

Towards Kyrgyz stop words

Ruslan Isaev

Computer Science department
Ala-Too International University
Ankara Street 1/8, Tunguch
720048 Bishkek, Kyrgyz Republic
E-mail: ruslan.isaev@alato0.edu.kg
ORCID iD: <https://orcid.org/0000-0003-4426-8837>

Gulzada Esenalieva

Computer Science department
Ala-Too International University
Ankara Street 1/8, Tunguch
720048 Bishkek, Kyrgyz Republic
E-mail: gulzada.esenalieva@alato0.edu.kg
ORCID iD: <https://orcid.org/0009-0000-9135-1671>

Ermek Doszhanov

Computer Science department
Ala-Too International University
Ankara Street 1/8, Tunguch
720048 Bishkek, Kyrgyz Republic
E-mail: ermek.doszhanov@alato0.edu.kg
ORCID iD: <https://orcid.org/0009-0002-4939-5683>

Abstract. The concept of stop words introduced by H. P. Lun in the mid-20th century plays a huge role in today's NLP practice. Stop words are used to reduce noisy text data, remove uninformative words, speed up text processing, and minimize the amount of memory required to store data.

The Kyrgyz language is an agglutinative Turkic language for which no scientific study of stop words has been previously published in English. In our study, we combined frequency analysis with rule-based linguistic analysis. First, we found the most frequently

Submitted: 06/09/2023. **Accepted:** 23/11/2023

Copyright © 2023 Ruslan Isaev, Gulzada Esenalieva, Ermek Doszhanov. Published by Vilnius University Press

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

used words, set a threshold, and removed words below the threshold. This way we got a list of the most frequently used words. Then we reduced the list by excluding from the list all words that do not belong to the category of function words of the Kyrgyz language. Finally, we got a list of 50 words that can be considered stop words in the Kyrgyz language. In our analysis, we used a single corpus of sentences collected and posted as an open source project by one of the local broadcasters.

Keywords: stop words, Kyrgyz language, frequency analysis, Turkic stop words, NLP

Introduction

The concept of stop words was firstly introduced by Hans Peter Luhn, the pioneer of computational linguistics, in 1958. His concept of ‘stop lists’ explained such words as noisy data, which can be neglected (neither indexed nor searchable) by the computational machine. The concept of stop words plays huge role in Natural Language Processing (NLP). As a part of general NLP, the stop words help to remove noisy or irrelevant information from the data, focus on the more meaningful or informative words in the text, speed up the processing time, and reduce the amount of storage needed for storing the data. Removing stop words increase accuracy and efficiency of such NLP techniques, as Topic Modelling, Sentiment Analysis, Information Retrieval, or Feature Selection (Ladani et al. 2020, 466).

Literally, stop words are words that are commonly used in a language but are generally considered to be of little value or significance in the context of text analysis or NLP tasks. Usually such words are removed from texts before any analysis or processing takes place. Here are some examples of stop words in English: *the, and, or, a, an, of, to, in,* and *is* (Ladani et al. 2020, 466).

The purpose of this study is to provide a list of stop words adapted for use in the Kyrgyz language. Since the concept of stop words gives a lot of freedom in interpreting and working with them, the list of stop words presented in this article is not universal.

1 Literature review

1.1 Rules and strategies for including to stop words

The pattern for stop words’ removal, suggesting to remove them before training a model, arises some debates. For instance, there are some works, proposing to remove stop words after model inference (removing stop words after the process of training a model) (Schofield et al. 2017, 435), or even training models without removing stop words (Cordeiro et al. 2004, 137).

The rules for considering words as stop words are derived from the stop words' general meaning: the words that are commonly used but do not carry much meaning in a sentence. The meaning is broad, but there are some general characteristics, which can be taken as stop words' properties (Sadeghi et al. 2014, 479):

- they have little meaning if they are used separately;
- they appear many times in a text;
- they are necessary for the construction of the language;
- they are general words and not particularly used in a certain field;
- they are not used as a search keyword;
- they never form a full sentence when used alone.

Along with this, the numbers (1, 2, 3, etc), emojis, and symbols like (@, #, &, %, *) have to be considered as stop words. Some texts may contain foreign words, these words can be also considered as stop words (Al-Shargabi et al. 2011, 2). In this study, we automatically exclude these symbols from the stop words' list.

