



• Turkic languages (SOV, agglutinative, vowel harmony)



heavy



Indiana University jonwashi@indiana.edu

UiT Norgga Árktalaš Universitehta francis.tyers@uit.no

Special thanks to: Tolgonay Kubatova Aida Sundetova Ağarahim Sultanmuradov

Jonathan North Washington Ilnar Salimzyanov Francis M. Tyers Kазан (Идел буе) федераль университеты ilnar.salimzyan@gmail.com

DESIGNING FINITE-STATE MORPHOLOGICAL TRANSDUCERS

FOR KYPCHAK LANGUAGES

 			Glos	S					
Кудай		жаратканынын	баарына	карап,			жакшы		
Құдай	Өзінің	жаратқандарының	бәріне	қарап,		өте	жақсы	екенін	көрді.
Аллаһ	Yзе	яраткан	нәрсәләргә	карап,	аларның	бик	яхшы	икәнен	күрде.
Аллагь	Овзю	яратгъан	затлагъа	къарап,	олар	бек	яхшы	экенин	гёрген.
God	own-his	created	[everything/thing-s]-to	looked.at,	they/their	very	good	being	saw.

'God looked at everything he had created and saw that it was very good.' (Bible, Genesis 1:31)

	Kyrgyz	Kazakh	Tatar	Kumyk		······································	put	
	\dmtkmz\	/qazaq/	/totar/	/qumuq/	Kyrgyz (kir)	Kazakh (kaz)	Tatar (tat)	Kun
lassification	Eastern	Southern	Northern	Western	Кудай Өзү жаратканынын баарына карап, өтө жакшы экенин көрдү.	Құдай Өзінің жаратқандарының бәріне қарап, өте жақсы екенін көрді.	Аллаһ Үзе яраткан нәрсәләргә карап, аларның бик яхшы икәнен күрде.	Алла къар
opulation of s	speakers					Құдай <n><nom> Өз<prn><ref><px3sp><gen></gen></px3sp></ref></prn></nom></n>	Аллаh <n><nom> Yз<prn><ref><px3sp><nom></nom></px3sp></ref></prn></nom></n>	Алла 0ьз<
umber	3M	8M-12M	5.4M	430K	жарат <v><tv><ger_past><px3sp><gen> баары<prn><qnt><px3sp><dat></dat></px3sp></qnt></prn></gen></px3sp></ger_past></tv></v>	жарат <v><tv><ger_past><pl><px3sp><gen>бәрі<pre>prn><qnt><px3sp><dat></dat></px3sp></qnt></pre></gen></px3sp></pl></ger_past></tv></v>	ярат <v><tv><gpr_past> нәрсә<n><pl><dat></dat></pl></n></gpr_past></tv></v>	ярат зат<
rimary	Kyrgyzstan	Kazakhstan	Tatarstan	Dagestan	кара <v><tv><gna_perf> ,<cm></cm></gna_perf></tv></v>	қара <v><tv><gna_perf> ,<cm></cm></gna_perf></tv></v>	кара <v><tv><gna_perf> ,<cm></cm></gna_perf></tv></v>	къар , <cm< td=""></cm<>
econdary	China, etc.	China, Mongolia	Bashqortostan		—— өтө <adv></adv>	—————————————————————————————————————	алар <pre>prn><pers><p3><pl><gen> бик<adv></adv></gen></pl></p3></pers></pre>	олар бек<
xternal influe	nces				жакшы <adj> э<cop><ger_past><px3sp><acc></acc></px3sp></ger_past></cop></adj>	жақсы <adi> e<cop><ger past=""><px3sp><acc></acc></px3sp></ger></cop></adi>	яхшы <adv> яхшы<adj> и<cop><ger_past><px3sp><acc></acc></px3sp></ger_past></cop></adj></adv>	ЯХШЬ Э<СО
Iongolic	moderate	moderate	light	light	κθρ<ν> <tv><ifi><p3><sg>.<sent></sent></sg></p3></ifi></tv>	κθρ<ν> <tv><ifi><p3><sg>.<sent></sent></sg></p3></ifi></tv>	κγp <v><tv><past><p3><sg>.<sent></sent></sg></p3></past></tv></v>	гёр< . <se< td=""></se<>
ghuz	_	_	light	moderate		Tag	set	

..... Morphological transducers.....

