

İki Düzeyli Özbek Morfolojisi

Ahmet Eralp KÜÇÜK^{1*}, Atakan KURT²

¹ Jeoloji Mühendisliği Bölümü, Mühendislik Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye

² Jeoloji Mühendisliği Bölümü, Mühendislik Fakültesi, Pamukkale Üniversitesi, Denizli, Türkiye

*¹ kucukea@gmail.com, ² atakan.kurt@istanbul.edu.tr

(Geliş/Received: 15/12/2021;

Kabul/Accepted: 10/02/2022)

Öz: Biçimbilimsel (morfolojik) belirsizlik giderme, sözdizimsel ayrıştırma ve makine çeviri sistemlerinin temelini oluşturan morfolojik analiz, doğal dil işlemedeki ana konulardan birisidir. İki düzeyli morfoloji, biçimbilimsel olarak karmaşık dillerin analizi ve üretimi için ilk genel model olarak bilinir. Bu makalede, Özbekçe'nin iki düzeyli morfolojik analizi, Nuve biçimbilimsel çözümleyici ile ayrıntılı olarak sunulmaktadır. Nuve, biçimbilimsel analiz için kullanılan açık kaynak kodlu bir kütüphanedir. Nuve Studio, www.nuvestudio.com internet adresinden erişilebilmekte olup, bu çalışma için geliştirilmiştir. Uygulama aşamasında, 30.000 kelime, kök kelime sözlüğü olarak kullanılmıştır. Çağdaş Özbekçe (yeni latin alfabesiyle) 26 iki düzeyli yazım (ortografik) kuralı ile tanımlanmıştır. Özbekçe kelime yapılarının biçim dizgeleri (morfolojik) fiil ve isim paradigmaları için sonlu durum makineleri ile gösterilmiştir. Fonolojik ve morfolojik özel durumlar, kısıtlamalar ve istisnalar örneklerle verilmiştir. Bu çalışma, sözcüksel, morfolojik ve sözdizimsel açıdan birçok ortak yönü bulunan Özbekçe ile diğer Türk dilleri arasında, Özbekçe dil işleme ve morfoloji tabanlı makine çevirisi için öncü bir çalışma olacaktır.

Anahtar kelimeler: Özbek Morfolojisi, Doğal Dil İşleme, Hesaplamalı Dilbilim, Morfolojik Analiz, Ortografik Kurallar.

Two Level Uzbek Morphology

Abstract: Morphological analysing is one of the main issues in natural language processing and which is base for morphological disambiguation, syntactic parsing and machine translation systems. Two level morphology is known as the first general model for the analysis and generation of morphologically complex languages. This paper gives a detailed two level morphological analysis of Uzbek via Nuve Studio morphological analyser. Nuve is an open source library for morphological analysis. Nuve Studio is developed for this study, which is available on www.nuvestudio.com. In the implementation, 30.000 words are used as root word lexicon. The orthographic rules of contemporary Uzbek (in new Latin alphabet) are encoded using 26 two level rules. Morphotactics of Uzbek word structures have been encoded as finite state machines for verbal and nominal paradigm. Phonological and morphological special cases, limitations and exceptions are given with examples. This study will be pioneering study for Uzbek language processing and morphology based machine translation between Uzbek and other Turkic languages which have a lot in common from lexical, morphological and syntactic aspects.

Key words: Uzbek Morphology, Natural Language Processing, Computational Linguistics, Morphological Analysis, Orthographic Rules.

1. Giriş

Özbekçe, diğer Türk dilleri gibi sondan eklemeli ve biçimbilimsel (morfolojik) açıdan zengin bir dildir. Dilin alfabesi tarih boyunca değişime uğramıştır. 1930'lu yıllara kadar Arap alfabesi, 1930 ve 1940 yılları arasında Latin alfabesi kullanılmıştır. 1940'lardan sonra Kiril alfabesi kullanılmaya başlanmıştır. Çağdaş Özbek dilinde Latin alfabesi resmi alfabesi olarak kabul edilmesine rağmen Kiril alfabesi halen yer yer kullanılmaktadır. Bu çalışmada Özbekçenin morfolojik tanımı için Latin alfabesi (1993[1]'den sonra kabul edilen alfabesi) kullanılmıştır.

İki düzeyli morfoloji bir dilin morfolojisini tanımlamak ve analiz etmek için kullanılmaktadır. Bu makalede, Özbek morfolojisini tanımlamak için Koskenniemi'nin iki düzeyli morfoloji modeli [2] kullanılmıştır. Bu model bilgisayarlı dilbiliminde bir dilin morfolojisini tanımlamak için kullanılan ilk genel modeldir. Bu modele göre, lexical (yapısal) ve surface (yüzeysel) formlar (biçimler) bir kelimeyi temsil etmek için kullanılır. Bir kelimenin yüzeysel formu, doğal bir dildeki normal kullanımındaki gösterim biçimi ile aynıdır. Sözcüksel form ise kelimenin eklerini (morfeplerini) gösterir. Ortografik (yazım-ımla) kuralları ise bu iki form arasındaki dönüşümü tanımlayan kurallardır. Biçim dizgesi (Morfolojik) kuralları ise kelimedeki bulunan eklerin sıralamasını gösterir ve sonlu durum otomatu (finite state automata (FSA)) olarak bu modelde kodlanır.

