

**A Comparison of Modern Turkic Languages (Turkish, Azerbaijani, Kazakh, Kyrgyz, Uzbek)
in Terms of Most Frequently Used 1000 Words**

En Sık Kullanılan 1000 Kelime Açısından

Modern Türk Dilleri (Türkçe, Azeri, Kazak, Kırgız, Özbek) Üzerine Bir Karşılaştırma

Hilal Oytun Altun

Dr, Jagiellonian University, Krakow, Poland

hilaloytun@gmail.com

Abstract

Lexical comparisons between Turkic languages provide useful information in the context of mutual agreement and language learning. Since the corpus studies on Turkic languages other than Turkish are still in the development stage, it seems to be a solution to use frequency lists in English which is the dominant language of the Internet for interlingual comparisons of those languages. Besides, in linguistic studies which tested the Zipf's statistical theory, it has been shown that the word frequencies are similar in different languages. On the other hand, the misleading of false friends (deceptive cognates), which are common among Turkic languages, can be minimized by reference to a third language. Despite the existence of false equivalents, it is seen that the most frequent 1000 words in contemporary Turkic languages, may provide at least beginners level mutual intelligibility for learners. And it should be added that the alphabets and the scripting systems also can support mutual intelligibility among Turkic speaking peoples, if produced soundly.

Keywords

Most frequent words, Turkic languages, Comparative lexicography, corpus.

Özet

Türk dilleri arasındaki sözlük bilgisi karşılaştırmaları, karşılıklı anlaşma ve dil öğrenme bağlamında faydalı bilgiler sağlamaktadır. Türkiye Türkçesi dışındaki Türk dilleri üzerine yapılan derlem çalışmaları hâlen gelişim aşamasında olduğundan, bu diller arasında yapılacak karşılaştırmalarda internetin hâkim dili olan İngilizce'ye ait sıklık listelerini kullanmak bir çözüm gibi görünmektedir. Ayrıca, Zipf'in istatistik teorisini test eden dilbilimsel çalışmalarda, en sık kullanılan kelime listelerinin farklı dillerde benzer olduğu ortaya konmuştur. Öte yandan, Türk dilleri arasında yaygın olan yalancı eşdeğerlerin (yanıltıcı kökendeşler) yanıltıcılığı üçüncü bir dile referansla en aza indirilebilir. Yalancı eşdeğerlerin varlığına rağmen, çağdaş Türk dillerinde en sık kullanılan 1000 kelimenin dil öğrenenler için en az başlangıç düzeyinde karşılıklı anlaşılabilirlik sağlayabildiği görülmektedir. Ayrıca, doğru tasarlanmaları hâlinde, alfabe ve yazı sistemlerinin de Türkçe konuşan halklar arasında karşılıklı anlaşmayı destekleyebilecekleri de eklenmelidir.

Anahtar kelimeler

En sık kelimeler, Türk dilleri, Karşılaştırmalı sözlük bilgisi, derlem(korpus).

Introduction

Natural language processing is based on corpora and examines the texts statistically using the methods of corpus linguistics. The data provided from the corpora was used in this paper, what it is tried to do here is not corpus linguistics however. The data taken from the corpus linguistics was served as a material for lexicographic comparison of different languages for this study. Thus, this paper does not offer a model for comparative lexicography studies; instead it'll be tried to make a lexicographical map of selected modern Turkic languages in order to see If the relations between those languages on the lexicographical level are meaningful for learners. For sake of brevity and applicability, this study has been limited to 5 Turkic languages and most frequent 1000 words. The languages listed in the title have been selected to study for practical reasons. First of all, the corpus linguistics which led the way by Turkish among the Turkic languages is a fairly new field of study. While Kazakh and Azerbaijan walk on the footsteps of NLP studies in Turkish, Uzbek and Kyrgyz are at the very first stages. As for Turkmen and Uighur teher aren't many digitalised materials to study on; and as an interesting fact, even Google translate doesn't offer translation service for the latter two languages while Hosa, Hmong or Igbo languages can be found on Google translate languages list. It is nearly impossible to find any digitalised material in many other Turkic languages (Tatar, Karakalpak, Chuwash, etc.) which have far less speakers than the languages mentioned above. The studies conducted in Turkic languages mainly in the area of machine translation (Orhun, 2017,p.496).

Other limitations of my study are related to the content. The lexicographic comparison in this paper is based on the 1000 most frequent words appeared in English corpus. It can be questioned why lists from the selected languages weren't used in this paper. The answer is; as it is explained above there's corpus available which offers decent amount of data in Turkish and its frequency lists mostly match with the English equivalents; which the same pattern can be observed in many other languages as it will be shown below.

Creating reliable frequency lists requires a great deal of digitalised texts from various sources and various fields, which lack in other Turkic languages.

Statistical appearance of a corpus varies depending on the specialties of the texts in it. The largest freely available corpus (Global Web-based English) is made of 2 billion words and as it can easily be assumed it is in English; as seen from the table below, numbers are compatible with its usage rating on the internet.

<https://www.internetworldstats.com/stats7.htm>

Top Ten Languages Used in the Web - April 30, 2019 (Number of Internet Users by Language)					
TOP TEN LANGUAGES IN THE INTERNET	World Population for this Language (2019 Estimate)	Internet Users by Language	Internet Penetration (% Population)	Internet Users Growth (2000 - 2019)	Internet Users % of World (Participation)
English	1,485,300,217	1,105,919,154	74.5 %	685.7 %	25.2 %
Chinese	1,457,821,239	863,230,794	59.2 %	2,572.3 %	19.3 %
Spanish	520,777,464	344,448,932	66.1 %	1,425.8 %	7.9 %
Arabic	444,016,517	226,595,470	51.0 %	8,917.3 %	5.2 %
Portuguese	289,923,583	171,583,004	59.2 %	2,164.8 %	3.9 %
Indonesian / Malaysian	302,430,273	169,685,798	56.1 %	2,861.4 %	3.9 %
French	422,308,112	144,695,288	34.3 %	1,106.0 %	3.3 %
Japanese	126,854,745	118,626,672	93.5 %	152.0 %	2.7 %
Russian	143,895,551	109,552,842	76.1 %	3,434.0 %	2.5 %
German	97,025,201	92,304,792	95.1 %	235.4 %	2.1 %
TOP 10 LANGUAGES	5,193,327,701	3,346,642,747	64.4 %	1,123.0 %	76.3 %
Rest of the Languages	2,522,895,508	1,039,842,794	41.2 %	1,090.4 %	23.7 %
WORLD TOTAL	7,716,223,209	4,386,485,541	56.8 %	1,115.1 %	100.0 %