- Take a surface form, and produce valid lexical form(s)
- Take a lexical form, and produce valid surface form(s) 'алдым' ↔ ал<v><tv><ifi><p1><sg>, алд<n><px1sg><nom>

..... Transducers for Turkic languages

- Turkish (Çöltekin, 2010 & 2014; Oflazer, 1994)
- Crimean Tatar (Altıntaş, 2001)
- Turkmen (Tantuğ et al., 2006)
- Kazakh (Бекманова & Махимов, 2013)
- our Kyrgyz, Kazakh, Tatar, Kumyk: all GPL (=free and open)!Framework: HFST.....

- Reimplements Xerox FST formalisms (lexc & twol) • Also provides a wrapper around popular free/open-source FST
- toolkits: SFST, OpenFST, and Foma
- Part of Apertium Turkic project:
- http://wiki.apertium.org/wiki/Apertium_Turkic
- Transducers available live at turkic.apertium.org
- Source code available from apertium's svn repo
- Turkic RBMT mailing list (>25 subscribers): apertium-turkic@lists.sourceforge.net Feel free to post in any language!
- See our papers in LREC proceedings (2012: Kyrgyz, 2014: Kazakh, Tatar, Kumyk)
- And feel free to contact the authors any time!

Kyrgyz	(kir)		Kazakh (kaz)		Tatar (tat))	Ku	myk (kum)	
	Өзү жаратканынын ба өтө жакшы экенин к		Құдай Өзінің жаратқ қарап, өте жақсы ек			е яраткан нәрсәла бик яхшы икәнен п		агь Оьзю яратгъю рап, олар бек ях	эн затлагъа кшы экенин гёрге
03 <prn жарат< баары< кара<v ,<cm> — өтө<аd жакшы< э<сор></cm></v </prn 	adj> <ger_past><px3sp><a <tv><ifi><p3><sg></sg></p3></ifi></tv></a </px3sp></ger_past>	t>	Құдай <n><nom> θз<prn><ref><px3sp> жарат<v><tv><ger_pa бәрі<prn=""><qnt><px3s қара<v=""><tv><gna_per ,<cm=""> — өте<adv> жақсы<adj> e<cop><ger_past><px көр<v=""><tv><ifi><p3> .<sent> .</sent></p3></ifi></tv></px></ger_past></cop></adj></adv></gna_per></tv></px3s></qnt></ger_pa></tv></v></px3sp></ref></prn></nom></n>	st> <pl><px3 p><dat> f> 3sp><acc></acc></dat></px3 </pl>	Ssp> <gen> яра́т<v>< нәрсә<n> кара<v>< ,<cm> алар<prn бик<adv> яхшы<adj; и<cop><ge< th=""><th>ref><px3sp><nom> tv><gpr_past> <pl><dat> tv><gna_perf> ><pers><p3><pl><q< th=""><th>0ьз яра зат къа ,<0 ола бек яхш асс> э<0 гёр</th><th>arb<n><nom> s<pre>s<pre>prn><ref><px3spre>s<pre>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n< pre=""><pre>c<n<<pre>c<n< pre=""><pre>c<n<<pre>c<n< pre=""><pre>c<n< p=""><pre>c<n< p="">c<n< p=""><pre>c<n< p=""><pre>c<n<n<n>c<n<n<n<n>c<n<n<n<n>c<n<n<n<n<n< th=""><th>st> erf> <pl><nom> x3sp><acc></acc></nom></pl></th></n<n<n<n<n<></n<n<n<n></n<n<n<n></n<n<n></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></n<<pre></pre></n<></n<<pre></pre></n<></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pl></n></pl></n></pl></n></pl></n></pl></n></pl></n></pl></n></pre></px3spre></ref></pre></pre></nom></n></th></q<></pl></p3></pers></gna_perf></dat></pl></gpr_past></nom></px3sp></th></ge<></cop></adj; </adv></prn </cm></v></n></v></gen>	ref> <px3sp><nom> tv><gpr_past> <pl><dat> tv><gna_perf> ><pers><p3><pl><q< th=""><th>0ьз яра зат къа ,<0 ола бек яхш асс> э<0 гёр</th><th>arb<n><nom> s<pre>s<pre>prn><ref><px3spre>s<pre>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n< pre=""><pre>c<n<<pre>c<n< pre=""><pre>c<n<<pre>c<n< pre=""><pre>c<n< p=""><pre>c<n< p="">c<n< p=""><pre>c<n< p=""><pre>c<n<n<n>c<n<n<n<n>c<n<n<n<n>c<n<n<n<n<n< th=""><th>st> erf> <pl><nom> x3sp><acc></acc></nom></pl></th></n<n<n<n<n<></n<n<n<n></n<n<n<n></n<n<n></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></n<<pre></pre></n<></n<<pre></pre></n<></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pl></n></pl></n></pl></n></pl></n></pl></n></pl></n></pl></n></pre></px3spre></ref></pre></pre></nom></n></th></q<></pl></p3></pers></gna_perf></dat></pl></gpr_past></nom></px3sp>	0ьз яра зат къа ,<0 ола бек яхш асс> э<0 гёр	arb <n><nom> s<pre>s<pre>prn><ref><px3spre>s<pre>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pl>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n><pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n<<pre>c<n< pre=""><pre>c<n<<pre>c<n< pre=""><pre>c<n<<pre>c<n< pre=""><pre>c<n< p=""><pre>c<n< p="">c<n< p=""><pre>c<n< p=""><pre>c<n<n<n>c<n<n<n<n>c<n<n<n<n>c<n<n<n<n<n< th=""><th>st> erf> <pl><nom> x3sp><acc></acc></nom></pl></th></n<n<n<n<n<></n<n<n<n></n<n<n<n></n<n<n></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></pre></n<></n<<pre></pre></n<></n<<pre></pre></n<></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></n<<pre></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pre></n></pl></n></pl></n></pl></n></pl></n></pl></n></pl></n></pl></n></pre></px3spre></ref></pre></pre></nom></n>	st> erf> <pl><nom> x3sp><acc></acc></nom></pl>
			Intropoitivo		O		ntonco		Vorbal advor
<n></n>	Noun	<iv></iv>	Intransitive	<nom></nom>	'Nominative'	<sent> Se</sent>		<gna_perf></gna_perf>	Verbal adverl
< V >	Verb	<tv></tv>	Transitive	<gen></gen>	Genitive	<past> Pa</past>	ıst (General)		(Perfect)
<pre><pre><pre><pre><pre><pre><pre><pre></pre></pre></pre></pre></pre></pre></pre></pre>	Pronoun	<p3></p3>	Third person	<acc></acc>	Accusative	<ifi> Pa</ifi>	ist	<gpr_past></gpr_past>	Verbal adject
<det></det>	Determiner	<pl><</pl>	Plural	<dat></dat>	Dative	(.	Eyewitness/R	ecent)	(Past)
<adj></adj>	Adjective	<ref></ref>	Reflexive	<qnt></qnt>	Quantifier	<px3sp> 3r</px3sp>	d person poss.	. <ger past=""></ger>	Verbal noun
to due	Adverb	<ners></ners>	Personal	<cm></cm>	Comma		Singular/Plura		(Past)

•	Example: morphologically distinct adjective classes (not doc-
	umented elsewhere)