Bir dili analiz etmek için, o dilin gramer kurallarına, sözlüğe (lexicon) ve morfolojik çözümleyiciye ihtiyaç duyulmaktadır. Bu çalışmada, Özbek morfolojisini analiz etmek için açık kaynak kodlu morfolojik çözümleyici

* Sorumlu yazar: kucukea@gmail.com. Yazarların ORCID Numaraları: ¹ 0000-0001-8614-1282, ² 0000-0002-9549-8475

ve makine çevirisi kütüphanesi olan Nuve kullanılmıştır. Nuve Studio (www.nuvestudio.com) bu kapsamdaki çalışmalar için geliştirilmiş ve uygulamalarda kullanılmıştır. Nuve dilden bağımsız bir kütüphanedir. Bir dilin Nuvede tanımlanması için suffix (ek) ve root (kök) sözlüklerine, morphotactics (morfotaktik) ve orthographic (ortografik) kurallar dosyalarına ihtiyaç vardır. Özbek morfolojisinin Nuve uygulamasında tanımlanması için 30.000 kök (root) kelimededen oluşan ve 370 ekten (suffix) oluşan kökler ve ekler sözlükleri kullanılmıştır. Çağdaş Özbekçe (kabul edilen yeni Latin alfabelerinden oluşan) için 26 tane iki düzeyli ortografi kuralı tanımlanmıştır. Diğer Türk dilleri gibi Özbekçe’de genellikle eklerin sıralanmasında bir düzene sahiptir. Örneğin “kitoblarimdagi” (kitaplarımdaki) kelimesindeki ekler aşağıdaki gibi gösterilebilir.

kitob + lar + im + da + gi

Örnekte yer alan ‘+’ işareti eklerin sınırlarını belirlemektedir. Bazen istisnai durumlardan dolayı biçimbilimsel analiz zor olabilir. Bu istisnai durumlar sonraki bölümlerde açıklanmıştır.

Bu çalışmanın amacı iki düzeyli Özbek morfolojisini tanımlamaktır. Çalışmamızın uygulamaları Nuve Studio kullanılarak yapılmıştır. Bu makalede bir sonraki bölümde benzer çalışmalar, sonraki bölümde iki düzeyli Özbek morfolojisi detaylı verilmiştir. Sonrasında, Özbek isim (nominal) ve fiil (verbal) biçim dizgesi tanımlamalarına değinilmiştir. Uygulama detayları, istisnai durumlar ve örnekler ile çalışma detaylandırılarak sonuç ve gelecek çalışmalar bölümleri ile makale sonuçlandırılmıştır.

2. Benzer Çalışmalar

Koskenniemi’nin iki düzeyli morfoloji modeli [2,3] birçok farklı sondan eklemeli dilin morfolojisinin tanımlanmasında kullanılmıştır. Birçok çalışmada farklı uygulama araçları kullanılarak bu model uygulanmıştır. Bask ve Svahili dilleri karmaşık morfolojiye sahip sondan eklemeli dillerdir. Bu iki dilin morfoloji çalışmalarında iki düzeyli morfoloji modeli kullanılmaktadır [4,5]. Svahili dilinin morfolojik analizi için SWATWOL uygulama aracı geliştirilmiştir. Japonca ve Korece de sondan eklemeli ve karmaşık morfolojiye sahip dillerdir. Japonca ve Korece’nin morfolojik analizi de iki düzeyli morfoloji modeli kullanılarak yapılmıştır [6,7]. Bu dillerin morfolojik analizinin uygulamasında sırasıyla, Koskenniemi’nin KIMMO ve bu uygulamanın yeni versiyonu PC-KIMMO morfolojik çözümleyici-üretici programları kullanılmıştır.

Türk dilleri de sondan eklemeli ve karmaşık morfolojiye sahip dillerdir. İki düzeyli morfoloji modeli Türk dillerinin biçimbilimsel analizinde de kullanılmıştır. Türkçe morfolojisi için bu alanda yapılan öncü çalışmalardan birinde 22 iki düzeyli yazım kuralları tanımlanarak PC-KIMMO programı ile uygulama gerçekleştirilmiştir [8]. Türk dillerinden Kazakça ve Kırgızca için gerçekleştirilen çalışmalarda iki düzeyli morfolojileri tanımlanmış ve uygulamaları sırasıyla Dilmaç [9] ve Nuve [10] programları ile gerçekleştirilmiştir.

