

THE LANGUAGE SITUATION OF AN ETHNIC COMMUNITY (ON THE MATERIAL OF THE TATAR LANGUAGE WRITTEN CORPUS)

Tauzikh Ibragimov,

Kazan Federal University,
18 Kremlevskaya Str., Kazan, 420008, Russia,
Tavzikh.Ibragimov@ksu.ru.

Mansur Saikhunov,

Institute of Computer Science, Tatarstan Academy of Sciences,
36A Levobulachnaya Str., Kazan, 420111, Russia,
6688000@gmail.com.

The problem of preserving a language as a way of seeing the world is becoming increasingly urgent under the conditions of globalization. This paper presents the results of the authors' attempt to penetrate the spiritual world of the nation by using the features of the Tatar language written corpus (The Corpus of Written Tatar). The Corpus in its modern version is a general totality of word forms. It enables the researcher to identify the frequency of their use, as well as restrictions on their compatibility with preceding and subsequent word forms. By choosing appropriate word forms and using the search engine of the Corpus, the researcher can obtain information both about the structure of the language and its speakers, and also reveal ethnic and cultural values of Tatars.

Key words: the Corpus of Written Tatar, search engine, linguistic community, word form, right context, left context, frequency of use.

In certain sociolinguistic researches, the language situation is defined as *"a result of the balanced interaction between linguistic and extra-linguistic factors"* [1]. Some scholars object to this definition. However, one cannot but agree with the statement made by I.A.Baudouin de Courtenay, founder of the Kazan Linguistic School, who said, *"The language exists and changes ... according to mental and sociological laws, and sociology is identified here with the so-called psychology of peoples"* [2].

The scientific works [3] based on the analysis of population census data for 1979 and 1989 reveal that *"... the more pronounced is the loss of ethnic language, the faster is the decline of language loyalty"*. According to R.Khakimov [4], the census of 1989 showed that *"... 1 million 130 thousand Tatars out of 6 million 654 thousand do not consider Tatar their mother tongue"*. What is the reason for the decline of loyalty to the mother tongue? Why is the language ceasing to act as an ethnic integration feature? What has changed: the language or the native speakers?

The present study of the language situation (the state of the Tatar language community) is based on the Corpus of Written Tatar developed by the authors. The Corpus now includes 116 million words (word tokens), or about 10 million sentences, which contain 1 million 520 thousand word forms.

Sixty per cent of the language material consists of texts brought from numerous websites, fiction texts are an approximate 34-35% share, and 4-5% – research literature in the humanities.

The style of fiction is represented by two or three paged extracts of the works by numerous Tatar writers. These extracts are included in the Tatar Book Stock and are promotional in nature. Thus, the materials of the Corpus do not bear any hallmark of any author's individual features, which enables us to consider the Corpus of Written Tatar in its present version as an electronic encyclopedia of texts. The Corpus is sufficiently balanced and adequately represents the modern Tatar language.

The search engine enables the researcher to find a given word, a list of words that can follow a given word (right context), and a list of words that can precede a given word (left context). In addition, the corpus manager provides all the words from the list with the frequency of their use in the given combination. Thus, the researcher has the opportunity to evaluate the application of word forms following the specified word.

The features of the search engine are not limited to the functions described above. Some of them will be considered during the presentation of our results. Full information about the Corpus can be found at <http://corpus.tatfolk.ru>, access mode: Free [5].

The written Tatar language has a long history. The first written records (*Qissai Yosif* ('*Tale of Yusuf*')) by Qol Ghali) date back to the beginning of the 18th century. The inherent cultural layer of the Tatar language was represented by borrowings from Arabic-Persian and Russian. For the Tatar language, the Arabic literature was a great source of words with a predominantly religious sense. The existing large layer of the ethical and moral vocabulary was introduced into Tatar namely with Islam. The Russian language enriched the Tatar lexis with the terms of Russian and European culture and science.

Over the centuries the language has developed a rich word-building system. The logical and grammatical resources of the language are sufficient to describe the syntax in terms of formal logic. As far as the phonetic system is concerned, it has undergone the same process of evolution as every other language, according to I.A.Baudouin de Courtenay, following the principle of energy saving or simply for the convenience of pronunciation. Thus, the Tatar language has all the necessary and polished tools to fulfill its main – communicative – function. Taking this into account, it is rather puzzling that today, one out of five ethnic Tartars does not recognize the Tatar language as their mother tongue [4].

Usually, the problems associated with the change of language (linguistic denationalization, language shift) are studied in ethnolinguistics by questioning the representative portion of the population.

