

**Türk Lehçeleri Arasında Yapılacak Makine Çevirisine Kısa Bir Bakıř:**  
**Başkurt Morfolojisinin İki Seviyeli Tanımı Örneğinde<sup>1</sup>**

Dr. Kalmamat KULAMSHAEV<sup>2</sup>

**Özet**

*Adriyatik'ten Çin seddine kadar uzanan topraklarda Türk dilini konuşan halkların yaşadığı bir gerçektir. Bu coğrafyada yaşayan topluluklar, herşeyiyle Türk Dünyasını oluşturur. Türk Dünyası üzerine yapılan çalışmalar, önce batılı bilginler tarafından yürütüldüyse de günümüzde Türk Dünyası mensuplarınca çok çeşitli boyutlarda yürütüldüğü aşikardır. 1990'lı yıllardan bu yana dil, tarih, kültür vb. alanlarda gerek Türkiye'de, gerekse Türk Dünyasında bu tür çalışmaların var hızıyla devam ediyor olması elbette gurur vericidir. Bu çalışmaların, Oğuzhan Aydın'ın "Bağımsızlıklarının 25. yılında TÜRKSOY'un Türk Topluluklarına Yapmış Olduğu Katkılar" adlı yazısında belirttiği gibi, Türk Dünyası üzerine her alanda yapılan faaliyetlerin resmini çıkararak ilgili tüm birimlerin konuya yönelik planlamalarına büyük katkı sağlayacağı kesindir. Böylece, 1990'lı yıllardan itibaren yapılan faaliyetlerin içinde zayıf kalan yönün, Türk dilleri ve Türkçe üzerinde Doğal Dil İşleme alanında yapılan çalışmaların başlangıç seviyesinde olduğunu görüyoruz. Her ne kadar Türkçe için Türkiye'de önde gelen üniversitelerde Bilgisayar Mühendisliği Bölümlerinde bilimsel arařtırmalar yapılıyor olsa da, Türkî Dillerde yapılan çalışmalar dar bir kapsamda kalmıřtır. Bunun temel nedenlerinden bir tanesi bu dillerde yetkin arařtırmacıların olmaması ve Türk Lehçelerini öğreten Çağdař Türk Lehçeleri Bölümleri ile Bilgisayar Mühendisliği Bölümleri arasında bilimsel iş birliđi ve ortak proje çalışmalarının yeterince yapılamamıř olmasıdır. Bu bildiride, üzerinde çalışmakta olduğum daha büyük bir Başkurtça-Türkçe Makine Çevirisi projesi olarak düşündüğüm programın bir parçası olan, Oflazer'in iki seviyeli Doğal Dil İşleme açısından Başkurtçanın orthografisine/ imla yapısına ve morphotactic/ morfolojik çözümlenmesine kısa bir bakıř sunulacaktır.*

**Anahtar Kelimeler:** Türk Dünyası, Türk Dilleri, Makine çevirisi, Doğal dil işleme, Sonlu durum makinesi.

**A Brief Overview On The Machine Translation Among Turkic Dialects: In**  
**The Example Of Two Level Description Of Bashkir Morphology**

**Abstract**

It is a fact that the territories from Adriatic through The Great Walls of China are inhabited by Turkic speaking people. These communities with their everything belong to Turkish World.

<sup>1</sup> Bu makale 21–23 Ekim 2018 tarihleri arasında Antalya'da düzenlenen II. Uluslararası Sosyal Arařtırmalar ve Davranıř Bilimleri Sempozyumu'nda sunulan özet bildirinin geliştirilmiř halidir.

<sup>2</sup> Kırgız Cumhuriyeti Yazarlar Birliđi Üyesi, Türkolog, tercüman, řair, kalmamat@yahoo.com.tr

Although the studies on the Turkish World initially were conducted by the western scholars, it is known that currently different studies about its various aspects have been carried out by the members of the Turkish World. Certainly, it is a pride to see that from 1990s the researches in the fields of language, culture, history and etc. continue to be carried out at full steam as in Turkey, as well as in the Turkish World. Of course, as Oğuzhan Aydın mentioned in his writing “The Contribution of TURKSOY to Turkic Communities in its 25th Anniversary”, these works will make a great contribution to the planning of all the bodies regarding the subject by showing the big picture of all the works done in every field in Turkish World. Thus, we can see that the weakness of the studies carried out from 1990s up to nowadays is that the level of the studies in the field of Natural Language Processing of Turkic languages and Turkish is at a beginner level. Although different scientific researches are being carried out on Turkish by the Computer Engineering Departments of the leading universities in Turkey, the studies on Turkic Languages remain narrow. One of the reasons for this is the lack of the competent researchers on these languages and insufficiency of the cooperation and joint studies between Modern Turkic Dialects Departments and Computer Engineering Departments. In this paper, a brief review of the orthographic and morphotactic analysis of Bashkir in terms of Oflazer’s Two Level Language Processing, which is a part of the the program I am currently working on and planning to present as a bigger Bashkir –Turkish Machine Translation, will be introduced.

**Key Words:** Turkish World, Turkic Languages , Machine Translation, Natural Language Processing, Finite State Machine

## Giriř

Son zamanlarda Türk ve Türk dünyası kavramları büyük önem kazanan ve önem arz eden birer kavramlar haline gelmiřtir.

