

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

UYGURCADAN TÜRKÇEYE BİLGİSAYARLI ÇEVİRİ

**DOKTORA TEZİ
Murat ORHUN**

Anabilim Dalı : Bilgisayar Mühendisliği

Programı : Bilgisayar Mühendisliği

MAYIS 2010

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

UYGURCADAN TÜRKÇEYE BİLGİSAYARLI ÇEVİRİ

DOKTORA TEZİ

Murat ORHUN

(504022206)

Tezin Enstitüye Verildiği Tarih : 02 Mart 2010

Tezin Savunulduğu Tarih : 10 Mayıs 2010

Tez Danışmanı : Prof. Dr. Eşref ADALI (İTÜ)

Diğer Jüri Üyeleri : Prof. Dr. A .Coşkun SÖNMEZ (YTÜ)

Yrd. Doç. Dr. Şule Gündüz ÖĞÜDÜCÜ (İTÜ)

Prof. Dr. A. Melek ÖZYETGİN (Ankara Üniv.)

Yrd. Doç.Dr.Gülşen Cebiroğlu ERYİĞİT (İTÜ)

MAYIS 2010

ÖNSÖZ

Tez çalışmam sırasında, çok yoğun olmasına karşın her zaman desteğini ve ilgisini benden esirgemeyen hocam Prof. Dr. Eşref Adalı ve Dr. Ahmet Cüneyd Tantığ'a teşekkür ve saygılarımı sunmayı bir borç bilirim.

Bu tez çalışması için çeşitli Uygurca kaynakları sağlayarak yardımcı olan Doğu Türkistan Vakfı'na teşekkürlerimi sunuyorum.

Bu tezi İstanbul Bilgi Üniversitesi Bilgisayar Bilimleri Bölümü'nde araştırma görevlisi olarak çalışırken hazırladım. Bu sürede bana sağladığı kolaylıklardan dolayı tüm bölüm çalışanlarına teşekkürlerimi sunuyorum.

Son olarak benim için her şeye katlanan fedakâr anneme, kardeşlerime sevgi ve saygılarımı sunuyorum.

Mart 2010

Murat ORHUN

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	iii
İÇİNDEKİLER.....	v
KISALTMALAR.....	ix
ÇİZELGE LİSTESİ.....	xi
ŞEKİL LİSTESİ.....	xiii
ÖZET.....	xv
SUMMARY.....	xvii
1. GİRİŞ	1
1.1 Bilgisayarlı Çeviri	2
1.2 Önceki Çalışmalar	3
1.2.1 Batı diller ile ilgili yapılan çalışmalar	4
1.2.2 Türk dilleri ile ilgili yapılan çalışmalar	5
1.2.3 Akrafa ve bitişken diller için karma yöntemi	7
1.2.4 Uyurca ile ilgili yapılan çalışmalar.....	9
1.3 Tezin Amacı.....	9
1.4 Tezin Bilimsel Katkısı	10
1.5 Tezin Bölümleri.....	11
2. BİLGİSAYARLI ÇEVİRİ TEKNİKLERİ	13
2.1 Bilgisayarlı Çeviride Kullanılan Yöntemler	14
2.1.1 Kural tabanlı çeviri yöntemi	14
2.1.1.1 Doğrudan aktarma yöntemi.....	14
2.1.1.2 Sözdizimsel aktarma yöntemi	15
2.1.1.3 Dilden bağımsız çeviri yöntemi.....	16
2.1.2 Derlem tabanlı çeviri yöntemi.....	17
2.1.2.1 İstatistiksel çeviri yöntemi	17
2.1.2.2 Örnek tabanlı çeviri yöntemi.....	17
2.2 Yakın Diller Arasında Bilgisayarlı Çeviri	18
2.2.1 Çevrilecek sözcüklerin biçimbilimsel çözümlemesi	19
2.2.2 Belirsizliklerin giderilmesi.....	20
2.2.3 Çeviri kuralları	20
2.2.4 Kök sözcüklerin çevrilmesi ve tümcenin oluşturulması	21
2.3 Biçimbilimsel Çözümleme.....	22
2.3.1 Sonlu durumlu dönüştürücüler	23
2.3.2 İki düzeyli biçimbilimsel çözümleme.....	25
2.3.2.1 Yazım kuralları	26
2.3.2.2 Bitiştirme kuralları.....	28
2.4 Çeviri Sonucunun Değerlendirilmesi	30
2.4.1 BLEU	31
2.4.2 WER	32
3. UYGUR DİLİ	35

3.1 Uygur Abeceleri.....	35
3.1.1 Orhun – Yenisey abecesi	35
3.1.2 Eski Uygur abecesi.....	36
3.1.3 Çağatay abecesi.....	37
3.1.4 Latin abecesi	37
3.1.5 Uygur Arap abecesi.....	37
3.1.6 Uygur bilgisayar abecesi	38
3.2 Uygurca Harflerdeki Uyumluluk	40
3.3 Uygurca İsimlerin Biçimbilimsel Çözümlemesi.....	42
3.3.1 Çoğul ekleri	42
3.3.2 İyelik ekleri.....	45
3.3.3 Durum ekleri.....	47
3.4 Uygurca Sıfatların Biçimbilimsel Çözümlemesi	51
3.5 Uygurca Zarfların Biçimbilimsel Çözümlemesi	54
3.6 Uygurca Eylemlerin Biçimbilimsel Çözümlemesi	56
3.6.1 Uygurca eylemlerin çeşitleri.....	56
3.6.1.1 Geçişli eylem ve geçişsiz eylem	57
3.6.1.2 Hareket eylemi ve durum eylemi	57
3.6.1.3 Bağımsız eylem ve yardımcı eylem	58
3.6.2 Eylemlerin sabit sözcük şekilleri	61
3.6.2.1 Eylemlerden oluşan isimler	61
3.6.2.2 Eylemlerden oluşan zarflar	62
3.6.2.3 Eylemlerden oluşan sıfatlar	63
3.6.2.4 Eylemlerden oluşan amaç kipi	63
3.6.3 Eylemlerin haber şekilleri.....	64
3.6.3.1 Eylemlerin şahıs sınıfı	64
3.6.3.2 Eylemlerin çekim (mood) sınıfı	64
3.6.4 Eylemlerin hal-zaman sınıfı.....	67
3.6.5 Eylemlerin özel şekilleri.....	70
4. UYGURCA İLE TÜRKÇENİN KARŞILAŞTIRILMASI VE ÇEVİRİ	
KURALLARI	73
4.1 Uygurca ile Türkçenin Söz Dizimi	73
4.2 Uygurca ile Türkçenin Zaman ve Kip Yönünden Karşılaştırılması	74
4.3 Uygurcadan Türkçeye Çeviri Kuralları.....	77
4.3.1 Kök sözcüklerin çevirilmesi	79
4.3.2 Sadece biçimbilimsel bilgilerinin değişmesi	81
4.3.3 Biçimbilimsel bilgilerin ve sözcüklerin aynı anda değişmesi	81
4.3.4 Eylemlerden dolayı bir önceki sözcüğün farklı ekler ile değişmesi	82
4.3.5 Tümce anlamına göre biçimbilimsel bilgilerin değişmesi	83
4.3.6 Birden fazla sözcüğün bir sözcük ile değişmesi	86
4.3.7 Bir sözcüğün birden fazla sözcük ile değişmesi	88
4.3.8 “İdi” yardımcı eyleminin çevirilmesi.....	91
5. BELİRSİZLİKLERİN GİDERİLMESİ	95
5.1 Biçimbilimsel Belirsizliğin Giderilmesi.....	95
5.1.1 Eylemlerde belirsizliğin emir kipine göre giderilmesi.....	96
5.1.2 Gelecek zaman etiketine göre belirsizlik giderme-1.....	98
5.1.3 Gelecek zaman etiketine göre belirsizlik giderme-2.....	100
5.1.4 Sıfatlara göre zarfların silinmesi.....	101
5.1.5 Zarfların seçilmesi.....	102
5.1.6 Eylemlere göre zarfların seçilmesi.....	103

5.1.7 Tümce sonunda eylemlerin seçilmesi.....	104
5.1.8 Tümce başında gelen eylemlerin silinmesi	105
5.1.9 Şahıs eklerine göre belirsizliğin giderilmesi	105
5.1.10 Öznesiz tümcelerde belirsizliğin giderilmesi	106
5.1.11 Biçimbilimsel belirsizlik giderici ile ilgili sonuçlar	107
5.2 Sözcük Anlamsal Belirsizliğin Giderilmesi.....	109
5.2.1 Sözlük ve bilgi tabanlı yöntemler.....	110
5.2.2 Denetimli yöntemler	113
5.2.3 Kısmen denetimli yöntemler	113
5.2.4 Denetimsiz yöntemler.....	114
5.3 Uygurcada Anlamsal Belirsizliğin Giderilmesi	114
6. UYGURCADAN TÜRKÇEYE BİLGİSAYARLI ÇEVİRİ YÖNTEMİ.....	125
6.1 Karma Yöntemin Kullanılması	126
6.1.1 Aktarma modelleri.....	127
6.1.2 İDM oluşturulması.....	127
6.2 Belirsizlik Giderme Yönteminin Kullanılması	128
7. ÇEVİRİ SİSTEMİNİN UYGULANMASI	133
7.1 Karma Model Tabanlı Çeviri Sisteminin Uygulanması	133
7.2 Belirsizlik Giderme Yöntemi ile Çeviri Sisteminin Uygulanması.....	137
8. SONUÇLAR.....	141
8.1 Sonuçlar	141
8.2 Çeviri Örnekleri.....	145
8.3 Değerlendirmeler.....	147
KAYNAKLAR.....	151
EKLER.....	161
ÖZGEÇMİŞ	179

KISALTMALAR

UKIJ	: Uygur Kompuyuter İlmy Jemiyeti
TDK	: Türk Dil Kurumu
SDD	: Sonlu Durum Dönüştürücü
BÇ	: Bilgisayarlı Çeviri
ASCII	: American Standard Code for Information Interchange
BLEU	: Bilingual Evaluation Understudy
TR	: Türkçe
UY	:Uygurca
WER	: Word Error Rate
IBM	: International Business Machine
İDM	: İstatistiksel Dil Modeli
WSD	: Word Sense Disambiguation
TÜBİTAK	: Türkiye Bilimsel ve Teknik Araştırma Kurumu
BOS	: Bilgisayarca Okunan Sözlük

ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1 : İki düzenli biçimbilimsel kurallar ve örnekler.	27
Çizelge 3.1 : Uygurcanın iyelik ekleri.	45
Çizelge 3.2 : Uygurcanın durum ekleri.	48
Çizelge 3.3 : Dolaysız hüküm bağlayıcısı.	58
Çizelge 3.4 : Dolaylı hüküm bağlayıcısı.	59
Çizelge 3.5 : Anlatma hüküm bağlayıcısı.	59
Çizelge 3.6 : Sübjektif tahmini hüküm bağlayıcısı.	60
Çizelge 3.7 : Objektif tahmini hüküm bağlayıcısı.	60
Çizelge 3.8 : Eylemlerin emir-talep çekim ekleri.	66
Çizelge 3.9 : Eylemlerin varsayım-itiraz çekim ekleri.	66
Çizelge 4.1 : Uygurca ile Türkçenin zaman ve kiplerinin karşılaştırılması.	76
Çizelge 5.1 : Biçimbilimsel belirsizlik giderici sınaama sonuçları.	109
Çizelge 5.2 : İngilizcede “ash” sözcüğünün tanımı.	111
Çizelge 5.3 : Sözükle anlamının hesaplanması.	112
Çizelge 5.4 : Örnek 1 ile ilgili sınaama sonuçları.	120
Çizelge 5.5 : Örnek 2 ile ilgili sınaama sonuçları.	121
Çizelge 5.6 : Örnek 3 ile ilgili sınaama sonuçları.	122
Çizelge 5.7 : Uygurca anlamsal belirsizlik giderici sınaama sonuçları.	123
Çizelge 8.1 : Eğitim derlemi ile ilgili istatistikler [32].	141
Çizelge 8.2 : Karma yöntem ve yeni yöntemin İDM Tip- I başarımı.	142
Çizelge 8.3 : Karma yöntem ve yeni yöntemin İDM Tip- II başarımı.	143
Çizelge 8.4 : Karma yöntem ve yeni yöntemin İDM Tip- III başarımı.	143
Çizelge 8.5 : Karma yöntem ve yeni yöntemin İDM Tip- IV başarımı.	144
Çizelge 8.6 : Karma yöntem ve yeni yöntemin İDM Tip -V başarımı.	144
Çizelge 8.7 : Çeviri örnekleri-I.	145
Çizelge 8.8 : Çeviri örnekleri-II.	146
Çizelge 8.9 : Çeviri örnekleri-III.	146
Çizelge 8.10 : Çeviri örnekleri-IV.	147

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1 : Kural tabanlı çeviri yöntemleri- Vauquous üçgeni.	15
Şekil 2.2 : Yakın diller arasında bilgisayarlı çeviri aşamaları.	19
Şekil 2.3 : Sonlu durumlu makinelerin eklere göre çalışması.	23
Şekil 2.4 : Sonlu durumlu dönüştürücüler.	24
Şekil 2.5 : Biçimbilimsel dönüştürücü.	24
Şekil 2.6 : Uygurcadan Türkçeye sonlu durumlu dönüştürücü.	25
Şekil 2.7 : Xerox Lexc için sözcük yapısı.	29
Şekil 3.1 : Orhun-Yenisey yazıtları [42].	36
Şekil 3.2 : Eski Uygur yazıtları [42].	36
Şekil 3.3 : Çağatay abecesi ile yazılan yazıt [42].	37
Şekil 3.4 : Latin abecesi ile yazılan metin[42].	38
Şekil 3.5 : Uygur Arap abecesi.	38
Şekil 3.6 : Uygur Arap abecesi ile Latin abecesinin karşılaştırılması [108].	39
Şekil 3.7 : İsimlerin sonlu durumlu makineler ile çözümlenmesi.	50
Şekil 3.8 : Uygurcada eylemlerin zaman kategorisi.	68
Şekil 3.9 : Eylemlerin sonlu durumu geçişleri -1.	69
Şekil 3.10 : Eylemlerin sonlu durumu geçişleri -2.	69
Şekil 3.11 : Eylem özeklerinin sonlu durum geçişleri.	72
Şekil 5.1 : Biçimbilimsel belirsizlik giderici.	107
Şekil 5.2 : Tümcelerin biçimbilimsel çözümlenmesi.	108
Şekil 5.3 : Asıl Lesk algoritmasının sözde kodu [127].	112
Şekil 5.4 : Anlamsal belirsizlik giderici.	116
Şekil 6.1 : Uygurca sözcüğün Türkçe karşılığının doğrudan oluşturulması ([32]'den uyarlanmıştır).	126
Şekil 6.2 : Karma yöntemi ([32]' den alınmıştır).	129
Şekil 6.3 : Belirsizlik giderme yöntemi([32]'den uyarlanmıştır).	130
Şekil 7.1 : Karma model ile Uygurcadan Türkçeye çeviri sistemi ([32] 'den uyarlanmıştır).	134
Şekil 7.2 : Belirsizlik giderme yöntemi ile çeviri sistemi.	138

UYGURCADAN TÜRKÇEYE BİLGİSAYARLI ÇEVİRİ

ÖZET

Bilgisayarlı çeviri yapay zekâ bilim alanının bir alt dalı olan Doğal Dil İşlemenin (DDİ) önemli bir konusudur. Bilgisayarlı çeviri, bir doğal dili başka bir doğal dile çevirme işlemidir. Bilgisayar teknolojisinin gelişmesi ve bilgisayarın yaygınlaşmasıyla bilgisayarlı çeviri alanındaki araştırmalar yaygınlaşmış ve birçok çeviri sistemleri geliştirilmiştir.

Bu tezin amacı Uygurcadan Türkçeye bilgisayarlı çeviri konusunda gerekli kuramsal altyapıyı oluşturmak ve bunun üzerinde kurulacak yazılımlar ile çeviri işlemini gerçekleştirmektir. Bir kaynak dilden hedef dile çeviri yapabilmek için, özellikle bitişken diller için ilk yapılması gereken çalışma dillerin biçimbilimsel çözümleyicilerinin tasarlanmasıdır. Türkçenin biçimbilimsel çözümü daha önceden yapıldığından, bu çalışmada sözkonusu bu çözümleyici doğrudan kullanılmıştır. Uygurcanın dil yapısı araştırıldıktan sonra Uygurca için bir biçimbilimsel çözümleyici geliştirilmiştir. Uygurcadan Türkçeye çeviri yapabilmek için bu tez kapsamında çeşitli çeviri kuralları tanımlanmıştır. Belirsizliklerin giderilmesi için ise biçimbilimsel ve anlamsal belirsizlik giderici geliştirilmiştir.

Bu tezde akraba ve bitişken diller için önerilen karma yöntemi üzerine belirsizlik giderme çalışmaları da ele alınarak Uygurca metinden Türkçe metne bilgisayarlı çeviri yapabilen bir sistem geliştirilmiştir. Farklı yapılarda bulunan 240 adet tümce, önerilen bu sistem üzerinde sınanmış ve sistemin elde ettiği başarı BLEU ölçeğine göre değerlendirilmiştir.