It worth to mention, that there is no definitive list of stop words for any language, and the decision to include a word as a stop word may vary depending on the specific application and context. This is why it's better to follow the general strategy: identify words that do not provide much meaningful information in the context of the certain corpus and can be safely removed without affecting the accuracy of the analysis.

1.2 Stop words' identification techniques

Today researchers propose different sophisticated methods for stop words identification: clustering algorithms in machine learning, CF-IDF, the Mutual Information method (MI), methods based on Zipf's Law, TRBS (Kaur et al. 2018, 208). However, even a simple frequency-analysis of word occurrences in a corpus provides better results for the tasks as Information Retrieval: top 10–20 frequent stop words can reduce by 20–30% the size of tokens for processing (Sadeghi et al. 2014, 476).

Along with the above methods, some authors propose to use combinations of different methods. For example, they propose the combination of a statistical model (frequency-analysis), and an information model, where they take any token (stop word) as a signal, and look at how much information or entropy is carried by the signal (Zou et al. 2006, 1012–1013).

One important note is that there is no a universal approach for identifying stop words in a corpus. The best approach depends on the specific characteristics of

the corpus and the language being analyzed. It's often necessary to experiment with different techniques to find the most effective one for a particular corpus.

1.3 Stop words in Turkic languages

Turkic languages are a group of agglutinate languages, spoken by more than 160 million people, including today about thirty languages. The languages of this family include Turkish (72mln), Uzbek (24mln), Azerbaijani (23mln), Kazakh (13mln), Uigur (10.5mln), Tatar (5.4mln), Turkmen (5mln), Kyrgyz (4.4mln), and other languages. Kyrgyz language speakers live in Kyrgyzstan, Uzbekistan, and China (Xinjiang) (Turkic Languages).

Turkic sets of stop words can vary across different dialects and regions of Turkic languages. Additionally, there are variations in the way Turkic languages are written, such as the use of the Cyrillic alphabet in Kazakh and Kyrgyz, or the use of the Latin alphabet in Turkish (Turkic Languages). There are several tools and libraries helping to work with stop words of Turkic languages (stop words-tr library for Turkish, stop words-uz library for Uzbek, etc).

Most of the existing models and methods used to identify stop words, are suitable for European natural language families. However, they can't successfully cover the problem of stop words' identification for agglutinate Turkic language family. In Turkic languages, many stop words are 'masked', and require different techniques for adding to stop lists, such as collocation and bigram methods (Madatov *et al.* 2022, 1).

Additionally, such libraries, as NLTK and SpaCy, provide options for creating custom lists of stop words. It's important to notice, that lists of stop words are flexible, and might vary depending on specific needs, particular application or project.

2 Methodology

In this research, we use a hybrid approach, involving two steps: 1) Frequency-based approach: identification the most common words in a corpus of Kyrgyz sentences with a threshold of 4000 occurrences; 2) Linguistic-based approach: identification words based on their grammatical function (POS identification), then dividing them into content and function words. Finally, we present a list of most frequent function words, as a stop words' list for the Kyrgyz language.

2.1 Frequency-based approach

First, we conducted a frequency analysis to determine the distribution of words in Kyrgyz texts. The corpus of Kyrgyz texts of the Kloop media was tokenized and lowercased. Then, using the FreqDist method of the NLTK library, the frequency of words was found. Having received the 200 most common words, we decided to set the frequency threshold at 4000 word occurrences in the corpus. So, we received 165 words to be subsequently analyzed in the next stage. The list of frequencies of Kyrgyz words can be found in the appendix to this article.

2.2 Linguistic-based approach

Secondly, we have divided the list of most frequent words into two categories: content and function words. Function words are the class of words playing rather a grammatical role in the text (closed-class words): prepositions, conjunctions, determiners, qualifiers, pronouns, interrogatives, numerals, etc. Content words are the class of semantically richer words, having a greater semantic meaning in the text (open-class words): nouns, verbs, adjectives, adverbs. (Bell et al. 2009, 92).