In this research work and while studying changes in the linguistic consciousness of the Tartars beyond the connection with its social component, i.e., from the perspective of the above-mentioned statement by I.A.Baudouin de Courtenay (much depends on the psychology of the people), we proceed from the following points:

1. It is important to find out what is happening to the linguistic community; what has caused more than 20% of Tartars to be disappointed with the environment in which they were born and are living, and the history and culture of their ancestors;

2. It is impossible to understand the psychology of the respondents, who have rejected their mother tongue, if they are asked: "What caused your transition into another language?" even in case they have a response form;

3. Therefore, as we see it, the texts, which are written on the topic of the day by people with great experience, namely in articles published in news-

papers and magazines, as well as people's destinies described in the works of the "engineers of human souls" can help to reveal the problem.

L.Bloomfield, one of the most prominent linguists of the 20th century, said at the end of his book *Language*, "*It is only a prospect, but not hopelessly remote, that the study of language may help us toward the understanding and control of human events*" [6: 559].

Bloomfield's words give reassurance that the analysis of the Corpus materials will lead to an understanding of what is happening in the ethnic community. Moreover, at present there is no other means of applying a solution to this problem.

Both the frequency of use of word forms and their compatibility give an indication of the changes taking place in the language and in the community. From this perspective, we will consider several examples.

Over the last years, we can observe the active use of the words *khäräm* ('*an unfairly obtained product or wealth*') and *khäläl* ('*an honestly, fairly obtained product or wealth*') in the Tatar language. These words penetrated into Tatar and other Turkic languages with Islam. The reason is that the religion is more concerned about the afterlife of faithful people, and, therefore, religious leaders had to say and write more about how to beware the forbidden product, i.e., *khäräm*. However, as the analysis shows, the word *khäräm* is found 614 times in the Corpus and the word *khäläl* is used 8489 times, i.e. what they should have spoken about more often is used 13 times less frequently. It can be explained only by the fact that *khäläl* product can be bought and sold.

Some confusion arises over the fact that in the Corpus the word *belem* ('*knowledge*') is found at a frequency of 70791 times, and the word *akcha* ('*money*') is used 79522 times. It appears that society cares more about money than about broadening their knowledge.

The Corpus of Written Tatar is a set of texts. The frequency of use and compatibility of the words (word forms) of the text are influenced by linguistic and extralinguistic restrictions. Not taking into account extralinguistic restrictions (based on the condition of the problem statement), the basic linguistic forms of restrictions are as follows:

- restrictions conditioned by the topic of a text;
- voice of a verb form;
- verb valency;
- word semantics;
- modal words;
- postpositions (prepositions).

Although the relationship between the words in the Tatar language is sufficiently revealed in morphology, it is often difficult to define the meaning of certain phrases. In order to solve this problem, the search engine of the Corpus is fitted with the function of extending the left or right context of a given word. We will observe this function in the following example.

Let us assume that a researcher is interested in information on the frequency with which the word form *dogasy* ('prayer') is used in the language, as well as with which words it enters the possessive relation, i.e., its left context data.

Resorting to the Corpus of Written Tatar, the researcher will get the following results:

– The word *dogasy* ('prayer') is found in the Corpus 609 times;

– The words that fill the left context, i.e., the words that clarify whose prayers the text of the Corpus refers to, are as follows: {*ana* ('mother') – 60, *äbilär* ('grandmothers') – 15, *änkäy* ('Mommy') – 4, *ananyñ* ('mother's') – 2, *analar* ('mothers') – 2, *ata-ana* ('parents') – 2, *äti-änise* ('parents') – 2}.

Having grouped the words that are similar in meaning and summed the frequency of their use, we can see that *äni dogasy* ('mother's prayer') is used 68 times, *äbilär dogasy* ('grandmothers' prayer') can be found 15 times and *äti-äni dogasy* ('parents' prayer') – 5 times.

According to the above-mentioned data, children are more attached to their mother, i.e., they are more in need of a maternal prayer. But a father's prayer is not mentioned even once. Why is that? Have fathers ceased to pray? Or do children not need fathers anymore? The first answer sounds more probable. Strangely enough, in the pre-revolutionary period, fathers were the most literate, it was mainly the fathers who visited mosques and only fathers performed Hajj.

Is it possible that they (fathers) are among those determinants that are expressed in the left context of a given word with the help of pronouns {*anyñ, alarnyñ*}, or adjectives *izge* – ('holy')?

The Corpus manager enables the researcher to become familiar with the sentences containing a given word, and a word form with an implicit meaning. Supposing the research interest is focused on sentences including the phrase *izge dogasy* ('holy prayer'). To find them, the researcher should use the word *izge* and study all the examples. This helps to discover the presence of two phrases *ätkäyñeñ izge dogasy* ('daddy's holy

prayer') and one phrase *äbilärebezñeñ izge dogasy* ('holy prayer of our grandmothers').