MACHINE TRANSLATION FROM UYGHUR TO TURKISH

SUMMARY

Machine translation is a subfield of artificial intelligence, more precisely, natural language processing. It is based on computer technology which uses software to translate one natural language to another. As computer technology grew, it became popular and many machine translation systems have been developed as a result.

The main goal of this dissertation is to develop a theoretical infrastructure and implement a machine translation software system which may translate Uyghur to Turkish. Implementing a machine translation for agglutinative languages requires the development of a morphological analyzer for source and target languages. A morphological analyzer for Turkish has already been developed, similar technology was applied to this system. For Uyghur Language, a morphological analyzer was developed with natural language technology which emphasizes Uyghur grammar. Translation rules were defined according to the structure of Uyghur and Turkish languages in order to perform translations. Additionally, to reduce ambiguity, morphological disambiguation and word sense disambiguation methods were implemented.

In this dissertation, a machine translation system which translates Uyghur to Turkish was implemented based on similar models which were designed for related languages. The model was improved through the application of disambiguation models. BLEU scores were made to validate system performance, wherein 240 differently structured sentences were translated from Uyghur to Turkish.

1. GİRİŞ

Dil insanlar arasında iletişim sağlayan doğal bir araçtır. Dil deyince genelde ilk akla gelen şey insanlar arasında iletişim kurmakta kullanılan sözcüklerdir. İnsanlar arasında iletişim kurmada kullanılan bir başka yöntem de, işaret dilidir. Yani insanlar vücutlarını kullanarak da birbirleri ile iletişim kurabilirler. Ancak en etkili iletişim aracı kelimler ile yapılanıdır. Doğal olarak insanlar farklı bölgelerde yaşar ve farklı diller ile iletişim kurarlar. Bu nedenle kullanılan dillerin yapıları da farklıdır. Genel olarak, birbirleri ile yakın bölgede yaşayan toplumlarda konuşulan diller, birbirlerinden daha uzak bölgede yaşayanlarınkine göre daha çok benzemektedir ve bazı kelimler ortak olarak kullanılmaktadır. Örneğin Orta Asya'da yaşayanlar farklı milletlerden insanlar olsa da, kullandıkları dil yapısı ve bazı sözcükler ortaktır. Türkçe ve Uygurca dil yapısı benzemekte ve ortak çok sayıda sözcük içermekle birlikte, Japonca ve Uygurca dil benzerliğine rağmen, ortak sözcük içermemektedir. Aynı şekilde Japonca ve Çince de belli oranda ortak karakterler kullanılsada (Japoncada kendisine göre abecesi var iken, Çin dilinin abecesi yoktur), sözcük yapıları birbirine tamamıyla benzememektedir. Farklı toplumlardan insanların iletişim kurabilmeleri için, her iki dilde aynı anda konuşabiliyor olmaları gerekmektedir. Çünkü bir dili kullanabilmek, o dilin yapısını, dil bilgisini, sözcüklerini de bilmeyi gerektirir. Bu nedenle bilgisayar farklı diller arasında çeviri işlemini yapan bir araç olarak kullanılmaya çalışılmıştır. İlk yıllarda bilgisayarlar çok pahalı olduğundan bilgisayarlı çeviri fikrine pek olumlu bakılmamıştır. Teknolojinin, özellikle bilgisayar teknolojisinin gelişmesiyle, diller üzerindeki bilgisayar destekli araştırmalar hız kazanmıştır. Ne yazık ki Uygurca ile ilgili olarak bu alanda yapılan çalışmalar çok azdır. Türk dilleri içinde, en çok araştırma Türkiye Türkçesi üzerinde yapılmıştır. Bu nedenle Türkiye Türkçesi üzerinde yapılan çalışmalar ve araştırmalar diğer Türk dilleri için de kaynak oluşturmaktadır. Bu tez Uygurcadan diğer Türk dillerine çeviri yapmak üzere gerçekleştirilen ilk çalışmadır. Dolayısıyla bu tezin ileride Uygurca için yapılacak doğal dil ve bilgisayarlı çeviri işlemleri için bir temel oluşturarak, diğer Türk dilleri üzerinde yapılacak çalışmalara katkıda bulunması hedeflenmiştir.

1.1 Bilgisayarlı Çeviri

Bir dili bir başka dile bilgisayarlı çevirebilmek için, öncelikle üzerinde işlem yapılacak dillerin araştırılması gerekmektedir. Genel olarak bir dili çözümlemek amacıyla yapılan bilgisayarlı çalışmalara Doğal Dil İşleme(DDİ) adı verilmektedir. Doğal Dil İşlemi ise yapay zeka ve dilbilimi çalışmalarının bir alt dalı olup yoğun olarak mühendislik bilgisi ve dilbilimine hakim olmayı gerektirmektedir. Çünkü bir dili çözümleyebilecek sistemleri geliştirebilmek, o dil hakkında doğru bilgiye sahip olmakla mümkündür. 1900'lü yılların ilk yarısına kadar, farklı diller arasındaki çeviri işlemleri sadece insanlar tarafından yapılırken günümüzde bu alanda bilgisayarlar da kullanılmaktadır. Elektronik bilgisayarların üretilmesiyle, bilgisayarla çeviri işleminin temeli 19'cu yüzyılın ikinci yarısından itibaren başlamıştır [1]. Bilgisayar teknolojisinin ve yazılımların gelişmesi, doğal dil araştırmalarını ve bunun bir alt dalı olan bilgisayarlı çeviri çalışmalarının gelişimine olanak sağlamıştır.

Bilgisayar deyince, günümüzde, İnternette bağlanmak, başkaları ile yazılı, sözlü, görüntülü iletişim kurmak, bunun içinde tuş takımı, mikrofon, kamera, fare kullanmak anlaşılmaktadır. Ancak internet ortamındaki kişilerin dilleri farklı ise yukarıdaki araçları kullanarak iletişim kurmak imkansızdır. İnternet ortamındaki bir kişi sadece kameraya bakarak işaret dilini kullanabilir fakat etkin bir iletişim kuramaz. Aynı şekilde, elektronik ortamda kayıtlı, farklı dilde yazılmış bir dosya, bir başka dile çevirilmek istendiğinde yukarıda bahsedilen basit iletişim yöntemleri yeterli olamaz.

Doğal dil işlemi, sözcüklerle ilgili olarak biçimbilimsel çözümleme, tanımlama, düzeltme, tümce yapısı araştırma, çeviri işlemleri gerçekleştirme gibi pek çok alanı kapsamaktadır. Bilgisayarlı çeviri ise, Doğal Dil İşleminin en yaygın olarak çalışılan ve araştırılan konuları arasında gelmektedir. Teknolojilerin gelişmesiyle, farklı topluluklar arasında iletişim geçmiş yıllara göre daha çok önem kazanmıştır. Bu nedenle bilgisayarlı çeviri ile ilgili çalışmalar ağırlık kazanmaktadır. Günümüzde bilgisayarlı çeviri alanında, kısıtlı da olsa başarılı sayılabilecek sistemler geliştirilmiş ve kullanılmıştır. Örneğin, İngilizceden Japoncaya bildiri başlığı çevirmek üzere geliştirilen sistemin başarısı %98 olarak gerçekleştirilmiş ve bilimsel araştırmalarda aktif olarak kullanılmıştır [2]. Kanada'da hava tahmini ile ilgili bilgileri İngilizceden Fransızcaya çevirmek için METEO adında bilgisayarlı sistem geliştirilmiştir [3]. Bu

sistem 1981 yılından 30 Aralık 2001'e kadar Kanada Çevre Kurumu tarafından kullanılmıştır [4-5]. Bu sistemin kapasitesi 1977 yılında, günde ortalama olarak 7500 kelime iken, 1980'li yıllara doğru ortalama 80,000'e ulaşmıştır. Yani yılda ortalama olarak İngilizceden Fransızcaya 30,000 sözcük çevirilmiştir. Bu sistem Kanada Çeviri Kurumunun iş yükünü %91 oranında hafifletmiştir [6]. Bu nedenle BÇ ile ilgili çalışmalar gün geçtikçe önem kazanmaktadır ve farklı çeviri yöntemleri üzerine araştırmalar yapılmaktadır.

1.2 Önceki Çalışmalar

Bilgisayarlı çeviri 50 yıllık kısa bir geçmişe sahip olmasına rağmen çok önemli başarılar kat etmiş ve çeşitli çeviri sistemleri geliştirilmiştir. Ancak genel amaçlı ve yüksek başarılı sonuç veren bir sistem geliştirilememiştir. Örneğin, METEO çeviri sistemi sadece hava tahmini ile ilgili bilgiler üzerinde çalışır. Kısıtlı ya da özel amaçlı olarak geliştirilen bazı bilgisayarlı çeviri sistemlerinin temel özelliği kural tabanlı olmasıdır ve bu yöntem 1980'li yılların sonuna kadar yaygın olarak kullanılmıştır [7]. SYSTRAN [8-9], EUROLANG [10], kural tabanlı olarak geliştirilen ünlü bilgisayarlı çeviri projelerine örnektir. 1989 yılından sonra, kural tabanlı çeviri yöntemlerinin yerine derlem tabanlı çeviri yöntemleri tercih edilmiştir [11-13]. Bilgisayarlı çeviride büyük gelişme ise, istatistiksel yöntemin geliştirilmesi ile olmuştur [7]. IBM çalışanları stokastik yöntemi kullanarak ses tanıma projesi üzerinde elde ettiği başarıları, bilgisayarlı çeviri üzerine uygulamışlar ve Candide sistemini geliştirmişlerdir [14]. Bu sistemin temel özelliği hiçbir kural kullanmadan tamamı ile istatistiksel olarak çeviri yapmasıdır [14]. IBM'in araştırmacıları kendi araştırmalarında İngilizce ve Fransızca sözcükler içeren Canadian Hansar derlemi kullanmışlardır [15]. Şimdiye kadar çalışılan istatistiksel bilgisayarlı çeviri sistemleri ise: Candide [16], LINGSTAT [17], Bilingual Tiling [18], PANGLOSS [19-20], SBTG [21].

Şimdiye kadar geliştirilen çeviri sistemleri başarımlı yönünden değerlendirme yaparken, genel olarak birbirine yakın ve akraba diller arasında geliştirilen sistemlerin başarımlı, yapısal olarak çok farklı olan dillere göre daha üstün olmuştur. Örneğin, aynı dil ailesine ait olan İngilizce ve Almanca arasında geliştirilen çeviri sistemi, farklı dil ailesinde olan İngilizce ve Japonca arasındakinden daha başarılı olmuştur [22-24].

Farklı diller arasında geliştirilen sistemlerin başarımının düşük olmasının temel nedeni ise, hedef ve kaynak dillerin tümcelerinin farklı olmasıdır. Bu nedenle farklı diller arasında çeviri sistemi geliştirmek için, her iki dilin tümce yapısının araştırılması gerekir. Farklı diller arasında çeviri yaparken bir başka olumsuz etken ise, bazı olayların anlatım şeklinin de farklı olması ve her ne kadar benzeri şekilde aktarma yapılsa da, hedef dilde tıpkı kaynak dilde olduğu gibi algılanmamasıdır. Örneğin, Uygurcada ikinci şahıs ile ilgili olarak, “sen”, “siz”, “sili”, “sizler”, “senler” gibi beş farklı sözcük vardır ve farklı anlamlar içerir. “Sen” genel olarak, samimi olan kişiler arasında kullanılırken, “siz” ise saygı göstermek amacıyla kullanılır. “Sili” ise daha çok nezaketli bir tabir ile konuştuğunu belirtir. Ancak aynı anda “sitem” yapma anlamı da taşır. “Senler” sözcüğü ise karşısındaki kişiye çok sert bir tabir ile konuştuğunu belirtir. Ancak İngilizcede sadece tek sözcük , “you” vardır ikinci şahıs ile ilgili olarak. Eğer İngilizcede, tıpkı Uygurcada olduğu gibi anlam veren bir yapı oluşturulmak istenirse, “you” sözcüğü için bazı yardımcı eylemlerin eklenmesi gerekir. Bu nedenle, benzer ya da yakın diller arasında geliştirilen çeviri sisteminin, farklı diller arasında geliştirilen sistemlere göre daha kolay ve başarılı olduğunu anlaşılır.

Bu tezde, aynı dil grubunda bulunan Uygurca ile Türkçe arasında çeviri sistemi geliştirileceğinden, kaynak taraması olarak, birbirine yakın diller için geliştirilen BÇ sistemleri araştırılmıştır. Yakın diller ile ilgili yapılan BÇ çalışmaları aşağıdaki bölümlerde özet şeklinde anlatılmıştır. Ancak yakın diller için çeviri sisteminde kullanılan yöntemler tezin ikinci bölümünde ayrıntılı olarak anlatılmıştır.

1.2.1 Batı diller ile ilgili yapılan çalışmalar

Batı dilleri, özellikle yakın diller arasında yapılan ilk çeviri sistemi Çekçeden Rusçaya çeviri yapan RUSLAN sistemidir [25]. Bu sistem tek yönlü, yani Çekçeden Rusçaya çeviri yapabilmekteydi. Bu sistemi geliştirmek için kullanılan yöntemler özetlendiğinde, biçimbilimsel çözümleyici, birleştirici, sözdizimsel ayrıştırıcı ve birleştirici çalışmaları yapılmış, sistemin başarımı ise %40 olarak hesaplanmıştır. Sistemin başarımının düşük olmasının temel nedenleri ise, çeviri sözcüğün az olması ve tanımlanan yapı çözümleme kurallarının yetersiz olmasındandır. Bu sistem kural bazında çalışmaktadır ve kaynak dil Çekçenin biçimbilimsel çözümleyicisi tasarlanırken, Rusça için birleştirici yani Çekçeden aktarılan sözcük ve biçimbilimsel

bilgilerden görünen biçimde sözcük oluşturma modeli bulanmaktadır. Batı dilleri arasında yapılan bir başka çalışma ise, Çekçe ile Slovakça arasında geliştirilen ČESILKO sistemidir [26]. Bu iki dilin yapısı Rusçaya göre daha çok benzemektedir. Bu nedenle RUSLAN sistemine göre daha basit yöntemler kullanılmış olup sözcük bazında çeviri yapılmaktadır. Yani bu çeviri sisteminde tümce yapı araştırması yapılmaya gerek duyulmamıştır. Bundan dolayı bu çeviri sisteminin gerçekleştirilmesinde, kaynak dil olan Çekçenin biçimbilimsel çözümleyicisi, buna karşılıklı olarak Slovakçanın biçimbilimsel üreticisi tasarlanmıştır. Ayrıca Çekçe sözcükler üzerinde biçimbilimsel çözümleme yaparken, oluşabilecek biçimbilimsel belirsizliklerin gidermek için modeller geliştirilmiştir. Bunlar hariç, genel ve özel amaçlı aktarım sözlüğü olmak üzere iki çeşit aktarım sözlüğü kullanılmıştır. TRADOS yazılımı ile hesaplanan sistem başarımı %90 olup, RUSLAN sistemine göre çok yüksek çıkmıştır [27].

ČESILKO sisteminin RUSLAN sistemine göre başarımın yüksek olmasının temel nedenlerinden biri, biçimbilimsel belirsizlik giderme çalışmasının yapılması ve aktarım sözcük sayısının fazla olmasıdır. Biçimbilimsel çözümleme ve bu çözümlemeler ile ilgili çalışmalar sadece BÇ çalışmalarının değil, aynı anda tüm DDİ çalışmalarının en önemli bileşenidir.

Yukarıda anlatılan çalışmalar özetlendiğinde, birbirine yakın batı dilleri arasında BÇ sistemi geliştirilirken, her iki dil ile ilgili biçimbilimsel çözümleyicinin tasarlanması, belirsizliklerin giderilmesi ve aktarım sözlüklerinin oluşturulmasıdır.

1.2.2 Türk dilleri ile ilgili yapılan çalışmalar

Türk dilleri ile yapılan bilgisayarlı çeviri sistemi ilk olarak Azerice ile Türkçe arasında gerçekleştirilmiştir [28]. Bu sistem sözcük tabanında çalışmaktadır ve Azericeden Türkçeye çeviri yapmaktadır. Sistemde kullanılan diller birbirine benzediğinden yani özellikle cümle yapısı benzediğinden dolayı çevirilecek sözcüklerin söz diziminde herhangi bir değişiklik yapılmamıştır. Öncelikle çevirilecek Türkçe sözcükler üzerinde biçimbilimsel çözümleme yapılmış, sonra elde edilen kök sözcük ve ekler Azerice karşılıkları ile değiştirilerek Azerice sözcükler oluşturulmuştur. Türkçe kök sözcüklerin Azerice karşılığı bulunurken, birden fazla karşılığı olan kök sözcüklerin çevirilmesinde oluşacak anlamsal belirsizliklerin giderilmesi için, daha önceden oluşturulmuş sözlüğe anlamsal bilgiler eklenmiştir.