Türk dillerinden biri olan Özbekçe de diğer Türk dilleri gibi sondan eklemeli ve karmaşık morfolojiye sahip bir dildir. Çağdaş Özbek dilinin morfolojisi, morfolojik analizi ve çözümlemesi üzerine sınırlı sayıda çalışma yapılmıştır. Sınırlı sayıda olan bu çalışmalardan bazıları Özbekçe için doğal dil işleme konusunda önemli bir yere sahiptir. Bir dil için biçimbilimsel analizin yapılabilmesi için o dilin yazım kurallarına ihtiyaç vardır. Çağdaş Özbekçe’nin yazım kuralları 24 Ağustos 1995 tarihinde Özbekistan Bakanlar Kurulu’nun 339 sayılı kararı ile yayımlanmıştır [11]. Bu günümüz Özbekçesi için son derece önemli bir çalışmadır. Modern Özbekçe için tanımlanan bu kurallar dil çalışmalarında temel referans kabul edilmiştir. Özbek morfolojisi, biçimdizgesi ve morfonemik kurallarla sunulmuştur. Tanımlanan Özbek morfolojisinin çözümlenmesi ve analizi için prolog tabanlı UZMORPP sistemi kullanılarak uygulaması yapılmıştır [12]. Özbekçe kelimelerin biçimbilimsel analizi FST metot ve yaklaşımıyla, kök kelime tanımları ve kelime sıralaması sağlanmıştır [13]. Özbekçe’nin iki düzeyli biçimbilimsel analizi FST teknoloji tabanlı biçimbilimsel çözümleyici ile uygulanmıştır [14].

3. İki Düzeyli Özbek Morfolojisi

İki düzeyli morfoloji, bir dildeki kelimelerin morfolojik tanımlanması için kullanılan dilden bağımsız bir yaklaşım sunan, genel bir modeldir[2]. Bu modele göre bir dilin analizi yapılırken kurallar ve sözlükler bir çözümleyici aracılığıyla birleştirilir ve dilin morfolojisi bu yaklaşımla modellenir ve analiz edilir. İki düzeyli morfoloji yaklaşımında ortografik kurallar iki düzeyli kurallar olarak tanımlanır. Yapım ve çekim eklerinin sıralamasını gösteren morfotaktik ise sonlu durum makineleri (Finite State Machine – FSM) kullanılarak tanımlanır.

İki düzeyli morfolojide kelimelerin sözcüksel (lexical) ve yüzeysel(surface) biçimde iki gösterimi bulunmaktadır. Yapısal form kelimenin biçimbilimsel yapısını ve kelimeye eklenen çekim ve yapım eklerinin

özelliklerini ve sırasını gösterir. Yüzeysel form da kelimenin çekim ve yapım eklerini aldıktan sonra ortaya çıkan yazım biçimini gösterir. İki düzeyli yazım kuralları (ortografik kurallar) bu iki form arasındaki dönüşümü gösterir [3,8]. İki düzeyli yazım kuralları aşağıda tanımlanmıştır.

Özbekçede 1993 yılından [1] günümüze Latin alfabesi kullanılmaktadır. Özbek alfabesi 29 harf ve apostrof (‘) karakteri [15] ile 30 karakterden oluşmaktadır. Özbek alfabesi 6 ünlü(a, e, i, u, o, o’) ve 23 ünsüz harften (b, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, z, y, x, g’, sh, ch, ng) oluşmaktadır[15,16]. Özbekçe ünlü harfleri Tablo 1’de ve Özbekçe ünsüz harfleri Tablo 2’de fonetik özellikleriyle gösterilmiştir [17].

Tablo 1. Özbekçe sesli harfler.

Ünlüler	Yuvarlak			Düz		
	Açık	Yarı Kapalı	Kapalı	Açık	Yarı Kapalı	Kapalı
Arka Sıralı	o	o’	u			
Ön Sıralı				a	e	i

Tablo 2. Özbekçe sessiz harfler.

Süreksiz	Tonsuz	Dudaksıllar	Dişsel	Damaksıl	Art Damak	Küçük Dil	Gırtlaksal
		Tonlu	p	t		k	q
Yarı Patlayıcı	Tonsuz	b	d		g		
	Tonlu			ch			
Sızıcı	Tonsuz	f	s	sh	x		h
	Tonlu		z			g’	
Geniz		m	n		ng		
Akıcı		v	l,r	y			

Özbekçe, Türk dilleriyle ünlü-ünsüz düşmesi, ünlü-ünsüz türemesi ve ünlü-ünsüz değişimi gibi konularda benzerlik gösterir. Birçok konuda Türk dilleriyle benzerlik gösteren Özbekçe’de düzenli bir ünlü ve ünsüz uyumu bulunmamaktadır [18]. Bu makalede Özbekçe dil ve yazım kuralları büyük bir titizlikle incelenerek, iki düzeyli yazım (ortografik) kuralları Özbek morfolojisini tanımlamak için kullanılmıştır. Yazım kurallarının daha sade ve kolay tanımlanması için harf kümelerinden oluşan meta-harfler kullanılmıştır. Tanımlamalarda kullanılan meta-harfler aşağıdaki gibidir.

$C = \{b, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, z, y, x, g', sh, ch, ng\}$, $C_s = \{f, s, t, k, q, ch, sh, l, n, s, h, y\}$, $V = \{a, e, i, u, o, o'\}$, $V_{bo} = \{o, u\}$, $V_{fu} = \{a, e, i\}$, $V_{de} = \{o, i, u\}$, $D = \{d, t\}$, $Q = \{q, k\}$, $G = \{q, k, g, g'\}$, $I = \{u, i\}$, $Y = \{a, y\}$, $E = \{a, i\}$.

