‘heap, pile’ 3,72010; **düzen** ‘order’ 3,67181; **yatak** ‘bed’ 3,66817; **dünya** ‘world, universe, earth’ 3,66814; **çatal** ‘fork’ 3,65547; **fitil** ‘wick, cord’ 3,65185; **kabak** ‘courgette, pumpkin’ 3,65113; **resim** ‘picture’ 3,65113; **ciğer** ‘lungs’ 3,64503; **düşük** ‘low’ 3,63775; **hanım** ‘khanim, khanum, mistress’ 3,63775; **yavaş** ‘slow’ 3,63775; **kadın** ‘woman’ 3,63208; **sinir** ‘nerve’ 3,63069; **örnek** ‘sample, model’ 3,62468; **içeri** ‘inside’ 3,61627; **parti** ‘party’ 3,61121; **aşağı** ‘bottom, bottom part’ 3,60523; **çukur** ‘pit, depression, excavation’ 3,60292; **canlı** ‘live’ 3,60009; **zaman** ‘time, period’ 3,60001; **küçük** ‘small’ 3,59901; **karın** ‘belly’ 3,59790; **yağlı** ‘fatty, oily’ 3,59698; **karar** ‘solution’ 3,59618; **sıcak** ‘heat’ 3,59618; **doğru** ‘Straight’ 3,59595; **güzel** ‘beautiful’ 3,59459; **kanlı** ‘bloodied, in blood’ 3,59459; **çıkış** ‘exit’ 3,59387; **kızıl** ‘bright red, red’ 3,59387; **takım** ‘group, company, circle of persons, team’ 3,59208; **demir** ‘iron’ 3,59184; **kâğıt** ‘paper’ 3,59184; **şeker** ‘sugar’ 3,59033; **gedik** ‘slit, crevice, crack’ 3,59025; **tarak** ‘comb’ 3,58869; **duman** ‘smoke’ 3,58814; **kenar** ‘Edge’ 3,58507; **çevre** ‘circumference’ 3,58049; **kalınlık I** ‘thick’ 3,58049; **kanat** ‘wing’ 3,58049; **telli** ‘fibrous’ 3,58049; **kulak** ‘ear’ 3,57817; **hazır** ‘ready’ 3,57411; **kırık I** ‘broken’ 3,56953; **salma** ‘let, let go’ 3,56953; **pamuk** ‘cotton’ 3,56897; **kesme** ‘slaughter’ 3,56742; **oğlan** ‘boy’ 3,56742; **dalga** ‘wave’ 3,56347; **idare** ‘management, guide’ 3,55407; **tulum** ‘waterskin’ 3,55407; **bebek** ‘infant, baby’ 3,54606; **damla** ‘drop’ 3,54606; **rahat** ‘rest, tranquility’ 3,54606; **çalım** ‘boasting, bragging, arrogance’ 3,53271; **cephe** ‘facade’ 3,53271; **ortak** ‘partner, companion, accomplice’ 3,53271; **toprak** ‘land’ 3,50124.

Since the purpose of PAL is to identify the cores of the lexical-semantic system of Turkish language, the consideration excludes lemma-phrases, which are means of secondary nomination, and only one-word lemmas are taken

into account. Are we not distorting the real picture of the lexical-semantic system of language? No: the “Frequency Dictionary of Turkish Written Language” made on a sample of 1 million word-uses and numbering 22,693 words (Göz, 2003) contains 3,863 composite nominations – 17% of the total dictionary. At the same time, the total frequency of these nominations is 30,480 word-use. Consequently, composite nominations, covering only 3% of the Turkish text, are low-use peripheral units of Turkish vocabulary, and their exclusion from consideration cannot significantly affect the selection of the core of the Turkish language lexical and semantic system.

The author of the “Frequency Dictionary of Turkish Written Language” writes in the foreword: “This dictionary was created twice. The first study from early 1997 to the end of 1999 based on written publications was abandoned. Groups of 2 or 3 words (e.g. *acil servis, a qk hava sinemas*) were counted as one unit there, while in the “Turkish Language Spelling Guide” of the Turkish Language Association (TLSG TLA) they were counted as the independent words *acil, servis, a’k, hava, sinema*. Therefore, we decided to start working again” (our translation. – Z. D., A. K.) (Göz, 2003). Recent borrowings from the English language and Greek-Latin internationalisms (for example, *know-how, stand-by, post-scriptum*) were also excluded from further consideration. Turkish words written with a hyphen (for example, *sıfat-fiil* ‘gram. the participle’ or *tink-tank* ‘spoken bosses’) were taken into account during the analysis.