Ayrıca, sözcüklere ek olarak, ilgili kavramsal özellikler eklenmiştir. Bu kavramsal özelliklerin, tümcede etkin olan kavram ile uyuşup uyuşmadığına bakılarak belirsizlik giderilmiştir. Bunların dışında birliktelik çizelgeleri hazırlanarak yapıların sıklığına göre belirsizlik giderilmiştir. Bu sistem kural bazında çalışmaktadır.

Türk dilleri için geliştirilen bir başka bilgisayarlı çeviri sistemi ise, Kırım Tatarcısı ile Türkçe arasında gerçekleştirilmiştir [29]. Türkçeden Kırım Tatarcasına çeviri sisteminde 5300 adet sözcük içeren bir çeviri sözlüğü kullanılmıştır. Geliştirilen sistem sadece tek yönlü, yani Türkçeden Kırım Tatarcasına çeviri yapacak şekilde çalışmaktadır. Tatarcadan Türkçeye çeviri işlemi gerçekleştirilmemiştir. Türkçe sözcüklerin biçimbilimsel çözümlemesi için Oflazer'in Türkçe için geliştirmiş olduğu biçimbilimsel çözümleyicisi kullanılmıştır [30]. Kırım Tatarcısı için sonlu durumlu makinelere göre çalışan bir biçimbilimsel çözümleyici tasarlanmıştır [31]. Çeviri işleminin gerçekleştirilmesi ise, çeviri sözlüğüne bakılarak yapılmaktadır. Türkçe sözcükler üzerinde biçimbilimsel çözümleme yapıldıktan sonra elde edilen kök sözcükler, Tatarca kök sözcüklere çevrilmiştir. Daha sonra bu kök sözcükler ve Türkçe sözcükler, çözümlendiği anda oluşan ek bilgiler birleştirilerek Tatarca sözcükler elde edilmiştir. Bu sistem sözcük bazında çalışmaktadır ve kural tabanlıdır.

Bu sistem geliştirilirken, biçimbilimsel çözümlemeye oluşan belirsizlikleri gidermek konusunda herhangi bir çalışma yapılmamıştır. Türkçe, kendi yapısı açısından belirsizlik içeren bir dildir. Bu nedenle, söz konusu sistem olası tüm çözümlemeleri aktarmaktadır. Ayrıca, sözcük aktarılırken, sözcükler üzerindeki anlamsal belirsizliği gidermek amacıyla bir çalışma yapılmamıştır. Çünkü Türkçedeki bir sözcüğe, Tatarcada birden fazla sözcük karşılık gelebilmektedir.

Türk dilleri için, en son bilgisayarlı çalışma ise, Tantuğ'un İstanbul teknik üniversitesi bilgisayar mühendisliği bölümünde, doktora tez çalışmasında önerdiği Karma¹ yöntem kullanarak geliştirdiği Türkmenceden Türkçeye çeviri yapan BÇ sistemidir [32]. Bu yöntemde tüm bitişken diller için, BÇ sisteminin geliştirilmesinde kullanılabilecek genel bir çözüm yöntemi önerilmiştir. Ayrıca önerilen sistemde, Türkmenceden Türkçeye bir bilgisayarlı çeviri uygulaması geliştirilmiştir ve

¹ Tezin aslı adı "Akraba ve Bitişken Diller Arasında Bilgisayalı Çeviri için Karma Bir Model" dir [32]. Atıfta bulunurken, kolay olması için "Karma yöntem" olarak kısaltılmıştır.

herhangi Türk dillerinden Türkçeye çeviri geliştirildiğinde, Türkçe ile ilgili yapılan çalışmaları hiçbir değişiklik yapmadan kullanabilecek bir alt yapı oluşturulmuştur. Bu alt yapı Uygurcadan Türkçeye çeviri sisteminde kullanılacağından aşağıdaki bölümde daha ayrıntılı olarak anlatılmıştır.

1.2.3 Akriba ve bitişken diller için karma yöntemi

Karma yöntem yakın diller arasında, özellikle sözdizimi benzeyen ve bitişken diller için önerilen çeviri yöntemidir [32]. Bu yöntemde, çalışılacak diller bitişken diller olduğundan, öncelikle işlem yapılacak kaynak ve hedef diller ile ilgili biçimbilimsel çözümleyicilerin tasarlanması önerilir. Sonra sözcük yapıları aktarma, belirsizlik giderme işlemleri sonucu hedef dilde çeviri sözcüğü üretilir. Yalnız bu sistemde, hedef dilde tümce üretirken, koşullara göre çalışan kurallar tanımlanır. Ayrıca belirsizlik gidermede istatistiksel yöntem kullanılır. Bu nedenle Karma model olarak adlandırılmıştır. Tantuğ önerdiği bu Karma yöntemi Türkmenceden Türkçeye geliştirdiği BÇ sistemi ile kanıtlamıştır ve sistemin başarımını BLEU değerlerine göre ölçmüştür. Tantuğ'un önerdiği Karma yöntemin uygulanması, Türkmenceden Türkçeye geliştirdiği BÇ sistemi üzerinden anlatılmıştır.

Bu sistemde, genelde tüm bitişken diller ile ilgili BÇ işlemi yaparken, karakter düzeltme, biçimbilimsel çözümleyici tasarlama, çeviri kuralları tanımlama gibi işlemlerin olduğundan hariç, en belirgin farkı biçimbilimsel çözümlemeleri aktarmak için kullanılan farklı yöntemi, sonra bu yöntemler ile bağlantılı olarak İDM'nin hesaplanmasıdır. Aşağıdaki alt başlıklarda bu iki yöntemin nasıl çalıştığı anlatılmıştır.

1. Aktarım fonksiyonu: Burada Türkmence sözcüklerin Türkçeye aktarılmasında üç farklı yöntem kullanılmıştır. Birincisi temel yöntemdir. Çevirilecek diller yakın diller olduğundan doğrudan sözcük bazında aktarma yapılmıştır. Ancak işlem yapılacak diller bitişken diller olduğundan, sözcükler üzerinde biçimbilimsel çözümlemeler yapılmıştır. Böylece sözcük ve biçimbilimsel bilgilerin aktarılması sağlanmıştır. İkinci yöntemde ise, çevirilecek dil, Türkçenin çoklu sözcük içermesi dikkate alınarak çoklu sözcük grup modeli, temel model üzerine eklenmiştir. Üçüncü yöntemde ise, Türkten Türkmenceye kök aktarma işleminden sonra oluşan söz dizimlerinde, Türkçe genel dil bilgisine göre düzenleme yapılmıştır.

2. Hedef dilde istatistiksel model: Türkmençe sözcükler üzerinde biçimbilimsel çözümleme yapıldıktan sonra, oluşan tüm çözümlemeler, Türkmenceden Türkçeye kök aktarma sözcüğünden yararlanılarak Türkmençeye aktarılmıştır. Biçimbilimsel bilgi içeren bu sözcüklerin hepsi istatistiksel dil modeline gönderilmiştir. İstatistiksel modelin çalışabilmesi için bir Türkçe eğitim derlemi oluşturulmuştur. Dolayısıyla, sözcükler ve sözcüklerin çözümlenmesinde elde edilen biçimbilimsel bilgilere göre olasılığı en yüksek olan tümce yapısı bulunmaya çalışılmıştır. Bu nedenle, İDM bileşeni, aynı anda, biçimbilimsel ve anlamsal belirsizlik giderici görevini üstlenmiştir.

İDM çalışma sonucu elde edilen tümce yapısı, yapısal düzeyde Türkçe için oluşturulan genel kuralların uygulanmasından sonra, Türkçenin biçimbilimsel çözümleyicisinden sentez edilerek Türkçe sözcükler elde edilmiştir. Türkmençe ile Türkçenin sözdizimi aynı olduğundan dolayı, çeviri sonucu oluşan sözcüklerin üzerinde sözdizimsel düzenleme yapılmamıştır.

Bu çalışmada, Türkmençe sözcüklerin biçimbilimsel çözümlemesini yapmak için, Tantuğ Türkçenin biçimbilimsel çözümleyicisini geliştirmede kullanılan aynı etiketleri kullanarak Türkmençe için biçimbilimsel çözümleyici geliştirmiştir [33]. Türkmençede bulunan ancak Türkçede bulunmayan durumların çözümlenmesinde özel etiketler kullanmıştır ve daha sonra çeviri kuralları tanımlayarak Türkçeye aktarılmasını sağlamıştır. Türkçenin biçimbilimsel çözümleyicisi için, Oflazer tarafından geliştirilen çözümleyici kullanılmıştır [30]. Akra ve bitişken diller için, özellikle Türk dilleri için, uygun bir yöntem olduğu, Türkmenceden Türkçeye çeviri için geliştirilen sistem ile kanıtlanmıştır [34]. Bu yöntemde başka dillerden Türkçeye çeviri sistemi geliştirirken, sadece kaynak dil ile ilgili çalışılması yeterli bulunmaktadır. Yapılması gereken tek şey ise, kaynak dilin biçimbilimsel çözümleyicisi tasarlanırken, Türkçenin biçimbilimsel çözümleyicisinde kullanılan etiketlerin kullanılması, çeviri kuralların tanımlanması ve bir kök aktarma sözcüğünün geliştirilmesi yeterli bulunmaktadır. Bu nedenle, Türkçe ilgili birçok çalışmadan soyutlandırılmış bulunmaktadır. Ayrıca bu tez kapsamında geliştirilen çeviri çerçevesi, bu çalışmaların yapılacağı çerçeveyi sunmaktadır.

Bu tez kapsamında, Uygurcadan Türkçeye çeviri sistemi geliştirileceğinden, bu Karma yöntem için kullanılan altı yapı kullanılmıştır. Böylece Türkçe ile yapılan önceki çalışmalardan verimli bir şekilde yararlanma sağlanmıştır. Bu yöntemin, bu

tez için nasıl uyarlandığı, tezin 6 ve 7'ci bölümlerinde örnekler ile ayrıntılı olarak anlatılmıştır.

1.2.4 Uygurca ile ilgili yapılan çalışmalar

Uygurca ile ilgili olarak, doğal dil çalışmaları pek fazla bilinmemektedir ya da henüz başlangıç aşamasındadır [35-41]. Türk dilleri içinde, en çok araştırılan dil ise Türkiye Türkçesidir. Bundan dolayı Türkçe için yapılan çalışmaları Uygurcaya nasıl uygulanacağı üzerinde araştırmalar yapılmıştır. Uygurca ile Türkçe aynı dil ailesine ait olmasına karşın, Uygurcada kullanılan zaman kipleri, sesli ve sessiz harflerin gruplandırılması, sözlüklerin oluşturulması, eklerin eklenme kuralları gibi birçok yönden farklılık göstermektedir [42]. Bundan dolayı diğer Türk dilleri için geliştirilen biçimbilimsel çözümleyici ve aktarma kuralları gibi araçları doğrudan kullanmak imkânsızdır. Bu nedenle Uygurca için ayrı bir çalışma yapılması gerekmektedir.

Uygurca için şimdiye kadar yapılan BÇ ise Japoncadan – Uygurcaya çeviri sistemidir [43]. Bu çalışmada Uygurca ile Japoncanın söz diziminin aynı olması avantajından yararlanılarak, sözcükten sözcüğe çeviri yapılmış, fakat belirsizlik konusunda hiçbir çalışma mevcut değildir.

1.3 Tezin Amacı

Bu tezin amacı, DDİ ve bilgisayarlı çeviri yöntemleri kullanarak Uygurca bir metni Türkçe bir metne çeviren bir sistem geliştirmektir. Bu nedenle kaynak dil olan Uygurca ve hedef dil olan Türkçe üzerinde yapılan çalışmalar araştırılmıştır. Ancak şimdiye kadar Uygurca ile ilgili DDİ alanında ya da bilgisayarlı çeviri alanında ciddi çalışmalar yapılmamıştır. Sadece Uygur abeceleri üzerinde araştırmalar yapılmıştır [35-36]. Uygurca eylemler araştırılmış ve Uygurca derlem geliştirmek için çalışmalar başlatılmıştır [37-41]. Tüm Türk dilleri arasında, doğal dil çalışmaları ile ilgili olarak en çok araştırma yapılan dil ise Türkiye Türkçesidir. Örneğin, Türkçe için yapılan bilgisayar ortamındaki ilk çalışma Aydın Köksal'ın 1976 yılında, Türkçe sözcüklerin biçimbilimlerinin bilgisayarla çözümlemesi üzerine Hacettepe Üniversitesi'nde tamamlamış olduğu doktora tezidir [44]. Daha sonra Türkçe ile ilgi pek çok çalışma yapılmış ve diğer Türk dilleri için temel oluşturmuştur [45-55]. Özellikle İstanbul Teknik Üniversitesi bilgisayar mühendisliği bölümünde bulunan DDİ araştırma

grubu Türkiye Türkçesi için aktif çalışmalar yapmaktadır [56]. Bu araştırma grubu, Türk Dil Kurumu(TDK) ile Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) iş birliğiyle projeler geliştirmiş ve geliştirmeye devam etmektedir [57-59].

Türk dilleri grubunda bulunan diller, konuşulduğu bölgelerden dolayı birbirine çok benzerlik göstermektedir. Örneğin bir Uygur bir Özbek'in konuştuklarını rahat bir şekilde anlayabilmektedir. Aynı durum Kazakça ve Kırgızca için de geçerlidir. Genel olarak tüm Türk dillerinde ortak konuşulan sözcükler azımsanmayacak kadar çoktur. Özellikle, Özbekçe ile Uygurca yaklaşık yüzde 95'in üzerinde birbirine benzemektedir. Kullanılan sözcükler ve tümce yapısı da aynıdır. Bazı kaynaklar ise Özbekçe ve Uygurca'yı aynı dilin farklı lehçeleri olarak tanımlamaktadırlar. Özbekistan'ın Sovyetler yönetiminde, Uygurların ise Çin yönetiminde kalmaları nedeniyle son yıllara doğru her iki dilde farklı sözcükler oluşmuştur. Uygurcaya Çince, Özbekçeye de Rusça sözcükler eklenmiştir. Ancak söz dizimi yapısı korunmuştur.

Bu tezde, öncelikle Uygurcanın dil yapısı incelenmiş, Türkçe ve diğer diller üzerinde yapılmış bilgisayarlı çeviri çalışmaları araştırılmış, bunların Uygurcaya nasıl uygulanabileceği üzerinde durulmuştur. Türkçe ve Uygurca benzer diller olmakla birlikte Türkçe için geliştirilen yöntemlerin doğrudan kullanılmayacağı kanaatine varılmış, bu nedenle Uygurcaya özgün araştırmalar yapılmıştır.

Ancak Uygurcada daha önce yapılmış olan doğal dil çalışmalarının sınırlı olması, araştırmalarımızı zorlaştırmaktadır.

1.4 Tezin Bilimsel Katkısı

Bu tez çalışama kapsamında, Uygurca ile ilgili olarak ilk defa bir biçimbilimsel çözümleyici geliştirilmiştir. Yani bu demek oluyor ki, ilk defa Uygurca üzerinde resmi bir doğal dil çalması yapılmıştır. Çünkü biçimbilimsel çözümleyici tüm bitişken diller için yapılması gereken ilk çalışmadır. Böylece öncelikle Uygur dilinin bundan sonraki DDİ çalışmaları için bir temel atılmıştır. Ayrıca Türk dilleri üzerinde DDİ çalışmaları yaparken, kaynak olabilecek bir çalışma yapılmıştır. Sonra bu biçimbilimsel çözümleyicinin çözümlediği sözcükler ile ilgili belirsizliği gidermek amacıyla, belirsizlik giderme çalışmaları yapılmıştır. Böylece, ilk defa bu

tez kapsamında Uygurca sözcüklerin biçimbilimsel belirsizlik oranı resmi rakamlar ile verilmiştir. Yani 1.44 olarak hesaplanmıştır. Uygurca ile yapılan bu belirsizlik ile ilgili çalışmalar Türkiye Türkçesi ile yapılan çalışmadan sonra, Türk dilleri içersinde yapılan ikinci çalışmadır.

Bu tez çalışmasında bulunan en önemli bulgu ise, anlamsal belirsizlik giderme yönteminin bitişik dillere nasıl uygulanacağı ile ilgilidir. Uygurcada henüz bilinen bir derlem olmadığından, bilgi tabanlı anlam belirsizlik giderme yöntemi ve Lesk algoritması kullanılmıştır. Ancak Uygurca bitişken dil olduğundan, sözcüklerin görünen biçimi üzerinden hesaplama yaparken yanlış sonuçlar elde edileceğinin kanaatine varılmıştır. Bu nedenle Lesk algoritmasını sözcüklerin görünen biçimi üzerinden değil, kök sözcükler üzerinden hesaplama yapıldığında daha doğru sonuç alınabileceği kanıtlanmıştır.

Tezin bir başka bilimsel katkısı ise, Tantığun geliştirdiği çeviri çerçevesinde çalışabilecek şekilde tasarlanmasıdır. Bu nedenle Türkçe ile ilgili fazla çalışma yapmaya (çeviri kuralları hariç) gerek kalmadan, daha önce yapılan çalışmalardan yararlanılması sağlanmıştır.

Tezin diğer bilimsel katkıları ayrıntılı olarak tezin son bölümünde verilmiştir.