Günümüzde, *Adriyatik’ten Çin seddine* kadar uzanan topraklarda Türk dilini konuşan halkların yařadığı bir gerçektir. Bu coğrafyada yařayan topluluklar, herřeyiyle Türk Dünyasını oluşturur. Türk Dünyası üzerine yapılan çalışmalar, önce batılı bilginler tarafından yürütüldüyse de, günümüzde Türk Dünyası mensuplarınca çok çeřitli boyutlarda arařtırmaların yapıldığı aşıkardır. 1990’lı yıllardan bu yana dil, tarih, kültür vb. alanlarda gerek Türkiye’de, gerekse Türk Dünyasında Türklük bilgisi üzerine yapılan çeřitli çalışmaların var hızıyla devam ediyor olması elbette gurur vericidir. Bu tür çalışmaların saptanması, Aydın (2017)’de belirtildiği gibi, Türk Dünyası üzerine her alanda yapılan *faaliyetlerin resmini çıkararak ilgili tüm birimlerin konuya yönelik planlamalarına* büyük katkı sağlayacağı kesindir. Böylece, 1990’lı yıllardan itibaren yapılan faaliyetlerin içinde zayıf kalan yönün, *Türk dilleri ve Türkçe üzerinde Doğal Dil İşleme alanında yapılan çalışmaların başlangıç seviyesinde olduğunu* görüyoruz. Her ne kadar Türkçe için Türkiye’de önde gelen üniversitelerde Bilgisayar Mühendisliği

Bölümlerinde bilimsel araştırmalar yapılıyor olsa da, Türkî Dillerde yapılan çalışmalar dar bir kapsamda kalmıştır.

Aslında “Kyrgyz Orthography and Morphotactics with Implementation in NUVE.” başlıklı bir bildiriye, konuyla ilgili dar kapsamda da olsa yapılan çalışmalardan şu şekilde bahsetmektedir:

Türk dilleri arasında çeviri amaçlı bir çok projeler ve çalışmalar yapılmıştır. Onlardan bir tanesi Azerbaycanca–Türkçe arasında Hamzaoğlu tarafından gerçekleştirilmiştir. Bir diğer çevri sistemi ise Kırım Tatarcası ve Türkçe arasında Altıntaş tarafından geliştirilmiştir. DILMAC başlangıçta Türkmençe-Türkçe için üretilmiş olan bir Türk Dilleri makine çeviri sistemidir. Tantuğ tarafından geliştirilen Türkmençe-Türkçe bir başka makine çevirisi de mevcuttur. DILMAC içerisinde Uygurca-Türkçe makine çevirisi Keskin tarafından yapılmıştır. (GÖRMEZ Z. v., 2016)

Ancak Başkurtça-Türkçe üzerine benzeri bir projenin olmadığı görülmektedir; dolayısıyla bu bildiriye, üzerinde çalışmakta olduğum daha büyük bir Başkurtça-Türkçe Makine Çevirisi projesi olarak düşündüğüm programın bir parçası olan, Oflazer’in iki seviyeli Doğal Dil İşleme (NLP) açısından Başkurtçanın orthografisine/ imla yapısına ve morphotactic/ morfolojik çözümlenmesine kısa bir bakış sunulacaktır.

### **Başkurtlar**

Başkurt etnonimi IX.-XI. yy’daki kaynaklarda *başkard*, *başgird*, *başkerd*, *baskart*, *bacgard*, *başcart*, *bascert* vb. şekillerde geçmektedir. Bir efsaneye göre, *Ergenekon* destanında olduğu gibi, düşmanın soykırımına uğrayıp, dağda mahsur kalan Türk boyunu, bir *kurt baş* (kılavuz) olmuş ve yol göstererek kurtarmıştır. Bu kurtuluştan sonra, kurtulan Türk boyu, *Başkurt* adıyla anılmaya başlamıştır (ЮЛДАШЕВ, 1997.).

Başkurtlar, günümüzde Rusya Federasyonu Başkurdistan Özerk Cumhuriyeti başta olmak üzere Rusya’nın çeşitli bölgelerinde yaşayan Türk halklarıdır.

1989 istatistiklerine göre, Başkurtların nüfusu 1 milyon 449 bin olup, Başkurtistan’da Ana dilini konuşanların sayısı ise 1 milyon 75 bin kişidir (a.g.k.).

### **Başkurt Kiril Alfabesi**

Başkurtların İslamiyetin kabul edilmesine kadar Run alfabesini, İslamiyetin kabul edilmesiyle birlikte Arap alfabesini, Sovyetlerin kuruluşundan sonra 1928-1939 yılları arasında Latin, 1940’dan bu yana Kiril asıllı Başkurt alfabesini kullandığını biliyoruz.

**1. Tablo: Kiril Asıllı Bařkurt Alfabeti ve Latince Transkripsiyonu**

Kiril	Latin	Kiril	Latin	Kiril	Latin	Kiril	Latin
А а	a	И и	i	Р р	r	Ш ш	sh
Б б	b	Й й	y	С с	s	Щ щ	řç
В в	v	К к	k	Ç ç	ř / ř	Ъ ъ	'
Г г	g	К к	k/q	Т т	t	Ы ы	I
Ғ ғ	ğ/γ	Л л	l	У у	u, w	Ь ь	'
Д д	d	М м	m	Ү ү	ü, w	Э э	é, e
З з	z / ź	Н н	n	Ф ф	f	Ә ә	ä
Е е	é, ye-, e	Ң ң	ñ	Х х	ħ/x	Ю ю	yu
Ё ё	ö, yo-	О о	ö	Һ һ	h	Я я	ya
Ж ж	j	Ө ө	ö	Ц ц	ts		
З з	z	П п	p	Ч ч	ch		