İki düzeyli kuralları tanımlarken kullanılan bazı semboller ve anlamları şunlardır; * önceki harfin sıfır ya da daha fazla tekrarlanması, ? önceki harfin sıfır ya da bir kez tekrarlandığını, 0 yüzeyde görünmediği, @ kelime sonunu, \$ sonraki harfin yüzeyde düştüğünü, + ek sınırlarını gösterirken, [a | b | c] ya da [abc] ise a veya b veya c anlamına gelmektedir.

Bu bölümde iki düzeyli kurallar örnekler ile açıklanmıştır. Her kuraldan sonra verilen örneklerin ilk satırında kelimenin sözcüksel (lexical) formu, ikinci satırda kelimenin yüzeysel (surface) formu gösterilmiştir. Uygulama aşaması 26 iki düzeyli yazım kuralı ile tamamlanmış olup, sadece en çok kullanılan 12 iki düzeyli yazım kuralı bu makaledeki sayfa kısıtlaması nedeniyle sunulabilmektedir.

1. Bu kurala göre gövdeye eklenen ektaki ünlü, gövdenin son ünlüsüne göre değişim göstermektedir. Buna göre, gövde son ünlüsü e veya i ise ve $A = \{a, o\}$ meta harfini içeren bir ek alırsa, A yüzeyde a olarak gösterilir. Gövde son ünlüsü o veya u ise A yüzeyde o olarak gösterilir. Bu kural için bazı istisnalar bulunmaktadır. $bu+(n)daqa \rightarrow bundaqa$ ve $shu+(n)daqa \rightarrow shundaqa$ örneklerinde bu istisnalar gösterilmektedir.

A:a => V:[ei]C*+C*_

A:o => V:[ou]C*+C*_

ovun+chAQ	V(teselli ol) + JVD_CHAQ	ovunchoq	avuntu
kekir+dAQ	V(geğir) + NVD_DAQ	kekirdak	gırtlak

2. Son ünlüsü o, i veya u olan bir gövdeye E={a, i} meta harfini içeren bir ek gelirse, E yüzeyde i olarak gösterilir. Son ünlüsü o' olan bir gövde, E meta harfini içeren ek alırsa E yüzeyde a olarak gösterilir.

E:i => V:[oiu]C*+C*_

E:a => V:o'C*+C*_

uch+Er	V(uç) + CAUSR_ER	uchir	birisini / bir şeyi uçurmak
o't+Ger	V(geç) + CAUSR_ER	o'tkar	bir şeyi / birisini geçirmek

3. Son ünlüsü u veya e olan bir gövdeye I={u, i} meta harfini içeren bir ek gelirse, I yüzeyde u olarak gösterilir. Son ünlüsü i olan bir gövde, I meta harfini içeren ek alırsa I yüzeyde i olarak gösterilir.

I:u => V:[eu]C*+C*_

I:i => V:iC*+C*_

kekir+IQ	V(geğir) + NVD_IQ	kekirik	geğirme
uz+IQ	V(koparmak) + NVD_IQ	uzuq	kopuk
yet+IQ	V(yet) + JVD_IQ	yetuk	yetişkin

4. C_s sert sessizlerinden biri ile biten gövdeye D={d, t} ile başlayan bir ek geldiğinde D meta harfi yüzeyde t sessizi ile gösterilir. C_s sert sessizlerinden biri ile bitmeyen gövdelere eklenen ekin başındaki D meta harfi d sessizi ile gösterilir.

D:t => Cs+ _

D:d => C-Cs+ _

ayt+Dir	V(söyle) + CAUSDIR_DIR	ayttir	birine söylet
qil+Dir	V(yap) + CAUSDIR_DIR	qildir	birinin yapmasını sağlatmak

5. Son ünlüsü o veya u olan gövdelere Q={q, k} meta harfini içeren bir ek gelirse, Q meta harfi yüzeyde q olarak gösterilir. Gövdenin son ünlüsü a, e veya i harflerinden biri ise ve Q meta harfini içeren bir ek alırsa Q meta harfi yüzeyde k olarak gösterilir. Bu kuralın istisnaları bulunmaktadır. Örneğin chush kelimesi Q_ir ekini aldığı anda yüzeyde chushk_ir olarak görünür. Bu örnekte Q meta harfi q olması gerekirken yüzeyde k olarak görülmektedir.

Q:q => V:[ou]C*+C?_C*

Q:k => V:[aei]C*+C?_C*

o'kir+IQ	V(bağır) + NVD_IQ	o'kirik	ağlama
solin+chAQ	V(kon) + JVD_CHAQ	solinchak	kulak memesi

6. k veya g ünsüzleri ile biten gövdelere G={q, k, g, g'} meta harfiyle başlayan bir ek gelirse, G meta harfi yüzeyde k olarak gösterilir. Gövdenin son ünsüzü g' veya C_s sert sessizlerinden biri ise, ekin başında bulunan G meta harfi yüzeyde q olarak gösterilir. e, k veya g harflerinden birini barındıran gövdeye G meta harfiyle başlayan bir ek gelirse G meta harfi yüzeyde g olarak gösterilir. o, q, x veya g' harflerinden birini barındıran gövdeye G meta harfiyle başlayan bir ek gelirse G meta harfi yüzeyde g' olarak gösterilir.

