

Uygur Dili ve Makineli Çeviri

Murat ORHUN¹, Eşref ADALI², A.Cüneyd TANTUĞ²

¹ İstanbul Bilgi Üniversitesi, Bilgisayar Bilimleri Bölümü, İstanbul

² İstanbul Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, İstanbul

muratmehmet@cs.bilgi.edu.tr, {adali,cuneyd}@cs.itu.edu.tr

Özet: Bu makalede, Uygur dilinin gramer yapısı, Uygur dilinde kullanılan alfabeler hakkında kısaca bilgi verilecektir. Uygur dili üzerinde yapılan son araştırmalar ve gelişmeler özetlenecektir. Özellikle, Uygur dili ile ilgili yapılan Doğal dil çakışmaları ve Makineli çeviri üzerinde elde edilen sonuçlar özetlenecektir. Makineli çeviride kullanılan yöntemlerin Uygurca için nasıl kullanılabileceği hakkında yaptığımız araştırmalarımız ve önerimiz verilecektir. Sonunda, Uygurca ile Türkçe arasında makineli çevrinin nasıl yapılabilceği hakkında sistem yapısı verilecektir.

Anahtar Kelimeler: Makineli Çeviri, Uygurca Gramer, Uygur dili, Kural Tabanlı Çeviri.

Uygur Language And Machine Translation

Abstract: This paper gives some information about the structure of the Uygur Grammar and Uygur Alphabets. Summarizes the resent researches about the language and some results achieved so far. Mainly, we discuss the researches and achievements about the Natural Language Processing and the Machine Translation of the Uygur Language, which are done so far. We explain some rules which used for Machine Translations in general, and give our suggestions about these rules which could be used for the Uygur Language. At last, we suggest a system which implements the machine translation between Turkish and Uygur.

Keywords: Machine Translations, Uygur Grammar, Uygur Language, Rule Based Translation.

1. Giriş

Uygur dili yaygın olarak, Doğu Türkistan'da kullanılmaktadır. Bunlar Hariç, Orta Asya, Afganistan, Türkiye gibi Ükelere bile birçok kişiler tarafından kullanılmaktadır. Uygurlar günümüze kadar birçok alfabe kullanmıştır. Günümüzde Uygurlar yaşadığı bölgelere göre farklı alfabeler kullanılmaktadır. Orta Asya da yaşayan Uygurlar Kırıl alfabetesini kullanırken, Afganistan da yaşayanlar ise, Araf alfabetesini kullanmaktadır.

Doğu Türkistan'da yaşayan Uygurlar ise, 80 - yıllara kadar Latin Alfabetesini temel alan alfa-be kullanırken, günümüzde ise Araf alfabetesini

temel alan 32 harften oluşan alfabe kullan-maktadırlar. Araf Alfabetesine geçiş yakın tarihlerde gerçekleştirildiğinden dolayı, birçok kişiler, halen Latin Alfabetesini temel alan Uygur alfabetesini kullanmaktadır. Resmi yazış-malarda, okullarda, Araf alfabetesini temel alan alfabe kullanılırken, bilgisayarlı yazışmalarda ise, Latin alfabeli yazışmalar yasal olması bile kullanılmaktadır.

Bilgisayar kullanımının yaygınlaşmasıyla, Latin alfabetesini standartlaşma çabaları yapılmıştır. Bu işlemleri yürütmek için UKIJ[1] adında bir kurum kurulmuştur. Hatta Araf alfabeti ile Latin alfabeti arasında unikodlu[2] çevri yapan ücretsiz yazılımlar geliştirilmiştir [1]. Uygur-

cada kullanılan Araf alfabesi ile Latin alfabesinin karşılaştırılması Tablo1’de verilmiştir [1]. Bu makalede gösterilen örneklerde Latin alfabesinde kullanılan harfleri temel almıştır.

2.1 Uygur Alfabesinde Kullanılan Harfler

Uygur dili alfabesi, 8 sesli, 24 sessiz toplam 32 harften oluşmaktadır.