Modern humanity has entered the era of globalization. We can witness the process of the unification of cultures which results in the loss of the unique recognition of the world and language. These circumstances are definitely reflected in the life of any nation, including Tatars.

Within the framework of the Corpus, we have analyzed changes in social relations that occurred over the past 20-25 years. The analysis data are given in Table 1.

Table 1. Types of social relations
Mönäsäbät törläre ('types of relations')

	<i>Säüdü-ik"tisad</i> ('trade and economic relations')	<i>Khalykara</i> ('interethnic relations')	<i>Gailädäge</i> ('family relations')	<i>Kanunlashtyrylgan</i> ('relations controlled by the law')	<i>Duslyk</i> ('friendly relations')	<i>Partnerlyk mönäsäbätläre</i> ('partnership relations')	<i>Khezmat mönäsäbäte</i> ('work relations')	<i>Tolerantlyk</i> ('tolerance')	<i>Köndäshlek</i> ('competition, rivalry')
F	303	62	203	74	113	160	86	30	14

According to Table 1, the number of relation forms between members of society has increased. However, modern forms of relations unite societies rather than people themselves. Thus, interethnic relations are usually presented as treaties between governments; partnership relations are agreements between small and medium-sized enterprises; and tolerance takes place between ethnic groups, the state and religions, as well as between owners of businesses and various institutions. The relations of competition/rivalry do not involve the participation of a person. It seems that people can manifest themselves as personalities only at home, in their family.

Under these conditions, what do people believe in? Table 2 shows the data of the left context, i.e., words that answer the questions 'who?' and 'what?' and can be combined with verbs *yshanu* ('believe, trust') and *yshana* ('believes, trusts'). In these phrases the verb is "a determinant", and the noun in the dative case is "determined". In accordance with the word order peculiar to the Tatar language, the verb is placed at the end of a sentence, and consequently, after its determinants.

Table 2 summarises the combined results of the analysis of the left contexts of the word forms

yshanu and *yshana*, that occur in the Corpus 1779 and 4096 times respectively.

Table 2. The data of the Corpus analysis, based on the question “Who or what do Tatar speakers believe/trust?”

	<i>yshanu</i> ('believe, trust') 1779		<i>yshana</i> ('believes, trusts') 4096	
<i>Dingä</i> ('believe in religion')	288	160	48	46
<i>Üz-tizenä</i> ('believe in oneself')				
<i>Khorafatlarga</i> ('believe in superstitions')				
<i>Sitizgä</i> ('trust words'), <i>kitapka</i> ('trust a book')				
<i>Mogžizalarga</i> ('believe in miracles')				
<i>Kilächäkkä</i> ('believe in the future')				
<i>Yuk-barga</i> ('believe in tall stories')				
<i>Khakimiyaikä</i> ('trust the government')				
F	288	160	48	46

As you can see, in spite of everything, people still believe in themselves and in their future. However, the lack of belief in friendship and in science is rather confusing.

One of the important moments in the problem of determining the state of the language and ethnic community is identifying the ethnic and cultural values and cultural orientation of the people. This topic was recently touched upon in one of our previous works [7: 154-159]. At that time, the Corpus contained only 45 million words, and the search engine did not fulfill the operation of context expansion of a given word and did not provide researchers with a list of the word forms within its context. Currently, the Corpus comprises 116 million words, which certainly enhances its representativeness and balance.

Table 3 provides a list of “cultural values” with numerical characteristics, i.e., the indicators of the level of preference shown by the Tatar language speakers (numbers in bold indicate the quantity of uses of a given word form in the Corpus).

Table 3. The list of ethnic cultural values and their ethnicity consolidating rates

<i>Etnokyyymätälär</i> (‘ethnic values’)	<i>Etnokyyymätälär</i> (‘ethnic values’)					Overall
	<i>Tugryk</i> (‘loyalty’)	<i>Tugrylykly</i> (‘loyal’)	<i>Ikhtiram</i> (‘respect’)	<i>Khörmät</i> (‘honour’)	<i>Üsterü</i> (‘development’)	
	1955	1428	5100	16457	32782	
<i>Tel</i> (‘language’)	7	3	59	176	1246	1491 14,6%
<i>Tugan il</i> (‘motherland’)	77	49	5	21	464	616 6,0%
<i>Khezmät</i> (‘work’)	123	80	25	87	1282	1597 15,6%
<i>Millät</i> (‘nation-’)	20	54	33	233	243	583 5,7%