NOTES: (1) Top Ten Languages Internet Stats were updated in April 30, 2019. (2) Internet Penetration is the ratio between the sum of Internet users speaking a language and the total population estimate that speaks that specific language. (3) The most recent Internet usage information comes from data published by [Nielsen Online](#), [International Telecommunications Union](#), [GfK](#), and other reliable sources. (4) Population estimates are based mainly on figures from the [United Nations Population Division](#) and local official sources. (5) For definitions, methodology and navigation help, please see the [Site Surfing Guide](#). (6) These statistics may be cited, stating the source and establishing an active link back to [Internet World Stats](#). Copyright © 2019, Miniwatts Marketing Group. All rights reserved worldwide.

(Internet world stats, 2019)

Despite the potential drawbacks, using English frequency lists as a reference point helped to eliminate the confusion which might be caused by deceptive cognates or "false friends of translator" in other words. There are some corpus studies to detect deceptive cognates in a couple of language pairs; Turkic languages aren't among them yet however. For example; "saat" means both "watch" and "hour" in selected Turkic languages while "el" might have different meanings; and a monolingual corpus wouldn't tell which one is which, without referring the collocations. Still as a side note here, I should admit that a perfect study would compare original frequency lists of selected languages; and of course with filtration model for deceptive cognates which hasn't been found for Turkic languages yet.

With all the shortcomings listed above to get a more or less reliable result there had to be made some arrangements. First of all, the lists were formed quite ambiguity-free with eliminated false friends and filtered functional words. Comparing the frequency lists which contain functional words, might be misleading due to the difference between grammatical structures of given languages. It was inevitable to revise the lists in selected Turkic languages to be sure such functional words like "to, on, was", etc. in English wouldn't cause confusion. Second, the words have been double checked with monolingual dictionaries as well as multilingual dictionaries in order to avoid the translation mistakes.

As expected, the basic vocabulary derived from proto-Turkic still live in Turkic languages with their original meanings. 150 of most frequent 1000 words are common both semantically and phonetically in selected

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

languages. Those are pronouns, numbers, simple verbs as well as words from basic vocabulary of everyday life such as agriculture, geography, anatomy, etc.

The words are written in their original alphabets. As for Kazakh that is undergoing a process of renewing the alphabet, the words are written with Cyrillic alphabet.

At this point I should add that; as it will also be seen clearly from my study one more time; alphabet defines a nation's identity among world cultures. It can serve mutual intelligibility and add to shared cultural identity like in Chagatai and Ottoman literary traditions; or it can cause alienation among related nations like it did during Soviet era. The lexicologic comparisons among Turkic languages also underline the importance of the alphabet beyond being only a tool one more time. It can clearly be seen from the table below that the same words look how different in different characters even if they are all in Latin: “diş” (tooth), “üç” (three), “ağız” (mouth), “saç” (hair) in Tr. and Azeri, but in Uzbek: “tish”, “uch”, “og’iz”, “soch”, etc.

The common basic vocabulary which is shared by five Turic languages from the most frequent 1000 words:

Tr	Az	Kz	Kr	Uz	En
bir	bir	бір	бир	bir	one
iki	iki	екі	эки	ikki	two
ikinci	ikinci	екінші	экинчи	ikkinchi	second
üç	üç	үш	үч	uch	three
üçüncü	üçüncü	үшінші	үчүнчү	uchinchi	third
dört	dörd	төрт	төрт	to’rt	four
beş	beş	бес	беш	bash	five
altı	altı	алты	алты	olti	six
yedi	yeddi	жеті	жети	yetti	seven
sekiz	səkkiz	сегіз	сегиз	sakkiz	eight
dokuz	doqquz	тоғыз	тоғуз	to’qqiz	nine
on	on	он	он	o’n	ten
yirmi	iyirmi	жиырма	жыйырма	yigirma	twenty
yüz	yüz	жүз	жүз	yuz	hundred
bin	min	мың	мын	ming	thousand
ben	mən	мен	мен	men	I
benim	mənim	мениң	менин	mening	my
biz	biz	біз	биз	biz	we
bizim	bizim	біздің	биздин	bizning	our
kim	kim	кім	ким	kim	who
kimin	kimin	кімнің	кимдин	kimning	whose
o	o	ол	ал	u	he
onlar	onlar	олар	алар	ular	they
siz	siz	сіз	сиз	siz	you
öz	özünü	өзін	өз	o’z-o’zini	self
açık	açıq	ашық	ачык	ochiq	open
ağır	ağır	ауыр	оор	og’ir	heavy
ağız	ağız	ауыз	ооз	og’iz	mouth
beyaz, ak	ağ	ақ	ак	oq	white
anne	ana	ана	эне	ona	mother
ay	ay	ай	ай	oy	month