Sesli harfler ise:

A, , E, İ, O, Ö, U, Ü

Sesli harfler, telaffuz edilirken, dil yerine göre üç türe ayrılır.

1. Dil aldı sesli harfler: , Ö, Ü
2. Dil Arka sesli harfler: A, O, U
3. Dil Ortası sesli harfler: E, İ

Sesli harfler, telaffuz edilirken dudak şekline göre iki türe ayrılır.

1. Yuvarlak dudaklı sesli harfler: O, U, Ö, Ü
2. Yuvarlak dudaklı olmayan sesli harfler: A, , E, İ

Uygurcada kullanılan sessiz harflar ise:

B, P, T, J, CH, X, D, R, Z, J, S, SH, GH, F, Q, K, G, NG, L, M, N, H, W, Y

Sessiz haralarda iki türe ayrılırlar ünlü, ünsüz diye.

1. Ünlü sessiz harfler: B, J, D, R, Z, J, GH, G, NG, L, M, N, H, W, Y
2. Ünsüz sessiz harfler: P, T, CH, X, S, SH, F, Q, K

Uygurcada tüm sesli harfler ünlü sessiz türüne dâhil edilebilinilmektedir [3].

Uygurca başka Türkî dilleri gibi bir bitişen dildir. Bir kök sözcüğe eklerin eklenmesi ile farklı anlam veren sözcükler oluşturulabilmektedir. Köklere ekler eklenirken, sesli ve sessiz uyumu sağlanması gereklidir. Bunlar hariç, bazen harflerin telaffuz edildiği andaki sese göre, bazı eklerin eklenme şeklide değişmektedir. Bundan dolayı Uygurcada eklerin eklenmesi başka Türk dillerine göre daha karmaşıktır. Özellikle, Farsça, Rusça, Çince ve Arapça sözcüklerin içermesi, kuralların daha çok karmaşık olmasına neden oluşturmaktadır.

2.2 Uygurcada Sesli Harflerin Uyumu

Uygurcada, sözcüklerin son hecesinde yer alan sesli harflerin, dil aldı ya da dil arkası olmasına göre ekler ekleniyorlar. Yani son hecelerde yer alan sesli harfler belirtici rolü üsteleniyor.

Mesela: kitap + lar
mektep+ lar

Ama son hecelerde yer alan sesli harfler, dil orta türüne dâhil ise, farklı durumlar söz konusu oluyor. Böyle durumlarda, bu heceleri oluşturan sessiz harflere, hatta tüm sözcüğe bakılması gerekiyor[4], [5].

Mesela: pilik +ler
it+ lar

Uygurcada, sözcüklere ekler eklendikten sonra, kök sözcüğünde bulunan bazı sesli ve sessiz harflerde da değişiklikler oluyor.

1.Sözcükleri oluşturan sesli harflerin sırası “ - İ “, “ Ü - E - İ “, “ E - İ - İ “ şeklinde ise, ekler eklenirken, son hecelere değil, bir önceki eklerde yer alan “” harfine bakılacaktır.

2. “A,” ile harfleri ile oluşan, bir ve çok heceli sözcüklere “ı” harfi ile oluşan ekler eklenerek, o sözcüklerde, eskiden yer alan “A,” harfleri

“E, İ” lere değişirse bile, yeniden ekler eklenecekse, o sözcüklerin eski orijinal durumu göz önüne alınacaktır.

3. Yuvarlak dudaklı sesli harfler ile oluşan tek heceli sözcüklere ve, son hecesi yuvarlak dudaklı sesli harf ile sonlanan çok heceli sözcüklere ekler eklenirken, eklerde yuvarlak dudaklı sesli harflardan oluşan ekler eklenecektir.

Mesela: qol + um
yüz +üm

4. Yuvarlak dudaklı olmayan sesli harfler ile oluşan tek heceli sözcüklere ve son hecesi yuvarlak dudaklı olmayan sesli harf ile sonlanan çok heceli sözcüklere ekler eklenirken, eklerde yuvarlak dudaklı olmayan sesli harflardan oluşan ekler eklenecektir.