<i>ality’</i>							
<i>Din</i> (‘religion’)	47	28	18	135	112	340	3,3%
<i>Tugan-tumachalyk</i> (‘family ties’)	18	12	48	268	369	715	7,0%
<i>Hönäri mädäniyat</i> (‘professional culture’)	2	7	29	75	1251	1364	13,4%
<i>Äkhlylyk</i> (‘morality’)	7	25	72	279	336	719	7,0%
<i>Rukhi mädäniyat</i> (‘spiritual culture’)	388	45	63	213	122	831	8,1%
<i>Gomumkeshelek ideallary</i> (‘universal human values’)	25	0	23	77	41	166	1,6%
<i>Khakimiya atributlary</i> (‘state attributes’)	58	17	14	78	0	167	0,8%
<i>Säüdä – ik’tisad</i> (‘trade and economy’)	0	0	0	0	1621	1621	15,88%

Table 3 (the last row) contains “trade and economic relations” as one of the ethnic values. According to formal characteristics (what is most discussed in society and what people try to develop to a greater extent), it deserves inclusion in this list. At the same time, society does not show loyalty to, respect for, or honour to “trade and economic relations”, which is identified by zeros in the corresponding cells. It seems that while interviewing the Tatar language speakers, the question “which of these values unite you with your people?” would have very few affirmative responses indicating “trade and economic relations” [8: 223-239]. Thus, although “trade and economic relations” occupy the minds of a large number of people today, they still do not have the status of an ethnicity consolidating value. According to Table 3, the most significant ethnicity consolidating characteristics are inherent in such attributes of ethnic groups as language, work and professional culture.

Ethnolinguistics considers the language as the major factor in moulding the psyche of an ethnic group. The language enables peoples to accumulate the experience of previous generations. Grammatical and semantic systems have a great impact on forming the structure of native speakers’ thinking. The Corpus, as a kind of encyclopedia of electronic texts, provides great opportunities for linguists to reconstruct their language history, their ancestors’ way of life, and to understand their philosophy. Further research will definitely reveal new preferences of the ethnicity consolidating

value of the Tatars and enable us to update the results of our observations.

References

1. *Stepanov G.V.* Tipologiya yazykovykh sostoyaniy i situatsiy v stranakh romanskoj rechi. M.: Nauka, 1976. 141 s. (in Russian)
2. *Boduen de Kurten I.A.* Yazyk i yazyki. Izbrannye trudy po obshchemu yazykoznaniyu. M.: Nauka, 1963. T.2. 94 c. (in Russian)
3. *Belikov V.I.* Krysin L.P. Sotsiolingvistika. M., 2001. 439 s. (in Russian)
4. *Khakimov R.* Ne privilegii, a zashchita. Yazykovaya politika v Respublike Tatarstan. Kazan', Izd. "Margarif", 1999. 342 s. (in Russian)
5. Pis'mennyy korpus tatarskogo yazyka [Elektronnyy resurs] / Saykhunov M.R., Ibragimov T.I. – Kazan', 2014. URL: <http://corpus.tatfolk.ru> (in Russian) (date of use 15.05.2015)
6. *Bloomfield L.* Language. – London and Aylesbury: Compton Printing Ltd., 1973. 566 p. (in English)
7. *Ibragimov T.I., Saykhunov M.R.* K postroeniyu strukturno-funktsional'noj modeli tselestnoy orientatsii tatarskogo etnosa (po materialam pis'mennogo korpusa tatarskogo yazyka) // Yazyki Rossii i stran blizhnego zarubezh'ya kak inostrannye. Materialy Mezhdunarodnoj nauchno-prakticheskoy konferentsii (28-29 noyabrya 2013g.) Kazan', 2013, s.154-159. (in Russian)
8. *Guboglo M.N.* Integriruyushchaya funktsiya yazyka// Sotsiolingvisticheskie problemy razvitya yushchikhsya stran. M.: Nauka, 1975. S. 223-239. (in Russian)

ЯЗЫКОВОЕ СОСТОЯНИЕ ЭТНИЧЕСКОЙ ОБЩНОСТИ НА МАТЕРИАЛЕ ПИСЬМЕННОГО КОРПУСА ТАТАРСКОГО ЯЗЫКА

Тавзих Ибрагимович Ибрагимов,
Казанский федеральный университет,
Россия, 420008, г.Казань, ул.Кремлевская, д.18,
Tavzikh.Ibragimov@ksu.ru

Мансур Рифкатович Сайхунوف,
Институт информатики АН РТ,
Россия, 420111, г.Казань, ул.Левобулачная, д.36А,
6688000@gmail.com

Человечество сегодня живет в условиях глобализации, когда стираются границы национальных различий. Проблема сохранения языка как самобытной системы видения мира превратилась в одну из самых актуальных проблем современности. В нашей статье мы раскрыли потенциал Письменного корпуса татарского языка в деле изучения языка как средоточия духовной культуры нации.

В современной версии Письменный корпус отвечает всем требованиям репрезентативности языка, позволяя выявить возможности стилей языка через их сравнительный анализ. Корпус позволяет выявить частоту употребления каждой из словоформ, а также ограничения в их сочетаемости с предшествующими и последующими словоформами. Используя поисковую систему Корпуса для выбора надлежащих словоформ, исследователь получает информацию как о структуре языка, так и о его носителях, а также об этнокультурных ценностях татарского народа.