G:q => @[g'Cs]+ _

G:k => @[kg]+ _

G:g => C*V*[ekg]V*C*+ _

G:g' => C*V*[oqxg']V*C*+ _

turt+Gi	V(dürt) + NVD_GI	turtki	dürtü
yul+Gich	V(yol) + NVD_GHCH	yulg'ich	hırsız

7. Ünlü ile biten gövdelere ünlü ile başlayan bir ek geldiğinde, ekin başındaki ünlü yüzeyde düşer.
V:0 => V+_

bola+Am	N(çocuk) + NNI_POSS1S_AM	bolam	çocuğum
---------	--------------------------	-------	---------

8. Sessiz ile biten bir gövdeye n sessizi ile başlayan bir ek geldiğinde, ekin başındaki n ünsüzü yüzeyde düşer.
n:0 <=> C+_

men+ning	Pron(ben) + NNI_GEN_NING	mening	benim
----------	--------------------------	--------	-------

9. Bazı kelimelerde, gövdenin son ünlüsü, ünlüyle başlayan bir ek aldığında yüzeyde düşer. Bazı durumlarda, ünsüzle başlayan bir ek alan gövdenin ünsüz harfi düşer.
V:0 => _C*+V
C:0 => _+C*

burun+Am	N(burun) + NNI_POSS1S_AM	burnim	burnum
suv+GAR	N(su) + VND_GAR	sug'or	sulamak

10. Gövdenin son ünsüzü g veya g' harflerinden biriye ve gövde G meta harfiyle başlayan bir ek alırsa gövdenin sonundaki g ünsüzü k ünsüzüne ve g' ünsüzü q ünsüzüne dönüşür. Bazı Rusça kökenli kelimeler örneğin biolog(ek aldıktan sonra yüzeyde biologga olarak yazılır) bu kurala uymaz.
g:k => _+G
g':q=> _+G

teg+Gani	V(değ) + AVD_GANI	tekkani	değmek için
bog'+Ga	N(bahçe) + NNI_DAT_GA	yulg'ich	bahçeye

11. Özbekçede iki ünlü yan yana gelmez. Bazı durumlarda, ünlü ile biten gövdelere ünlü ile başlayan bir ek geldiğinde y ünsüzü iki ünlü arasına eklenir.

parvo+Imiz	N(ilgi) + NNI_POSS1S_IM	parvoyimiz	ilgimiz
------------	-------------------------	------------	---------

12. Ünsüzle biten gövdelere ünsüzle başlayan bir ek geldiğinde ekin başındaki ünsüz yüzeyde düşer.
C:0 => C+_

bosh+si	N(baş) + NNI_POSS3S_SI	boshi	onun başı
---------	------------------------	-------	-----------

4. Uygulama

Bu çalışmada Özbek morfolojisinin iki düzeyli tanımı detayları ile sunulmuştur. Özbek morfolojisinin iki düzeyli tanımının uygulanması Nuve Studio kullanılarak yapılmıştır. Nuve Studio[19,20], Nuve'nin yeni bir uygulama arayüzü olarak geliştirildi. Nuve[21] açık kaynak kodlu, biçimbilimsel çözümleme ve makine çevirisi için kullanılan doğal dil işleme kütüphanesidir.

Nuve Studio'da bir dilin morfolojik olarak tanımlayabilmek için kökler ve ekler sözlüğü ile yazım ve biçim dizgesi (morfolojik) kuralları tanımlanmalıdır. Özbekçe için hazırlanan dosyalar http://www.nuvestudio.com/downloads/uz_UZ/uz_UZ.zip linkinden indirilebilir.

5. Özel Durumlar ve İstisnalar

Bu bölümde, uygulama esnasında ortaya çıkan, Özbek morfolojisinin bazı özel durumları ve istisnaları örneklerle sunulmuştur. Özel durumlar ve istisnalar için özel kurallar tanımlanmıştır. Özel durumların ve istisnaların örnekleri kelimenin biçim bilimsel analizi ve ek yapısını kapsamaktadır. Özel durumlar, kelimenin cümle içerisindeki kullanımını ve özel durumu göstermek için cümleler ile sunulmuştur. Kelimeler, beklenen ve istisnai durumun gerçekleştiği ya da gerçekleşmediği kelimenin doğru gösterimi ile örneklendirilmiştir. Özel durumlar ve istisnalar örneklerle aşağıda detaylandırılmıştır.

1. Bazı sayılara +dan veya +lab ekleri eklenince gövde zarfa dönüşür. Bu sayılara eklenen aynı ek bazı durumlarda gövdenin zarfa dönüşmeyerek sayı olarak kalmasına neden olur. +dan çekim eki istisnai durumlarda yapım eki gibi davranır. Aşağıdaki örneklerin ilkinde +dan ekini alan sayı gövdesi zarfa dönüşürken, ikinci örnekte sayı olarak kalmıştır.