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

balık	balıq	балық	балык	balıq	fish
boya	boya	бояу	боек	bo'yoq	paint
boyun	boyun	мойын	моюн	bo'yin	neck
bugün	bu gün	бүгін	бүгүн	bugun	today
bulut	bulud	бұлт	булут	bulut	cloud
burun	burun	мұрын	мурун	burun	nose
buz	ice	мұз	муз	muz	ice
dağ	dağ	тау	тоо	tog'	mountain
demir	dəmir	темір	темир	temir	iron
deniz	dəniz	теңіз	деңиз	dengiz	sea
deri	dəri	тері	тери	teri	skin
dil	dil	тіл	тил	til	language
diş	diş	тістері	тиштер	tish	tooth
dolu	dolu	толық	толук	to'liq	full
elma	alma	алма	алма	olma	apple
ev	ev	үй	үй	uy	home
gemi	gəmi	кеме	кеме	kema	ship
giriş	giriş	кіру	кирүү	kirish	entry
göl	göl	көл	көл	ko'l	lake
görünüş	görünüş	көрініс	көрүнүш	ko'rinish	view
göz	göz	көз	көз	ko'z	eye
güç	güc	күш	күч	kuch	force
güçlü	güclü	күшті	күчтүү	kuchli	strong
gümüş	gümüş	күміс	күмүш	kumush	silver
gün	gün	күн	күн	kun	day
ısı	istilik	ыстық	ысык	issiq	heat
incir	əncir	інжір	анжыр	anjir	fig
baş / kafa	baş / qafa	бас	баш	bosh	head
kan	qan	қан	кан	qon	blood
kanat	qanad	қанат	канат	qanot	wing
siyah / kara	qara	қара	кара	qora	black
karşı	qarşı	қарсы	каршы	qarshi	against
kısa	qısa	қысқа	кыска	qisqa	short
kış	qış	қыс	кыш	qish	winter
kız	qız	қыз	кыз	qiz	girl
kol	qol	қол	кол	qo'l	arm
kulak	qulac	құлақ	кулак	quloq	ear
kum	qum	құм	кум	qum	sand
oğul	son	ұл	уул	o'g'il	son
orta	orta	орта	орто	o'rta	middle
oyun	oyun	ойын	оюн	o'yin	game
oyuncu	oyunçu	ойыншы	оюнчу	o'yinchi	player
oyuncak	oyuncaq	ойыншық	оюнчук	o'yinchoq	toy
ölü	ölü	өлі	өлгөн	o'lgan	dead
ölüm	ölüm	өлім	өлүм	o'lim	death
ördek	ördək	үйрек	өрдөк	o'rdak	duck
saç	saç	шаш	чач	soch	hair
sarı	sarı	сары	сары	sariq	yellow

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

soğuk	soyuq	суық	суук	sovuq	cold
söz	söz	сөз	сөз	so'z	word
su	su	су	суу	suv	water
süt	süd	сүт	сүт	sut	milk
taş	daş	тас	таш	tosh	stone
tepe	təpə	төбө	дөбө	tepa	hill
tuz	duz	тұз	туз	tuz	salt
uyku	yuxu	ұйқы	үйкү	uyqu	sleep
uzunluk	uzunluk	ұзындық	узундук	uzunlik	length
var	var	бар	бар	bor	have
yakın	yaxın	жақын	жанында	yaqin	near
yaş	yaş	жас	жашы	yosh	age
yaz	yay	жаз	жай	yoz	summer
yazı	yazı	жазу	жазуу	yozish	writing
yeni	yeni	жаңа	жаңы	yangi	new
yeşil	yaşıl	жасыл	жашыл	yashil	green
yıl	il	жыл	жыл	yil	year
yıllık	illik	жылдық	жылдык	yillik	annual
yıldız	ulduz	жұлдыз	жылдыз	yulduz	star
yukarıda	yuxarıda	жоғарыда	жогоруда	yuqorida	above
yumuşak	yumşaq	жұмсақ	жумшак	yumshoq	soft
yol	yol	жол	жол	yo'l	way
arasında	arasında	арасында	арасында	orasida	among
bütün	bütün	бүтін	бүт	butun	whole
eğer	əgər	егер	эгер	agar	if
geç	gec	кеш	кеч	kech	late
için	üçün	үшін	үчүн	uchun	for
karşı	qarşı	қарсы	каршы	qarshi	against
yok, hayır	yox	жоқ	жок	yo'q	no
başka	başqa	басқа	башка / дагы	boshqa	else

Basic verbs:

Tr	Az	Kz	Kr	Uz	En
ağlamak	ağlamaq	жылау	ыйлоо	yig'lamoq	cry
akmak	axmaq	ағу	агуу	oqim	flow
almak	almaq	алу	алуу	olish	take, get
başlamak	başlamaq	бастау	баштоо	boshlash	begin
bilmek	bilmək	білу	билүү	bilish	know
demek	demək	деу	дейт	deb	say
doldurmak	doldurmaq	толтыру	толтуруу	to'ldirish	fill
durmak	durmaq	тұру	туруу	turmoq	stand
gelmek	gəlmək	келу	келген	kelish	come
girmek	girmək	кіру	кирүү	kiritish	enter
gitmek	getmək	кету	кетти	ketish	gone
giyinmek	geyinmək	киину	кийинүү	kiyinish	dress
görmek	görmək	көрү	көрүү	ko'rish	see
görünmek	görmək	көрину	көрүнүү	ko'rinish	seem
göstermek	göstərmək	көрсету	көрсөтүү	ko'rsatish	show

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

gülmek	gülmək	күлү	күлүү	kulish	laugh
kalmak	qalmaq	калу	калуу	qolish	stay
kesmek	kəsmək	кесу	кесүү	kesish	cut
korkmak	qorxmaq	қорқу	коркуу	qo'rqish	afraid
koymak	qoymaq	қою	коюу	qo'yish	put
okumak	oxumaq	оқу	оқуу	o'qish	read
olmak	olmaq	болу	болуу	bo'lish	be / become
oturmak	oturmaq	отыру	отуруу	o'tirish	sit
oynamak	oynamaq	ойнау	ойноо	o'ynash	play
öğrenmek	öyrənmək	үйрену	үйрөнүү	o'rganish	learn
öldürmek	öldürmək	өлтіру	өлтүрүү	o'ldirish	kill
ölmek	ölmək	өлү	өлүү	o'lish	die
saklamak	saxlanılır	сақтау	сактау	saqlash	keep
satıcı	satıcı	сатушы	сатуучу	sotuvchi	seller
satmak	satmaq	сату	сатуу	sotish	sell
sormak	sormaq	сұрау	суроо	so'rash	ask
vermek	vermək	беру	берүү	berish	give
yağmur	yağış	жаңбыр	жаан	yomg'ir	rain
yazmak	yazmaq	жазу	жазуу	yozish	write
yemek	yemək	жеу	жеу	yemoq	eat
yeterli	yetər	жеткілікті	жетишет	yetarli	enough

There is also common cultural vocabulary in selected languages from Turkic origin and borrowings.