Ключевые слова: Письменный корпус татарского языка, поисковая система, языковая общность, словоформа, правый контекст, левый контекст, частотность употребления.

В отдельных социолингвистических работах состояние языка определяется как «некий итог сбалансированного взаимодействия языковых и внеязыковых факторов» [1]. Такое определение вызывает у отдельных ученых возражение, но вместе с тем нельзя не согласиться с утверждением основателя Казанской лингвистической школы И.А.Бодуэна де Куртенэ:

«Язык существует и изменяется ... по законам психическим и социологическим, причем социологию мы отождествляем с так называемой психологией народов» [2].

В литературе [3] на основании анализа данных переписей населения за 1979 и 1989 гг. отмечается такая закономерность: «... чем сильнее выражена утрата этнического языка, тем быст-

рее идет снижение языковой лояльности». Согласно [4], по переписи населения 1989 года «...из 6 млн. 654 тыс. татар 1 млн. 130 тыс. не считают свой язык родным». В чем причина снижения лояльности к родному языку? Почему язык перестает играть роль этноинтегрирующего признака? Изменился язык или изменились носители языка?

В настоящей работе изучение состояния языка (состояния татарской языковой общности) проводится на базе разработанного авторами Письменного корпуса татарского языка. Корпус к данному моменту включает 116 млн. слов (словоупотреблений) или около 10 млн. предложений, в которых насчитывается 1 млн. 520 тысяч словоформ. Языковой материал более чем на 60% состоит из текстов, составляющих содержание многочисленных WEB сайтов, на долю стиля художественной прозы приходится приблизительно 34-35% текстов и 4-5% научной литературы гуманитарного профиля.

Стиль художественной прозы представлен двух-трехстраничными отрывками из произведений большого числа татарских писателей. Эти отрезки включены в Татарский книжный фонд и носят рекламный характер. Материал Корпуса по этой причине не несет отпечатка каких бы то ни было индивидуальностей, что позволяет рассматривать Письменный корпус в настоящей его версии как электронную энциклопедию текстов. Корпус в достаточной степени сбалансирован и адекватно представляет современный татарский язык.

Поисковая система дает возможность исследователю найти заданное слово; перечень слов, способных следовать за заданным словом (правый контекст), а также перечень слов, способных предшествовать заданному слову (левый контекст). Корпусный менеджер, помимо того, каждому слову из списка приписывает частоту его употребления в данном сочетании. Таким образом, исследователю предоставляется возможность оценить употребляемость данной словоформы в языке, а также употребляемость следующих за заданным словом словоформ.

Возможности поисковой системы не ограничиваются указанными функциями. Некоторые из них будут рассмотрены в процессе изложения полученных нами результатов. Полные сведения о Корпусе можно найти по адресу <http://corpus.tatfolk.ru>, режим доступа: свободный [5].

Письменный татарский язык имеет многовековую историю. Первые письменные памятники («Кысса-и Йусуф» Кул Гали) относятся к началу XIII века. Неотъемлемым культурным пластом татарского языка явились заимствования из арабо-персидского и русского языков. Через арабскую письменность в татарский язык проникли преимущественно слова религиозного содержания. Существующий и поныне большой пласт этико-нравственной лексики был привнесен именно исламом. Через русский язык татарская лексика обогатилась терминами русской и европейской культуры и науки.

За многовековую историю язык выработал богатую словообразовательную систему. Логико-грамматические ресурсы языка являются достаточными для описания синтаксиса в терминах формальной логики. Что касается фонетической системы, то она в процессе своей эволюции прошла тот путь, по которому, согласно И.А.Бодуэну де Куртенэ, следует каждый язык, подчинясь принципу экономии энергии или просто удобопроизнесению. Таким образом, татарский язык обладает необходимым и отшлифованным инструментарием для выполнения своей основной – коммуникативной – функции. На этом фоне вызывает некоторое недоумение тот факт, что сегодня каждый пятый татарин не признает его родным языком [4].

Обычно проблемы, связанные со сменой языка (языковая денационализация, сдвиг языка), в этносоциолингвистике изучают путем проведения анкетирования репрезентативной части населения.

В данной работе при исследовании изменений в языковом сознании татар вне связи с его социальной составляющей, то есть в ракурсе приведенного высказывания И.А.Бодуэна де Куртенэ (многое зависит от психологии народа), мы исходим из следующих моментов:

1. Важно выяснить, что происходит с языковой общностью; чем вызвано разочарование более 20% татар и неприятие ими той среды, в которой они родились и жили (живут), истории, культуры своих предков.