Mesela: ügen + is
en +is

5: “A “ harfi ile oluşan tek heceli sözlere “i” ile başlanan ekeler eklendiğinde, o hece açık hece halinde kalır. Ve “A” harfi ise, “i” harfi ile değişecektir.

Mesela:
Al + ip = ilip

6: “E “ harfi ile oluşan tek heceli sözcüklere ekler eklenirken, “E” harfi ise “i” harfine değişecektir.

Mesela: Kel + er = Kiler

7: Bazı iki heceli sözcüklerin ikin hecesindeki yüksek sesli harfler, o sözcüklere ekler eklenirken, kendi vurgusunu kayıp eder. Yani o sesli harf kayıp olacaktır.

Mesela: köngül + i = köngli
Burada “ü” harfi kayıp oldu.

2.3 Uygurcada Sessiz Harflerin Uyumu

1. “D, B, H” den başka ünlü sessiz harfler ile sonlanan sözcüklere, ünlü sessiz harfler ile başlanan ekler, ünsüz harfler ile sonlanan sözcüklere, ünsüz harfler ile başlanan ekler ekleneceklerdir.

Mesela: baj + din
al+ dim

2. Son hecesi açık hece olan sözcüklere, birinci, ikinci şahıs ekleri eklendiğinde, eklerin

Mesela: imla +yim
imla + ying
imla + si
su +süyüm
su + süyüng

2.4 Uygurcada Bağımsız ve Türetilmiş Sözcüklerin Yazılması

Uygurcada bir birinden bağımsız iki ya da daha fazladan sözcükler birleşerek, yeni anlam veren sözcükler oluştururlar. Bunlara Türetilmiş sözcükler denir. Türetilmiş sözcükleri oluşturan bağımsız sözcükler, bazen hiç bir değişime uğramadan bir-biriyle birleşebilirken, bazı durumlarda değişmektedir.

Mesela: aş + tahta = aştahta
(burda değişme olmadı)
beng + bash = bengwash

(burada, ikinci sözcükte gecen “b” harfi ise “w” ile değişti.)

Yani, ikinci sözcük “b” ile başlanan bazı sözcüklerde (hepsinde değil) , “b” harfi ise “w” ile değişecektir , “w” sesi ile okunacaktır.

Araf Alfabeti		Latin Alfabeti		Türkçe Anlamı
Harfler	Örnekler	Harfler	Örnekler	
ئا	ئات	a	at	at
ئە	ئەينەك	e	eynek	ayna
ب	بېلىق	b	beliq	balık
پ	پاڧا	p	paqa	kurbağa
ت	توشقان	t	toshqan	tavşan
ج	جۇۋا	j	juwa	
چ	چەينەك	ch	cheynek	çaydan
خ	خوراز	x	xoraz	horoz
د	داپ	d	dap	def
ر	راۋاپ	r	rawap	revap
ز	زەنجىر	z	zenjir	zincir
ژ	ژۇرنال	j	jurnal	jurnal
س	سائەت	s	saat	saat
ش	شىر	sh	shir	aslan
غ	غاز	gh	ghaz	gaz
ف	فونتان	f	fontan	fıskiye
ق	قوي	q	qoy	koyun
ك	كالا	k	kala	inek
گ	گۈل	g	gül	gül
ڭ	ياڭاق	ng	yangaq	ceviz
ل	لامپا	l	lampa	ampul
م	مۈشۈك	m	müshük	kedi
ن	نان	n	nan	ekmek
ھ	ھارۋا	h	harwa	
ئو	ئورغاق	o	oghaq	orak
ئۇ	ئۇۋا	u	uwa	yuva
ئۆ	ئۆي	ö	öy	ev
ئۈ	ئۈزۈم	ü	üzüm	üzüm
ۋ	ۋېلسپېت	w	welespit	bisiklet
ئې	ئېيىق	é	éyiq	ayı
ئى	ئىت	i	it	it
ي	يولۋاس	y	yolwas	kaplan

Tablo1. Uygurcada kullanılan Arap alfabesi ile Latin alfabesinin karşılaştırılması.