Biçimbirimler(Morphemes)	Yapısal(Lexical)	Kelime	Cümle
bir+dan	SAYI(bir)+ÇIKMA(dan) (zarf – aniden)	birdan	Birdan kelip goldi. (Birden gelip kaldı.)
bir+dan	SAYI(bir)+ÇIKMA(dan) (sayı – bir)	birdan	Birdan birni kamaytsa nol goladi. (Birden bir çıkarsa sıfır kalır.)

2. İşaret zamirlerine eklenen bazı hal ekleri(yönelme, bulunma ve çıkma) ve bazı çekim eklerinin arasına n sessizi ve ni harfleri türer. Üçüncü çoğul şahıs ekleri işaret zamirlerine eklenmesi durumunda istisnai olarak ni türemesi gözlenmez.

Biçimbirimler(Morphemes)	Yapısal(Lexical)	Beklenen	Doğru
shu+da	ZAMİR(shu)+BULUNMA (da)	shuda	shunda(şunda)
bu+ga	ZAMİR(bu)+YÖNELME (ga)	buga	bunga(buna)
shu+si	ZAMİR(shu)+3.TEKİL ŞAHIS İYELİK EKİ(si)	shusi	shunisi(
shu+lar+i	ZAMİR(shu)+ÇOĞUL(lar)+ 3.ÇOĞUL ŞAHIS İYELİK EKİ(i)	shunilari	shulari(şunları)

3. Özbekçe’de geçmiş zaman eki sadece +di ekidir. Geçmiş zaman eki ünsüz uyumundan dolayı herhangi bir ses değişimine uğramaz.

Biçimbirimler(Morphemes)	Yapısal(Lexical)	Beklenen	Doğru
ket+di	FİİL(ket)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS	ketti	ketdi(gitti)
top+di	FİİL(top)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS	topdi	topdi(bulundu)

4. Özbekçede bazı kelimelerin yazılışı aynı fakat anlamları farklıdır. Farklı eklerin bu gövdelere eklenmesi sonrası yazılışları aynı olsa bile farklı özel durumlara ve anlamlara sahip olabilir. Ayrıca, aynı ekler eklense bile bu gövdelerin anlamları ve morfolojik yapısı farklı olabilir. Bu kelimelerin hangisinin ne anlama geldiği ancak cümle veya konuşma dilinde anlaşılabilir olup, morfolojik belirsizliğe yol açmaktadır. Bu kelimelerin cümle içinde doğru morfolojik çözülmesi için morfolojik belirsizlik giderme yöntemlerinin kullanılması gerekmektedir. Uygulama esnasında farklı kurallar tanımlanarak, bu kelimeler için özel durumlar oluşturulmuştur.

Biçimbirimler(Morphemes)	Yapısal(Lexical)	Sonuç
bo'l+di	FİİL(bo'l)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS	bo'ldi(böldü)
bo'l+di	FİİL(bo'l)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS	bo'ldi(oldu)
bek+mi	İSİM(bek)+3.TEKİL ŞAHIS+SORU(mi)	bekmi?(Bey mi?)
bek+mi	SIFAT(bek)+3.TEKİL ŞAHIS+SORU(mi)	bekmi?(Kapalı mı?)
qot+ma	FİİL(qot)+F.İ.YAPIM EKİ(ma)	qotma(sıska)
qot+ma	FİİL(qot)+OLUMSUZLUK(ma)	qotma(sert olma)
qot+ma	FİİL(qot)+F.F.YAPIM EKİ(ma)	qotma(setleşme)
qotma	İSİM(qotma)	qotma(sıska)

5. Bir gövdenin son harfi g, g', k ve q sessizlerinden biriye ve +Gan geçmiş zaman veya +Ga hal eklerinin biri gövdeye eklenirse, ekin ilk sessizi ve gövdenin son sessizi k veya q sessizine dönüşür. Gövdenin son harfleri g, g', k ve q sessizlerinden biri değilse bu durum gerçekleşmez. Özbekçe'de normalde, eğer bir gövde g, g', k ve q sessizlerinden biriyle biter ve G sessizi ile başlayan bir ek alırsa, sadece G sessizi k ve q sessizine dönüşür.

Biçimbirimler(Morphemes)	Yapısal(Lexical)	Beklenen	Doğru
teg+gan	FİİL(teg)+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	teggan	tekkan(değmiş)
bo'g'+gan	FİİL(bo'g')+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	bo'g'gan	boqqan(besledi)
tik+gan	FİİL(tik)+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	tikgan	tikkan(dikilmiş)
uq+gan	FİİL(uq)+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	uqgan	uqqan(uyanmış)
ket+gan	FİİL(ket)+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS	ketgan	ketgan(gitmiş)
bosh+ga	İSİM(bosh)+YÖNELME(ga)	boshga	boshga(başka)

6. Bir gövdenin son harfi y, r, sh ve n sessizlerinden biriye ve +Dir ettirgen fiil eki eklenirse, ekin ilk sessizi t(tir) olarak görünür. Eğer gövde herhangi bir başka sessizle biterse, bu durum gerçekleşmez. Gövdenin son harfinin l sessizi olması durumunda, +Dir eki tir olarak görünür. Sadece bir istisna kelimedede bu durum görünmez. Kel(gel) kelimesi ve istisnai durumu aşağıda sunulmuştur.