### Bařkurtça

Bařkurt dili, Ural-Altay dil ailesinin bir kolu olan Türk Dili'nin Kıpçak grubunun Kıpçak-Bulgar alt grubunda yer alır. Bařkurtçaya Kıpçak lehçeleri arasında en yakın olanı Tatarca'dır. Eklerin varyantlılıđı açısından Kazakça ve Kırgızca'ya, dudak uyumu açısından yine Kırgızcaya yakındır (a.g.k.). Sonuç olarak Bařkurtça sondan eklemeli bir dil olup, yeni kelimelerin türetilmesi için köklere ekler gelerek yapılır. Ekler köklere gelirken belirli ses deđişiklerine uğrar. Elbette ses deđişmeleri geliřigüzel rastgele deđişiklikler deđildir. Her dilde olduđu gibi Bařkurtçanın da kendine özgü ses yapısı, belirli kanunları, kuralları vardır: kök ve eklerin bir araya gelmesi, ünlü ünlü uyumu, ünlü ünsüz uyumu, ünsüz ünsüz uyumu vb.

Dilin aslında iki görünümü vardır: birincisi eki almadan önceki, diđeri ek aldıktan sonraki görünümü. Mesela Türkçede sonu /p/, /ç/, /t/, /k/ ünsüzleriyle biten kelimelere ünlüyle bařlayan bir kelime geldiđinde istisnaları dıřında bu ünsüzlerde tonlulařma meydana gelecektir.

O halde, sözcüklerin, ek almadan önceki görünümü, sözcük(sel), ek aldıktan sonraki görünümü ise yüzey(sel) görünümleri olarak karřımıza çıkmakta, bu da iki seviyeli morfolojik modellemeyi gerektirmektedir (OFLAZER, 1994; bk. Ayrıca: GÖRMEZ Z. v., GÖKGÖZ, 2011; ÖZTOPRAK, 2015). Konuyla ilgili makalelerde, "İki seviyeli morfoloji, dođal dillerin morfolojik kurallarını kodlamak için dilden bađımsız bir yöntem, olduđu "Bu kodlama kelimelerin sözcük ve yüzey olarak iki biçimde temsil edildiđi; İki seviyeli kuralların da dönüşümü tanımladıđı (ÖZTOPRAK, 2015; bk. Ayrıca: OFLAZER, 1994; GÖRMEZ Z. v., GÖKGÖZ, 2011) belirtilmektedir. Burada

kök ve eklerin, seslerin sıralanışındaki dilin doğal işleyişindeki koşulları<sup>3</sup> sağ ve sol koşul olarak belirtilerek, hangi koşulda ne tür dönüşüm olacağı kuralı tanımlanmıştır.

Oflazer (1994) Türkçenin fonetik kurallarını 22 iki-seviyeli kurallara göre kodladığını, şekil yapısını da fiil yapımı, isim yapımı ve çekim örnekleriyle, sonlu durum makinelerine uygun olarak (finite-state) kodlanabilirliğini örneklerle kanıtlamıştır. Kodlamalarda elbette bizim bildiğimiz gramer kurallarından biraz farklı yöntemler kullanıldığını söyleyebiliriz. Ama yine de bu kodlamalar dilin fonetik ve morfolojik yapısının doğal işleyişine uygun olarak düzenlenmesi gerekir.

## Başkurtça Sesler

### Ünlüler Ve Ünlülerin Sınıflandırılması

Kiril asıllı Başkurt alfabesinde ünlüler için kullanılan 10 harf vardır: *a, ə, o, ø, y, ɣ, e, ə, u, ɯ*. Yuldaşev (1997) ünlüleri şu şekilde sınıflandırmıştır:

**2. Tablo. Başkurtça Ünlülerin Sınıflandırılması.**

Çenenin duruma göre	Teşekkül noktasına göre →	Sıra			
		Ön		Arka	
	Dudağın durumuna göre →	Düz	Dudak	Düz	Dudak
Dar		и (i)	Ү (ü)		у (u)
Yarı dar		е (é)		ы (ı)	
Yarı geniş			ø (ǔ)		о (ö)
Geniş		ə (ä)		а (a)	

### Ünsüzler Ve Sınıflandırılması

Başkurt Türkçesinde 26 ünsüz vardır. Bu ünsüzlerin, Yuldaşev (1997)'de yapılan sınıflandırılması, biraz değişiklik yapılarak aşağıdaki tabloda gösterilmektedir:

**3. Tablo. Başkurtça Ünsüzlerin Sınıflandırılması.**

TEŞEKKÜL TARZINA GÖRE (TEMAS DERESESİNE)	TEŞEKKÜL NOKTASINA GÖRE				
	Dudak	Ön damak			

<sup>3</sup> Örneğin: /t/ ve /d/ ile başlayan eklerin ne zaman /t/ ne zaman /d/ olacağı geleneksel dil öğretiminden de bildiğimiz üzere, geldikleri kelime tabanının, fıstıkçışahap'tan biriyle bitmesi koşuluna; /p/, /ç/, /t/, /k/ ile biten kelimelerde bu son seslerin yumuşamasının istisnaları dışında yumuşası, bu seslerle biten kelimelerden sonra ünlü ile başlayan bir morfemin gelmesi koşuluna bağlıdır.