- Other Turkic transducers: 0-derivation (overgenerates)
- Our approach: categorisation (generates/analyses only correct forms)

Type	Gloss	<adj>(<comp>)</comp></adj>	<adj>(<comp>)<subst></subst></comp></adj>	<adj>(<comp>)<advl></advl></comp></adj>
A1 A2 A3 A4	ʻgood' ʻold' ʻdead' ʻbasic'	яхшы (яхшырак) иске (искерәк) үле (—) төп (—)	яхшы (яхшырак) иске (искерәк) үле (—) — (—)	яхшы (яхшырак) — (—) — (—) — (—)

	letters	values	examples
kaz	и, у, ю	/wej, aw, jaw/ /wej, we, jev/	қиюд <mark>а</mark> 'chopping down' киюд <mark>е</mark> 'getting dressed'
tat	e	э / С _ /j/+ы /j/+э	дәресләр 'lessons' еллар 'years' егетләр 'boys'
		/ø, y/ / C _	гюнлер 'days' гёзлер 'eyes'
kum	ë, ю	/jø, jy/	юреклер 'hearts' ёнкюлер 'darlings'
		/jo, ju/	юлдузлар 'stars' ёллар 'roads'
adjus	counted-f st rules fo	for words get a or harmony-for	ver majority of examples harmony-forcing character cing characters
adjus	counted-factorial counted for the state of t	for words get a or harmony-for Loar	harmony-forcing character
adjus Lette	counted- st rules for ers that re "back" ve	for words get a or harmony-fore Loar present front verseits in Russia	harmony-forcing character cing characters words owels in native words may repre-
adjus Lette	counted- st rules for ers that re "back" vo native we елдін 'со	for words get a r harmony-ford Loar present front versent sin Russiant Production of the Europe Ford example	harmony-forcing character cing characters words owels in native words may repre-
Lette sent kaz tat kum	counted-fat rules for that re "back" vo native we елдің 'со галимнәр сёзлер 'м	for words get a r harmony-fore Loar Loar present front verse in Russia ord example Fountry's' For scientists' a cords' cords'	harmony-forcing character cing characters words owels in native words may repre- an words Russian word example Назарбаевтын 'Nazarbayev's'
LEXI	counted-fat rules for that re "back" vo native we елдің 'со галимнәр сёзлер 'м	for words get a r harmony-fore Loar present front versent sin Russia and cords'	harmony-forcing character cing characters words words an words Russian word example Назарбаевтын 'Nazarbayev's' ртистлар 'artists' амолётлар 'airplanes'
kaz tat kum solut :%{%	counted-fat rules for that re we ending ceзлер 'we cesnep 'we ces	for words get a r harmony-fore Loar Loar present front versions in Russia and example For scientists' a cords' carte continuation RUS	harmony-forcing character cing characters words words an words Russian word example Назарбаевтын 'Nazarbayev's' ртистлар 'artists' амолётлар 'airplanes'
kaz tat kum solut :%{ LEXI aptu	counted-fat rules for that re 'back' vo native we елдің 'со галимнәр сёзлер 'м ion: sepa CON N1-2%} N1; CON Nouser: арти	for words get a r harmony-fore Loar Loar present front versions in Russia and example Fountry's Artificial Continuation RUS	harmony-forcing character cing characters nwords owels in native words may represen words Russian word example Назарбаевтын 'Nazarbayev's' ртистлар 'artists' амолётлар 'airplanes' on lexicon (messy rules) ! "artist"