The list of words after removing all content words is provided below:

- **Conjunctions:** *менен* (menen) ‘with’, *жана* (jana) ‘and’, *эми* (èmi) ‘now’, *гана* (gana) ‘only’, *же* (je) ‘or’, *болсо* (bolso) ‘if so’, *катары* (katary) ‘as’, *эгер* (èger) ‘if’, *анткени* (antkeni) ‘because’;
- **Pronouns:** *ал* (al) ‘she/he/it’, *бул* (bul) ‘this’, *анын* (анын) ‘her/his’, *алар* (alar) ‘they’, *өз* (òz) ‘-self/own’, *алардын* (alardyn) ‘their’, *аны* (any) ‘him/her/it’, *мен* (men) ‘I/me’, *биз* (biz) ‘we’, *ага* (aga) ‘to her/him’, *андан* (andan) ‘from her/him’, *ошол* (oşol) ‘from her/him’, *экенин* (ekenin) ‘that’s/of that’, *буга* (buga) ‘to this’, *бардык* (bardyk) ‘(in) all’, *мындан* (myndan) ‘from this’, *өзү* (òzù) ‘her/him/it-self’, *биздин* (bizdin) ‘our’;
- **Prepositions:** *боюнча* (boûnça) ‘by/according’, *үчүн* (ùçün) ‘for’, *да* (da) ‘too/also’, *бирок* (birok) ‘but/however’, *кийин* (kijin) ‘after’, *тууралуу* (tu-uraluu) ‘about’, *чейин* (čejin) ‘till/up to’, *айтымьында* (ajtymynda) ‘according to’, *дагы* (dagy) ‘still/again’, *анда* (anda) ‘then’, *карата* (karata) ‘in relation to’, *тарабынан* (tarabınan) ‘by/from’, *ылайык* (ylajyk) ‘according (to)’, *жөнүндө* (žònùndò) ‘about’, *беру* (beri) ‘since’, *аркылуу* (arkyluu) ‘through’;
- **Numerals:** *бир* (bir) ‘one’, *эки* (èki) ‘two’, *биринчи* (birinçi) ‘first’, *үч* (üç) ‘three’, *бири* (biri) ‘one (of)’;
- **Interjections/Determiners:** *эле* (èle) ‘just’, *ар* (ar) ‘each’.

There are two significant notes: 1) the main reason why we ended up with a list of only 50 words due to the fact that we set a frequency threshold of 4000 occurrences of a word in a corpus; 2) the English translations of the words above can be seen as different parts of speech depending on the context in which they are used; the same is true for Kyrgyz words: they are assigned one part of speech, but the actual POS depends on the context in which they are used.

3 Results

The list of 50 words presented in the part of *Linguistic-based approach*, can be considered as a list of Kyrgyz stop words, based on the corpus of Kyrgyz sentences from Kloop media. There are other lists of open-source Kyrgyz stop words: 1) the list of Turkish stop words translated into Kyrgyz and taken as a set of Kyrgyz equivalents of stop words (both languages are Turkic); 2) the list of Kyrgyz stop words provided by the SpaCy library. While these lists of Kyrgyz stop words can be used, we have not been able to find any scientific basis or research to clarify the inclusion of these particular words into the lists of Kyrgyz stop words.

4 Discussion and conclusion

As it was mentioned above, there are many sophisticated methods that help to define certain sets of stop words. Besides, there are specific strategies, advised for identification of Turkic language stop words. This analysis is carried out on the only corpus of Kyrgyz sentences, collected by the “Kloop” media (2020). It cannot reflect the entire set of Kyrgyz stop words. Other corpora, as the Kyrgyz news Corpus of the Leipzig Corpora Collection (2020), can be analyzed in future as well.

Moreover, in relation to Turkic languages, including Kyrgyz, having rich and diverse affixes’ nature, it might be reasonable to widen the ‘stop lists’ concept, and including **some affixes** into the ‘stop lists’ (Tukeyev *et al.* 2020, 4–5). In this method, the morphological analysis should be carried out before removing ‘stop words’ and ‘stop affixes’.

The main purpose of this study is to fill a gap in research in this area. The approach, proposed in this article, is not ideal. Thus, the list of stop words of the Kyrgyz language proposed in this study can be subsequently modified by more sophisticated methods and models. Finally, we have to mention that it is recommended to define stop words for each particular corpus.