GÖRE)			Çift dudak	Diş dudak	Dil üstü	Diş arası	Diş - diş eti	Ön damak	Orta damak	Orta dil	Dil ardi	Küçük dil	Boğaz	
Hava akışına göre	Patlamalı	sedalı	б (b)		т (t)		д (d)			г (g)				
		sedasız	п (p)						ч (ç)	к (k)		к (k/q)		
	Sızıcı	Tek fokuslu	sedalı	у~ү (w)	в (v)	з (z)	з (z)	с (s)					ғ (ğ/ğ)	һ (h)
			sedasız		ф (f)		ç (ç)					х (h)		
		Çift fokuslu	sedalı							ж (j)				
			sedasız						ш (ş)					
	sonor	Akıcı								й (y)				
		Patlayıcı	Geniz	м (m)						н (n)	ң (ñ)			
		Sürekli	Avurt						л (l)					
		Titrek							р (r)					

NOT: y'eleşen harfler olarak bildiğimiz, “Ё ё” harfi /y/ ve /o/, “Ю ю” harfi /y/ ve /u/, “Я я” harfi ise /y/ ve /a/ seslerinden oluşan birleşik sestir. Ayrıca çift ses olarak “Ц ц” harfi /t/ ve /s/, “Щ щ” harfi de /ş/ ve /ç/ seslerinden oluşmaktadır ve ödünç alınan kelimelerde kullanılmaktadır. Bunların dışında Rusça ve Rusça vasıtasıyla geçen kelimelerde görülen ayırma/ kalınlaştırma işareti “Ъ ъ” ile inceltme işareti “Ь ь”

## Ses Uyumları

### 4. Tablo. Ünlü Uyumu.

Kelimenin son hecesindeki ünlü	ekin ünlüsü	Örnek kelimeler
а, у, ы, ю, я	а, ы	барыу, барлыктан, кулдан, юғарыға, һораны
е, и, ү, ә	е, ә	күтәрелгәнән, күрсетеп, күмләкте, көлгә, сине, миндән, тәгәрәгәндәр
о	о, а	бороу, королдо, бойорғанда, коштокомы
ө	ө, ә	өйөнә, теләгән, өшөнә,

### 5. Tablo. Ünsüz Uyumu

Kelimenin son ünsüzü	Ekin başındaki ünsüz	Örnek kelimeler
-ж, -з, -л, -м, -н, -ң	д-, г-, ф-	багаждары, ихтыяждан, персонаждар, репортажды, тиражға, колхозда, билде, миндә, һиндә, туғанда, закондар
-з, -и, -й, -р, -у, -ү	з-, г-, ф-	беззекә, сүзгә, тигән, йәйгән, куйзы, торзо, асыузан, килеүгә
-к, -к, -п, -с, -ç, -т, -ф, -х, -ш	к-, к-, т-	халыкка, кунакта, асқыстағы, курста, хәрәфкә, шкапка, тарихты, оймахан, юлдашка

NOT: Ünsüz-ünsüz uyumuna isimden isim yapma eki olan –llk, soru eki –mI ve olumsuzluk eki -mA uymaz; ayrıca kilterew “getirmek” fiilinde görülen -DIr eki de bu bağlamda -her ne kadar kalıplaşmış olsa da- istisna olarak bu uyum dışındadır.

## İki Seviyeli Kurallar

Başkurtçada, görüldüğü gibi, ünlü uyumları da, ünsüz uyumu da güçlüdür. Bu da, bildirinin başında bahsedildiği gibi, Başkurtçada eklerin çok varyantlı olduğuna işaret etmektedir. Bu da, Başkurtçanın morfolojisinin iki seviyeli kodlanmasında, dilin fonetik ve morfonolojik yapısının doğal işleyişine uygun olarak, ünlü ve ünsüz uyumlarına göre değişen seslerin çok dikkat edilerek kodlanmasını, kodlanması için de seslerin özelliklerinin net olarak programa tanıtılmasını gerektirmektedir:

<b>Dar ünlüler</b>	: и /i/, ү /ü/, у /u/
<b>Yarı dar</b>	: е /é/, ы (ı)
<b>Yarı geniş</b>	: ө (ü), о (ö)
<b>Geniş</b>	: ә (ä), а (a)

**Sedalı konsonantlar** : б, г, ф, д, з, ж, з, й, л, м, н, ң, р, у, ү,

**Sedasız konsonantlar**: к, к, п, с, ç, т, ф, ц, х, ш, ш

Kiril asıllı alfabe olmasından dolayı Başkurt alfabesindeki çift sesi karşılayan harfler ise (ë /yo/, я /ya/, ю /yu/, ц /ts/, ш /şç/), bünyesindeki harflerin özelliklerine göre aşağıda resimde görüldüğü gibi sınıflandırılacaktır (bkz. 1. Resim). Yine hem ünlü y /u/, ү /ü/, hem ünsüzü (y, /w/ kalın, ү /w/ ince) karşılayan harfler ise sadece ünlüler kümesinde yer alacaktır. Programa gerektiği yerlerde eklenecektir.