The borrowings from Arabic and Persian which are common for selected Turkic languages:

Tr	Az	Kz	Kr	Uz	En
adam	adam	адам	адам	odam	man
araba	araba	арба	араба	arava	cart
bilim / ilim	elm	ғылым	илим	ilm-fan	science
cevap	cavab	жауап	жооп	javob	answer
cuma	cümə	жума	жума	juma	friday
çift	cüt	жүп	жүп	juft	pair
doğal / tabii	təbii	табиғи	табигый	tabiiy	natural
dünya /alem	dünya /aləm	дүние /әлем	дүйнө	dunyo /olam	world
harita	xəritə	карта	карта	xarita	map
hava	hava	ауа	аба	havo	air
her	hər	әр	ар	har	every
hesap	hesab	есеп	эсеб	hisob	account
hiç kimse	heç kim	ешкім	эч ким	hech kim	none
hizmet	xidmət	қызмет	кызмат	xizmat	service
hoşgeldiniz	hoş gəlmisiniz	қош келдіңіз	кош келдиңиз	xush kelibsiz	welcome
hükümet	hökumət	үкімет	өкмөт	hukumat	government
kâğıt	kağız	қағаз	кагаз	qog'oz	paper
kelime / söz	söz	сөз	сөз	so'z	word
kitap	kitab	кітап	китеп	kitob	book
kuvvet	qüvvə	қуат	кубат	quvvat	power
kütüphane	kitabxana	кітапхана	китепкана	kutubxona	library
makale	məqalə	мақала	макала	maqola	article

Özel Sayı "Polonya ve Türk Dünyası", Kasım 2019

mümkün	mümkün	мүмкін	мүмкүн	mumkin	possible
pazar	bazar	базар	базар	bozor	market
resmî	rəsmi	ресми	расмий	rasmiy	official
saat	saat	сағат	саат	soat	hour
saat	saat	сағат	саат	soat	clock
sebeup / neden	səbəb	себеп	себеп	sabab	cause
seyahat	səyahət	саяхат	саякат	sayohat	travel
siyaset	siyasət	саясат	саясат	siyosat	policy
siyasi	siyasi	саяси	саясий	siyosiy	political
sonuç / netice	nəticə	нәтижесі	жыйынтык	natijalar	result
şartlar	şərtlər	шарттар	шарттар	shartlar	conditions
tarih	tarix	тарихы	тарых	tarix	history
tecrübe	təcrübə	тәжірибе	тажрыйба	tajriba	experience
ümit	ümid	үміт	үмүт	umid	hope

The borrowings from western languages which are common for selected Turkic languages are universal words at the same time:

Tr	Az	Kz	Kr	Uz	En
aksesuar	aksesuar	аксессуар	аксессуар	aksesuar	accessory
atom	atom	атом	атом	atom	atom
banka	bank	банк	банк	bank	bank
bar	bar	бар	бар	bar	bar
bit	bit	биттік	бит	bit	bit
dizayn	dizayn	дизайн	дизайн	dizayn	design
domain	domain	домен	домен	domen	domain
elektrik	elektrik	электр	электрдик	elektr	electric
elektronik	elektronika	электроника	электроника	elektronika	electronics
eleman	element	элемент	элемент	element	element
enerji	enerji	энергия	энергия	energiya	energy
film	film	фильм	фильм	film	film
gaz	qaz	газ	газ	gaz	gas
hosting	hosting	хостинг	хостинг	hosting	hosting
internet	internet	интернет	интернет	internet	internet
kart	kart	карта	карта	karta	card
makine	maşın	машина	машина	mashina	machine
metal	metal	металл	метал	metall	metal
mıknatis	maqnit	магнит	магнит	magnit	magnet
milyon	milyon	миллион	миллион	million	million
model	model	модель	модель	model	model
modeller	modelləri	модельдер	моделдери	modellari	models
molekülü	molekul	молекуласы	молекуласы	molekulasi	molecule
müzik	musiqi	музыка	музыка	musiqa	music
organ	orqan	орган	орган	organ	organ
parti	party	партия	партия	partiya	party
posta	poçt	пошта	почта	pochta	mail
radio	radio	радио	радио	radio	radio
site	sayt	сайт	сайт	sayt	site

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

standart	standart	стандарт	стандарт	standart	standard
teknik	texniki	техникалық	техникалык	texnik	technical
teknoloji	texnologiya	технология	технология	texnologiya	technology
telefon	telefon	телефон	тел	telefon	phone
üniversite	universitet	университет	университет	universitet	university

Some words are unique for one language but common for others. Turkish have more words in this category, it differs from its relatives more, in other words. As it can be seen from the coloured cells which contain borrowings, as a result of the language revolution conducted during the early years of the Turkey Republic, Turkish has less borrowings than the others.