..... Acronyms and numerals

twol rules handle phonology for spelt-out words

			TAT T	C .	
Part of s	peech _		Number o	of stems	
		Kyrgyz	Kazakh	Tatar	Kumy
Noun		4582	2640	2795	256
Verb		1193	1470	1143	38
Adjective		1211	_754	816	21
Proper no	oun	5887	5701	5361	144
Adverb		312	171	177	6
Numeral	•	66	63	63	4
Conjunct	llon	77	46	45	1
Postposit Pronoun	.1011	50 51	50 32	43 28	1
Determin	16r	64	32	39	3
)	101	04			J
Total:		13749	11224	10737	484
		Test	corpora		
W	/ikipedia	News	Relig		
	ikipedia	azattyk.org		n + Bible	
	⁷ ikipedia ⁷ ikipedia	azattyq.org tat.tatar-infor		n + Bible n + New Te	actamont
ırar vv ı myk —	-	yoldash.etnos		sis + New 16	
us receiv	ving ≥ 1	analysis	of surface		C
ous receivantes de la receivante de la r	ving ≥ 1 nbiguity n found if n - of a fo		number of a corpus ses, % corr	analyses	for each
ous received fream and ace form recision Recall - 9 or a form	ving ≥ 1 nbiguity found if of a found of analogous n, by cor	analysis - average an analysed orm's analysel lyses providing againg	number of a corpus ses, % corr ded by trans ainst a gold	analyses ect sducer th standard	for each at are cor
us received fean and ace form recision second form or a form	ving ≥ 1 nbiguity found if of a found n, by cor	analysis - average an analysed orm's analysel orm's analysel orm's analyselyses providing against a second	number of a corpus ses, % corr ded by trans ainst a gold on results	analyses ect sducer th standard	for each at are cor
us received and and ace form a form a form	ving ≥ 1 nbiguity n found if n - of a found n, by cor Corpus	analysis - average nonalysed orm's analyselyses providing against a record of the control of the	number of a corpus ses, % correlated by translatinst a gold on results Coverage	analyses ect sducer th standard	for each at are cor
Jus received Alean and ace form Precision Recall - 9 or a form	ving ≥ 1 nbiguity found if of a found n, by cor	analysis - average nonalysed orm's analyselyses providing against a record of the control of the	number of a corpus ses, % correlated by translatinst a gold on results Coverage	analyses ect sducer th standard	for each at are cor
Jean and ace form ace form a form a form a form	ving ≥ 1 nbiguity n found if n - of a found n, by cor Corpus	analysis - average nonalysed orm's analyselves provious aparing against the Tokens of the Maring and the Maring against the Mar	number of a corpus ses, % correlated by translatinst a gold on results Coverage	analyses ect sducer th standard	for each at are cor
Is received to the second of t	ving ≥ 1 nbiguity found if of a for of analon, by cor Corpus Wikiped Average	analysis - average nonalysed orm's analysel provide a provide a provide a paring against the second	number of a corpus ses, % correded by translainst a gold on results Coverage	analyses ect sducer th standard (a) (b) (c) (c) (d) (d) (d) (d) (d) (d) (d) (d) (d) (d	for each at are cor d Amb.
Is received fream and the form of the form	ving ≥ 1 nbiguity found if of a for of analon, by cor Corpus Wikiped	analysis - average nonalysed orm's analysel provide a provide a provide a paring against the second	number of a corpus ses, % correded by translainst a gold on results Coverage 1 85.61	analyses ect sducer th standard (e) (e) (w) (h) (h) (h) (h) (h) (h) (h) (h) (h) (h	for each at are con
Is received from the form the	ving ≥ 1 nbiguity found if of a for of analon, by cor Corpus Wikiped Average Wikiped	analysis - average norm's analysed providing against a material ma	number of a corpus ses, % corrected by translainst a gold on results Coverage 1 85.61 ± 92.12 ± 100.000 m	analyses ect sducer th standard (e) (%) (b) (c) (4) (4) (4) (4) (5) (4) (4) (5) (6) (7) (7) (7) (7) (7) (7) (7) (7) (7) (7	for each at are cor Amb. 2.43