Başkurtçada daha önce bahsedildiği üzere, eklerin çok varyantları vardır. Bu da aynı ek olup, ünlü ve ünsüz uyumlarına göre değişen seslerin büyük harflerle kodlanmasını gerektirmektedir; ki, bu kodlanan harfler ortography.xml dosyasındaki alfabe tanımlamasına konulacak (bk. 1. Resim) ve kodlanan harflerin dönüşüm kuralları yazılacaktır. Mesela çokluk ekinin Başkurtçada ünlüsü iki, ünsüzü dört şekilli olmak üzere 8 varyantı vardır. Görüldüğü gibi hem ünsüzü hem ünlüsünde değişiklik vardır. O halde ortography.xml dosyasında kodlanan seslerin Başkurtça ünsüz uyumu çerçevesinde dönüşüm L kuralı, tanımlanacaktır; ayrıca ünlü uyumu çerçevesinde de dönüşüm A kuralı da yazılacaktır. Bu dönüşüm hangi eklerde görülüyorsa, suffix.txt dosyasında, gerekli eklerin karşısında (rules/dönüşüm sütununda) belirtilecektir (bk. 2. Resim).

### 1. RESİM. Ortography.xml dosyası

```

1 <?xml version="1.0" encoding="utf-8"?>
2
3 <!DOCTYPE orthography SYSTEM "../orthography.dtd">
4
5 <orthography lang="Bsk-BSK">
6
7 <alphabet>
8   <consonants>бвггдзжэйккклмнңппррстфхцчшшшLDGGQ</consonants>
9   <vowels>азэеиоюуүяәAUI</vowels>
10
11 </alphabet>
12

```

## 2. RESİM. suffix.txt dosyası.

id	lexicalForm	type	Grup	flags	rules	surfaces	Örnekler
1	IC_COGUL_LAR	LAR	I	COGUL	DONUSUM_L, DONUSUM_A		лар, лэр, дар, дэр, тар, тэр, зар, зэр
2	IC_SAHİPLİK_BEN_(U)m	U	I	IYELİK	DONUSUM_U, DUSME_UNLU	м, ым, ен, он, өң	балам, се
3	IC_SAHİPLİK_SEN_(U)n	U	I	IYELİK	DONUSUM_U, DUSME_UNLU	н, ын, өң, өң	балаң, сәс
4	IC_SAHİPLİK_O_(U)u	hU	I	IYELİK	DONUSUM_U, DUSME_UNLU	ы, е, о, ө, һы, һе, һо, һө	балаһы
5	IC_SAHİPLİK_BİZ_(U)buз	UGUз	I	IYELİK	DONUSUM_U, DUSME_UNLU	быз, бөз, бөз, бөз, ыбыз,	
6	IC_SAHİPLİK_SİZ_(U)Guz	UGUз	I	IYELİK	DONUSUM_U, DUSME_UNLU, DONUSUM_G	гыз, гэр, гэр,	гөз,

Böylece orthography.xml dosyasında değişen seslerinin dönüşüm kuralları tanımlanan ve sözcüksel düzeyde LAR olarak kodanan ek, surfaces/ yüzeysel kolonunda görülen dönüşüm varyantları olarak gerçekleşmesini sağlayacaktır, ancak kelimelerin de root.txt dosyasında işlenmiş olması gerekmektedir (bk. 3 Resim).

## 3. RESİM. Root.txt dosyası.

root	surfaces	lex active	Id	flags	rules	anlami
кеше		ISIM	isim			
кала		ISIM	isim			
дала		ISIM	isim			
таң		ISIM	isim			
кул		ISIM	isim			
көң		ISIM	isim			
китап	китап	ISIM	isim		YUMUSAMA_nб	
күлдәк	күлдәк	ISIM	isim		YUMUSAMA_xr	

**Sözcüksel** : китап+LAR isim(kitap)+ IC\_COGUL\_Lar  
**Yüzeysel** : китап0тар китаптар

Bu bütün benzer kodlu ekler için geçerli olacaktır.

Elbette bu bildiriye, Başkurtçanın imla yapısını (orthography) ve şekilsel işleyiş döngüsünü (morphotactic) sığdırmak mümkün değildir; dolayısıyla, üzerinde çalışmakta olduğum daha büyük bir Başkurtça-Türkçe Makine Çevirisi projesi olarak düşündüğüm programın bir parçası olan, Oflazer'in iki seviyeli Doğal Dil İşleme açısından Başkurtçanın orthografisine/ imla yapısı ile morphotactic/ morfolojik çözümlenmesine kısa bakış sunulabilmek için birkaç kural verilecektir.



### Yumuşama Kuralı

1. k:g <=> V \_\_ +:0 (@:0)V
2. q:ğ <=> V \_\_ +:0 (@:0)V
3. 3. p:b <=> V \_\_ +:0 (@:0)V

Bu kural, önceki morfemin к /k/, к /q/, п /p/ sesleri ile bitmesi, sonraki morfemin ünlü ile başlaması koşulunu gerektirmektedir (bk. ZAFER, 2011; ayrıca bk. OFLAZER, 1994). Bu kuralda, bir ekin gelmesiyle kelime tabanında eşzamanlı olarak gerçekleşen bir ses olayı tanımlanmaktadır.

#### YUMUSAMA\_кг

**Sözcüksel** : күлдәк+Уң ISIM (gömlek)+ IC\_SAHİPLİK\_SEN\_(U)n  
**Yüzeysel** : күлдег0ем күлдәгем (gömleğim)

#### YUMUSAMA\_кф

**Sözcüksel**: тарак+sU+минән ISIM (tarak)+IC\_SAHİPLİK\_O\_(s)U +  
IC\_HAL\_VASITA\_minän  
**Yüzeysel** : тараф00ы\_минән тарағы минән (tarağıyla, tarağı ile)

#### YUMUSAMA\_пб

**Sözcüksel** : китап+Um ISIM (kitap)+IC\_SAHİPLİK\_BEN\_(U)m  
**Yüzeysel** : китаб0ым китабым (kitabım)

Kelime tabanında gerçekleşen ses olayları için ses olayının adı ile değişen seslerin harf karşılıkları girilerek kural yazılır ve root.txt dosyasında adı geçen ses olayının gerçekleştiği kelimelerin surface/ yüzeysel kolonuna değiştiği hali yazılacaktır (bk. 3.Resim). Kelime tabanının son sesine göre gelen ekin de değişime uğradığı durumlar için dönüşüm kuralları tanımlanmıştır.