1.5 Tezin Bölümleri

Bu tezin bölümleri şu şekilde düzenlenmiştir. 2. Bölümde bilgisayarlı çeviri ile ilgili temel yöntemler tanıtılmış, bilgisayarlı çeviri başarı değerlendirme teknikleri anlatılmıştır. 3. Bölümde Uygur dili araştırılmış ve Uygurcanın biçimbilimsel çözümleyicisi tasarlanmıştır. 4. Bölümde Uygurca ile Türkçe karşılaştırılmış ve Uygurcadan Türkçeye bilgisayarlı çeviri sistemi için gereken çeviri kuralları tanımlanmıştır. 5. Bölümde Uygurcada biçimbilimsel belirsizlik durumu araştırılmış ve bu belirsizliği gidermek için kural tabanlı biçimbilimsel belirsizlik giderici geliştirilmiştir. Ayrıca, sözcüklerin anlamsal belirsizlikleri üzerinde çalışma yapılmış ve Uygurca için sözcüğün anlamsal belirsizliğini giderici geliştirilmiştir. 6. Bölümde ise Uygurcadan Türkçeye bilgisayarlı çeviri için kullanılan yöntemler anlatılmıştır. 7. Bölümde Uygurcadan Türkçeye bilgisayarlı çeviri uygulaması yapılmıştır. 8. Bölümde tez ile ilgili başarı sonuçları verilmiştir, ayrıca çeviri örnekleri ve bu

örneklerle ilgili deęerlendirme sonuçları verilmiřtir. Sonunda genel bir deęerlendirme yapılmıřtır.

Tezin ek bölümlerinde, Uygurcanın biçimbilimsel çözümleyicisi tasarımında kullanılan iki düzeyli kurallar, biçimbilimsel çözümleyici ile sınanan örnekler, Uygurcadan Türkçeye çeviri kuralları, biçimbilimsel çözümleyicide sözcükleri etiketlemek için kullanılan etiketler ve açıklamaları verilmiřtir.

2. BİLGİSAYARLI ÇEVİRİ TEKNİKLERİ

Bilgisayarlı çeviri, doğal dil alanının yoğun olarak çalışılan bir alt koludur. Bilgisayarlı çeviri, bir dilin bir başka dile bilgisayar yardımıyla çevrilmesi işlemidir. Ancak, nasıl ki bir dili bir başka dile insan gücüyle çevirirken, güçlüklerle karşılaşılırsa, bilgisayar kullanılırken de çeşitli zorluklar ile karşılaşmaktadır.

Çünkü bir insanın bir dili tam anlamıyla çevirebilmesi için aynı anda her iki dili de iyi bilmesi gerekir. Ancak bunun maliyeti çok yüksektir. Eğer birden fazla yerde, birden fazla yöne, aynı anda çeviri yapmak gerekirse, bu işlemi bir kişinin yapması mümkün olamaz. Oysa çok iyi tasarlanmış bir çeviri sistemi daha kullanışlı olur. Bu nedenle bilgisayarlı çeviri ile ilgili yoğun araştırmalar yapılmıştır. Çeviri yapılacak dillere göre, dil yapılarının farklı olduğundan dolayı farklı yöntemler önerilmiştir. Tüm bilgisayarlı çeviri çalışmalarının son hedefinde mükemmel, tam yetkin çeviri yapan sistemin tasarlanması bulunmaktadır.

Bir çeviri sisteminin tam yetkin olabilmesi için öncelikle hiçbir insan yardımı olmadan çeviri yapabilmesi, üretilen çevirinin anlaşılır, yani aslına uygun olması ve sistemin kısıtlama olmaksızın tüm konular üzerinde çalışabilmesi gerekmektedir. Görüldüğü gibi tam yetkin çeviri, ne bilgisayar, ne de insan tarafından yapılacak kadar kolaydır. Örneğin Çince'de bir insanın bir gazete okuyabilecek duruma gelmesi için en az orta okulu bitirmesi gerekmektedir. Ancak bu insanın kimya, ya da tıp gibi farklı konularda dergi okuyabilmesi için (burada anlamaktan söz edilmiyor), en az 1-2 yıl eğitim alması gerekmektedir. Bu nedenle, normal bir insanın tek başına gerçekleştiremediği tam yetkin çeviriyi, günümüz teknolojilerinden beklemek doğru olmaz [60]. Tam yetkin çeviri yapılamamasının temel nedenlerinden biri, dillerin kültür ile bağlantılı olmasıdır. Bir dilin konuşulduğu toplumun kültürü, bir başka toplumunki ile aynı olmayabilir. Hatta bu tür olaylar, sadece farklı dillerin konuşulduğu toplumlarda değil, aynı zamanda aynı dili konuşan toplumlarda bile yaşanır. Örneğin, Uygurcada “sili” sözcüğü saygı anlamına gelirken, yine Uygurcayın konuşulduğu Kaşgar şehrinde aynı zamanda sitem anlamında kullanılır. Bunu ancak Kaşgarda yaşayan insanlar fark eder.

Bunun ile birlikte, tam yetkin çeviri sistemi geliştirilemese bile, kısıtlı koşullar ile çalışan sistemler geliştirilebilmiş ve günümüzde kullanılmaktadır. Bazı durumlarda, Google'nin geliştirmiş olduğu çeviri sisteminde olduğu gibi, elde edilen çeviri sadece sonuçları okumakta ya da çevirisi yapılan metnin özetini elde etmekte kullanılmaktadır. Bazı çeviri sistemleri ise, dil üzerinde çeviri yaparken, destek amaçlı olarak kullanılmaktadır. Bu tarz sistemlerde, öncelikle metin çevrilir, sonra da bu metin üzerinde düzeltmeler yapılır. Bu tür sistemlerde, özellikle basit tümcelerin çevirisi yapıldığından işlem hızlanmış olur. Sadece belli konular üzerinde çeviri yapmak amacıyla sistemler de geliştirilmiştir. Bu tür sistemlerin kullanılacağı alanda, sözcükler ve tümce yapısı belli olduğundan dolayı diğer sistemlere göre daha başarılıdır. Örneğin Météo çeviri sistemi İngilizce ile Fransızca arasında sadece hava tahmini bilgilerini çevirmek için geliştirilmiştir [3].

2.1 Bilgisayarlı Çeviride Kullanılan Yöntemler

Bilgisayarlı çeviri için kullanılan yöntemler genel olarak kural tabanlı yöntemler ve derlem tabanlı yöntemler olmak üzere ikiye ayrılır. Kural tabanlı yöntemler insanlar tarafından tanımlanan kurallara göre çeviri işlemleri gerçekleştirirken, derlem tabanlı yöntemler ise çeşitli makineli öğrenim algoritmaları kullanarak çeviri yapma bilgilerini öğrenir.

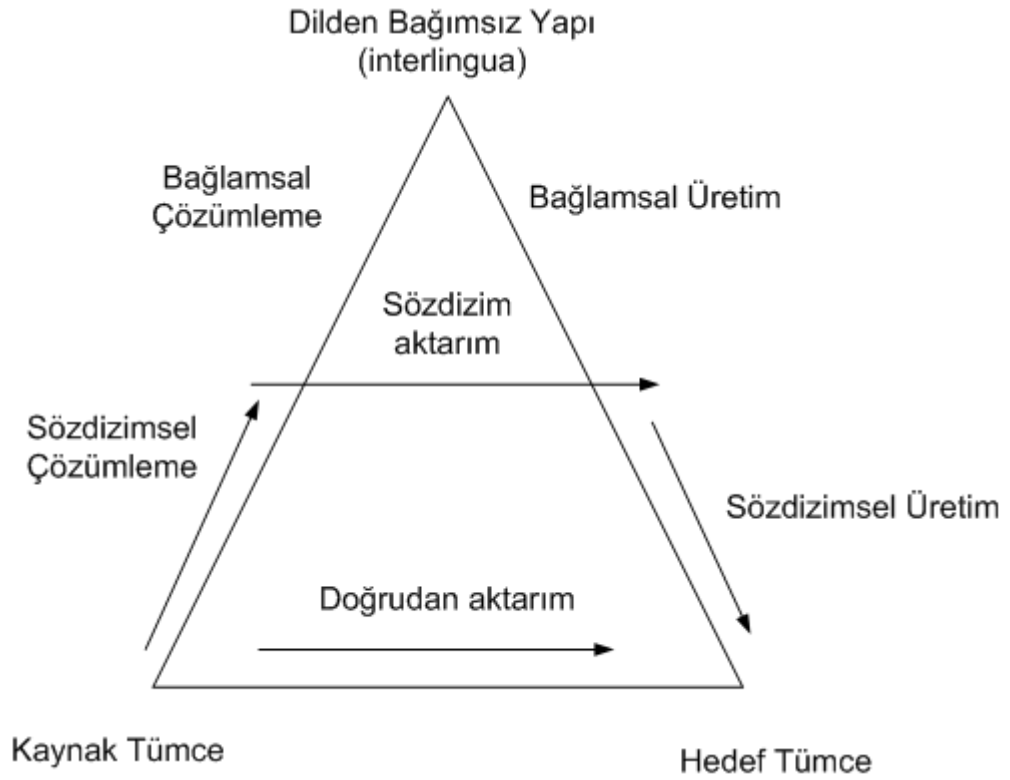
2.1.1 Kural tabanlı çeviri yöntemi

1980 yıllarına kadar geliştirilen bilgisayarlı çeviri sistemlerinin çoğu kural tabanlıdır. Bunlar arasında da en çok kullanılan yöntemler ve onların arasındaki farklar Şekil 2.1'de gösterildiği gibi Vauquous üçgeni ile özetlenebilir.

2.1.1.1 Doğrudan aktarma yöntemi

Bu yöntem bilgisayarlı çeviri çalışmalarında kullanılan en eski yöntemlerden biridir. Çeviri yapılacak kaynak sözcükler üzerinde, çevirilecek dillere göre, biçimbilimsel analiz yaptıktan sonra, fazladan ara işlemler gerektirmeden hedef dile aktarılır. Aktarma işlemleri için sözcükler kullanılır. Bu sistemin iyi yönü geliştirilmesi kolay olmasıdır. Olumsuz yönü ise, sözcük aktarımında oluşan belirsizlik sorunun çözümlenememesidir. Örneğin Uygurcada “at” sözcüğü Türkçede “at”, “ad”, “fırlat” gibi üç farklı anlama karşılık gelmektedir. Bu sözcüklerin bazıları sıfat, bazıları

eylem ve bazıları ise isimdir. Dolayısıyla farklı anlamıyla çevirildiğinde tümcenin anlamı da değişir. Bir başka olumsuz yönü de, çeviri sonucu oluşturulan sözcükleri doğru sıralayamamasıdır (söz dizimi farklı diller için). Doğrudan çeviri yöntemi çalışırken, kaynak dildeki bir sözcüğün doğru bir şekilde hedef dile aktarılabilmesi için, hedef dildeki sözcükten bir önce ya da bir sonra gelen sözcükleri de kontrol edilir. Ancak hedef sözcükten bir önce ya da bir sonraki sözcükte belirsizlik olursa, sistemin başarısı da düşük olur. Doğrudan çevirme yöntemi ile geliştirilen en ünlü çeviri sistemi ise SYSTRAN [61-62] dır.



Şekil 2.1 : Kural tabanlı çeviri yöntemleri- Vauquous üçgeni.

2.1.1.2 Sözdizimsel aktarma yöntemi

Bu yöntem kaynak dil ile hedef dilin sözdizimine göre çalışır. Kaynak dildeki bir tümce hedef dile aktarılmadan önce, hedef dilde aynı anlamı verecek sözdizimi yapı şeklinde çözümlenir ve aktarılır. Sonra doğrudan aktarma yönteminde olduğu gibi bir aktarım sözlüğü kullanılır ve hedef dil çevirisi üretir. Yapı aktarma yöntemine göre çalışan kural tabanlı yöntemler 1980’li yılların sonuna kadar yaygın olarak kullanılmıştır. Bu yöntemle çok sayıda çeviri sistemi geliştirilmiştir. Yapı aktarma yöntemiyle geliştirilen sistemlere örnek olarak SUSY [63], GETA [64] verilebilir. Ancak bu yöntem sözdizimsel ayrıştırma yapabilmek için, bu işlemleri yapan

kişilerin hedef ve kaynak dilinin cümle yapısını, sözdizimini çok iyi bilmelerini zorunlu kılmaktadır. Bu da fazla insan emeği demektir. Genelde yapı aktarma yöntemi birbirine yakın diller üzerinde kullanıldığında başarı daha yüksek olmaktadır [65]. Bunun nedeni ise sözdiziminin aynı olması, söz ve sözcüklerin aynı olmasıdır. Bundan dolayı, çeviri kuralı yazarken ya da tasarlarırken daha kolaylık sağlamaktadır. Birbirinden farklı diller arasında çeviri yaparken, oluşabilecek sözdizimsel karmaşıklık azaltmak için aktarılabilecek sözdizimsel yapılara anlamlar ekleme yöntemi kullanılmıştır ve çeviri sistemleri geliştirilmiştir [66-67].

2.1.1.3 Dilden bağımsız çeviri yöntemi

Bazen birden fazla dil arasında aynı anda çeviri yapmak gerekmektedir. Örneğin Avrupa Birliği'nde birçok dosyanın aynı anda üye ülkelerin dillerine çevrilmesi gerekmektedir. Her dil arasında çeviri yapabilen bir sistemin geliştirilmesi çok zor bir işlemdir. Bunun için dilden bağımsız anlamsal yöntem en uygun çözümdür. Dilden bağımsız çeviri sisteminin çalışma yönteminde, öncelikle çevirilecek kaynak dil, tüm dili temsil edebilecek bir ara dile, yani temsili bir dile çevrilir. Bu ara dilden, çevirilecek diğer dillere çeviri metinleri oluşturulur. Genelde ara dil bir bilgi tabanlı sistemden oluşmaktadır. Dolayısıyla kaynak dillerin hepsinin bu uzman sistem tarafından tanınması ve yorumlanması gerekmektedir. Bu sistemin avantajı, çeviri yaparken işlem yükünü hafifletmesidir. Çünkü tüm dillerden bağımsız bir çeviri sistemi geliştirilmiştir. Bu sisteme yeni bir dil eklemek için, dilden bağımsız olarak, geliştirilen dile göre çeviri sistemi oluşturmak yeterlidir. Çünkü diğer dillerde dilden bağımsız olarak geliştirilen dile göre çeviri yaptığından aynı anda diğer diller de çeviri yapabilmektedir. Şu an geliştirilmiş dilden bağımsız bir dil, yani bir aradilin olmaması ve tüm dilleri kapsayacak bir ara dil geliştirmenin zorluğu bu sistemin zayıf yönüdür. Çünkü tüm dillerin yapısı ve anlatım biçimi tamamıyla farklıdır. Bazı dillerde geçen söz ya da sözcüklerin karşılığı diğer dilde hiç yoktur. Geliştirilen ara dilde tüm bu problemlerin çözülmesi gerekmektedir. Genelde tam geliştirilmiş dilden bağımsız bir dilin tasarlanması imkânsızdır. Geliştirilmiş bir sisteme yeni bir dilin eklenmesi sadece aradil ile geliştirilecek dil arasında çeviri yapmaktan ibaret değildir. Bu sisteme dâhil edilmiş tüm dillerin yapılarını da hesaba katmak gerekmektedir. Dilden bağımsız anlamsal çeviri yöntemleri ile geliştirilen çeviri sistemlerine örnek olarak Rosetta [68], KBMT [69] gösterilebilir.

2.1.2 Derlem tabanlı çeviri yöntemi

1980’li yıllarının sonuna doğru, araştırmacılar, kural tabanlı çeviri sisteminin, doğal dil çalışmaları ile ilgili bazı problemler karşısında çözümsüz kaldığının farkına varmışlardır. Üstelik yüksek hızlı ve yüksek bellekli bilgisayarların üretilmesi ve elektronik ortamdaki verilerin çoğalmasıyla doğal dil araştırmaları derlem tabanlı çalışmalara yönelmiş ve buna paralel olarak derlem tabanlı çeviri sistemleri üzerinde çalışmalar yapılmıştır. Derlem tabanlı çeviri sistemleri genel olarak istatistiksel çeviri sistemleri ve örnek tabanlı çeviri sistemi diye ikiye ayrılır.

2.1.2.1 İstatistiksel çeviri yöntemi

Bu yöntem, daha önceden karşılıklı olarak çevirilen derlemler kullanarak, hedef ve kaynak dillerde söz ve sözcüklerin eşleştirilmesini istatistiksel hesaplamalar yoluyla çeviri yapma yöntemidir [70]. Bundan dolayı çok sayıda çeviri tümceleri içeren büyük bir derlem olmadan istatistiksel çeviriden kaliteli çeviri üretilmesi zordur [71-72]. Candide çeviri sisteminde Fransızca ile İngilizce arasında çeviri yapmada üç milyon çift tümce içeren Hansard derlemi kullanılmıştır [14-15]. Türk dilleri için geliştirilen istatistiksel çeviri sistemi Türkmence ile Türkçe arsında geliştirilmiştir² [32]. Bu sistemde bir eğitim derlemi kullanılmıştır. İstatistiksel yöntemin iyi yönü ise, sistem geliştiricilerin dilbilgisi ile uğraşmamasıdır. Çünkü sistemde herhangi bir kural tanımlanmaz, sadece söz ve sözcüklerin kullanım sıklığı istatistiksel bilgilere göre karar verilir. İstatistiksel çeviri yönteminin kalitesi derlem ile doğrudan ilgilidir. Ancak, Türkiye Türkçesi dışındaki Türk dillerinde derlem bulunmamaktadır ya da geliştirilme aşamasındadır [40]. Bu nedenle, Uygurcadan Türkçeye geliştirilecek çeviri sisteminde kural tabanlı çeviri yöntemi seçilmiştir.