2.5 Yardımcı Sözcüklerin Yazılması

“al , qal, chal, sal, bol, kəl, “ gibi filler, dolaylı geçmiş zaman anlamında kullanıldığında , “l” harfi ortalıktan kalkacaktır.

Mesela: al+di -> Aptu
qal+di -> qaptu,
sal + di -> saptu

Ama bu filler, geçmiş zamanda, dolaysız anlamında kullanılırsa, “l” harfi olduğu gibi kullanılacaktır.

Mesela:
eliwaldi. seliwaldi.

Burada, harflerin düşüp ya da düşmeyeceği ise, sözcüğün anlamına göre yapılmaktadır.

Bunlar hariç, Çince, Rusça den gelen sözcükler içinde özel yazama kuralları kullanılmaktadır.

3.1 Makineli Çeviri

Makineli çevri ise, Doğal Dili çalışmalarının bir alt dalıdır. Genelde bilgisayara dayanır. Çeşitli yazılımlar kullanarak, bir doğal dili bir başka doğal dile çevrime işlemidir.[6].

Bilgisayar teknolojisinin gelişmesi ve bilgisayarın yaygınlaşmasıyla, Makineli Çevri baya yaygınlaşmıştır. Şimdiye kadar Türk dilleri için yapılan çalışmalar Türkiye Türkçesi üzerinde yapılmıştır [7][8][9][10].

En son yapılan çıkışma ise, Türkiye Türkçesi ile Türkmence arasında geliştirilen çevri sistemidir. Aynı anda bu çevri sistemi İstanbul Teknik Üniversitesi Doğal Diller Çalışma Grubunda, Tüm Türk dilleri arsında Makineli çevri geliştirme ponjesinde geliştirilmektedir. Bundan daha önce, ZEMBİREK[11] adında açık kaynaklı biçimbirimsel çözümleyici çalış-

ması bulunmaktadır. Bu yazılım ise TÜBİTAK tarafından geliştirilmiş olan, açık kaynaklı bir sistemdir. Bu sistem Şimdi Türkçe ve Azerice, Diller üzerinde çözümleme yapabilmektedir.

Dolaysıyla, Uygurca üzerinde araştırma yapmak gerçekten çok önemlidir. Ama Uygurca için yapılan doğal dil ile ilgili az sayıda çalışma bulunmamaktadır.[12].

Yapılmış olan çalışmalarda[2], belirsizliği giderme hakkında çalışma yapılmamıştır. Uygurcada kullanılan kelimeler ve yapısı yönünden karmaşık kelimeler içerin bir dildir. Eş anlamlı bir çok kelimeleri içermektedir.

Mesela:

yüz	,isim, nesnelerin yüzü
yüz	,sayı, 100
yüz	,fiil, suda yüz
yüz	,fiil kafasını yüz

3.2 İki Düzeyli Biçimbirimde Kural Türleri

Uygurca çok ekler içerin ve eklerin eklenmesi ile yeni sözcükler üreten bilen bir dildir. Bundan dolayı, Uygurcacinın biçimbirimsel yapısını analiz etmek, için, Biçim birimsel çözümlemede en çok kullanılan yöntemlerden[13], [14], [15], [16], iki düzeyli biçimbirimsel kural yöntemi kullan ılınabilir[17] .

İki düzeyli biçimbirimde kural yöntemi ise, yapısal(lexicon) biçim ve görünen(surface) biçimden oluşuyor. Görünen biçimde bazı kurallar uygulanarak, yazıların görünme şekli kontrol edile bilinir. Bunun ili ilgili sonlu durum-otomat tabanlı yazılımlar geliştirilmiştir. Mesela, XEROX[19], KIMMO[20] . Bu yazılımlar herhangi bir dilden bağımsızdır.