### Dönüşüm Kuralları

#### DONUSUM\_L

1. L:д <=> [ л | м, н | н | з | ж ] (':') +:0 \_\_
2. L:т <=> [ б | в | г | д | к | к | п | с | с | т | ф | х | h | ц | ч | ш | ш ] (':') +:0
3. L:з <=> L:т<=> [ з | й | р | у | ү | и | ю ] (':') +:0

Yukarıdaki dönüşüm kuralının tamamına göre, sözcüksel L'nin dönüşümü, belirtilen koşullarda yüzeyde, sırasıyla L=>д, L=>т, L=>з olarak, diğer durumlarda L=>л olarak gerçekleşeceği anlaşılmaktadır.

<b>Sözcüksel</b> : һан+ЛАр	ISIM (sayı)+IC_COGUL_LAr
<b>Yüzeysel</b> : һан0дар	һандар (sayılar)
<b>Sözcüksel</b> : мәктәп+ЛАр	ISIM (mekterp)+ IC_COGUL_LAr
<b>Yüzeysel</b> : мәктәп0тәр	мәктәптәр
<b>Sözcüksel</b> : тоз	ISIM (tuz)+ IY_SIFAT_sUz+IC_COGUL_LAr
<b>Yüzeysel</b> : тоз0һоз0зар	тозһоззар (tuzsuzlar, beceriksizler)
<b>Sözcüksel</b> : һоло+ЛАр	ISIM(yulaf)+ IC_COGUL_LAr
<b>Yüzeysel</b> : һололар	һололар (yulaflar)

Bu kural, DONUSUM\_L koşulları doğrultusunda dönüşüm yapan ekler için de uygulanabilir. Mesela:

<b>Sözcüksel</b> : мин+ЛА	ZAMIR(ben)+IC_HAL_BULUNMA_LA
<b>Yüzeysel</b> : мин0дә	миндә
<b>Sözcüksel</b> : кош+ЛА	ISIM (kuş)+IC_HAL_BULUNMA_LA
<b>Yüzeysel</b> : кош0та	кошта (kuşta)
<b>Sözcüksel</b> : күз+ЛА	ISIM (göz) IC_HAL_BULUNMA_LA
<b>Yüzeysel</b> : күз0зә	күззә (gözde)
<b>Sözcüksel</b> : кала+ЛА	ISIM(şehir)+IC_HAL_BULUNMA_LA
<b>Yüzeysel</b> : кала0ла	калала (şehirde)
<b>Sözcüksel</b> : кеше+ЛА	ISIM (kişi)+IC_HAL_BULUNMA_LA
<b>Yüzeysel</b> : кеше0лә	кешелә (kişide)

#### DONUSUM\_I

I : ы => [ а | о | у | ы | ю | я ] (‘:’) +:0

I : е => [ ә | е | и | ө | ү ] (‘:’) +:0

Yukarıdaki dönüşüm kuralının tamamına göre, sözcüksel I'nın dönüşümü, belirtilen koşullarda yüzeyde, sırasıyla I=>ы, I=>е olarak gerçekleştirileceğini anlatmaktadır. Bu kuralı ortography.xml dosyasındaki tanımlanması, 4. Resimde gösterilmektedir.

<b>Sözcüksel</b> : юк +һIn+Gаһ	ISIM(yok)+IY_FIIL_sIn+ FC_ZAMAN_GORULEN_UZAK_GECMIS_GAn +SAHIS_ZAMIR_BEN_mUn
<b>Yüzeysel</b> : юк0һын0ған0мын	юкһынғанмын (sıkılmışım, özlemişim)

#### 4. RESİM. DONUSUM\_I kuralı otrografy.xml.

```
<rule id="DONUSUM_I" phase="2">
  <description>Önceki ekin son sesli harfine göre I => u,e, olur </description>
  <transformation morpheme="This" action="Replace" operandOne="I" operandTwo="u" flag="all" >
    <conditions flag="Or">
      <condition morpheme="Previous" operator="LastVowelEquals" operand="aoyuoo" />
    </conditions>
  </transformation>
  <transformation morpheme="This" action="Replace" operandOne="I" operandTwo="e" flag="all" >
    <conditions flag="Or">
      <condition morpheme="Previous" operator="LastVowelEquals" operand="eemey" />
    </conditions>
  </transformation>
</rule>
```

#### Düşme Kuralları

##### DUSME\_UNLU

A | U : 0 => V+:0\_\_@\*

Bu düşme kuralının tamamına göre, sözcüksel A | U kodlarının karşılıkları sesler, belirtilen koşullarda yüzeyde düşer. Düşen sesler suffix.txt dosyasında parantez içine alınarak kodlanacaktır (bk. 2. Resim).

**Sözcüksel** : кала+Uн ISIM (şehir)+IC\_SAHİPLİK\_SEN\_(U)n  
**Yüzeysel** : кала00н калаң (şehrin).

##### DUSME\_UNSUZ

s : 0 => V+:0\_\_@\*

Bu düşme kuralının tamamına göre, sözcüksel (s), belirtilen koşullarda yüzeyde düşer.