2.1.2.2 Örnek tabanlı çeviri yöntemi

Bu yöntem benzeri sözcük dizimlerine dayanarak çeviri yapma yöntemidir [22]. Çeviri yaparken kaynak tümceler sözcük gruplarına ayrılır. Bu şekilde, genel olarak hedef ve kaynak dilinde birebir karşılıklı sözcük gruplarının çevirisinden oluşan derlemler geliştirilmiş olmaktadır. Yani tümceler üzerine çeviri yaparken, öncelikle kaynak dili derleminde bulunabilecek sözcük grubu seçilir.

² Bu sistemde ayrıca, değişik kurallarda tanımlanmıştır. Bu nedenle Türkmencenden Türkçeye çeviri için geliştiren bu yöntem “karma yöntem” adı verilmiştir [32].

Bu sözcük grubu derlemde bulunduğuna göre bu sözcüklerin hedef dilde çevirisi var demektir. Eğer sözcük grubu aranırken, tam benzeyen sözcük grubu bulunamazsa, buna neden olan sözcükle derlemde daha önce eklenen sözcükler arasında benzerlik karşılaştırılması yapılır. Bu nedenle, soruna neden olan sözcüğe en çok benzeyen sözcük grubu kullanılır. Bu durum, bir söz birden fazla anlam içerdiğinde yaşanmaktadır. Bu yöntemin iyi bir avantajı ise, derleme yeni sözcük grubunun eklenmesidir. Yani bir çeşit makineli öğrenme yöntemi olmasıdır. Bu yöntem derlem tabanlı yöntemler arasında, istatistiksel yöntemden sonra en yaygın kullanılan yöntemdir ve birçok uygulama geliştirilmiştir [73-75].

2.2 Yakın Diller Arasında Bilgisayarlı Çeviri

Dillerin yakınlığı ifadesiyle, öncelikli olarak dillerin söz dizimin birbirine belli bir oranda benzemesi anlaşılmalıdır. Örneğin Çince ile Japoncanın ortak kullanılan çok sayıda karakteri³ olmasına karşın, bu iki dil yakın dil olarak kabul edilemez. Çünkü söz dizimleri birbirinden tamamen farklıdır. Ayrıca Japoncanın abecesi⁴, vardır ve söz dizimi Türk dilinin söz dizimine benzer. Bu nedenle, Japonca Türk dillerine göre yakın dil olarak kabul edilir. Dolayısıyla yakın diller arasında çeviri sistemi tasarlarken, dillerin sözdizimi üzerine araştırma yapmaya gerek yoktur, sadece yapısal biçimde araştırılması yeterlidir [22]. Eğer geliştirilecek çeviri sisteminde, kural tabanlı sistem kullanılırsa, tanımlanması gereken kural sayısı, farklı diller için gerek olandan daha az ve basit olur [22], [29], [32].

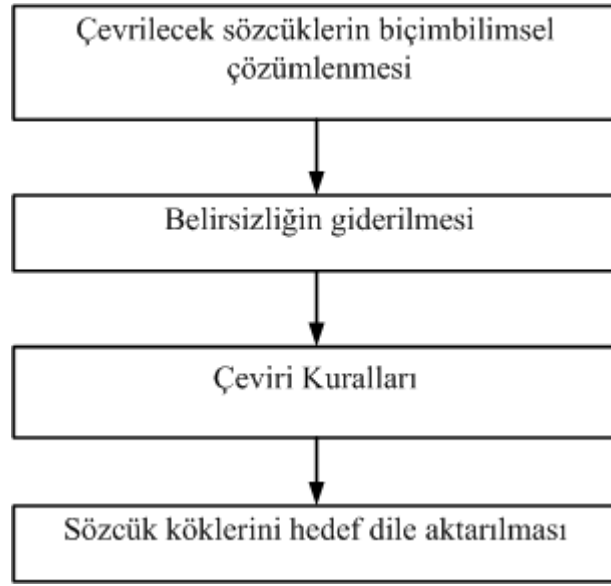
Yakın diller arasında kural tabanlı çeviri sistemi geliştirilirken, gerekli çalışma aşamaları Şekil 2.2’de gösterilmiştir. Yakın diller arasındaki çeviri aşamalarının çalışması aşağıdaki bölümlerde örneklerle anlatılmıştır⁵. Uygurca ile Türkçe bitişken dillerdir ve bir sözcük sonsuz ek alabilir. Bu nedenle kural tabanlı çeviri sistemlerinde tarif edilen doğrudan aktarma yöntemi olduğu gibi kullanılamaz.

Genel olarak yakın diller arasındaki çeviri işlemleri aşağıdaki adımlardan oluşur:

³ Japoncada yazılarda eski Çince ile yazılan karakterler (Kanji) bulunmaktadır. Bu karakterler anlam olarak Çin dili ile aynı anlam taşır. Ancak, farklı ses ile okunur.

⁴ Japoncada abecesi bulunan bir dildir. Yazı yazarken, Kanji (Çince karakterler) ile beraber yazılabilir.

⁵ Bu tezde Uygurcadan Türkçeye çeviri sistemi geliştirileceğinden, yakın diller olarak Uygurca ile Türkçeye atıfta bulunulmuştur.



Şekil 2.2 : Yakın diller arasında bilgisayarlı çeviri aşamaları.

2.2.1 Çevrilecek sözcüklerin biçimbilimsel çözümlenmesi

Sadece yakın diller değil, herhangi bir dil üzerinde doğal dil çalışması yaparken, ilk yapılması gereken şey, sözcükler üzerindeki biçimbilimsel çalışmalardır. Bilgisayarlı çeviri yaparken bir sözcüğün doğru aktarılabilmesi için o sözcüğün tüm yönleri ve olası tüm anlamlarıyla çözümlenmesi gerekir. Söz konusu çalışılacak diller bitişken dil olduğunda, biçimbilimsel çözümlemelerde oluşan her bir biçim çok önemlidir. Çünkü bitişken dillerde eklerin eklenmesiyle yeni sözcükler oluşur ve bu oluşan sözcükler kök sözcük ya da eklerin eklenmesinden önceki sözcüklere göre, tamamiyle farklı görevler üstlenir. Eklemeler yoluyla sözcük oluşturulmak istendiğinde, sayısız sözcük söz konusu olabilir [30], [76].

Örneğin, “ishligentuq” sözcüğü 5 adet biçimden oluşur⁶. “iş(isim) + li (eylem, 2-şahıs) + gen (bitmiş şimdiki zaman) + tu (geçmiş zaman) + q (birinci şahıs çoğul eki).

Biçimbilimsel çözümleme ise bir sözcüğün kök kısmından başlayarak, o sözcüğü oluşturan tüm biçimleri çözümleme anlamına gelir. Burada sözcüğün kökü “iş” yani bir isim iken, eklenen eklerin etkisi ile sözcük eyleme dönüşmüştür. Birçok kişinin bir işi beraber yaptığını belirtmektedir. Bundan dolayı, bilgisayarlı çeviri

⁶ Burada Uygurca için tasarlanmış bir biçimbilimsel çözümleyici çalıştırıldığında, Uygurcanın dil bilgisine göre çözümlenmesi gereken biçimler verilmiştir.

geliştirilirken, özellikle bitişken diller için, öncelikle biçimbilimsel çözümleyicinin geliştirilmesi çok önemlidir ve ilk yapılması gereken çalışmalardır [32]. Biçimbilimsel çalışmalar sadece bilgisayarlı çeviri işleminin ön çalışması değil, aynı anda doğal dil çalışmalarının en önemli alanıdır [77-82] . Bu nedenle, biçimbilimsel çözümleyiciler için sürekli yeni yöntemler geliştirilmektedir ve araştırmalar yapılmaktadır. Örneğin Türkçe, Tatarca ve Türkmençe için kural tabanlı ve leksikondan oluşan biçimbilimsel çözümleyiciler tasarlanmıştır [30-31], [33]. Türkçe için geliştirilen bir başka biçimbilimsel çözümleyici leksikon kullanmadan, son eklerin elenmesi yöntemi ile geliştirilmiştir [83].

2.2.2 Belirsizliklerin giderilmesi

Biçimbilimsel çözümleyici bir sözcük ile çözümleme yaptığında, olası tüm bilgileri çözümler. Birçok dilde, genelde bir sözcük aynı anda birden fazla olaylar için kullanılmasından dolayı, bazı sözcükler birden fazla anlam taşır. Örneğin, Uygurcada “at” sözcüğü, üç farklı anlam taşır [84-85]. Birincisi “hayvan” olan at. İkincisi “atmak” eyleminin emir biçimidir. Üçüncü anlamı ise “isim” dir. Yani bir insana verilmiş isimdir. Bu nedenle her hangi bir Uygurca tümcede yer alan “at” sözcüğü Türkçeye çevrilirken, doğrudan “at” ya da “ad” olarak çevrilebilir. Dolayısıyla burada “at” sözcüğü için belirsizlik söz konusudur. Genelde belirsizlikler, biçimbilimsel belirsizlik ve anlamsal belirsizlik olmak üzere iki şekilde yorumlanır. Her iki çeşit belirsizliği gidermek için kural tabanlı ve derlem tabanlı çözümleme yöntemleri geliştirilmiştir.

Sonuç olarak belirsizlik giderme çalışmaları, diller birbirine yakın olsa bile, çalışılması gereken önemli konular arasında gelir. Bu nedenle belirsizlik giderme konusunda ciddi çalışmalar yapılmıştır. Ancak tüm belirsizlikleri çözümleyen tam yetkin çözümleyiciler henüz geliştirilememiştir [86-88]. Bu tez çalışmasında, biçimbilimsel ve anlamsal belirsizlikleri gidermek için kural tabanlı yöntemler kullanılmış ve tezin 5. bölümünde kapsamlı olarak anlatılmıştır.

2.2.3 Çeviri kuralları

Diller birbirine yakın ve söz dizimleri birbirine benzese de, hedef dilde aynı anlam veren tümce üretebilmek için, kök sözcüklerin tümcedeki sırasının değiştirilmemesi bazen eklerinin yeri değiştirilmesi gerekir. Çünkü bitişken dillerde eklerin eklenmesi ile sözcüklerin anlamı ve aynı anda tümcelerdeki görevi de değişir. Örneğin Türkçe

ile Kırım Tatarcası birbirine o kadar yakın olmasına karşın, sözcüklere eklenen ekler farklı olabilir. Bundan dolayı Türkçeden Kırım Tatarcasına çeviri sistemi geliştirilirken farklı düzeyde çeviri kuralları tanımlanmıştır [29]. Türk dilleri ailesinde bulunan ve en son geliştirilen bir başka çeviri sistemi Türkmençe ile Türkçe arasında geliştirilmiştir. Bu sistemde kök aktarmadan önce ve sonra çalışan çeviri kuralları tanımlanmıştır [32]. Çeviri kuralları her sistem için zorunlu değildir. Örneğin istatistiksel çeviri sistemlerinde hiç kural tanımlanmaz ve istatistiksel hesaplamalara göre sonuç elde edilir. Ama yüksek başarımlar elde etmek için, istatistiksel yöntemler ile kural tabanlı çeviri sistemleri beraber kullanılabilir [32]. Bu tez kapsamında geliştirilen sistem ile ilgili çeviri kuralları tezin 4.bölümünde anlatılmıştır.

2.2.4 Kök sözcüklerin çevrilmesi ve tümcenin oluşturulması

Kural tabanlı çeviri sistemlerinde, çeviri kuralları tanımlandıktan sonra, en son kaynak sözcüğün kökü hedef sözcüğün köküne çevirilir. Kök sözcüklerin çevrilmesi genelde kök sözcüklerden oluşmuş bir sözlük aracılığı ile yapılır. Bundan dolayı sözcüklerin biçimbilimsel çözümlenmesi yapılırken, kök sözcük kısmına kadar çözümlemek zorunludur. Yani sözcüklerin kökü kesinlikle bulunmak zorundadır. Bundan dolayı bitişken dillerde kök bulma işlemi çok önemlidir ve Türkiye Türkçesinde kök bulmak için farklı çalışmalar yapılmıştır [52], [77]. Hedef dildeki sözcükler oluşturulduktan sonra, Kaynak dildeki sırasına göre yerleştirilerek, hedef dilin tümcesi elde edilir. Ancak hedef dilin yapısına göre, kök aktarmadan sonra, görünen biçimde doğru sonuç almak için, ayrı kuralların tanımlanması gerekebilir. Örneğin Uygurcada bir soru sözcüğü⁷, çözümlendikten sonra tek bir biçimbilimsel çözümleme üretir. Bu çözümleme Türkçeye aktarılırken tek bir çeviri sözcüğü ile aktarılır. Ancak Türkçede görünen düzeyde karşılığı oluşturulurken, Türkçenin dil bilgisine göre soru sözcükleri (mı, mi, mu, mü) bir önceki sözcük ile ses uyumu yapmak zorundadır ve bunun için ayrı kurallar tanımlanması gerekir [32].

⁷ Uygurcada soru ek bir önceki sözcüğe eklenerek yazılır ancak Türkçede ayrı yazılır. Bu neden ile soru sözcükleri için aktarma kuralları tanımlanmıştır. Örnekler için 4. bölüme bakılması tavsiye edilir.

2.3 Biçimbilimsel Çözümleme

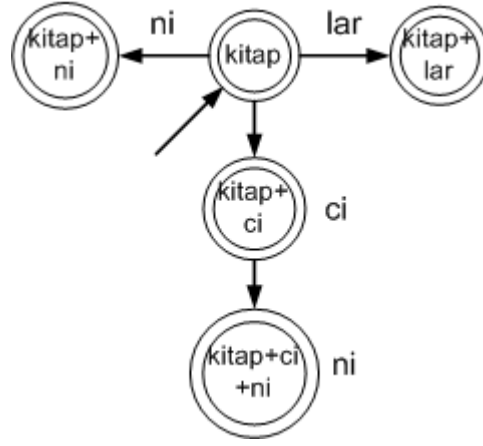
Biçimbilimsel çözümleme, bir sözcüğün tüm ayrıntılarıyla araştırılması anlamına gelir. Örneğin bir sözcüğün kökü, bu kökün dâhil olduğu kategori, aldığı ek veya ekler, eklerin eklenmesi ile sözcük türünde oluşan yeni kategori gibi. Eğer bu sözcüğün aynı anda birden fazla anlamı varsa, söz konusu anlamların her biri ile ilgili biçimbilimsel çözümlemelerin üretilmesi gerekir. Biçimbilimsel çözümleme yapılırken, yani bir sözcükle ilgili oluşturucu özellikler çözümlenirken, o dilde tanımlı sözcükler ve eklere göre işlem yapılır. İşlem yapılacak dil bitişken dil olursa, eklenebileceği ekler sonsuz sayıda olduğundan, bir sözcük ile üretililecek bilgiler de sonsuz olur. Bilimsel çalışmalarda sözcüklerle ilgili bilgi taşıyan alt çözümlemelere “biçem” adı verilmektedir. Biçemleri göstermek için birbirinden farklı etiketler kullanılır. Türk dilleri için genel olarak Türkiye Türkçesinin biçimbilimsel çözümleyicisi için kullanılan etiketler kullanılmıştır [30].

Bir dil ile ilgili biçimbilimsel çözümlemelerin doğruluğu, geliştirilen biçimbilimsel çözümleyici ile doğrudan ilgilidir. Tüm işlemler o sözcük ile ilgili çözümlenen bilgilere göre çalışır.

Biçimbilimsel çözümleyici tasarımında en yaygın kullanılan yöntem ise sonlu durumlu makinelerdir. Özellikle bitişken diller için, sonlu durumlu makineler en uygun yöntemdir. Çünkü bitişken dillerde, sözcükler herhangi bir ek eklediğinde sözcüğün anlamı da ona uygun olarak değişir. Eğer geçersiz bir ekleme yapılırsa anlamsız bir sözcük oluşur. Ekleme yoluyla bir başka sözcük ya da yapı oluşturulması, sonlu durumlu makinelerin çalışma yöntemine benzer. Çünkü sonlu durumlu makinelerde giriş bilgilerine⁸ göre bir durumdan bir başka duruma geçiş yapılabilmektedir. Eğer geçiş sonucu, son duruma ulaşabilirse, o zaman girilen bilgiler bu makineler tarafından tanımlanmış sayılır, yani bu makineler için geçerli sayılır. Eğer makineler bir durumdan bir başka duruma geçer, son duruma ulaşamazsa, o zaman yanlış geçiş yapılmış kabul edilir ve buna neden olan bilgi otomatlar tarafından tanınmaz.