Is received fean and ace form recision form form form form form form form form	ving ≥ 1 nbiguity n found in n - of a for of analon, by cor Corpus Wikiped Average Wikiped News	analysis - average reading analysed orm's analysed orm's analysed orm's analysed orm's analysed orm's analyses provided analyses provided analyses provided analyses provided analyses provided analysed orm's analysed	number of a corpus ses, % corrected by translainst a gold on results Coverage 1 85.61 ± 92.12 ± 100.000 m	analyses ect sducer th standard	for each at are cor Amb. 2.43 2.88
Alean an ace form ace form a f	ving ≥ 1 nbiguity found if found if of a for of analon, by cor Corpus Wikiped Average Wikiped News Religion Average	analysis - average non analysed orm's analysed orm's analysed orm's analysed orm's analyses provided analyses provided analyses provided analyses provided analyses provided analyses provided analysed orm's analysed orm's analyses provided analyses prov	number of a corpus ses, % corrected by translainst a gold on results Coverage 1 92.12 ± 92.49 ± 90.07 ±	analyses ect sducer th standard (a) (b) (b) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	for each at are cor at are cor and 2.43 2.88 2.88 2.63 2.64
us received lean and ace form recision Recall - 9 or a form recision and recision and recision and recision are a form recision are a form recision and recision are a form recision and recision are a form recision and recision are a form recision are a fo	ving ≥ 1 nbiguity n found if n - of a for No of analon, by cor Corpus Wikiped Wikiped News Religion	analysis - average non analysed orm's analysed orm's analysed orm's analysed orm's analyses provided analyses provided analyses provided analyses provided analyses provided analyses provided analysed orm's analysed orm's analyses provided analyses prov	number of a corpus ses, % corrected by transfainst a gold on results Coverage 1 92.12 ± 92.49 ± 90.07 ± 86.35	analyses ect sducer th standard (e) (%) (b) (4) (4) (4) (5) (7) (6) (7) (7) (7) (7) (8) (8) (9) (1) (8) (9) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	for each at are cor at
us received lean and ace form recision Recall - 9 or a form 54474)	ving ≥ 1 nbiguity found if f	analysis - average in analysed orm's analysis lyses providing against the second sec	number of a corpus ses, % corrected by transfainst a gold on results Coverage 1 92.12 ± 92.49 ± 90.07 ± 86.35 ± 89.75	analyses ect sducer th standard (a) (b) (b) (c) (c) (d) (d) (d) (d) (d) (d) (d) (d) (d) (d	for each at are cor and 2.43 2.88 2.63 2.64 2.24
Alean and ace form recision Recall - 9 or a form form form form form form form form	ving ≥ 1 nbiguity found if of a for of analon, by cor Corpus Wikiped Wikiped News Religion Average Wikiped News Religion Average	analysis - average in analysed orm's analysis orm's	number of a corpus ses, % corrected by transfainst a gold on results Coverage 1 92.12 ± 92.49 ± 92.49 ± 86.35 ± 89.75	analyses ect sducer th standard (e) (w) (b) (c) (d) (d) (d) (d) (d) (d) (e) (e) (f) (e) (f) (e) (f) (e) (f) (f) (f) (f) (f) (f) (f) (f) (f) (f	for each at are cor at and at are cor at and at are cor at and at are cor at
Is received fean and ace form fecision form form form form form fazakh f	ving ≥ 1 nbiguity n found in n - of a for % of analon, by cor Corpus Wikiped Wikiped News Religion Average Wikiped News Religion Average	analysis - average in analysed orm's analysis orm's	number of a corpus ses, % corrected by transfainst a gold on results Coverage 1 92.12 ± 92.49 ± 92.49 ± 92.49 ± 86.35 ± 89.75 ± 91.25 ± 89.12	analyses ect sducer th standard (e) (a) (b) (c) (c) (d) (d) (d) (d) (e) (e) (f) (e) (f) (f) (e) (f) (f	for each at are cord at 2.43 2.88 2.63 2.64 2.24 2.30 2.24