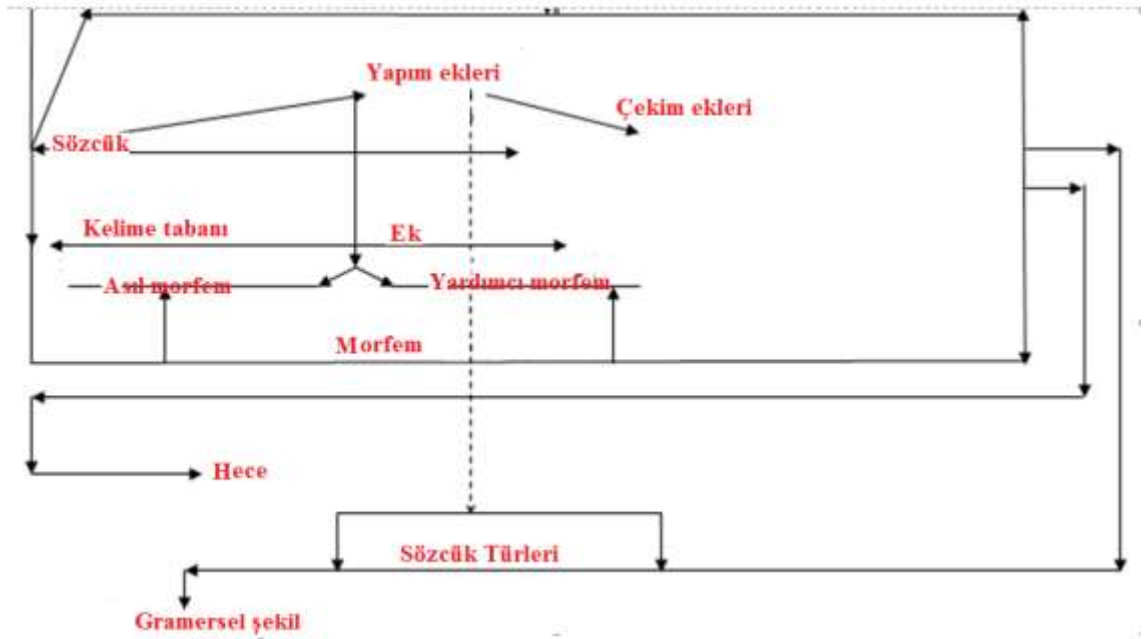
**Sözcüksel** : күз+sU ISIM (göz)+IC\_SAHİPLİK\_O\_(s)U  
**Yüzeysel** : күз00e күзе (gözü)

#### Başkurtçanın Şekilsel İşleyiş Yapısı

Kelimeler; kök ve eklerden oluşur. Kök ile ekin her biri, birer morfedir. Morfemler ise, asıl ve yardımcı morfemler olarak ikiye ayrılır. Asıl morfemler kelime tabanıdır, yardımcı morfemler ise yapım ve çekim ekleridir. Her morfedir bir **gramatikal şekli** oluşturur. Her gramatikal şeklin gramatikal anlamı vardır. Mesela **“dil”** asıl morfedir ve bir **gramatikal şekildir**, **gramatikal anlamı** da **isimdir**; Çekim eki, asıl morfemlere gelerek, sadece gramatikal şeklini değiştirirken, yapım ekleri ise gramatikal şeklin

gramatikal anlamlarına da etki ederler; Mesela, **dil** kelimesinin, **+de** hâl eki aldığındaki gramatikal anlamı yine **isimdir**, sadece **bulunma hâli ile çekimlenmiş isimdir**; -e yapım eki gelince ise, gramatikal anlamı **fiil** oluyor. Artık, dile- fiili olarak, kendisinden sonra fiil çekim ve fiil yapım ekleri alacaktır. Bildiri konusunun bir yönünü oluşturan, Başkurtçadaki bu işleyiş döngüsü, sonlu durum makinelerine uygun iki seviyeli kurallarla tanıtılmış, test edilmiştir.

### 5. RESİM. Morfemlerin sıralanışının şematik gösterimi.



5. Resim’de şemada gösterildiği gibi yapım ekleri ise asıl morfemlere gelerek, yeni bir gramatikal şekil oluşturur. Yeni gramatikal şekil demek yeni bir gramatikal anlamdır. Bunları, sonlu durum makinalarına tanımlamak için her bir gramatikal şekli, yani morfemi, tanımlayan birer “id” tanımlanacaktır; Örneğin: ISİM: isimlerin genel “id”sidir; SIFAT: sıfatların genel “id”sidir; FİİL: fiillerin genel “id”sidir vb.. Yardımcı morfemlere de fonksiyonlarına göre IC (isim çekim eklerin genel “id”si), IY (isimden isim yapan eklerin genel “id”si), FC (fiil çekim eklerinin genel “id”si) vb. “id”ler verilmesi gerekir (bk. 2. ve 3. Resim). Çekim ve yapım eklerinin ise, önce grulandırılmaları gerekir; HAL (ad durum eklerinin grup id’si), İYELİK (sahiplik eklerinin grup id’si), İY\_ISİM (isimden isim yapım eklerinin grup id’si), FC\_ZAMAN (bildirme kiplerinin grup id’si), ZAMIR\_IYELİK (iyelik menşeli şahıs eklerinin grup id’si) vb. olarak Grup id’lerinin de kodlanması gerekir. Yani her gramatikal şeklin

gramatikal anlamları için birer “id” oluşturacak ve bu gramatikal şekillerin listelendiği .xml ve .txt dosyalarında bu “id”ler kullanılacaktır. (bk. 2. ve 3 Resim).