Tr	Az	Kz	Kr	Uz	En
<i>arkadaşlar</i>	dostlar	достар	достор	do'stlar	friends
<i>benzer</i>	oxşar	ұқсас	окшош	o'xshash	similar
<i>bilgisayar</i>	kompyuter	компьютер	компьютер	kompyuter	computer
<i>bir diğer/başka</i>	başqa	басқа	башка	boshqa	another
<i>bulmak</i>	tapmaq	табу	табуу	topish	find
<i>çalışma</i>	işləyirik	жұмыс істеу	иштеп	ishlaydi	working
<i>çelik</i>	polad	болат	болот	po'lat	steel
<i>çiçek</i>	gül	гүл	гүл	gul	flower
<i>değerlendirme</i>	reyting	рейтинг	рейтинг	reyting	rating
<i>dergi</i>	jurnal	журналы	журнал	jurnal	journal
<i>devlet başkanı</i>	prezident	президенті	президент	prezident	president
<i>dönüş</i>	qayıtmaq	қайтару	кайра	qaytish	return
<i>genellikle</i>	adətən	әдетте	адатта	odatda	usually
<i>gereklikler</i>	tələblər	талаптар	талаптар	talablar	requirements
<i>indir</i>	yükləyin	жүктеу	жүктөп алуу	yuklab olish	download
<i>iyi</i>	yaxşı	жақсы	жакшы	yaxshi	good
<i>katılmak</i>	qoşulmaq	қосылу	кошулуу	qo'shilish	join
<i>kaynaklar</i>	resursları	ресурстар	ресурстары	resurslar	resources
<i>keci</i>	pişik	мысық	мышык	mushuk	cat
<i>kemik</i>	sümük	сүйек	сөөк	suyak	bone
<i>koku</i>	iy	иіс	жыт	hid	smell
<i>köle</i>	qul	құл	кул	qul	slave
<i>köpek</i>	it / köpek	ит	ит	it	dog
<i>ne zaman</i>	haçan	қашан	качан	qachon	when
<i>okul</i>	məktəb	мектеп	мектеп	maktab	school
<i>örnek</i>	misal	мысал	мисал	misol	example
<i>sıcak</i>	isti	ыстық	ысык	issiq	hot
<i>sokak</i>	küçə	көше	көчө	ko'cha	street
<i>şu anda</i>	hazırda	қазіргі уақытта	азыркы	hozir	currently
<i>tekne</i>	qayıq	қайық	кайык	qayıq	boat
<i>yıkama</i>	yuma	жуу	жуунуу	yuvish	wash
<i>haklar</i>	hüquqlar	құқықтар	укуктары	huquqlar	rights
<i>kablosuz</i>	simsiz	сымсыз	зымсыз	simsiz	wireless
<i>kalp</i>	ürək	жүрек	жүрек	yurak	heart
<i>liman</i>	port	порт	порт	port	port

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

<i>mevcut</i>	hazırda	қазіргі	азыркы	hozirgi	present
<i>müdür</i>	menecer	менеджер	менеджер	menejeri	manager
<i>pazarlama</i>	marketing	маркетинг	маркетинг	marketing	marketing
<i>şartlar</i>	şartları	терминдер	шарттары	shartlar	terms
<i>zaman</i>	vaxt	уақыт	убакыт	vaqt	time

The words unique for Azeri:

Tr	Az	Kz	Kr	Uz	En
altın	<i>qızıl</i>	алтын	алтын	oltin	gold
bölüm	<i>şöbə/ bölmə</i>	бөлім	бөлүм	bo'lim	department / section
orman	<i>meşə</i>	орман	токой	o'rmon	forest
sormak	<i>soruşmaq</i>	сұрау	суроо	so'rmok	ask
yalnız	<i>tək</i>	жалғыз	жалгыз	yolg'iz	lone
bölüm	<i>fəsil</i>	бөлім	бөлүм	bo'lim	chapter
dükkan	<i>mağaza</i>	дүкен	дүкөн	do'kon	shop
gerekir	<i>lazımdır</i>	керек	керек	kerak	must
hata	<i>səhv</i>	қате	ката	xatolik	error
içinde	<i>daxilində</i>	ішінде	ичинде	ichida	within
mülk	<i>əmlak</i>	мүлік	мүлк	mulk	property
spor	<i>idman</i>	спорт	спорт	sport	sports
tertip	<i>nizam</i>	тәртіп	тартип	tartib	order

The words unique for Kazakh:

Tr	Az	Kz	Kr	Uz	En
at	at	<i>жылқы</i>	ат	ot	horse
baz	bazası	<i>негіз</i>	база	bazasi	base
bunlar	bunlar	<i>мыналар</i>	булар	bular	these
düşman	düşmən	<i>жау</i>	душман	dushman	enemy
grup	qrup	<i>топ</i>	группа	guruh	group
iş	iş	<i>жұмыс</i>	иш	ish	job
komşu	qonşu	<i>көрші</i>	коңшу	qo'shni	neighbor
uzun	uzun	<i>ұзақ</i>	узун	uzun	long
video	video	<i>бейне</i>	video	video	video
yarım	yarım	<i>жарты</i>	жарым	yarim	half
ağustos	avqust	<i>тамыз</i>	август	avgustda	august
arşiv	arxivlər	<i>мұрағаттар</i>	архив	arxivlar	archives
asker	əsgər	<i>солдат</i>	аскер	askar	soldier
dünya	dünya	<i>әлем</i>	дүйнө	dunyo	world
kredi	kredit	<i>несие</i>	кредит	kredit	credit
okyanus	okean	<i>мұхит</i>	океан	okean	ocean
park	park	<i>саябақ</i>	парк	park	park
reklâm	reklam	<i>жарнама</i>	рекламалоо	reklama	advertising
şartlar	şartları	<i>терминдер</i>	шарттары	shartlar	terms
şehir	şəhər	<i>қала</i>	шаары	shahar	city
versiyon	versiya	<i>нұсқасы</i>	версия	versiyasi	version
yaşam	yaşamaq	<i>өмір сүру</i>	жашоо	yashash	living

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

The words unique for Kyrgyz:

Tr	Az	Kz	Kr	Uz	En
arasında	arasında	арасында	<i>ортосунда</i>	orasida	between
ayak	ayaq	аяқ	<i>таман</i>	oyoq	foot
davranmak hareket	hərəkət et	эрекет	<i>иш</i>	harakat qilish	act
filmler	filmlər	фильмдер	<i>тасмалар</i>	filmlar	movies
gelecek	gələcək	келешек	<i>болочок</i>	kelajak	future
imla / yazım	imla	емле	<i>жазуу эрежеси</i>	imlo	spelling
kalın	qalın	қалың	<i>жоон</i>	qalin	thick
kolej	kollec	колледж	<i>оокуу жай</i>	kollej	college
konferans	konfrans	конференция	<i>жыйындар</i>	konferensiya	conference
kuş	quş	құс	<i>чымчык</i>	qush	bird
öğretmek	öyrətmək	үйрету	<i>окутуу</i>	o'rgatish	teach
ölçmek	ölçmək	өлшеу	<i>ченөө</i>	o'lchov	measure
segmenti	segment	сегмент	<i>кесинди</i>	segment	segment
sınıf	sinif	сынып	<i>тап</i>	sinf	class
fonksiyon	funksiyası	функциясы	<i>милдети</i>	funktsiyasi	function
hafta	həftə	апта	<i>жума</i>	xafta	week
kamera	kamera	камера	<i>фотоаппарат</i>	kamera	camera
metin	mətn	мәтін	<i>текст</i>	matn	text
mevsim	mövsüm	маусым	<i>мезгил</i>	mavsum	season
şarkı sözleri	sözləri	сөздері	<i>текст</i>	qo'shiq so'zlari	lyrics