Since the purpose of the study is the cores of the lexical-semantic system of language, the consideration excludes words that do not carry the actual lexical semantics and are not primary names and verbs, including numerical, adverbs, pronouns, imitatives, predicatives and function words. In Turkish, there is a kind of parts of speech syncretism of adjectives and adverbs that differ not formally, but by their position in the sentence (compatibility). Words used not only as adverbs, but also as adjectives (e.g. *hızlı* ‘fast, impetuous, choppy’; ‘fast, impetuously, choppy’; ‘strongly, with all their might’) were included into the database, their adverbial meanings were taken into account.

We do not have statistics on the parts of speech in the Turkish dictionary, but in the dictionaries of the Russian standard language noun adverbs are 1.58% of the dictionary (Obratnyy slovar..., 1974, p. 944). In small Romanesque-Russian dictionaries, the representation of adverbs is as follows: in Romanian – 3%, in Italian – 2%, in Portuguese and French – 1% each and in Spanish – 0.48% [Titov, 2002, p. 186]. It is unlikely that in Turkish dictionaries these proportions are significantly different. So the exclusion of adverbs hardly damages the selection of the lexical-semantic core of the Turkish language.

Lexical semantics is concentrated in nouns, adjectives and verbs; adverbs borrow it from them through suffixation, reduplication, isolation, etc. Thus, the exclusion of adverbs from the lexical core in the parametric analysis of Turkish vocabulary cannot distort the lexical system of the Turkish language, also because the lexical semantics of the adverb is not independent, but is derived from names and verbs by which it will be presented. The adverb “*ok* 1) a lot, 2) very, 3) long, 4) more than...” does not contain lexical semantics, performing LF (lexical function) Magn [Melchuk, Zholkovskiy, 1984]. Although this function is called “lexical”, it actually carries a grammatical meaning and refers not to vocabulary, but to the grammar of the language. In grammar it is impossible to do without it, in vocabulary – it is possible. Its antonym – the word *az* 1) insufficient, insignificant, meager, minuscule, 2) containing / having a small amount of something, 3) little, a little, 4) less” is taken into consideration, but not because it performs LF AntiMagn, but because it has the lexical meaning of the adjective: ‘insufficient, insignificant’. The range of parts of the speech is entirely determined by the interpretations taken by R.R. Yusipova in her dictionary. Predicatives *var* ‘there is, there are’, *yok* ‘there is not, none’, *gerek* ‘necessary’, *lâzım* ‘necessary’ have not lexical, but grammatical meanings: of the presence-absence or meaning of modality. It is illogical and impractical to include them in the lexical and semantic core of the language. Since the core of the lexical-semantic system is an appellative vocabulary, proper names (*onyms*, as opposed to common names – *appellatives*) are excluded from consideration, including ethnonyms – names of peoples, names of months, days of the week, letters, notes, etc.

## Conclusion

So, we have analyzed the largest (25,097 words) and most modern of the Turkish-Russian dictionaries (Yusipova, 2005) by parametric analysis – PAL and received verifiable and therefore objective information about the system organization of the Turkish vocabulary and the role of each of the full-meaning words of the source dictionary in the organization of the lexico-semantic system of the Turkish language.

How is it customary to describe vocabulary? Let’s take for example a textbook on the Lexicology of the English language [Kharitonchik, 1992]. Let’s look at the Table of Contents “Lexical units of language” (word, native and borrowed vocabulary), Here – “Meanings of lexical units” (aspects and types of meanings). “Polysemy” (intraverbal derivational – epidigmatic – connections of meanings). “Homonymy” (connections of values by a random coincidence of the form). “Semantic connections of words” (paradigmatic connections – synonymy, antonymy, hypo-hyperonymy). “Word formation” (inter-word derivational connections: word-formation nest, word-formation paradigm, word-formation category). This also includes “Methods of word formation”, i.e. the creation of inter-word derivational connections (affixation, conversion, word composition). “Compatibility of lexical units” (syntagmatics – rules of word compatibility, phraseology) [Kharitonchik, 1992, pp. 228-229].

As we can see, the system-forming connections (syntagmatic, paradigmatic and epidigmatic) are described. It is even pointed out that these connections correlate with the frequency of words: frequency words are ambiguous, short words have more meanings than long ones, frequency words are native, neutral). In conclusion, it says: “it is possible to identify layers of vocabulary in which the intended correlations are the most probabilistic and form a bundle of interdependencies, the most obvious and clearly traceable. These are the most stable layers of vocabulary, which in linguistics have been described as the “main vocabulary” of the language. <...> It seems appropriate not to reject the concept of the basic vocabulary, but to conduct research in which to experimentally establish the signs of the units that make up it” [Kharitonchik, 1992, p. 224].