Her dilini kendi yapısına göre kurallar tanımlana bilinir, alfabesi verilebilir. Bu kurallar iki yönü çalışmaktadır. Yani, verdigimiz ya-

zılım ve biçimbirimsel kurallar göre normal kelimeler üretebilirken, aynı anda, verdiğimiz düzgün bir kelimenin biçim birimsel yapısının çöze bilmektedir.

Bu yazılımlardan Ya da bunun tarzda yazılımlar geliştirerek, farklı diller aralarında makineli çeviri sistemi geliştirilmiştir. [21],[22][23][24][25]. Genelde İki düzenli çözümleyicilerde kullanılan kurallar altta verilmiştir

$a:b \Rightarrow LC_RC$

Yapısal gösterimdeki bir ‘a’ sesi, kendinden önce ve sonra belirtilen bağlamlar varsa (LC -

left context ve RC – right context) görünen biçimde ‘b’ sesine dönüşebilir, (ancak bu dönüşüm zorunlu değildir).

$a:b \Leftarrow LC_RC$

Yapısal gösterimdeki bir ‘a’ sesi, kendinden önce ve sonra belirtilen bağlamlar varsa (LC

-left context ve RC – right context) görünen biçimde ‘b’ sesine mutlaka dönüşür (bu dönüşüm zorunludur, koşullar sağlandığında gerçekleşmelidir).

$a:b \Leftrightarrow LC_RC$

Yapısal gösterimdeki bir ‘a’ sesi, kendinden önce ve sonra belirtilen bağlamlar varsa (LC-left context ve RC – right context) görünen biçimde ‘b’ sesine mutlaka dönüşür, (bu dönüşüm zorunludur) ve başka hiçbir bağlamda bu dönüşüm olmaz.

$a:b / \Leftarrow LC_RC$

Yapısal gösterimdeki bir ‘a’ sesi, kendinden önce ve sonra belirtilen bağlamlar varsa (LC - left context ve RC – right context) görünen biçimde ‘b’ sesine kesinlikle dönüşemez.

Tablo 2 de iki düzenli biçimbirimsel kural karşılaştırılmıştır ve örnekler verilmiştir.

$a:b \Rightarrow l_r$	lar lbr	Lar lar	lbr lbr	xay xby (olamaz)
$a:b \Leftarrow l_r$	lar lbr	Lar lar(olamaz)	lbr lbr	xay xby
$a:b \Leftrightarrow l_r$	lar lbr	Lar lar(olamaz)	lbr lbr	xay xby(olamaz)
$a:b / \Leftarrow l_r$	lar lbr(olamaz)	Lar Lar	lbr lbr	xay xby

Tablo 2. İki düzenli biçimbirimsel kurallar ve örnekler[17]

3. Uygurca Cümlelerin Yapısı ve Türkçe Cümleler ile Karşılaştırılması

Uygurca ile Türkçe cümle yapısı yanı sıra söz dizimi olarak yönünden benziyor, aynı anda çok sayıda farklar bulunmaktadır. Türkçe ile Orta Asya da konuşulan Türkçenin aynı lehçe olup olmadığı yönde tartışmalar bulunmaktadır. [26] .

Altta, Uygurca ve Türkçe cümleler verilmiştir.

Men bazargha kettim.

Ben pazara gittim.

Sen su ichting mu?

Sen su içtin mi?

Qoylar taghqa ketti.

Koyunlar dağa gitti.

Bu cümleler bir-birine çok benziyor. Hatta bazı kelimeler aynı anlam veriyor .

Ama bazen, tam tersin durumlar söz konusu oluyor.

Men bazargha badim ve sebze aldım.

Ben pazara gittim ve havuç aldım.

Bundan dolayı, Uyurca ile Türkçenin cümle yapısı ne kadar benzese bile, doğrudan çeviri yapmak çok yanlış sonuçlar getirebilmektedir. Dolayısıyla, biçim birimsel analiz yaptıktan sonra, kesin olarak belirsizlik, karşılıkları gidermek çalışması yapılması gerekmektedir.