Data sources

The Kloop Media corpus of Kyrgyz texts <https://github.com/kyrgyz-nlp/kloop-corpus>
Transliteration tool (ISO 9 mode) <https://www.transliteration.com/transliteration/en/kyrgyz/iso-9/>
Kyrgyz stop words by SpaCy. https://github.com/explosion/spaCy/blob/master/spacy/lang/ky/stop_words.py
Leipzig Corpora Collection. Kyrgyz news corpus based on material from 2020. https://corpora.uni-leipzig.de?corpusId=kir_news_2020
Turkic Languages. <http://www.languagesgulper.com/eng/Turkic.html>
Turkish stop words translated into Kyrgyz. <https://www.kaggle.com/datasets/crocuta/kyrgyz-language-stopwords>

References

- Al-Shargabi, Bassam, Waseem Al-Romimah & Fekry Olayah. 2011. A comparative study for Arabic text classification algorithms based on stop words elimination. *Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applications (ISWSA'11)*. Association for Computing Machinery, New York, NY, USA, Article 11, 1–5. <https://doi.org/10.1145/1980822.1980833>
- Bell, Alan, Jason Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60, 92–111. <https://doi.org/10.1016/j.jml.2008.06.003>
- Cordeiro, João & Pavel Brazdil. 2004. Learning Text Extraction Rules, without Ignoring Stop Words. *Pattern Recognition in Information Systems 2004*, 128–138. Retrieved from <https://www.di.ubi.pt/~jpaulo/publications/PRIS2004.pdf>
- Kaur, Jashanjot & Preetpal Buttar. 2018. A Systematic Review on stop word Removal Algorithms. *International Journal on Future Revolution in Computer Science & Communication Engineering* 4 (4), 207–210. Retrieved from <http://www.ijfrcsce.org/index.php/ijfrcsce/article/view/1499/1499>
- Ladani, Dhara & Nikita Desai. 2020. Stop word Identification and Removal Techniques on TC and IR applications: A Survey. *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 466–472. <https://doi.org/10.1109/ICACCS48705.2020.9074166>
- Madatov, Khabibulla, Shukurla Bekchanov & Jernej Vičič. 2022. Dataset of stop words extracted from Uzbek texts. *Data in Brief* 43, 108351, 1–7. <https://doi.org/10.1016/j.dib.2022.108351>
- Sadeghi, Mohammad & Jesús Vegas. 2014. Automatic identification of light stop words for Persian information retrieval systems. *Journal of Information Science* 40, 476–487. <https://doi.org/10.1177/0165551514530655>

- Schofield, Alexandra, Måns Magnusson & David Mimno. 2017. Pulling Out the Stops: Rethinking stop word Removal for Topic Models. *Conference of the European Chapter of the Association for Computational Linguistics*, 432–436. <http://dx.doi.org/10.18653/v1/E17-2069>
- Tukeyev, Ualsher, Aidana Karibayeva & Zhandos Zhumanov. 2020. Morphological segmentation method for Turkic language neural machine translation. *Cogent Engineering* 7, 1856500, 1–15. <https://doi.org/10.1080/23311916.2020.1856500>
- Zou, Feng, Fu Lee Wang, Xiaotie Deng, Song Han & Lu Wang. 2006. Automatic construction of Chinese stop word list. *Proceedings of the 5th WSEAS International Conference on Applied Computer Science, Hangzhou, China, April 16–18, 2006*, 1010–1015. Retrieved from <https://www.cs.cityu.edu.hk/~lwang/research/hangzhou06.pdf>

Appendix

Index	Kyr	Translit	Eng	Freq
1	менен	menen	with	71302
2	жана	jana	and	59732
3	ал	al	she/he/it	56863
4	деп	dep	saying	44065
5	боюнча	boûnča	by/according	40321
6	бул	bul	this	35208
7	бир	bir	one	33546
8	үчүн	ùçùn	for	31914
9	эле	èle	just	26122
10	болгон	bolgon	was/has been	24342
11	да	da	too/also	24247
12	анын	anyn	her/his	22929
13	бирок	birok	but/however	22528
14	эмес	èmes	not	18599
15	болуп	bolup	being	17835
16	кийин	kijin	after	16500
17	жок	žok	no	16133
18	тууралуу	tuuraluu	about	14763
19	чейин	čejin	till/up to	14720
20	керек	kerek	needed	14600
21	бар	bar	have/is	13815
22	алып	alyp	taking	13786
23	алар	alar	they	13069
24	деди	dedi	said/told	12865