Is received fean and ace form fecision for a form form fazakh faz	ving ≥ 1 nbiguity found if fou	analysis - average in analysed orm's analysed orm's analysed orm's analysed orm's analysed orm's analyses provided anal	number of a corpus ses, % corrected by translating a gold on results Coverage 1 92.12 ± 92.49 ± 92.49 ± 92.49 ± 91.25	analyses ect sducer th standard (**) (**) (**) (**) (**) (**) (**) (**	for each at are cord Amb. 2.43 2.88 2.88 2.63 2.64 2.24 2.24 2.26
lean and ace form recision recision for a form form form form form form form form	ving ≥ 1 nbiguity n found in n of a for No of analon, by cor Corpus Wikiped News Religion Average Wikiped News Religion Average News Religion Average	analysis - average rance and analysed corm's analysis provided analyses provided an	number of a corpus ses, % corrected by translating a gold on results Coverage 1 92.12 ± 92.49 ± 92.49 ± 92.49 ± 91.25	analyses ect sducer the standard (a) (b) (b) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c	for each at are cord at 3 2.43 2.88 2.63 2.64 2.24 2.24 2.26 1.53
Viean and acceptance form acceptance form a form acceptance form acceptance form acceptance for a form accepta	ving ≥ 1 nbiguity n found in n of a for No of analon, by cor Corpus Wikiped Average Wikiped News Religion Average News Religion Average News Religion Average	analysis - average rance and analysed corm's analysis provided analyses provided an	number of a corpus ses, % corrected by translating a gold on results Coverage 1	analyses ect sducer th standard (**) (**) (**) (**) (**) (**) (**) (**	for each at are cord at are cord and 2.43 2.88 2.63 2.64 2.24 2.30 2.24 2.26 1.53 1.53 1.53
Ace formate formate for a formate form	ving ≥ 1 nbiguity n found in n - of a for % of analon, by cor Corpus Wikiped News Religion Average Wikiped News Religion Average News Religion Average proofed un	analysis - average reading analysed orm's analysed orm's analysed orm's analysed orm's analysed orm's analyses provided	number of a corpus ses, % corrected by translating a gold on results Coverage 1	analyses ect sducer th standard (**) (**) (**) (**) (**) (**) (**) (**	for each at are cord at are cord and 2.43 2.88 2.63 2.64 2.24 2.30 2.24 2.26 1.53 1.53 1.53
Viean an ace form recision Recall - 9 or a form 154474) Cazakh 150260) Cumyk 150300) Clected & Jangu	ving ≥ 1 nbiguity found if found if for a for of a for of analom, by cor Corpus Wikiped Average Wikiped News Religion Average News Religion Average Proofed un uage	analysis - average in analysed orm's analysed organization of the second organization of the second organization organiza	number of a corpus ses, % corrected by translating a gold on results Coverage 1	analyses ect sducer th standard (a)	for each at are cord at are cord 2.43 2.88 2.63 2.64 2.24 2.26 1.53 1.53 1.53 corpora call (%)
Alean an ace form recision Recall - 9 or a form 154474) Cazakh 150547) Catar 150260) Cumyk 150300) Clected & part 150300	ving ≥ 1 nbiguity found if found if for a for of a for of analom, by cor Corpus Wikiped Average Wikiped News Religion Average News Religion Average Proofed un uage	analysis - average reading analysed orm's analysed orm's analysed orm's analysed orm's analysed orm's analyses provided	number of a corpus ses, % corrected by translating a gold on results Coverage 1	enalyses ect sducer th standard (e (%) (e (%) (fe (%)	for each at are cord 2.43 2.88 2.63 2.64 2.24 2.26 1.53 1.53 1.53 corpora





- Disambiguation, more stems
- Clean up Kyrgyz transducer
- Machine translation between these languages
- Bring our other Kypchak transducers to comparable perfor-
- Qaraqalpaq, Bashqort, Nogay, Crimean Tatar
- Other Turkic lgs.: Uzbek, Uyghur, Chuvash, Yakut, Tuvan,