The words unique for Uzbek:

Tr	Az	Kz	Kr	Uz	En
niye	niyə	неге	неге	<i>nima uchun</i>	why
kuyruk	quyruq	қўйрық	куйрук	<i>dum</i>	tail
de	de	де	де	<i>ham</i>	de
enerji	enerji	энергия	энергия	<i>quvvat energiya</i>	energy
et	ət	ет	эт	<i>go'sht</i>	meat
koruma	qorunması	қорғау	коргоо	<i>himoya qilish</i>	protection
sol	sol	сол	сол	<i>chap</i>	left
biletler	biletlər	билеттер	билеты	<i>chiptalar</i>	tickets

The list below shows common words of southwest (Az, Tr) versus east and southeast (Kz, Kr, Uz) Turkic languages:

Tr	Az	Kz	Kr	Uz	En
ada	ada	арал	арал	orol	island
altında	altında	астында	астында	ostida	under
bak	baxın	қараң	кароо	qara	look
boyunca	boyunca	бірге	бирге	bilan birga	along
düşünce	düşünce	ой	ой	o'y	thought
büyüme	böyümə	өсу	өсүш	o'sishi	growth
birlikte	birlikdə	бірге	бирге	bilan birga	together
çiplak	çılraq	жалаңаш	жылаңач	yalang'och	nude

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

<i>çok</i>	<i>çox</i>	өте / көп	көп	ko'p	much
<i>çubuk</i>	<i>çubuq</i>	таяқ	таяк	tayoq	stick
<i>dalga</i>	<i>dalğa</i>	толқын	толкун	to'lqin	wave
<i>değişim</i>	<i>dəyişiklik</i>	өзгеріс	өзгөртүү	o'zgarish	change
<i>demek</i>	<i>demək</i>	айту	айтуу	aytish	say
<i>diğer</i>	<i>digər</i>	басқа	башка	boshqa	other
<i>düşünmek</i>	<i>düşünmək</i>	ойлау	ойлоо	o'ylayman	think
<i>eğlence</i>	<i>əyləncə</i>	ойын-сауық	оюн-зоок	o'yin-kulgi	entertainment
<i>el</i>	<i>əl</i>	қолмен	кол	qo'l	hand
<i>gece</i>	<i>gecə</i>	түн	түн	tun	night
<i>geçmek</i>	<i>keçmək</i>	өту	өтүү	o'tishi	pass
<i>geçmiş</i>	<i>keçmiş</i>	өткен	өткөн	o'tgan	past
<i>genç</i>	<i>gənc</i>	жас	жаш	yosh	young
<i>geniş</i>	<i>geniş</i>	кең	кең	keng	wide
<i>gitmek</i>	<i>getmək</i>	барыңыз	барып,	borish	go
<i>göndermek</i>	<i>göndərmək</i>	жіберіңіз	жиберүү	yuborish	send
<i>hangi</i>	<i>hansı</i>	кайсы	кайсы	qaysi	which
<i>ışık</i>	<i>işıq</i>	жарық	жарык	yorug'lik	light
<i>ileri</i>	<i>irəli</i>	алға	алдыга	olg'a	forward
<i>kapı</i>	<i>qapı</i>	есік	эшик	eshik	door
<i>karşısında</i>	<i>qarşı</i>	қарама-қарсы	карама каршы	qarama-qarshi	opposite
<i>kaybetti / yitirdi</i>	<i>itirilmiş</i>	жоғалған	жоголгон	yo'qolgan	lost
<i>kırmızı</i>	<i>qırmızı</i>	қызыл	кызыл	qizil	red
<i>kısa</i>	<i>qısa</i>	қысқа	кыска	qisqa	short
<i>korku</i>	<i>qorxu</i>	қорқыныш	коркунуч	qo'rqinch	fear
<i>ödeme</i>	<i>ödəyirəm</i>	төлеуге	төлөө	to'lash	pay
<i>ödeme</i>	<i>ödənış</i>	төлем	төлөө	to'lov	payment
<i>ön</i>	<i>ön</i>	алдыңғы	алдыңкы	oldida	front
<i>pay</i>	<i>pay</i>	үлес	үлүшү	ulushi	share
<i>sağlanan / temin edilen</i>	<i>təmin edildi</i>	берілген	берилген	berilgan	provided
<i>satın almak</i>	<i>satın al</i>	сатып алу	сатып алуу	sotib oling	buy
<i>seçmek</i>	<i>seçin</i>	таңдау	тандоо	tanlash	choose
<i>son</i>	<i>son</i>	жақында	жакынкы	yaqinda	recent
<i>sonra</i>	<i>sonra</i>	кейін	кийин	keyin	after
<i>toprak</i>	<i>torpaq</i>	топырақ	жер	yer	soil
<i>üzerinde</i>	<i>üzərində</i>	үстінде	үстүндө	ustida	over
<i>vergi</i>	<i>vergi</i>	салық	салык	soliq	tax
<i>yumurta</i>	<i>yumurta</i>	жұмыртқа/ тұқым	жумуртка/ тукум	tuxum	egg
<i>adım</i>	<i>addım</i>	қадам	кадам	qadam	step
<i>kadın</i>	<i>qadın</i>	әйел	аял	ayol	woman
<i>pamuk</i>	<i>pambıq</i>	мақта	пахта	paxta	cotton
<i>penceresi</i>	<i>pəncərə</i>	терезе	терезе	deraza	window
<i>şimdi</i>	<i>indi</i>	қазір	азыр	hozir	now
<i>tel</i>	<i>tel</i>	сым	зым	sim	wire
<i>temiz</i>	<i>təmiz</i>	таза	таза	toza	clean
<i>ucuz</i>	<i>ucuz</i>	арзан	арзан	arzon	cheap