This is exactly what we have done by means of parametric analysis based on the material of the Turkish-Russian dictionary (Yusipova, 2005). As Z.A. Kharitonchik rightly points out, “The **main peculiarity of the system of lexical units... lies in the very inventory of nominative means... of the language and the relations that are established between them**” (highlighted by us. – Z. D., A. K.) [Kharitonchik, 1992, p. 226].

In vain we would look in the textbook [Kharitonchik, 1992] for a systematic description of this inventory, which is needed not only for teaching English. It is needed in order to isolate the lexical cores of languages with a size of about 1000 units, to make possible the typological lexicology of the languages of the world and the historical lexicology of each of the languages with a sufficiently long written tradition.

Our parametric description of the Turkish vocabulary and its result – obtaining the parametric core of the Turkish vocabulary – is a step towards the comparative lexicology of the Turkic languages. Moreover, it is a contribution to the lexical typology of the languages of the world.

Parametric analysis of the vocabulary of the Turkish-Russian dictionary (Yusipova, 2005) made it possible to carry out a systematic stratification of the Turkish vocabulary and obtain 4 systemic cores: Small – 140 words, Middle – 630 words, Large – 3234 words and the core of the Dictionary – 6861 words.

Now we can answer a question that has not even been asked before: which word of the Turkish language is the most important (systemically)? This is a lexico-semantic *dominant*. The *dominant* feature of the lexico-semantic system of the Turkish language according to the dictionary (Yusipova, 2005) was the word *iş* ‘job, work’ with I-weight – 3,957, and the *vice-dominant* – the word *üst* ‘the upper part, the top’ with I-weight – 3,948.

The near-term perspective of the study is to select the parametric core of the Kyrgyz language, the further one is to select the parametric cores of vocabulary of other Turkic languages, represented by Turkic-Russian dictionaries.

#### ACKNOWLEDGEMENTS

We are grateful to Elena Markovna Napolnova (Özyeğin Üniversitesi Yabancı Diller Yüksek Okulu,

İstanbul, Türkiye), who drew our attention to “Frequency Dictionary of Turkish Written Language” by İlyas Göz (Göz, 2003) for providing access to the dictionary.

#### REFERENCES

- Bugaev V.P., 2006. *Parametricheskii analiz tiurkskogo slovaria* [Parametric Analysis of Turkish Vocabulary]. Voronezh. 80 p.
- Burlak S.A., Starostin S.A., 2005. *Sravnitelno-istoricheskaya lingvistika* [Comparative-Historical Linguistics]. Moscow, Akademiya Publ. 432 p.
- Kharitonchik Z.A., 1992. *Lexikologiya angliyskogo yazyka* [Lexicology of the English Language]. Minsk, Vysshaya shkola Publ. 229 p.
- Kononov A.N., 1997. *Tureckiy yazyk v mire: Tiurkskie yazyki* [Turkish Language in the World: Turkic Languages]. Moscow, Indrik Publ., pp. 394-411.
- Kretov A.A., 2011. Problemy kvantitativnoy leksikologii slavianskikh yazykov [Problems of Quantitative Lexicology of Slavic Languages]. *Voprosy yazykoznaniiya* [Topics in the Study of Language], no. 1, pp. 52-65.
- Kretov A.A., Voevudskaya O.M., Merkulova I.A., Titov V.T., 2016. *Edinstvo Evropy po dannym lexiki* [Unity of Europe According to Vocabulary]. Voronezh, Izd. dom VGU. 412 p.
- Kretov A.A., 2017. V.T. Titov i parametricheskii analiz lexiki [V.T. Titov and Parametric Analysis of Vocabulary]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkulturnaya kommunikatsiya* [Proceeding of Voronezh State University. Linguistics and Intercultural Communication], no. 4, pp. 5-9.
- Kretov A.A., Cherechecha A.D., 2020 *Teoreticheskie problemy lexiko-semanticheskoy tipologii (na primere kavkazskikh yazykov)* [Theoretical Problems of Lexico-Semantic Typology (On the Example of Caucasian Languages)]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkulturnaya kommunikatsiya* [Proceeding of Voronezh State University. Linguistics and Intercultural Communication], no. 1, pp. 6-15. DOI: <https://doi.org/10.17308/lic.2020.1/2724>
- Kretov A.A., Gasuns M.Yu., Leonchenko V.V., 2021. Parametricheskii analiz «Sanskritsko-russkogo slovarya» V.A. Kocherginoy [Parametric Analysis of the “Sanskrit-Russian Dictionary” by V.A. Kochergina]. Kogan A.I., Panin A.S., eds. *Problemy obshchey i vostokovednoy lingvistiki. Sochetaemost yazykovykh edinits i yazykovye modeli. Pamyati Z.M. Shalyapinoy (1946–2020)*