Index	Kyr	Translit	Eng	Freq
25	<i>эки</i>	èki	two	12328
26	<i>эми</i>	èmi	now	12269
27	<i>кыргызстандын</i>	kyrgyzstandyn	of Kyrgyzstan	11739
28	<i>өз</i>	ôz	-self/own	11409
29	<i>алардын</i>	alardyn	their	11090
30	<i>жылдын</i>	žyldyn	of a/the year	10731
31	<i>аны</i>	any	him/her/it	10669
32	<i>жылы</i>	žyly	year	10410
33	<i>жатат</i>	žatat	lies	10379
34	<i>айтымында</i>	ajtymynda	according to	10266
35	<i>дагы</i>	dagy	still/again	10186
36	<i>деген</i>	degen	said/told	10127
37	<i>мен</i>	men	I/me	10005
38	<i>сөз</i>	sôz	word	9916
39	<i>турган</i>	turgan	was standing	9858
40	<i>кыргыз</i>	kyrgyz	Kyrgyz	9808
41	<i>биз</i>	biz	we	9796
42	<i>билдирди</i>	bildirdi	has reported	9694
43	<i>ош</i>	oř	Osh	9688
44	<i>гана</i>	gana	only	9645
45	<i>каршы</i>	karřy	against	9458
46	<i>мамлекеттик</i>	mamlekettik	(belonging to) state	9311
47	<i>башка</i>	bařka	other	8684
48	<i>алган</i>	algan	has taken/received	8624
49	<i>жаткан</i>	žatkan	was lying/lied	8580
50	<i>болот</i>	bolot	will be	8481
51	<i>ошондой</i>	ořon DOJ	the same	8449
52	<i>анда</i>	anda	then	8313
53	<i>жаңы</i>	žañy	new	8278
54	<i>берген</i>	bergen	has given/gave	8223
55	<i>басма</i>	basma	printed	8166
56	<i>учурда</i>	uçurda	at the moment	8129
57	<i>же</i>	je	or	8076
58	<i>адам</i>	adam	human	7921
59	<i>эч</i>	eç	no-(thing)	7817
60	<i>үй</i>	üy	house	7639
61	<i>башкы</i>	bařkı	main	7588
62	<i>кандай</i>	kanday	how	7426
63	<i>ага</i>	aga	to her/him	7258

Index	Kyr	Translit	Eng	Freq
64	<i>андан</i>	andan	from her/him	7220
65	<i>иш</i>	iş	work	7172
66	<i>айтып</i>	aytıp	saying	7017
67	<i>бишкек</i>	bişkek	Bishkek	6998
68	<i>президент</i>	prezident	President	6989
69	<i>kloop</i>	kloop	kloop	6961
70	<i>ошол</i>	oşol	that	6935
71	<i>жол</i>	jol	way	6891
72	<i>келген</i>	kelgen	came	6840
73	<i>болсо</i>	bolso	if (will be) so	6832
74	<i>мүмкүн</i>	mümkün	maybe/perhaps	6806
75	<i>кыргызстан</i>	kırgızstan	Kyrgyzstan	6743
76	<i>көп</i>	köp	a lot/many	6690
77	<i>ар</i>	ar	each	6612
78	<i>мыйзам</i>	mıyzam	law	6591
79	<i>экегин</i>	ekenin	that's	6510
80	<i>сүрөт</i>	süröt	picture	6502
81	<i>биринчи</i>	birinçi	first	6424
82	<i>ата</i>	ata	father	6323
83	<i>мурдагы</i>	murdagı	former	6191
84	<i>башичысы</i>	başçısı	chief/head	6152
85	<i>карата</i>	karata	in relation to	6132
86	<i>кылмыш</i>	kılmıř	crime	6116
87	<i>калган</i>	kalgan	remained/other	6097
88	<i>тарабынан</i>	tarabınan	by/from	6079
89	<i>жолу</i>	jolu	way/time	6064
90	<i>укук</i>	ukuk	the right	6040
91	<i>үч</i>	üç	three	5985
92	<i>алуу</i>	aluu	getting/receiving	5984
93	<i>жогорку</i>	jogorku	high	5938
94	<i>миң</i>	miñ	thousand	5926
95	<i>чек</i>	çek	cheque	5884
96	<i>баш</i>	baş	head	5863
97	<i>кол</i>	kol	hand	5842
98	<i>буга</i>	buga	to this	5807
99	<i>кабыл</i>	kabil	accept	5797
100	<i>маалымат</i>	maalymat	information	5797
101	<i>бири</i>	biri	one (of)	5776
102	<i>билдирген</i>	bildirgen	reported	5756