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

<i>yüzde</i>	<i>yüzdə</i>	пайызбен	пайыз	foiz	percent
<i>destek</i>	<i>dəstək</i>	қолдау	колдоо	qo'llab-quvvatlash	support
<i>emin</i>	<i>əmin</i>	сенімді	ишенимдүү	ishonch	sure
<i>emirler</i>	<i>əmirləri</i>	бўйрықтар	буйруктар	buyurtmalar	orders
<i>eşyalar</i>	<i>şeylər</i>	заттар нерселер	нерселер	narsalar	things
<i>haber</i>	<i>xəbərlər</i>	жаңалықтар	жаңылык	yangiliklar	news
<i>hazır</i>	<i>hazır</i>	дайын	даяр	tayyor	ready
<i>idare</i>	<i>idarə</i>	басқару	башкаруу	boshqarmoq	govern
<i>insan</i>	<i>insan</i>	адам	адам	odam	human
<i>kitle</i>	<i>kütləvi</i>	масса	массалык маалымат	massa	mass
<i>mahkeme</i>	<i>məhkəmə</i>	сот	сот	sud	court
<i>mavi</i>	<i>mavi</i>	көк	көк	ko'k	blue
<i>mektup</i>	<i>məktub</i>	хат	кат	xat	letter
<i>mesajları</i>	<i>mesajlar</i>	хабарламалар	кабарлар	xabarlar	messages
<i>olaylar / hadiseler</i>	<i>hadisələr</i>	оқиғалар	оқуялар	voqealar	events
<i>pazartesi</i>	<i>pazartesi</i>	дүйсенбі	дүйшөмбү	dushanba	monday
<i>sayfa</i>	<i>səhifə</i>	бет	бет	bet	page
<i>şirket</i>	<i>şirkət</i>	компания	компания	kompaniya	company
<i>taze</i>	<i>təzə</i>	жаңа	жаңы	yangi	fresh
<i>teşekkürler</i>	<i>təşəkkür ediram</i>	рахмет	рахмат	rahmat	thanks
<i>ticaret</i>	<i>ticarət</i>	сауда	соода	savdo	trade
<i>vasitasiyla</i>	<i>vasitəsilə</i>	арқылы	аркылуу	orqali	through
<i>zengin</i>	<i>zəngin</i>	бай	бай	boy	rich

Uzbek, Azeri and Turkish have common words, both from Turkic origin and borrowings. Equivalent of those in Kyrgyz and Kazakh are also common generally.

Tr	Az	Kz	Kr	Uz	En
<i>çünkü</i>	<i>çünki</i>	өйткені	анткени	<i>chunki</i>	because
<i>gece</i>	<i>gecə</i>	түн	түн	<i>kecha</i>	night
<i>gibi</i>	<i>kimi</i>	сияқты	сыяктуу	<i>kabi</i>	like
<i>gidiyor</i>	<i>gedir</i>	бару	бара	<i>ketadi</i>	going
<i>gitti</i>	<i>getdi</i>	барды	барды	<i>ketdi</i>	went
<i>güneş</i>	<i>günəş</i>	күн	күн	<i>quyosh</i>	sun
<i>güzellik</i>	<i>gözəllik</i>	сұлулық	сүлуулук	<i>go'zallik</i>	beauty
<i>istiyorum</i>	<i>istəyirəm</i>	кала	каалоо	<i>istaymiz</i>	want
<i>kuru</i>	<i>quru</i>	құрғақ	куркак	<i>quruq</i>	dry
<i>küçük</i>	<i>kiçik</i>	кішкентай	кичине	<i>kichkina</i>	small
<i>öğrenme</i>	<i>öyrənmə</i>	оқу	оқутуу	<i>o'rganish</i>	learning
<i>toplamak</i>	<i>toplamaq</i>	жинау	жыйноо	<i>to'plash</i>	gather
<i>uzak</i>	<i>uzaq</i>	алыс	алыс	<i>uzoq</i>	far
<i>uzun</i>	<i>uzun müddətdir</i>	ұзақ	үзун	<i>uzoq</i>	long
<i>yağ</i>	<i>yağ</i>	май	май	<i>yog'</i>	oil
<i>yangın</i>	<i>yanğın</i>	от	от	<i>yong'in</i>	fire
<i>yardim</i>	<i>kömək</i>	көмек	жардам	<i>yordam</i>	help