4. Kısaltmalar

UKIJ Uyur Komputer İlmi Jemiyeti

5. Kaynaklar

[1]. <http://www.ukij.org>

[2]. <http://unicode.org>

[3]. Hamit T. Modern Uyur Grammar (Morphology). Yıldız Teknik Üniversitesi, Fen-Ed Fak. T.D.E Bölümü. İstanbul 2003.(İngilizce baskı).

[4]. Mirsultan Osmanof, “Hazırqı Zaman Uyghur Edebiy Tilining İmla ve Teleppuz Lughiti”. Shin Jiang Xeliq Neshiriyatı. Ocak 1997. (Uyurca baskı)

[5]. Yi Shen Xiu Gao Xhi Ju, “Uyghur Dili Grammatkası”. Çin Halik Cumhuriyeti Merkezi Milletler Üniversitesi Baskısı, Şubat 1998. (Çince baskı).

[6]. <http://www.aai.org/AITopics/html/machtr.html>

[7]. Tantı A:C, Adalı E, Oflazer K.,2006. “Computer Analysis of The Turkmen Language Morphology” Proceeding of the 5th International Conference on Natural Language Processing,Fin TAL 2006 , Turku, Finland.

[8]. Gülşen Eryiğit and Eşref Adalı. “An Affix Stripping Morphological Analyzer for Turkish”

[9]. İ. Hamzaoğlu, “Machine translation from Turkish to other Turkic languages and an implementation for the Azeri languages,” in Institute for Graduate Studies in Science and Engineering. vol. MSc Thesis İstanbul: Bogazici University, 1993.

[10]. K. Altıntaş, “Turkish to Crimean Tatar Machine Translation System,” in Bilgisayar Mühendisliği Bölümü. vol. MSc Ankara: Bilkent Üniversitesi, 2000.

[11]. <https://zemberek.dev.java.net/>

[12]. Muhtar Mahsut, Yasuhiro Ogawa, Kazeu Sugino, Katsuhiko Tuyama, ans Yasuyoshi Inagaki , “An Experiment on Japanese- Uighur Machine Translation and Its Evalutation” AMTA 2004, LNAI 3265,pp.208-216 2004

[13]. K. Koskenniemi, “Two-Level Morphology : A General Computational Model for Word Form Recognition and Production,” Department of General Linguistics, University of Helsinki 1983.

- [14]. L. Karttunen, "KIMMO : A General Morphological Processor," in Texas Linguistic Forum, Texas, USA, 1983, pp. 163-186.
- [15]. E. L. Antworth, "PC-KIMMO: A Two-Level Processor for Morphological Analysis," Summer Institute of Linguistics, Dallas, Texas, USA 1990.
- [16]. R. Sproat, Morphology and Computation: MIT Press 1992.
- [17]. Two Level Morphology.
<http://www.ling.helsinki.fi/~koskenni/essli-2001-karttunen/>
- [19]. PC-KIMMO.
<http://www.sil.org/pckimmo/pc-kimmo.html>
- [20]. <http://www.xrce.xerox.com/competencies/content-analysis/fst/>
- [21]. K. Oflazer, "Two-level Description of Turkish Morphology," Literary and Linguistic Computing, vol. 9, pp. 137-148, 1995.
- [22]. L. Karttunen and K. Wittenburg, "A Two-Level Morphological Analysis of English," in Texas Linguistic Forum, Texas, USA, 1983, pp. 217-228.
- [23]. R. Khan, "A Two-Level Morphological Analysis of Rumanian," in Texas Linguistic Forum, Texas, USA, 1983, pp. 253-270.
- [24]. K. Koskenniemi, "An Application of the Two-Level Model to Finnish," University of Helsinki Department of General Linguistics (1985. 104
- [25]. S. Lun, "A Two-Level Morphological Analysis of French," in Texas Linguistic Forum, Texas, USA, 1983, pp. 271-278.
- [26]. Ahmet Cüneyd Tantuğ, "Türk Dilleri Arasında Sonlu Durum Modeli Tabanlı Metin Aktarımı". Doktora Tezi. Mayıs 2007.