Biçimbirimler(Morphemes)	Yapısal(Lexical)	Beklenen	Doğru
aylan+Dir	FİİL(aylan)+ETTİRGEN(dir)	aylandir	aylantir(dönüştür)
chiqar+Dir	FİİL(chiqar)+ETTİRGEN(dir)	chiqardir	chiqartir(çıkarttır)
ko'pey+Dir	FİİL(ko'pey)+ETTİRGEN(dir)	ko'peydir	ko'peytir(çoktur)
kel+Dir	FİİL(kel)+ETTİRGEN(dir)	keldir	keltir(getir)
ol+Dir	FİİL(ol)+ETTİRGEN(dir)	oldir	oldir(öldür)

7. Bu istisna, +chAQ son ekiyle ilgilidir. Duruma göre ek +choq veya +chak olarak gözlemlenir. Bir kelimedede x,q ve g' sessizleri olduğunda, kelimedeki arka sıralı ünlüler kesinlikle arka sıralı ünlüler

gibi davranır. Diğer durumlarda, arka sıralı ünlüler kelimedeki ön sıralı ünlüler olarak sunulur. Bu durumlar aşağıda örneklendirilmiştir.

Biçimbirimler(Morphemes)	Yapısal(Lexical)	Sonuç
qo'zic+chAQ	İSİM(qo'zic)+İ.İ.YAPIM EKİ(chAQ)	qo'zichoq (kuzucuk)
o'rgim+chAQ	İSİM(o'rgim)+İ.İ.YAPIM EKİ(chAQ)	o'rgimchak(örümcek)

6. Örnekler

Bu bölümde, Nüve Studio uygulama ortamında test edilen örnekler ve sonuçları aşağıda sunulmuştur.

- Olmangiz

olma+ngiz	[İSİM(olma)+2.ÇOĞUL ŞAHIS İYELİK(ngiz)] (elmalarımız)
ol+ma+ngiz	[FİİL(ol)+F.İ.YAPIM EKİ(ma)+ 2.ÇOĞUL ŞAHIS İYELİK(ngiz)] (bir şey alma)
ol+ma+ngiz	[FİİL(ol)+OLUMSUZLUK(ma)+2.ÇOĞUL ŞAHIS EMİR(ngiz)] (alma)

- Bo'ldimi

bo'l+di+mi	[FİİL(bo'l)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS+SORU(mi)](Böldü mü?)
bo'l+di+mi	[FİİL(bo'l)+GEÇMİŞ ZAMAN(di)+3.TEKİL ŞAHIS+SORU(mi)](Böldü mü?)

- Bekmi

bek+mi	[İSİM(bek)+3.TEKİL ŞAHIS+SORU(mi)] (Bey mi?)
bek+mi	[SIFAT(bek)+3.TEKİL ŞAHIS+SORU(mi)] (Kapalı mı?)

- Sirdagi

sir+da+gi	[İSİM(sir)+BULUNMA(da)+İLGİ(gi)](sırdaki)
sir+da+gi	[İSİM(sir)+BULUNMA(da)+İLGİ(gi)](boyadaki)
sir+da+gi	[İSİM(sir)+BULUNMA(da)+İLGİ(gi)](peynirdeki)

- Keltirilgan

kel+tir+il+gan	[FİİL(kel)+ETTİRGEN(tir)+EDİLGEN(il)+GEÇMİŞ ZAMAN(gan)+3.TEKİL ŞAHIS] (Getirildi.)
kel+tir+il+gan	[FİİL(kel)+ETTİRGEN(tir)+EDİLGEN(il)+ SIFAT-FİİL(gan)] (getirilmiş)

- Qotma

qot+ma	[FİİL(qot)+F.İ.YAPIM EKİ(ma)](sıska)
qot+ma	[FİİL(qot)+OLUMSUZLUK(ma)](sert olma)
qot+ma	[FİİL(qot)+F.F.YAPIM EKİ(ma)](sertleşme)
qotma	[İSİM(qotma)](sıska)

7. Sonuç ve Gelecek Çalışmalar

Bu makalede, iki düzeyli Özbek morfolojisi detayları ile araştırılmıştır. İki düzeyli ortografi (yazım) kuralları tanımlanarak, Özbekçenin morfolojik çözümlemesi sağlanmıştır. Morfolojik çözümleme esnasında karşılaşılan özel durumlar ve istisnalar araştırılmış ve uygulama aşamasında tanımlanmıştır. Çalışmanın uygulaması bu çalışma için geliştirilen Nuve Studio uygulaması aracı ile yapılmıştır.