⁸ Bazı kaynaklarda belli bir abece üzerinde tanımlanmış giriş bilgisi veya parametreler olarak geçmektedir.

Örneğin: “kitap” sözcüğü bir isimdir, kök sözcüktür. Eğer “kitap” sözcüğüne “ci” eki eklenirse, bu sefer yeni bir isim oluşmuş olur ve “kitapçı” sözcüğü oluşur. Bu sözcük Uygurcada geçerli bir sözcüktür. Böylece bir geçerli durum (kitap) dan, yeni bir geçerli sonlu duruma (kitapçı) geçiş sağlanmış olunur. Eğer makineler “kitap” durumundayken, “lar” eki geldiğinde, bir başka geçerli duruma ulaşılır. Makineler “kitap+ci” durumundayken, “ni” ekinin eklenmesiyle, “kitap+ci+ni” durumuna geçer. Ancak bu durumdayken “lar” eki eklenmeye çalışıldığında makineler kabul etmez. Yani yeni bir duruma geçişe izin verilmez. Çünkü Uygurcada “ni” ile sonlanan sözcüklerden sonra başka ek gelmez. Ayrıca eklerin ekleme sırası da önemlidir ve Şekil 2.3’de gösterilmiştir. Örneğin “kitap” sözcüğüne “ci” sözcüğü eklenebilir. Ancak kitap+lar +ci şeklinde bir sözcük oluşturulamaz.



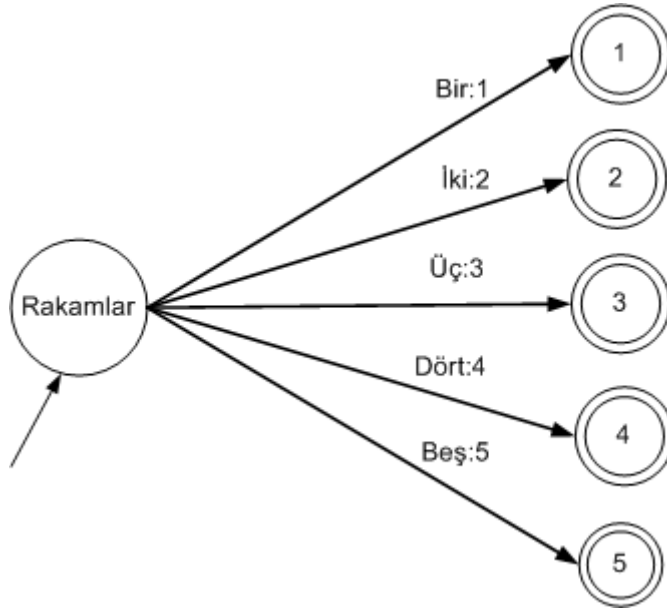
Şekil 2.3 : Sonlu durumlu makinelerin eklere göre çalışması.

Bu nedenle sonlu durumlu makineler doğal dil çalışmaları için çok faydalıdır ve bu alanda yaygın olarak kullanılmaktadır [89]. Sonlu durumlu makineler, doğal dil çalışmalarında en yaygın olarak durum dönüştürmede ve iki düzeyli biçimbilimsel çözümleyici tasarımlarında kullanılmıştır [32].

2.3.1 Sonlu durumlu dönüştürücüler

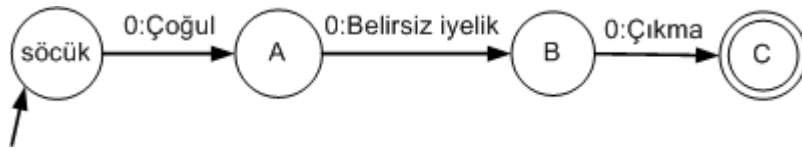
Sonlu durumlu makinelerde, bir durumdan bir başka duruma geçişleri o anki durumda girilen giriş bilgilerine göre değişir. Geçerli bir giriş bilgisi girildiğinde makineler son duruma ulaşır. Eğer durum geçişleri belli bir kurala göre gerçekleştirilirse ve son duruma ulaşabilirse, elde edilen çıktılar geçerli sonuç olarak kabul edilir [32]. Örneğin, harf ile yazılan sayıları rakamlara çeviren sonlu durumlu dönüştürücü Şekil 2.4’de gösterildiği gibi çalışır.

Bu makinelerde geiş kuralı bařlangı durumunda “bir”⁹ yazısı girildiğinde sayısal olarak “1” e dnüşürme kuralına göre alıřmaktadır. řekil 2.4’de gsterilen SDD’da geerli szcükler ise, “bir, iki, üç, drt ve beř” tir.



řekil 2.4 : Sonlu durumlu dnüşürücüler.

Bu szcükler dıřında her hangi bir giriř yapıldığında sonlu SDD da herhangi bir geiş saėlanmaz. Bu rnekten yararlanılarak, bir dilde geen szcükler iin SDD tanımlanabilir. Örneėin, “mekteplerden” (okullardan) szcüėü ile ilgili biçimbilimsel bilgiler: bu szcüėün kökü, “mektep”, ait olduėu szcük kategori “isim”, çoėul eki alır, iyelik eki almaz ve ıkma durum eki taşır. Bu bilgiler bir insana verildiğinde, “mekteplerden” szcüėünü oluřturabilir. Aynı iřlem SDD üzerine řekil 2.5’de gsterildiėi gibi tasarlanabilir.



řekil 2.5 : Biimbilimsel dnüşürücü.

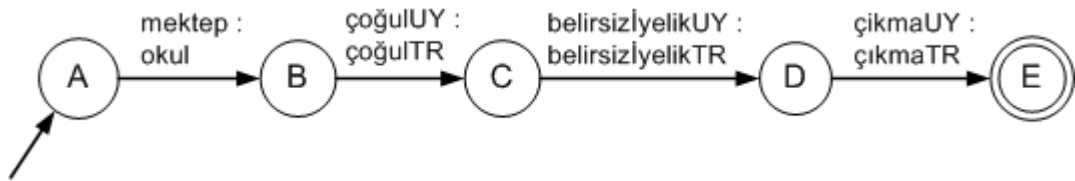
Bu biçimbilimsel dnüşürücü, “evlerden, daėlardan, baėlardan” gibi yapılarıdaki szcükle ilgili biçimbilimsel bilgileri, bu szcüklere dnüştürebilmeklerdir.

⁹ Bu rnekten, “bir” szcüėü tek bir karakter gibi tanımlanış özel karakter olarak kabul edilmiştir.

Şekil 2.5’de gösterilen “0:çoğul” anlamı ise, bu sözcükten sonra, “çoğul” eki¹⁰ alır anlamı verir.

2.3.2 İki düzeyli biçimbilimsel çözümleme

Bir sözcük ile ilgili biçimbilimsel bilgiler edinildikten sonra, o dil için tanımlanan biçimbilimsel dönüştürücüler kullanılarak, sözcükler oluşturulabilir. Bu nedenle biçimbilimsel dönüştürücüler bilgisayarlı çeviride çok önemlidir. Çünkü kaynak dilde, bir sözcük için biçimbilimsel çözümleme yapıldıktan sonra, hedef dil için geliştirilen biçimbilimsel dönüştürücüden geçirildiğinde hedef dilde çeviri sözcüğü oluşur. Bu nedenle öncelikle kaynak dilde sözcüklerin biçimbilimsel bilgisinin elde edilmesi gerekir. Örneğin Uygurcada bulunan “mekteplerden”¹¹ sözcüğünün Türkçeye çevirmek için geliştirilen sonlu durumlu dönüştürücü Şekil 2.6’de gösterildiği gibi tasarlanabilir.



Şekil 2.6 : Uygurcadan Türkçeye sonlu durumlu dönüştürücü.

Şekil 2.6’da gösterilen sonlu durumlu makinelerde, makinelerin başlangıç durumu A’ya Uygurca sözcük “mektep” sözcüğü girilirken, Türkçe karşılığı olan “okul” sözcüğüne dönüştürülür¹². Aynı sıra ile “mektep” sözcüğü ile ilgili biçemlerde Türkçe karşılığına dönüştürülür. Son durum E’ye geldiğinde “okullardan” sözcüğü oluşturulur.

Biçimbilimsel çözümleme için, özellikle bitişken diller için en yaygın olarak kullanılan yöntem, sonlu durumlu makineler temelinde kurulan iki düzeyli yöntemdir [76], [90-92]. Bu yöntemle, Japonca, İngilizce, Korece, Fransızca ve Rumence gibi birçok dilin biçimbilimsel çözümleyicileri geliştirilmiştir [93-97].

Türk dilleri için iki düzeyli biçimbilimsel çözümleyiciler, Türkçe, Türkmence ve Kırım Tatarcası için geliştirilmiştir [30-31], [33]. Genel olarak tüm bitişken dillerin

¹⁰ Uygurcada çoğul ekleri “lar/ler” dir. Burada “çoğul” sözcüğü her iki çeşit ek için geçerli olarak kullanılmıştır.

¹¹ Biçimbilimsel bilgilerin anlamı bir önceki örnekte anlatıldığından burada tekrar verilmemiştir.

¹² Burada Uygurca- Türkçe kök sözcüğü üzerinden aktarma yapılır.

biçimbilimsel çözümleyicisi aynı yönteme göre çalışır. Onları farklı kılan tek şey sözcükleri üreten kurallardır. İki düzeyli biçimbilimsel doğal dil çalışmaları, özellikle bitişken diller üzerinde bilgisayarlı çeviri işlemlerinde yaygın kullanmalarının temel nedeni iki yönlü çalışmasıdır. Bir sözcük normal yazıda kullanılan biçimde yani görünen biçim (surface representation) de çözümleyiciye gönderilirken, bu sözcük ile ilgili tüm biçimbilimsel özellikleri veren yapısal biçimler (lexical representation) oluşturmaktadır. Aynı anda, sözcüklerle ilgili yapısal biçimler girildiğinde, o sözcüklerin görünen biçimi oluşturulabilmektedir. Bu nedenle iki düzeyli çözümleyici olarak adlandırılır. İki düzeyli biçimbilimsel yöntemi ise yazılım ve bitleştirme kuralları diye iki bileşenden oluşur.

2.3.2.1 Yazım kuralları

Yazım kuralları, iki düzeyli biçimbilimsel yöntemin çalışmasını kolaylaştıran en önemli bileşendir. Çünkü bu bileşenle bir dilin yazım kuralları tanımlanabilir. Sonra geliştirilen biçimbilimsel çözümleyici kurallara göre sözcükler ile ilgili yapısal bilgilerden görünen biçim oluşturulur. Uygurca ve Türkçe gibi bitişken diller için, kök sözcüklere veya bir önceki sözcüklere eklerin nasıl bir kurala uyması gerektiği yazım kuralları ile tanımlanır. Örneğin Uygurcada “kitap” sözcüğüne, çoğul eki “lar” eklenirken, “qelem” sözcüğüne ise “ler” eki getirilir. Yani çoğul anlamında bir sözcük oluşturmak için, sözcüklere “lar” ya da “ler” çekim eki eklenir. Ses uyumuna genelde, sözcüklerin son hecesinde yer alan ünlü harflere bakılarak karar verilir. Bir başka örnek ise Uygurcada “men” sözcüğüne iyelik eki “niN” eklenirken, Türkçe anlamı ise “ben+im”, “men” sözcüğünde yer alan “e ” ünlü harfi ise, ”é” harfiyle değişmektedir. Bu çeşit değişimler genelde o sözcüklerin ön ve arkalarına bakılarak, belli kurallar çerçevesinde gerçekleşir. Örneğin Uygurcada bulunan çoğul ekleri lar/ler ile ilgili olarak ne zaman “lar” ne zaman “ler” ekleneceğine karar vermek için öncelikle bu eklerde değişen harfe bakılır. Burada “a/e” harfinde değişme görülmektedir. Sonlu durumlu makinelerde bu değişme “A:a” ve “A:e” ile gösterilebilir. Böylece Uygurcada tanımlanan çoğul eki tek ek “lAr” ile değiştirilmiş olur. Bu tek tek tanımlandıktan sonra, “A” harfinin “a” ve “e” ye değişmesi için kurallar tanımlanır. Bu kurallar da bir dilin yazım kurallarına göre tanımlanır. Bu nedenle Uygurcada “laf” ve “leb” eklenecek sözcükleri iki türe ayırmaktan ve tek-tek listesini oluşturmaktansa, tek “lAr” eki kullanılarak genel bir çoğul eki ekleme kuralı oluşturulabilir.

İki düzeyli biçimbilimsel çözümleyicilerde harflerin eklenmesi, silinmesi veya değişmesi gibi işlemlerin yapılması için, dört farklı kural tanıma yöntemi vardır ve bu yöntemler “=>”, “<=”, “<=>”, “/<=” ile belirtilir [76]. Bu yöntemlerin anlamı ve çalışma prensipleri Çizelge 2,1’de örnekle verilmiştir.

Çizelgede yer alan 1. kural “=>” şöyle çalışmaktadır: “l” ve “r” harfleri arasında eğer “a” harfi girilirse, o zaman “a” harfi “b” harfine dönüşür. Ama bu değişim her zaman zorunlu değildir. Bu kural için tanımlanan örnek I’de, giriş metni olarak, yapısal biçimde “lar” sözcüğü verilmiştir. Yani “a” harfi “l” ile “b” harfleri arasında gelmiştir. Değişme şartı sağlanmıştır. Bundan dolayı, sonuç olarak, görsel biçimde “lbr” üretilir.

Çizelge 2.1 : İki düzenli biçimbilimsel kurallar ve örnekler.

Kurallar- Bağlamlar		Örnek bağlamlar ve kuralların uygulandıktan sonraki sonuçları				
		Biçimler	Örnek I	Örnek II	Örnek III	Örnek IV
1	a:b => l _ r	Yapısal	lar	lar	lbr	xay
		Görsel	lbr	lar	lbr	xby (olamaz)
2	a:b <= l _ r	Yapısal	lar	lar	lbr	xay
		Görsel	lbr	lar(olamaz)	lbr	xby
3	a:b <=> l _ r	Yapısal	lar	lar	lbr	xay
		Görsel	lbr	lar (olamaz)	lbr	xby (olamaz)
4	a:b / <= l _ r	Yapısal	lar	lar	lbr	xay
		Görsel	lbr (olamaz)	lar	lbr	xby

Örnek II’de ise, giriş sözcüğü olarak, yapısal biçimde “lar” verilmiş, ama sonuç olarak, görsel biçimde “lar” üretmiştir. Yani bu kuralda, koşul sağlansa bile kesin olarak “a”nın “b” olarak değişmesi zorunlu olmadığından dolayı değişme olmamıştır. Dolayısıyla “lar” ise geçerli sonuçtur.

Örnek III’de ise, giriş sözcüğü olarak yapısal biçimde “lbr” verilmiştir. Bu kuralda değişmesi gereken harf ise “a”dır. Ama burada “b” gelmiş olup, tümcede değişmesi gereken “a” harfi yoktur. Dolayısıyla değişim olmaz. Bundan dolayı girilen sözcük olduğu gibi görsel biçimde görünür.

Örnek IV’de ise, giriş sözcüğü yapısal biçimde, “xay” verilmiştir. Tanımlanan kuralda ise, “a” nın “b” ile değişmesi için, “a” kesin olarak “l” ve “r” harfi arasında gelmek zorundaydı. Dolayısıyla “xar” girildiğinde, görsel biçimde “xbr” sonucunun oluşması imkansızdır. Dolayısıyla geçersizdir.

İkinci kural “<=>” nin anlamı ise, “l” ve “r” harfi arasında eğer “a” harfi girilirse, o zaman “a” harfi “b” harfine dönüşür ve değişmesi zorunludur. Koşul sağlanmasa da dönüşebilir. Bu nedenle bu kural 1. kuraldan farklı olarak örnek II de yapısal biçimde “lar” girişi girildiğinde, görünen biçimde “lar” biçimi üretilemez. Örnek IV de ise “xay” den “xby” üretilmiştir ve geçerlidir.

Üçüncü kural “<=>” nin anlamı ise, “l” ve “r” harfleri arasında eğer “a” harfi girilirse, o zaman “a” harfi “b” harfine kesin olarak dönüşür, değilse kesin olarak değişmez. Bu kural aslında birinci ve ikinci kuralların birleşmesidir. Bu nedenle Örnek II ve IV deki sonuçlar üretilemez, yani geçersizdir.

Dördüncü kural “/<=>” nin anlamı ise, “l” ve “r” harfleri arasında eğer “a” harfi girilirse, o zaman “a” harfi “b” harfine kesin olarak dönüşmez. Bu nedenle Örnek I de olduğu gibi değişim gerçekleşmez.