Mesela;

*байзыкынан* (zengininkinden) kelimesi, her birisi birer gramatikal anlama sahip üç gramatikal şekilden -yani üç morfemden- oluşmaktadır. Gramatikal anlamı, sırasıyla genel aitlik eki ve çıkma hali ile çekimlenmiş isimdir. Yani, *-зыкы* (-nın-ki) genel aitlik ekli *бай* “zengin” kelimesinin ayırılma hali ile çekimlenmiş halidir. Sonlu durum makinesinde bu çekimin gerçekleşmesi aşağıdaki gibi test edilmiştir ve kusursuz sonuç elde edilmiştir:

**Sözcüksel** : бай ISIM(zengin)+IC\_AITLIK\_NUQU  
+IC\_HAL\_AYRILMA\_DAn  
**Yüzeysel** : бай0зыкы0нан байзыкынан (zengininkinden)

İsimden fiil yarım eki almış тел “dil” isminin çoğul birinci şahıs görülen uzak geçmiş zaman çekiminin sonlu durum makinesinde çekimlenmesi de test edilmiş olup aşağıdaki gibi kusursuz şekilde gerçekleşmiştir.

**Sözcüksel** : тел ISIM (dil)+IY\_FIIL\_A+  
FC\_ZAMAN\_GORULEN\_UZAK\_GECMIS\_GAn  
+SAHIS\_ZAMIR\_BIZ\_bUz]  
**Yüzeysel** : тел0ә0гән0без теләгәнбез (dilemişiz)

## Sonuç

Bu çalışmada Oflazer’in uyguladığı iki seviyeli morfoloji kuralları çerçevesinde Başkurtça’nın morfonolojik yapısı incelenmiştir. Çalışmanın neticesinde Başkurtçanın morfonolojik yapısı, sonlu durum makinaları (FSM) için iki seviyeli kurallarla tanımlanması mümkün olduğu ispat edilmiştir. Başkurtçanın morfonolojik yapısından hareketle, sonlu durum makinaları için uygun imla kuralları oluşturulmuş olup, bunlar hazır çatı üzerinde denenmiş ve başarılı olduğu kanıtlanmıştır. Bu projenin nihayi hedefi Başkurtça ve Türkçe başta olmak üzere, diğer Türk lehçeleri arasında karşılıklı elektronik çeviri gerçekleştirmektir. Dolayısıyla Başkurtça’nın iki seviyeli morfonolojik kuralları oluşturulurken, Türkçe ve diğer Türk lehçeleri de göz önünde bulundurulmuştur. Dilin iki seviyeli morfonolojik kurallarının çıkartılmasında, elektronik çeviri programlarına yapay zeka kazandırılabilceği de ön görülmüştür. Çünkü sonlu durum

makinelerinde, Trk lehelerinin iki seviyeli morfonolojisinin tanımlanmıř olmasıyla kaynak ve hedef dillerdeki kelimeler, sadece szlkteki karřılıklarıyla eřleřtirilmeyecektir; aldıđı eklerin eřleřtirilmesine, hatta fonksiyonlarının bile eřleřtirilmesine imkan dođacaktır.

alıřmalarımı sonlu durum makinasında (NUVE) test edebilmem iin benimle bu yazılıma paylařan sayın Harun Reřit Zafer'e, bu programın online versiyonunu yapan deđerli Ahmet Eralp Kk'e teřekkr ederim.

## Kaynaka

- AYDIN, O. (2017). Bađımsızlıklarının 25.Yılında Trksoy'un Trk Topuluklarına Yapmıř Olduđu Katkıları. B *Uluslararası Sosyal Bilimler Kongresi: Trkistan Forumu III "Kreselleřme Srecinde Trk Dnyasının Geleceđi"* (str. 955-967).
- GKGZ, E. v. (2011). Two-Level Qazan Tatar Morphology. B *1st International Conference on Foreign Language Teaching and Applied Linguistics May 5-7 2011 Sarajevo* (str. 428-431). Saraevo.
- GRMEZ, Z. v. (2016). Kyrgyz Orthography and Morphotactics with Implementation in NUVE. *INTERNATIONAL CONFERENCE ON ENGINEERING AND NATURAL SCIENCE. SARAJEVO.*
- GRMEZ, Z. v. (b.d.). An Overview of Two Level Finite State Kyrgyz Morphology. *The 2. International Symposium on Computing in Science & Engineering (ISCSE)*.
- OFLAZER, K. (1994). Two-level Description of Turkish Morphology. *Literary and Linguistic Computing Volume9, Is-sue2*, 137-148.
- ZTOPRAK, S. v. (2015). An Implementation of Tatar Orthography Using The Nve Framework. B *3rd International Conference on Computer Processing In Turkic Languages (TURKLANG 2015)* (str. 127-137). Kazan: Akademi of Sciences of the Republic Tatarstan Press.

ZAFER, H. R. (2011). 1st International Conference on Foreign Language Teaching and Applied Linguistics May 5-7 2011 Sarajevo. *Two-Level Description of Kazakh Morphology*, (стр. 506-564). Sarajevo.

ЮЛДАШЕВ, А. (1997.). Башкирский язык. В *Языки Мира: Тюркские Языки* (стр. 206-224.). Бишкек.