Özel Sayı “Polonya ve Türk Dünyası”, Kasım 2019

<i>dakika</i>	<i>dəqiqə</i>	минут	мүнөт	<i>daqıqa</i>	minute
<i>a kadar</i>	<i>qədər</i>	дейін	чейин	<i>qadar</i>	until
<i>abone ol</i>	<i>abunə olun</i>	жазылу	жазылуу	<i>obuna bo'lish</i>	subscribe
<i>hakkında</i>	<i>haqqında</i>	туралы	жөнүндө	<i>haqida</i>	about
<i>hissetmek</i>	<i>hiss etmək</i>	сезу	сезүү	<i>his</i>	feel
<i>karar vermek</i>	<i>qərar</i>	шешу	чечүү	<i>qaror</i>	decide
<i>meşhur / ünlü</i>	<i>məşhur</i>	атақты	атактуу	<i>mashhur</i>	famous
<i>meyve</i>	<i>meyvə</i>	жеміс	жемиш	<i>meva</i>	fruit
<i>millet</i>	<i>millət</i>	ұлт	улут	<i>millat</i>	nation
<i>milli ulusal</i>	<i>milli</i>	ұлттық	улуттук	<i>milliy</i>	national
<i>renk</i>	<i>rəng</i>	түсі	түс	<i>rangi</i>	color
<i>teklif</i>	<i>təklif edirik</i>	ұсыныс	сунуш	<i>taklif</i>	offer
<i>tıbbi</i>	<i>tibbi</i>	медициналық	медициналык	<i>tibbiy</i>	medical
<i>devlet memleket</i>	<i>dövlət məmləkət</i>	мемлекет	мамлекет	<i>mamlakat davlat</i>	state
<i>ekonomik</i>	<i>iqtisadi</i>	экономикалық	экономикалык	<i>iqtisodiy</i>	economic
<i>hayvan</i>	<i>heyvan</i>	жануар	жаныбар	<i>hayvon</i>	animal
<i>insan</i>	<i>insan</i>	адам	адам	<i>inson</i>	human
<i>kişi</i>	<i>şəxs</i>	адам	адам	<i>shaxs</i>	person
<i>kral</i>	<i>kral</i>	патша	падыша	<i>qirol</i>	king
<i>mühendislik</i>	<i>mühəndislik</i>	инженерия	инженердик	<i>muhandislik</i>	engineering
<i>parola</i>	<i>parol</i>	құпия сөз	қупуя сөз	<i>parol</i>	password
<i>sayfa</i>	<i>səhifə</i>	бет	бет	<i>sahifa</i>	page
<i>şeker</i>	<i>şəkər</i>	қант	кант	<i>shakar</i>	sugar
<i>bahçe</i>	<i>bağ</i>	бақша	бакча	<i>bog '</i>	garden
<i>alınan</i>	<i>qəbul edildi</i>	алынған	кабыл алынған	<i>qabul qildi</i>	received
<i>alınmış</i>	<i>qəbul</i>	қабылданды	кабыл алынған	<i>olingan</i>	taken

etc.

There are common words more than listed above among the 1000 most frequent words. In order to interpret the numbers, framing the context would be useful by asking such a question: in which context this information is valuable? Or in other words; what will those numbers mean?

The Council of Europe put together a guideline named Common European Framework of Reference for Languages (CEFR) between 1989 and 1996 to describe the learners' language proficiency (2001). According to CEFR each level requires a certain amount of words. Common words between Turkic languages provide a lexical proficiency for the learners at beginners level.

Another information to give a context to most frequent words, it is needed to refer to Zipf, whose statistical theory became widely popular starting from 1930's, since corpus linguistics studies today also confirm Zipf's works. (Shuiyuan, Chunshan, Haitao, 2006)

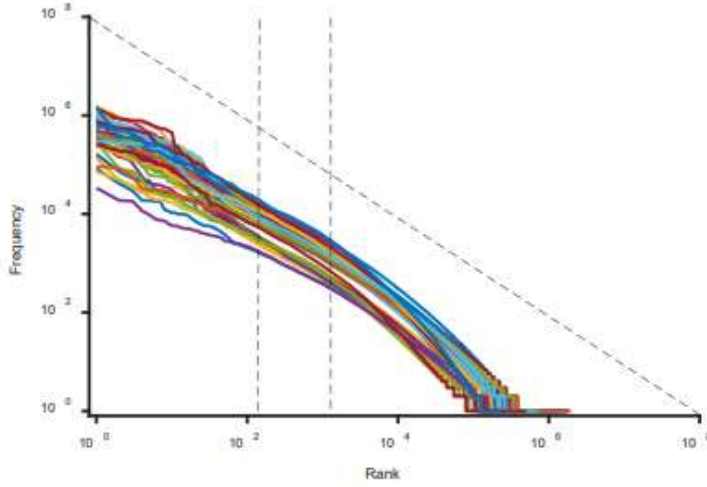


Figure 1. The frequency-rank curves found in 50 languages of Leipzig corpus. The upper segment of each curve (the segment before the 1st vertical line) is unsmooth; the middle segment of each curve (the segment between the 1st and the 2nd vertical lines) is smooth, with the gradient being roughly 1. The lower segment of each curve (the segment after the 2nd vertical line) is also smooth, but bends downward to deviate from the expected line.

Conclusion

As seen above most frequent 1000 words, based on English, are shared by at least 2 or 3 of the Turkic languages. For this study 940 words were compared in total from the most frequent 1000 words lists of English corpuses; this is a more or less number of content words left after functional words were excluded. More than 200 of Turkish words from the list aren't shared by other languages; it is because of the language policy of purification adopted in early republican era. It can be found borrowing equivalents of some of those words exist in Turkish; thus the numbers of unique words for Turkish might decrease. On the other hand Azerbaijan Turkish has fewer words which aren't shared by any other language. The number doesn't exceed 100, thanks to Arabic and Persian loanwords. There are more words which live in languages in topic still exist; but they weren't tagged as "common"; since they are either false friend or archaism. Nearly 200 words in Kazakh and Kyrgyz lists also aren't shared by other 3 languages in terms of meaning; but most of these words are false friends; or more precisely, deceptive cognates. Similar numbers are valid for Uzbek too.

As a conclusion; specially designed Cyrillic alphabets served the purpose of masking the shared vocabulary of Turkic languages during the last century. There's a chance now in the post-Soviet era, for the policy makers of those countries to take into consideration the transnational ties in accordance with their history and culture in the process of making new alphabets.

References

- Common European Framework of Reference for Languages: Learning, Teaching, Assessment (2001) Available at: <https://rm.coe.int/16802fc1bf> (Accessed: 21 April 2019)
- Global Web-based English, Available at: <https://www.english-corpora.org/glowbe/> (Accessed: 21 April 2019)
- Internet world stats (2019) 'Internet World Users by Language Top 10 Languages', Available at: <https://internetworldstats.com/stats7.htm> (Accessed: 21 April 2019)
- Orhun, M. (2017) 'Özbekçe Ünlü Harflerinin Eş Sesli Sözcüklerde Uygurcaya Bilgisayarlı Çevrilmesi', *Süleyman Demirel University Journal of Natural and Applied Sciences*, 21(2), 495-504.
- Shuiyuan YU, Chunshan Xu, Haitao LIU (2006) 'Zipf's law in 50 languages: Its structural pattern, linguistic interpretation, and cognitive motivation', Available at: <https://arxiv.org/ftp/arxiv/papers/1807/1807.01855.pdf> (Accessed: 21 April 2019)