- [Problems of General and Oriental Linguistics. Compatibility of Language Units and Language Models. In Memory of Z.M. Chaliapina (1946–2020)]. Moscow, Izd-vo RAN, pp. 287-301. DOI: 10.31696/978-5-907543-08-9-287-301
- Melchuk I.A., Zholkovskiy A.K., 1984. *Tolkovyy kombinatornyy slovar sovremennogo russkogo yazyka* [Explanatory Combinatorial Dictionary of Modern Russian]. Wien, Ges. zur Foerderung slawist. Studien. 992 p. (Wiener Slawistischer Almanach, Sonderband, 14).
- Merkulova I.A., 2018. *Lexicheskaya nukleologia slavianskikh yazykov: avtoref. dis. ... d-ra filol. nauk* [Lexical Nucleology of Slavic Languages. Dr. philol. sci. abs. diss.]. Voronezh. 35 p.
- Semenova I.D., 2018. *Parametricheskyy analiz lexiki karachaevo-balkarskogo yazyka na tyurkskom fone: avtoref. dis. ... kand. filol. nauk* [Parametric Analysis of the Vocabulary of the Karachay-Balkar Language on a Turkic Background. Cand. philol. sci. abs. diss.]. Moscow. 22 p.
- Titov V.T., 2002. *Obshchaya kvantitativnaya lexicologia romanskikh yazykov* [General Quantitative Lexicology of Romance Languages]. Voronezh, Izd-vo Voronezh. gos. un-ta. 240 p.
- Titov V.T., 2004a. *Chastnaya kvantitativnaya lexicologia romanskikh yazykov* [Private Quantitative Lexicology of Romance Languages]. Voronezh, Izd. dom VGU. 552 p.
- Titov V.T., 2004b. *Kritika lingvisticheskikh istochnikov kak razdela lingvisticheskogo prognostitsizma* [Criticism of Linguistic Sources as a Section of Linguistic Prognosticism]. *Problems of Linguistic Prognosticism*, no. 3, pp. 232-274.
- Voevodskaya O.M., 2015. *Kontseptsiya ideograficheskogo slovarya osnovnogo lexiceskogo fonda germanskikh yazykov* [Concept Ideographic Dictionary of the Main Lexical Fund of Germanic Languages]. Moscow, Nauka Publ., Unipress Publ. 343 p.

#### SOURCES AND DICTIONARIES

- Göz İ. *Yazılı Türkçenin kelime sıklığı sözlüğü*. Ankara, Turkish Language Institution, 2003. XV, 576 p.
- Obratnyy slovar russkogo yazyka* [Reverse Dictionary of the Russian Language]. Moscow, Entsyclopedia Publ., 1974. 944 p.
- Rybalchenko T.E. *Turetsko-russkiy i russko-turetskiy slovar* [Turkish-Russian and Russian-Turkish Dictionary]. Moscow, Russkiy yazyk Publ., 2001. 696 p.
- Shcherbinin V.G. *Kratkiy turetsko-russkiy slovar* [Brief Turkish-Russian Dictionary]. Moscow, Russkiy yazyk Publ., 1977. 405 p.
- Yusipova R.R. *Turetsko-russkiy slovar* [Turkish-Russian Dictionary]. Moscow, Russkiy yazyk Publ., 2005. X, 694 p.

#### Information About the Authors

**Zamira K. Derbisheva**, Doctor of Sciences (Philology), Professor, Department of Philology, Kyrgyz-Turkish Manas University, Chyngyz Aitmatov Campus (Djal), 720038 Bishkek, Kyrgyzstan, zamira.derbisheva@manas.edu.kg, <https://orcid.org/0000-0003-4333-4425>

**Alexey A. Kretov**, Doctor of Sciences (Philology), Professor, Department of Theoretical and Applied Linguistics, Voronezh State University, Lenina Sq, 10, 394077 Voronezh, Russia, kretov@rgph.vsu.ru, <https://orcid.org/0000-0002-1474-3177>

#### Информация об авторах

**Замира Касымбековна Дербисева**, доктор филологических наук, профессор кафедры филологии, Кыргызско-Турецкий университет «Манас», Кампус им. Ч. Айтматова (Джал), 720038 г. Бишкек, Кыргызстан, zamira.derbisheva@manas.edu.kg, <https://orcid.org/0000-0003-4333-4425>

**Алексей Александрович Кретов**, доктор филологических наук, профессор кафедры теоретической и прикладной лингвистики, Воронежский государственный университет, пл. Ленина, 10, 394077 г. Воронеж, Россия, kretov@rgph.vsu.ru, <https://orcid.org/0000-0002-1474-3177>