2. Проведение анкетирования среди отказавшихся от родного татарского языка (по вопросу «Чем обусловлен Ваш переход на другой язык?») даже с подсказкой форм ответов не позволит исследователю заглянуть в психологию анкетлируемых.

3. Тексты пишутся на злобу дня людьми с большим жизненным опытом. Поэтому статьи, напечатанные в газетах и журналах, а также

описанные в произведениях «инженеров человеческих душ» судьбы людей могут помочь, как нам представляется, раскрытию проблемы.

Один из видных лингвистов XX века Л.Блумфилд заканчивает свою книгу «Язык» словами: «И хотя сейчас это только мечта, можно надеяться, что в недалеком будущем изучение языка поможет нам понимать поступки людей и управлять этими поступками» [6: 559].

Слова классика обнадеживают – анализ материалов Корпуса приведет к пониманию происхождения в этнической общности. К тому же в настоящее время не существует иных способов решения задачи.

Частотность употребления словоформ и их сочетаемостей позволяет судить об изменениях, происходящих как в самом языке, так и в обществе. С этих позиций рассмотрим несколько примеров.

В последние годы в татарском языке активно употребляются слова *хәрәм* (нечестно добытый продукт или богатство) и *хәләл* (заработанный, честно добытый продукт или богатство). В татарский и другие тюркские языки эти слова вошли с исламом. Причина в том, что религия больше заботится о загробной жизни правоверных, и, следовательно, духовенство должно было больше говорить и писать на тему, как уберечься от запретного продукта, то есть от продукта «*хәрәм*». Однако, как показывает проведенный анализ, слово *хәрәм* в Корпусе употребляется с частотой 614, а слово *хәләл* – 8489, то есть то, о чем следовало бы говорить чаще, употребляется в 13 раз меньше. Объяснено это может быть только тем, что *хәләл продукт* продается и покупается.

Некоторое недоумение возникает и по поводу того, что в Корпусе слово *белем* (знание) встречается с частотой 70791, а слово *акча* (деньги) 79522 раз. Получается, что общество больше заботится о деньгах, чем о пополнении своих знаний.

Письменный корпус представляет совокупность текстов. На употребительность слов (словоформ) текста накладываются языковые и внеязыковые ограничения. Если не брать во внимание внеязыковые ограничения (по условию постановки задачи), то основные языковые формы ограничения сводятся к следующим:

- ограничения, накладываемые тематикой текста,
- залоговой формой глагола,
- валентностью глагола,

- семантикой слова,
- модальными словами,
- послелогам (предлогами).

Хотя связь между словами в татарском языке достаточно морфологизирована, зачастую трудно определить значение тех или иных словосочетаний. В поисковой системе Корпуса для решения этой задачи предусмотрена возможность некоторого удлинения левого или правого контекста заданного слова. Покажем это на конкретном примере.

Допустим, что исследователя интересуют данные о том, как часто встречается в языке словоформа *догасы*, а также с какими словами она вступает в отношение посессивности, то есть данные ее левого контекста.

Обращение к Письменному корпусу дает следующие результаты:

– Слово *догасы* встречается в Корпусе 609 раз.

– Среди слов, заполняющих левый контекст, то есть слов, уточняющих, о чьих молитвах идет речь в текстах Корпуса, значатся {*ана* (мать) – 60, *эбиләр* (бабушки) – 15, *әнкәй* (мамочка) – 4, *ананың* (матери) – 2, *аналар* (мам) – 2, *ата-ана* (родители) – 2, *эти-энисе* (родителей) – 2}.

Если сгруппировать близкие по значению слова и просуммировать частоты их употреблений, то окажется, что *эни догасы* – (материнская молитва) встречается с частотой 68, *эбиләр догасы* – (молитва бабушек) – 15 и *молитва родителей* с частотой 5.

Как следует из приведенных данных, дети сильнее привязаны к матери, то есть более нуждаются в материнской молитве. Но нет ни слова об отцовской молитве. Что это? Отцы перестали молиться или дети в них (отцах) не нуждаются? Скорее, первое: отцы перестали молиться. А ведь в дореволюционную пору именно отцы были наиболее грамотными, в основном, они посещали мечети и только они совершали хадж.

Возможно, они (отцы) находятся среди тех определителей, которые в левом контексте заданного слова выражены местоимениями {*аның, аларның*} или прилагательным *изге* – (святая)?

Менеджер Корпуса позволяет исследователю ознакомиться с предложениями, содержащими в своем составе заданное слово, а также словоформу неявного значения. Допустим, интерес представляют предложения, включающие словосочетание *изге догасы*. Для этого необхо-

димо задействовать слово *изге* и, обратившись к конкретным примерам, просмотреть их. Именно таким способом было обнаружено наличие в Корпусе двух словосочетаний *эткэйнең изге догасы* (святая молитва папочки) и одно словосочетание *эбиләрбезнең изге догасы* (святая молитва наших бабушек).