Özbekçenin biçimbilimsel çözümlemesinin uygulama aşamasında, 30.000 kök kelime ve 370 ek ile tanımlı kök ve ek sözlüğü kullanılmıştır. Uygulama için 26 iki düzeyli ortografi kuralı tanımlanarak bu kurallar doğrultusunda çözümlemeler yapılmıştır. İstisnai durumlar ayrıca işlenerek sözlüğe eklenmiş ve çözümlemeleri

sağlanmıştır. Bir kelimenin birden fazla biçimbilimsel çözümünün olması biçimbilimsel belirsizlik gidermenin konusu olarak bir sonraki aşamada belirsizlik giderme uygulamaları ile çözümlenecek olup, sözlükte tanımlanan kelimelerin istenilen çözümlerinin elde edildiği gözlemlenmiş olup, birden fazla çözümün birçok kelime için üretildiği ayrıca gözlemlenmiştir. Uygulama testlerinde tüm sözlük kullanılarak sağlanan testte 30.000 kelimenin 27.993 tanesi istenilen – doğru (true) ve 2.007 tanesi farklı – yanlış (false) çözümlenerek %93,31 oranında istenilen sonuca ulaşılmıştır. Farklı çözümler biçimbilimsel belirsizlik gidermenin konusu olup, biçimbilimsel belirsizlik giderme sonrası bu oranın artış göstereceği öngörülmektedir.

Sonuç olarak, Özbekçenin morfolojisinin karmaşık yapısına, özel ve düzensiz durumlar ile istisnalara rağmen uygulama başarılı olarak kabul edilebilir bir oranda istenilen sonuçları sağlamış olup, Özbekçenin biçimbilimsel çözümlemesi detayları ile anlatılmıştır. Özbekçenin biçimbilimsel belirsizlik giderme ve sözdizimsel çözümleme konuları gelecekte araştırılarak, Özbekçe ve Türkçe arasında makine çevirisi üzerinde çalışmalar yapılacaktır.

Kaynaklar

- [1] Üşenmez E. Modern Özbek Türkçesinin Oluşumu. *Turkish Studies (Elektronik)* 2017; 12(7): 463-476.
- [2] Koskenniemi K. Two-level model for morphological analysis. In: *IJCAL*.1983; 683-685.
- [3] Koskenniemi K. Two-level morphology: A General Computational Model for Word-Form Recognition and Production. University of Helsinki, Dept. of General Linguistics 1983; 11.
- [4] Agirre E, Alegria I, Arregi X, Artola X, de Ilaraza AD et al. XUXEN: A spelling checker/corrector for Basque based on Two-Level morphology. In: *Third Conference on Applied Natural Language Processing*. 1992; 119-125.
- [5] Hurskainen A. A Two-Level Computer Formalism for the Analysis of Bantu Morphology An Application to Swahili. In: *NJAS*. 1992; 1:87-119.
- [6] Alam YS. A Two-Level Morphological Analysis of Japanese. In: *Texas Linguistic Forum*. 1983; 22:229-252.
- [7] Kim DB, Lee SJ, Choi KS, Kim GC. A two-level morphological analysis of Korean. In: *COLING The 15th International Conference on Computational Linguistics*. 1994; 1:535-539.
- [8] Oflazer K. Two-level description of Turkish morphology. *Literary and linguistic computing* 1994; 9(2): 137-148.
- [9] Zafer HR, Tilki B, Kurt A, Kara M. Two-level description of Kazakh morphology. In: *Proceedings of the 1st International Conference on Foreign Language Teaching and Applied Linguistics (FLTAL)*. 2011; 560-564.
- [10] Yiner Z, Kurt A, Kulamshaev K, Zafer HR. Kyrgyz Orthography and Morphotactics with Implementation in NUVE. In: *International Conference on Engineering and Natural Science (ICENS)*, Sarajevo, Bosnia-Herzegovina. 2016; 1650-1658.
- [11] Principal Orthographic Rules For The Uzbek Language. The Uzbekistan Cabinet of Minister's Resolution No. 339., of August 24, 1995. Tashkent, Uzbekistan.
- [12] Matlatipov G, Vetulani Z. Representation of uzbek morphology in prolog. In: *Aspects of Natural Language Processing*. Springer, Berlin, Heidelberg, 2009; 83-110.
- [13] Atadjanov JA. Models of Morphological Analysis of Uzbek Words. *Кибернетика и программирование* 2016; 6:70-73.
- [14] Abdurakhmonova NZ. A Two-Level Morphological Analysis Of The Uzbek Corpus. *Иностранная филология. Социальная и национальная вариативность языка и литературы* 2019; 425-430.
- [15] Yusupova N. *Türkçe Özbekçe Sözlük*. Ankara, Türkiye: Türk Dil Kurumu, 2018.
- [16] Üşenmez E, Boltabayev S, Tuğlacı G. *Özbekçe-Türkçe Sözlük*. İstanbul, Türkiye: Türk Dünyası Vakfı, 2016.
- [17] Van Der Hulst H, Van De Weijer J. Topics in Turkish phonology. *Turkish Linguistics Today*. 1991; 11-59.
- [18] Üşenmez E. *Modern Özbek Türkçesi: (Giriş-Özbekistan-Dil Tarihi-Metinler-Gramer-Sözlük)*. İstanbul, Türkiye Akademik Kitaplar, 2012.
- [19] Nuve Studio (2021). Nuve Studio [online] <http://nuvestudio.com> [accessed 11. Month Year].
- [20] Küçük AE. Nuve Studio Source Code. <https://github.com/kucukea/nuve-studio>, 2021.
- [21] Zafer HR. Nuve : A Natural Language Processing Library for Turkish in C#. <https://github.com/hrzafer/nuve>, 2021.