2.3.2.2 Bitiştirme kuralları

Bitişme kuralları bir dilde bir sözcüğün nasıl oluşturulacağı ile ilgili kuralları içerir. Örneğin bir sözcüğe hangi eklerin ekleneceği, eklerin eklenme sırası ve sözcük türlerine göre eklenecek ekler gibi kurallar tanımlanır ve sonlu durumlu makineler üzerinde gerçekleştirilebilir. Örneğin Şekil 2.3’de isim kategorisinde bulunan sözcükler için, “çoğul, belirtme durum” eklerinin eklenmesiyle ilgili basit bir makine olarak kabul edilebilir. Mekinede gösterildiği gibi, Uygurcada bir sözcüğe belirtme durum eki “ni” eklendikten sonra çoğul eki eklenmez. Eğer bir çoğul eki eklenecekse, önce “çoğul” eki sonra “belirtme” eki eklenir. Yani ekleme sırası “sözcük + çoğul+belirtme eki” sırasına göre çalışır. Söz konusu diller bitişken dil olduğunda, bitişme kuralları daha çok önem kazanır. Genel olarak bitiştirme kuralları, sözcüklerin kök sözcükleri ve ekleri birleştiren bir sonlu durumlu makineler üzerinde gerçekleştirilir. Makinelerdeki durum geçişleri ise sözcüğü oluşturan eklere göre gerçekleşir ve hangi alt sözlükten (lexicon) devam edileceği belirtilir [98-100]. Bu tez çalışmasında biçimbilimsel çözümleyici geliştirmek için XEROX nin geliştirdiği araçlar kullanılmıştır. Bu nedenle XEROX kullanarak

Uygurca için geliştirilen sözcük yapısı Şekil 2.7’de verilmiştir¹³. Bu sözcük yapısına göre Uygurcada isimler ile ilgili dizimi şöyle gerçekleştirir: bir sözcük bu makinelerin başlangıç durumu “Root” dan okunur .Eğer isim ise “Nouns” alt sözlüğünde yer alır. Sonra bu sözcük “+Noun” olarak etiketlenir. Bu sözcük isim olduğundan çoğul eki alabilir ya da almaz. Sözcük durumuna göre “+A3pl/+A3sg” ile etiketlenir. İsimler çoğul eklerinden sonra iyelik ekleri alabilir . Eğer iyelik eki almazsa “+Pnon” ile belirtilir. İyelik eklerinden sonra durum eki alabilir. Örneğin, belirtme eki “ni” eklendiğinde “+Acc” ile etiketlenir ve son duruma ulaşır “#” .

```

LEXICON Root

NOUNS;
PROPER-NOUNS;
ADJECTIVES;
ADVERBS ;
CONNECTIVES;

LEXICON NOUNS

kitap  NOUN-POST  ;
qelem  NOUN-POST  ;

LEXICON NOUN-POST

+Noun : 0          NOUN-NOMINAL-ROOT;

LEXICON NOUN-NOMINAL-ROOT

+A3pl:IAr NOUN-PLURAL;
+A3sg:0   NOUN-PLURAL;

LEXICON NOUN-PLURAL

+Pnon:0      NOUN-POSSESSIVE ;
+P1sg:Hm     NOUN-POSSESSIVE ;

LEXICON NOUN-CASE-SUSSIXES

+Nom:0       CASE      ;
              ACCUSATIVE ;
              DATIVE   ;

LEXICON ACCUSATIVE

+Acc:ni      #;

```

Şekil 2.7 : Xerox Lexc için sözcük yapısı.

¹³ Burada örnek olarak açıklama yapmak için kısıtlamıştır ve sadece sözcük türleri ve durum ekleri ile örnekler verilmiştir.

Yukarıda anlatılan adımlardan Uyurca “kitaplarını” sözcüğü geçirildiğinde aşğıdaki biçimbilimsel çözümlemeler üretilir¹⁴:

kitaplarını: kitap+Noun+A3pl+Pnon+Acc

Genelde iki düzeyli biçimbilimsel çözümleyiciler çift yönlü, çözümleyici ve üretici olarak çalıştırılabilir. Bu nedenle bir sözcük ile ilgili biçimbilimsel çözümlemelerden, o sözcüklerin görünen biçimi oluşturulabilir.

2.4 Çeviri Sonucunun Değerlendirilmesi

Bilgisayarlı çeviri çalışmalarında, en zor işlemlerin biri, geliştirilen sistemin (yöntem veya algoritmaların) değerlendirilmesidir. Bilimsel çalışmaların tüm alanında, geliştirilen yeni bir yöntemden bahsedilirken, geliştirilen yöntemlerin başarımını sayısal sonuçlar ile göstermek ve kanıtlamak gerekmektedir. Bu nedenle, bilgisayarlı çeviri ile yeni bir sistem geliştirildiğinde, sistemin başarısını belli sayısal rakamlar ile göstermek zorunludur. Diller ile ilgili en büyük sorun ise, üzerlerinde matematik ya da fizik gibi doğrudan bir formül ile uygulanamıyor olmasıdır. Dil, kendi doğası gereği belli bir dereceye kadar belirsizdir.

Bu nedenle, doğal dil çalışmalarında doğrudan bir kriter uygulanamaz. Örneğin, bilgisayarlı çeviri ile değerlendirme yaparken, hiçbir zaman en iyi ve tek bir çözüm diye değerlendirme yapılamaz. Doğal olarak kaynak cümleye karşın birden fazla mükemmel çeviri üretilebilir. Çeviri cümleinin kalitesi ise, çeviri için kullanılan sözcüklere göre, hatta bazen aynı sözcükleri kullansa bile, sözcüklerin tümcede sırasına göre değişebilir .

Bazen insan oğlu bile iyi çevirilmemiş tümceler ile iyi çevirilen tümceleri fark edemeyebilir [101]. O zaman iyi çeviri nedir?. Bu soruya insanlar bile doğru cevap verememektedirler. Bir bilgisayarlı çeviri sisteminin başarısı söz konusu olduğunda, insanlar hep farklı kriterler üzerinden değerlendirme yaparlar, yani uzlaşılan bir kriter yoktur.

Genelde iyi bir çeviri ise üretilen çevirinin, hedef dil bilgisine göre düzgün ve kaynak tümcedeki anlamları doğru şekilde aktarmasıdır. Değerlendirme kalitesinin

¹⁴ Biçimbilimsel çözümlemelerde sadece son durumu ulaşabilen çözümlemeler geçerli sayılır. Birden fazla son duruma ulaşabilirse hepsi geçerli sonuç olur.

belli kiritlerler üzerinden, insan gücü ile yapılmasıdır. Ancak insanlar değerlendirme yaparken hata yapma olasılıkları yüksektir. Üstelik değerlendirme yapacak insanların çeviri yapılan dillerde derin dilbilgisine sahip olmaları gerekmektedir.

Bu nedenle bilgisayarlı çevirileri otomatik değerlendirme yöntemleri geliştirilmiştir. Geliştirilen bu yöntemler içinde en yaygın kullanılan ise BLEU [101], NIST[102], F ölçütü [103], METEOR [104], WER [105] dir. Bu yöntemler çeviri sistemin başarılarını hesaplarken, insanlar tarafından yapılan referans çeviri ile sistemin ürettiği çeviriler üzerinde değişik karşılaştırma yöntemlerine dayanır. Yukarıda sözü edilen başarımlar ölçme yöntemlerinden NIST, F ölçütü ve METEOR ise BLEU üzerinde değişiklik yapılarak geliştirilmiştir. Bu tez kapsamında bu sistemin başarımlarını, BLEU yöntemi ile değerlendirecektir. Bu nedenle alttaki bölümlerde BLEU yöntemi ile WER yöntemi kısaca tanıtılacaktır.

2.4.1 BLEU

BLEU ölçütü, bilgisayarlı çeviri sonuçlarını insanlar tarafından yapılan referans çevirileri üzerinden n-gram bazında yapılan benzerliğinin istatistiksel sonuçlarına göre hesaplar [101]. Doğal olarak bir tümce bir başka dile farklı biçimlerde çevrilebilir. Bu nedenle daha gerçekçi sonuçlar elde edebilmek için her hedef tümceye karşılık birkaç referans tümce oluşturulur. Bu nedenle bazı referans tümceleri sistemin üreteceği aday tümceden daha uzun olabilir ve sistemin yüksek sonuç vermesine neden olabilir. Bu durumu önlemek için ceza katsayısı uygulanır.

Birden fazla tümce ile çeviri yaparken, tüm sistemin başarısını sınamak için kesinliği değiştirilmiş P_n (2.1)'e göre hesaplanır. Formülde, C test sisteminin ürettiği aday çeviriler kümesidir. Aday çeviri kümesinde yer alan her tümcede geçen sözcükler, bu çeviri kümesi için oluşturulan referans tümcesindeki her tümce ile eşleştirilir. Bu eşleştirme n-gram bazında yapılır. Bir tümcenin birden fazla referans çevirisi olduğundan, aday tümcedeki sözcük referans tümcesinde birden fazla geçse de sadece bir sefer sayılır.

$$P_n = \frac{\sum_{S \in C} \sum_{ngram \in S} Adet_{eşleşen}(ngram)}{\sum_{S \in C} \sum_{ngram \in S} Adet(ngram)} \quad (2.1)$$

BLEU yöntemi birden fazla referans tümcesi kullanıldığında geri getirim değerini hesaplamak zorundadır ve genel olarak kesinlik değerine göre hesaplama yapar.

Sistemin başarımını hesaplanması, referans tümcelerden dolayı oluşacak hatayı önlemek amacıyla ceza puanı ağırlıklı olarak kesinlik (precision) ölçütüne dayanmaktadır. Birden fazla referans formül: (2.2)’ye göre hesaplanır. Formülde, r referans tümcenin uzunluğu, C aday tümcenin uzunluğudur.

$$BP = \begin{cases} 1 & \text{eğer } c > r \\ e^{1-r/c} & \text{eğer } c \leq r \end{cases} \quad (2.2)$$

Kesinlik değeri ve ceza puanı hesaplandıktan sonra BLEU puanı aşağıdaki gibi hesaplanır:

$$BLEU = BP \times \exp\left(\sum_{n=1}^N w_n \log P_n\right) \quad (2.3)$$

Formülde $w=1/4$ olarak hesaplanır. Genelde BLEU sonucu 0 ile 1 arasında değişir. Eğer sistem sonucu 1 olursa, sistemin ürettiği aday tümceden en az bir referans tümce ile birebir benzediği anlamına gelir.

2.4.2 WER

Hatalı sözcük oranı ses tanıma ve bilgisayarlı çeviri sistemlerinde yaygın kullanılan yöntemlerden biridir.

Geliştirilmiş sistemlerin başarısı değerlendirilirken, genel olarak yaşanan zorluk, sistemlerin ürettiği çözümler ile referans tümce sözcüklerin farklı olmasıdır. Hatalı sözcük oranı WER sözcük bazında hesaplanır. Sistemin ürettiği çeviri sözcüğü ile referansta geçen sözcüğün birbirine olan benzeme oranına (yakınlık oranı) göre hesaplanır.

İki sözcük arasındaki benzerlik oranına Levenshtian mesafesi denir [106]. Levenshtian mesafesi hedef sözcüğü kaynak sözcüğe benzetmek için, karakter ekleme, silme ve değiştirme için gereken en az işlem sayısıdır [107].

Örneğin, “kitten” ile “sitting” sözcükleri arasındaki Levenshtian mesafesi 3 tür, ve aşağıdaki gibi hesaplanmıştır.

1. kitten → sitten (“k” karakteri ise “s” karakteri ile değiştirilmiştir).
2. sitten → sittin (“e” karakteri ise “i” karakteri ile değiştirilmiştir).

3. sittin → sitting (“g” karakteri eklenmiştir)

Levenishtian mesafesi hesaplandıktan sonra, hatalı sözcük oranı aşağıdaki formülle hesaplanır.

$$WER = \frac{S + D + I}{N} \quad (2.4)$$

Formülde, S değiştirilen karakter sayısını, D silinen karakter sayısını, I eklenen sözcük sayısını ve N ise referansta bulunan toplam sözcük sayısını belirtir.

3. UYGUR DİLİ

Uygur dili Ural-Altay dilleri ailesinin Türk dilleri grubu içinde yer alır ve Uygur Türkleri tarafından kullanılmaktadır. Uygurlar Türk boyları içinde en eski boylardan biridir ve şu an yoğun olarak Doğu Türkistan’da yaşamaktadırlar. Doğu Türkistan dışında Orta Asya Türk cumhuriyetleri ve Türkiye’de çok sayıda Uygur yaşamaktadır. Uygurca ile diğer Türk dilleri arasında çok yakın benzerlikler bulunmaktadır. Genel olarak sözlük, yani Uygurcada kullanılan sözcükler, diğer Türk dillerine göre daha zengindir. Bundan dolayı, diğer Türk dilleri konuşulurken, Uygurca konuşabilen biri rahat anlayabilmektedir. Örneğin, Özbekçe, Türkmence, Kazakça ve Kırgızcada konuşulan sözcüklerin çoğunluğu günümüz Uygurcasında kullanılmaktadır.

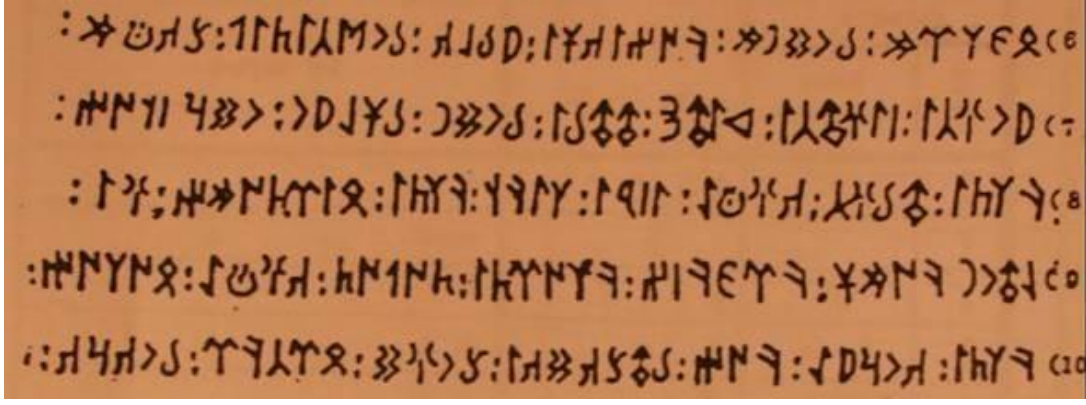
3.1 Uygur Abeceleri

Uygur Türkleri farklı tarihlerde kendine özgü abeceler üretmiş ve başka topluluklarda bu abeceleri kullanılmıştır. Aynı anda Uygurlar da başka toplumların abecelerini kullanmıştır. Bu bölümde Uygurların şimdiye kadar kullandıkları abeceler hakkında kısa bilgiler verilmiştir.

3.1.1 Orhun – Yenisey abecesi

Orhun-Yenisey abecesi 6–7’inci yüzyıllarda, Ordus, Orhun, Selenge civarında ve 840 – yıllarında sonra şimdiki Doğu Türkistan’da kullanılmıştı [42].

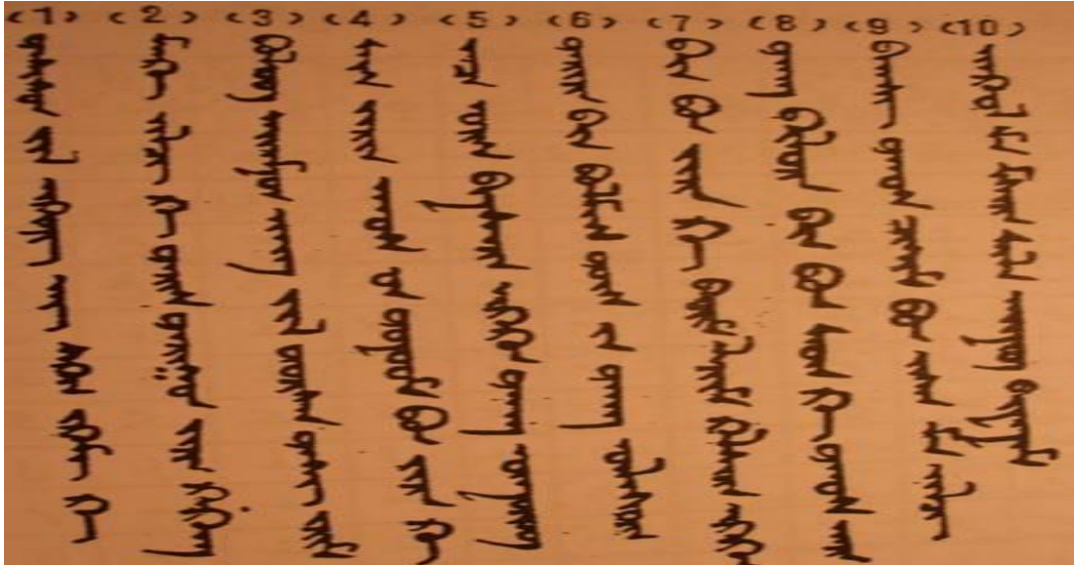
Bu yazı 1889 senesinde Orhun ve Yenisey civarında yapılan arkeolojik çalışmalarda keşfedilmiştir. Bu yazıtların bulunuşu ve Göktürk abecesinin çözülüşünden sonra, bu yazılarda kullanılan sistemlere “Runik Türk Abecesi” adı verilmiştir. Bu abecede yazılar, yukarıda aşağıya ve soldan sağa olmak üzere farklı iki yöntem ile yazılmıştır. Şekil 3.1’de Orhun – Yenisey abecesi ile yazılmış yazıtlardan bir örnek verilmiştir.