В эпоху глобализации происходит унификация культур, а вместе с ней человек утрачивает собственное, самобытное восприятие мира и языка. Указанные обстоятельства не могли не отразиться в жизни любой нации, в том числе и татар.

В настоящей работе в рамках Корпуса проанализированы изменения общественных отношений, которые произошли за последние 20-25 лет. Данные анализа приведены в таблице 1.

Таблица 1. Виды общественных отношений
Мөнәсәбәт төрләре (виды отношений)

Сәүдә-икътисад (товарно-экономические)	Халыкара (межэтнические)	Гаиләдәге (семейные)	Канунлаштырылган (регулируемые законом)	Дуслык (дружеские)	Партнерлык мөнәсәбәтләре (партнерские отношения)	Хезмәт мөнәсәбәте (трудовые отношения)	Толерантлык (толерантность)	Көнәшлек (конкуренция)
F 303	62	203	74	113	160	86	30	14

Как следует из таблицы 1, число форм отношений между членами общества увеличилось. Однако современные формы отношений объединяют не столько людей, сколько общества. Так, межэтнические отношения – это обычно договоры, заключенные между правительствами; партнерские отношения представляют соглашения между мелкими и средними предприятиями, толерантность – между этносами, государством и конфессиями, а также между собственниками предприятий и различных заведений. Такой вид отношений, как конкуренция, также не предполагает участия человека как такового. Похоже, что человек как личность может проявить себя только дома, в семье?

Во что человек в этих условиях верит? В таблице 2 приведены данные левых контекстов, то есть слов, отвечающих на вопросы «кому?» «чему?» и сочетающихся с глаголами *ышану* (верить) и *ышана* (верит).

«Определяемым» – существительное в дательном падеже. В согласии с характерным для татарского языка порядком слов, глагол находится в конце предложения, а следовательно, и после своих определителей.

Объединенные данные анализа левых контекстов словоформ *ышану* и *ышана*, которые встречаются в Корпусе 1779 и 4096 раз соответственно, суммированы в таблице 2.

Таблица 2. Данные анализа Корпуса по вопросу
«Кому, чему / во что верят носители татарского языка?»

Ышану (верить) 1779 / Ышана (верит) 4096

	Дингә (в религию)	Үз-үзенә (в самого себя)	Хорафатларга (в суеверия)	Сүзгә (слову), китапка (книге)	Могҗизаларга (в чудо)	Киләчәккә (в будущее)	Юк-барга (небылицам)	Хакимияткә (правительству)
F	288	160	48	46	38	46	22	10

Как видно, несмотря ни на что, человек все еще верит в себя и в свое будущее. Несколько смущает отсутствие веры в друзей, в науку.

Одной из важных тем в проблеме определения состояния языка и этнической общности является выявление этнокультурных ценностей и культурной ориентации народа. Эта тема была затронута недавно в работе [7: 154-159]. На тот момент объем Корпуса составлял лишь 45 млн. слов и поисковая система не предусматривала операцию расширения контекстов заданного слова, не предоставляла исследователю списка словоформ, входящих в его контексты. В настоящее время объем Корпуса насчитывает 116 млн. слов, что, безусловно, повышает его репрезентативность и сбалансированность.

В таблице 3 представлен перечень «культурных ценностей» с числовыми характеристиками, то есть показателями их предпочтительности с точки зрения носителей татарского языка (жирным шрифтом указано количество употреблений данной словоформы в Корпусе).

Таблица 3. Перечень этнокультурных ценностей и их этнообъединяющие показатели

Этнокультурные ценности (этноценности)	Тугрык (преданность)	Тугрыкы (преданный)	Ихтирам (уважение)	Хөрмәт (почесть)	Үстүр (развитие)	Общие
	1955	1428	5100	16457	32782	
Тел (язык)	7	3	59	176	1246	1491 14,6%
Туган ил (Родина)	77	49	5	21	464	616 6,0%
Хезмәт (труд)	123	80	25	87	1282	1597 15,6%
Милләт (национальность)	20	54	33	233	243	583 5,7%
Дин (религия)	47	28	18	135	112	340 3,3%
Туган-тумачалык (родственные связи)	18	12	48	268	369	715 7,0%
Һөнәри мәдәният	2	7	29	75	1251	1364 13,4%
Әхлакылык (нравственность)	7	25	72	279	336	719 7,0%
Руһи мәдәният (духовная культура)	388	45	63	213	122	831 8,1%
Гомумкешелек идеаллары (общечеловеческие идеалы)	25	0	23	77	41	166 1,6%
Хакимият атрибутлары (государственные атрибуты)	58	17	14	78	0	167 0,8%
Сәүдә – икътисад (торговля – экономика)	0	0	0	0	1621	1621 15,88%