Index	Kyr	Translit	Eng	Freq
103	<i>айтты</i>	ajtty	told/said	5739
104	<i>1</i>	1	1	5715
105	<i>эл</i>	èl	people/public	5691
106	<i>катары</i>	katary	as	5653
107	<i>мындай</i>	myndaj	this way/like this	5546
108	<i>күнү</i>	kùnù	day	5479
109	<i>болду</i>	boldu	was/happened	5467
110	<i>сом</i>	som	som	5318
111	<i>дейт</i>	dejt	says	5266
112	<i>каза</i>	kaza	died/death	5246
113	<i>ылайык</i>	ylajyk	according (to)	5203
114	<i>талап</i>	talap	require/demand	5151
115	<i>орун</i>	orun	place/seat	5145
116	<i>улам</i>	ulam	little by little	5142
117	<i>кеткен</i>	ketken	gone	5107
118	<i>болчу</i>	bolču	was	4904
119	<i>өткөн</i>	òtkòn	the past/passed	4902
120	<i>млн</i>	mln	mln	4889
121	<i>жөнүндө</i>	žònùndò	about	4878
122	<i>бардык</i>	bardyk	(in) all	4878
123	<i>бери</i>	beri	since	4877
124	<i>премьер</i>	prem'er	Prime Minister	4865
125	<i>келип</i>	kelip	come/arrived	4831
126	<i>атамбаев</i>	atambaev	Atambayev	4829
127	<i>коргоо</i>	korgoo	defend/protection	4806
128	<i>чыккан</i>	čykkàn	came out/was released	4804
129	<i>5</i>	5	5	4789
130	<i>аралык</i>	aralyk	distant	4784
131	<i>эң</i>	èñ	the most	4782
132	<i>кыргызстанда</i>	kyrgyzstanda	in Kyrgyzstan	4771
133	<i>эгер</i>	èger	if	4755
134	<i>10</i>	10	10	4750
135	<i>акча</i>	akča	money	4704
136	<i>берүү</i>	berùù	giving	4689
137	<i>жатышат</i>	žatyšat	(are) lying	4681
138	<i>жыл</i>	žyl	year	4656
139	<i>ичинде</i>	ičinde	inside	4636
140	<i>сот</i>	sot	the court	4608
141	<i>ала</i>	ala	take	4571

Index	Kyr	Translit	Eng	Freq
142	<i>азыр</i>	azyr	now	4566
143	<i>милиция</i>	miliciâ	the police	4566
144	<i>нече</i>	neče	few/several	4550
145	<i>киши</i>	kiši	man	4495
146	<i>өзү</i>	ôzù	her/him/it-self	4481
147	<i>чыгып</i>	čygyp	going/letting/coming out	4478
148	<i>атайын</i>	atajyn	specially/on purpose	4459
149	<i>анткени</i>	antkeni	because	4434
150	<i>шайлоо</i>	šajloo	election	4433
151	<i>депутат</i>	deputat	Deputy	4401
152	<i>эмне</i>	ëmne	what	4394
153	<i>жардам</i>	žardam	aid/help/assistance	4345
154	<i>мындан</i>	myndan	from this	4344
155	<i>иштеп</i>	ištep	working	4300
156	<i>биздин</i>	bizdin	our	4220
157	<i>шаардык</i>	šaardyk	municipal/urban	4220
158	2	2	2	4193
159	<i>улуттук</i>	uluttuk	national	4178
160	<i>сатып</i>	satyp	buying	4173
161	<i>кызматы</i>	kyzmaty	service	4161
162	<i>жакшы</i>	žakšy	good	4093
163	<i>айткан</i>	ajtkan	told/said	4090
164	<i>кызматкерлери</i>	kyzmatkerleri	service workers/officers	4040
165	<i>аркылуу</i>	arkyluu	through	4033