Таблица 3 (последняя строка) содержит в качестве одной из этноценностей «торгово-экономическое отношение». По формальным признакам (о чем больше говорят, стремятся развить в большей мере) оно заслуживает включения в число этноценностей. В то же время общество не проявляет к «торгово-экономическому отношению» ни преданности, ни уважения, ни почести (нули в соответствующих графах). Нам представляется, что при анкетировании носителей татарского языка на вопрос «Какая из указанных ценностей объединяет Вас с вашим народом?» утвердительных ответов с указанием «торгово-экономические отношения» было бы немного [8: 223-239]. Таким образом, на данный момент времени «торгово-экономическое отношение», хотя и

занимает умы большого числа людей, все еще не получило статуса этнообъединяющей ценности. По данным таблицы 3, наиболее значительными этноинтегрирующими характеристиками обладают такие признаки этноса, как язык, труд, профессиональная культура.

В этнолингвистике основным фактором формирования психики этноса считается язык. С помощью языка народы накапливают опыт предыдущих поколений. Под влиянием грамматической и семантической систем формируются структура мышления носителей языковой общности. Письменный корпус, являясь своеобразной энциклопедией текстов в электронном формате, предоставляет лингвистам большие возможности для восстановления своей истории, реконструирования быта своих предков, понимания их жизненной философии. Дальнейшие исследования, безусловно, выявят новые предпочтения и этнообъединяющие ценности татарского народа, позволят откорректировать результаты наших наблюдений.

Литература

1. Степанов Г.В. Типология языковых состояний и ситуаций в странах романской речи. М.: Наука, 1976. 141 с.
2. Бодуэн де Куртенэ И.А. Язык и языки. Избранные труды по общему языкознанию. М.: Наука, 1963. Т.2. 94 с.
3. Беликов В.И., Крысин Л.П. Социоллингвистика. М., 2001. 439 с.
4. Хакимов Р. Не привилегии, а защита. Языковая политика в Республике Татарстан. Казань, Изд. «Магариф», 1999. 342 с.
5. Письменный корпус татарского языка [Электронный ресурс] / Сайхун М.Р., Ибрагимов Т.И. Казань, 2014. URL: <http://corpus.tatfolk.ru>. (дата обращения 15.05.2015).
6. Блумфилд Л. Язык. М.: Прогресс, 1968. 606 с.
7. Ибрагимов Т.И., Сайхун М.Р. К построению структурно-функциональной модели ценностной ориентации татарского этноса (по материалам письменного корпуса татарского языка) // Языки России и стран ближнего зарубежья как иностранные. Материалы Международной научно-практической конференции (28-29 ноября 2013г.). Казань, 2013, С.154-159.
8. Губогло М.Н. Интегрирующая функция языка // Социоллингвистические проблемы развивающихся стран. М.: Наука, 1975. С. 223-239.

ТАТАР ТЕЛЕНЕҢ ЯЗМА КОРПУСЫ МАТЕРИАЛЛАРЫ АША МИЛЛИ ТЕЛ ТОРЫШЫН ӨЙРӘНҮ

Тәүзих Ибраһим улы Ибраһимов,
Казан федераль университеты,
Россия, 420008, Казан ш., Кремль ур., 18 нче йорт,
Tavzikh.Ibragimov@ksu.ru.

Мансур Рифкатъ улы Сайхунов,
Татарстан Республикасы Фәннәр академиясенең информатика институты,
Россия, 420111, Казан ш., Сул Болак ур., 36А нчы йорт,
6688000@gmail.com.

Бүгенге көндә кешелек глобализм, ягъни милли аермалыкларын югалту чорында яши. Бу шартларда халыкның телен дөньяны үзенчәлекле тану системасы буларак саклап калу көн тәртибенә куелган мәсьәләләрнең берсе булып тора. Тәкъдим ителгән мәкаләдә Татар теленең язма корпусы ярдәмендә милләтнең тел дөньясы ачыклана.

Хәзерге халәтendә Татар теленең язма корпусы репрезентативлык һәм төрле стильләрне телгә бәрабәр күләмдә чагылдыру таләпләренә тулысынча җавап бирә. Ул телчеләргә сүзформаларның кулланылу ешлыгын, шулай ук аларның үз чолганышындагы сүзформалар белән мөнәсәбәткә керү мөмкинлекләрен бәян итә, татар милләтенең дөньяны кабул итү үзенчәлекләре барлана.

Төп төшенчәләр: Татар теленең язма корпусы, мәгълүмат эзләү системасы, тел җәмгыяте, сүзформа, уң контекст, сул контекст, кулланылу ешлыгы.