Şekil 3.1 : Orhun-Yenisey yazıtları [42].

3.1.2 Eski Uygur abecesi

Eski Uygur Abecesi, Uygurlar Doğu Türkistan'a yerleştikten sonra 17. yüzyıla kadar kullanılmıştır.



Şekil 3.2 : Eski Uygur yazıtları [42].

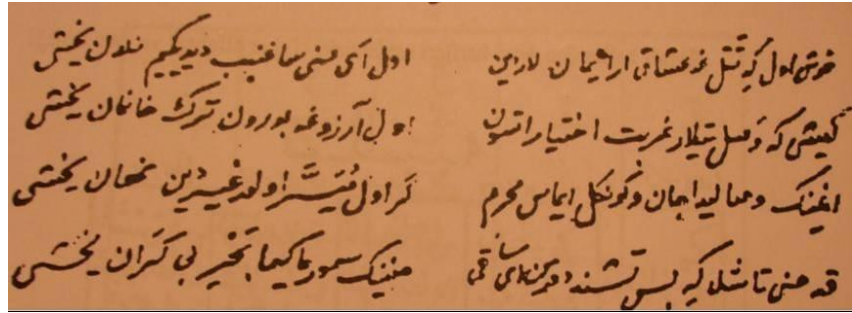
Şekil 3.2'de eski Uygur abecesi ile yazılan bir yazıt verilmiştir. Bu abece tarihte Moğollar ve Mançular tarafından da kullanılmıştır. Cengiz İmparatorluğunun resmi yazışmaları bu abece ile yazılmıştır [42].

Eski Uygur abecesinin temel özellikleri aşağıdaki gibi özetlenebilir:

- Yazılar sağdan sola doğru yazılmıştır.
- 4 ünlü ve 19 ünsüz olarak toplam 23 harften oluşmuştur.
- Bu abecede, b, p ve f gibi birbirine yakın sesler için aynı harf kullanılırken, t-d, s-z için de aynı harf kullanılmıştır [42].

3.1.3 Çağatay abecesi

Uygurlar 932'da İslamiyeti kabul ettikten sonra Arap abecesini kullanmaya başlamıştır. Özellikle İslam dininin yayılmasıyla 10. yüzyılın sonları ve 11. yüzyılın başlarından itibaren Eksî Uygur abecesinin yerine kullanılmaya başlanmıştır. Bu abece özellikle Cengiz hanın oğullarından Çağatay'ın Uygurları yönettiği zamanlarda daha çok geliştiğinden Çağatay abecesi adını almaktadır. Bu abecede Arap abecesinden alınmış 20-32'ye kadar harf kullanılmıştır. Şekil 3.3'de Çağatay abecesi ile yazılan bir yazıt verilmiştir.



Şekil 3.3 : Çağatay abecesi ile yazılan yazıt [42].

Günümüzde bu abece ile yazılan çok sayıda kitap bulunmaktadır. Bu abece ile yazılan belgeler üzerinde araştırma yapabilmek için Çağatay dili araştırma birimleri kurulmuştur.

3.1.4 Latin abecesi

Latin abecesi 1957–1983 yılları arasında kullanılmıştır. Bu abecenin kullanılmasında, o tarihlerdeki siyasi durumlar çok belirgin olmuştur. Özellikle, Uygurların Çin egemenliğine girmesi ve Çin'deki komünizm rejiminden dolayı Çağatay abecesinin terk edilmesi için baskı yapılmıştır. Şekil 3.4'te Latin abecesini temel alan Uygur abecesi ile yazılan yazıt verilmiştir.

3.1.5 Uygur Arap abecesi

Bu abece 1983 yılından Çin yönetiminin izin vermesi sonucu günümüze kadar kullanılmıştır. Bu abece Çağatay abecesinden daha farklı ve daha kolay bir şekilde yazılacak şekilde geliştirilmiştir. Şekil 3.5'te Uygur Arap abecesi ile yazılan bir metin örneği verilmiştir.

Kim qekilsa tupriqingoja jang ara at saliman,
Yengiman yawni qokum bu—ahdu-pormanin sanga.

Huddi Tarim dolkunidek jux urar seygum mening,
Ey guzal pahrin, qururim tondiki kanim sanga.

Bu yengi, xanlik saparda taq kabi tohpə koxux —
Məndiki küqlük irada baqı - hostanim sanga.

Şekil 3.4 : Latin abecesi ile yazılan metin[42].

ئوباما 1961-يىلى ئامېرىكا ھاۋايى شتاتىنىڭ ھونولۇلۇدا تۇغۇلغان، دادىسى كېنىدىن ئامېرىكىغا ئوقۇشقا كەلگەن ئوقۇغۇچى، ئانىسى كانساس شتاتلىق ئاق تەنلىك ئايال. شۇڭا سايلامدا غەلبە قىلغان ئوباما ئامېرىكا تارىخىدىكى تۇنجى ئافرىقىلىقلارنىڭ پۇشىدىن بولغان زۇڭتۇڭ، ئۇ كېلەر يىلى 1-ئاينىڭ 20-كۈنى ئاقسارايدا رەسمىي قەسەم بىرىپ ۋەزىپە تاپشۇرۇۋالدى. نۆۋەتتە جۇمھۇرىيەتچىلەر پارتىيىسىنىڭ زۇڭتۇڭ نامزاتى ماككائىن ئوباماغا تېلېفون بىرىپ ئۇنى تەبرىكلىدى. ئۇ يەنە قوللىغۇچىلىرىنى ئوبامانى قوللاپ، ئىختىلاپنى تۈگىتىپ، دۆلەتنىڭ گۈللىشى ۋە مۇقىملىقىنى بىرلىكتە قوغداشنى مۇراجەت قىلدى. ئامېرىكىنىڭ ھازىرقى زۇڭتۇڭى بۇشمۇ ئوبامانى قولايلىق ۋاقىتتا ئاقساراينى زىيارەت قىلىشقا تەكلىپ قىلدى.

Şekil 3.5 : Uygur Arap abecesi.

3.1.6 Uygur bilgisayar abecesi

Latin abecesinden Arap abecesine geçiş yakın tarihte gerçekteştirildiğinden, Latin abecesi, resmi olmasa da yine kullanılmaktadır. Özellikle İnternet teknolojisinin yaygınlaşmasıyla, internet ortamında yazılan Uygurca yazıların çoğu Latin abecesi ile yazılmaya başlanmıştır. Bazı internet siteleri Uygur Arap abecesi ve Latin abecesi kullanmak üzere farklı iki abecede yayınlanmaya başlamıştır. Ama Latin abecesi kullanırken bazı karışıklıklar da ortaya çıkmıştır. Örneğin, Türkçede “ş” harfinin karşılığı, “x” ve “sh” gibi iki farklı şekilde yazılmaya başlamıştır. Bunun gibi farklılıkları gidermek, yani internet ortamında, ortak bir Latin abecesi için kullanmak amacıyla UKIJ adlı bir kuruluş, gönüllü dil bilimcileri ve bilgisayar mühendisleri tarafından oluşturulmuştur [108]. Arap Abecesi ile Uygur bilgisayar abecesi arasında otomatik dönüşüm yapan yazılımlar geliştirilmiş ve ücretsiz temin edilmektedir. Bundan dolayı, Uygurca için ortak bilgisayar abecesi kullanmak zorunludur. Arap abecesi temelindeki Uygur abecesi için de “unicode” kullanmayı zorunlu kılmıştır [109], ve "unicode" kullanılabilmesi için çalışmalar yapılmıştır [35-

36]. Çünkü farklı kodlama kullanırken, otomatik çeviri yapan programların çalışma verimi çok düşük olmaktadır.

UKIJ'nin çabaları sonucu, Doğu Türkistan'da üretim yapan tüm bilişim firmaları da “unicode”lu Arap abecesi kullanmaya teşvik edilmektedir. Son 2–3 yılda yayınlanan Uygurca web sitelerinin, hepsi “unikod” kullanılarak hazırlanmıştır. Bu durum, gelecek yıllarda Uygur dili üzerinde bilgisayara dayalı araştırma yapabilmek için çok önemli temel hazırlamaktadır. Şekil 3.6'de Uygur Arap abecesi ve bu abecede kullanılan harflere uygun olarak kabul edilen Latince harfler karşılaştırılmıştır.

Araf Alfabeti		Latin Alfabeti		Türkçe Anlamı
Harfler	Örnekler	Harfler	Örnekler	
ا	ئات	a	at	at
ئ	ئەينەك	e	eynek	ayna
ب	بېلىق	b	beliq	balık
پ	پاقا	p	paqa	kurbaga
ت	توشقان	t	toshqan	tavşan
ج	چۈۋا	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	ياغاچ	ng	yangaq	ceviz
ل	لامپا	l	lampa	ampul
م	مۈشۈك	m	müshük	kedi
ن	نان	n	nan	ekmek
ھ	ھارۋا	h	harwa	
ئو	ئورغاچ	o	oghaq	orak
ئۇ	ئۇۋا	u	uwa	yuva
ئۆ	ئۆي	ö	öy	ev
ئۈ	ئۈزۈم	ü	üzüm	üzüm
ۋ	ۋېلىسپېت	w	welespit	bisiklet
ئې	ئېيىق	e	eyiq	ayı
ئى	ئىت	i	it	it
ي	يولۋاس	y	yolwas	kaplan

Şekil 3.6 : Uygur Arap abecesi ile Latin abecesinin karşılaştırılması [108].

Uygur bilgisayar abecesine bakıldığında, Uygur Latin abecesinden farklı olarak, ASCII karakterleri kullanılmıştır. Bunun faydası da, Uygur bilgisayar abecesi için, hiçbir tuş takımı kısıtlaması olmadan kullanma imkanı sağlamasıdır. Olumsuz yönü ise bazı harfler için çift karakter kullanılmış olmasıdır. Örneğin; “sh”, “zh”, ”ng”, ”gh” gibi.

Bu tez çalışmasında ve bu tez için geliştirilecek yazılımlarda, Uygur bilgisayar abecesi kullanılacaktır. Bu tezde Uygurcanın biçimbilimsel çözümleyicisi için XEROX sonlu durum makinesi kullanıldığından ve bu yazılım ASCII karakterlerine göre çalıştığından, Uygur bilgisayar abecesinde kullanılan çift karakterler tek karakterler ile temsil edilmiştir. Yani, ch →c ile, sh→S ile ng→N ile, gh→G ile, zh→Z ile, ö→O ile, ü⇒U ile, é →E ile temsil edilmiştir.

Böylece bu tezde kullanılan Uygur abecesi karakterleri büyük ve küçük olmak üzere tek karakterlerden oluşmaktadır.

a e b p t j c x d r z Z s S G f q k g N l m n h o O u U w E i y

Yani büyük ve küçük karakterler birbirinden tamamıyla farklıdır ve farklı karakteri temsil etmektedir. Dolayısıyla farklı okunmaktadır.

3.2 Uygurca Harflerdeki Uyumluluk

Uygurca bitişken bir dildir. Yani bir kök sözcüğe eklemeler yoluyla, bazen yeni sözcükler oluşurken, bazende sözcüklere ek anlam katılmaktadır. Örneğin; çoğul ve tekil, durum ekleri gibi. Ancak Uygurcada eklemeler belli kurallara göre yapılmaktadır [42], [110-113]. Bundan dolayı Uygurca harfler birkaç alt kümeye ayrılmıştır.

Uygur harfleri genel olarak, ünlü ve ünsüz diye iki ana gruba ayrılır.

Uygurcada 24 adet ünsüz harf vardır: b p t j c x d r z Z s S G f q k g N l m n h w y

Ünsüz harfler kendi içersinde ünlü ünsüz harfler ve ünsüz ünsüz harfler diye iki türe ayrılır.

- Ünlü ünsüz harfler: b j d r z Z G g N l m n h w y
- Ünsüz ünsüz harfler: p t c x s S f q k
- Uygurcada sekiz adet ünlü harf vardır: a e E i o O u U

Ünlü harfler telaffuz edilirken, dil yerinin değişmesine göre üç türe ayrılırlar.

- Dil altı ünlü harfler: e O U
- Dil arkası ünlü harfler: a o u
- Dil arası ünlü harfler: E i

Ünlü harfler telaffuz edilirken, dudak şeklinin değişmesine göre iki türe ayrılırlar.

- Yuvarlak dudaklı ünlü harfler: o O u U
- Yuvarlak dudaklı olmayan ünlü harfler: a e E i

Uygurcada ekler eklenirken, ünlü harfler belirginleşmektedir. Bu açıdan diğer Türk dilleri ile aynıdır. Ama Uygurcada bazı ünlü karakterlerin kesin bir şekli yoktur. Türkçede ünlü karakterlerin sadece kalın ve ince diye iki şekli varken, Uygurcada üç farklı şekli bulunmaktadır. Örneğin, dil arası ünlü harflerde olduğu gibi. Bu türdeki ünlü harfler kümesinde iki harf bulunmaktadır, “E” ve “i”. Bu harfler sözcüklerde geçen bazı ünsüz karakterlere göre değişmektedir. Bu nedenle ünsüz karakterler de birkaç farklı alt kümelere ayrılırlar.

Uygurca içerdiği sözcük yönünden diğer Türk dillerine göre, özellikle Türkiye Türkçesine göre daha karmaşıktır. Türkiye Cumhuriyeti kurulduktan sonra, Türkiye Türkçesi üzerinde yoğun bir çalışma yapılmıştır. Türk asıllı olmayan sözcüklerin çoğu Türk sözlüğünden çıkarılmış ve bu sözcükler yerine, Türkçe dilbilgisi ve Türkçe ses uyumlarına göre yeni sözcükler eklenmiştir. Yani Türkçe sözcüklerin çoğu ses uyumu kurallarına uyum sağlamaktadır.

Uygurcaya Arapça, Farsça, Rusça ve son zamanlarda ise Çince den pek çok sözcük girmiştir. Bundan dolayı, genel olarak, Uygurca sözcükler için tanımlanan bazı sözcükler için geçerli olamamaktadır [42], [110-113].

İki düzeyli biçimbilimsel çözümleyici kural tabanlı işlem yapan bir yöntemdir. Ancak Uygurcanın karmaşık yapısından dolayı, tanımlanan kurallar sayı ve kural yapısı yönünden diğer Türk dillerine göre daha karmaşıktır. Genel olarak Uygurcada, seslerin uyumu sözcüklerin son hecesinde yer alan ünlü sözcük tarafından belirlenir. Eğer son hecesindeki ünlü harf, “dil altı” grubuna ait ise bu sözcüğe eklenecek eklerin de, “dil altı” çeşidi eklenir. Eğer son hecesindeki ünlü harf, “dil arkası” grubuna ait ise bu sözcüğe eklenecek eklerin de, “dil arkası” çeşidi eklenir. Eğer son

hecesindeki ünlü harf, “dil arası” grubuna ait ise bu sefer bu sözcüklerde geçen diğer ünsüz harflere bakılır.

Uygurcada bir sözcük “dil arası” ünlü sözcüğü içerirse ve bu sözcük içerisinde “k, g” ünsüz harfleri varsa, o zaman bu “dil arası” ünlü karakterler “dil altı” ünlü harfi olarak kabul edilir ve eklerin “dil altı” çeşidi eklenir. Aksi durumlarda ise “dil arkası” çeşidi eklenir.

Uygurcada bazı özel sözcükler için sadece son heceye bakmak yeterli olmaz. Böyle durumlarda sözcüğün bir önceki hecesine de bakılması gerekir.

Uygurca yapı yönünden çok karmaşık olduğundan genel bir kural ile anlatmak imkânsızdır. Bu bölümün ilerleyen kısımlarında her durum, örnekler verilerek anlatılmıştır.

3.3 Uygurca İsimlerin Biçimbilimsel Çözümlemesi

İsim, bütün varlıkları ve kavramları karşılayan, varlıkların ve kavramların adları olan sözcüklerdir. İsimler dilde ya kök veya gövde olarak ya da işletme ekleri olarak kullanılırlar. Çokluk eki, iyelik ekleri, hal ekleri ve soru eki isim işletme ekleridir.

3.3.1 Çoğul ekleri

İsimlere çoğul eki eklenirken, isimlerin son hecesinde yer alan ünlü harfe bakılır. Dil altı ünlü harf var ise “ler”, dil arkası ünlü harf var ise “lar” eki alırlar.

Örnek:

kitap + lar

kalem + ler

qoy + lar

baG + lar

kOz + ler

Yukarıda gösterilen örneklerle bakarak sözlük (lexicon) düzeyinde, sözcüklerin son hecesinde geçen ünlü harfe “lar” ya da “ler” geleceğine karar verilebilir. Dolayısıyla, çokluk ekinin eklenmesi için, iki düzenli biçimbilimsel yöntemlerine göre şöyle bir kural tanımlanabilir.

Yapısal düzeyde bir “A” ünlü harfi tanımlanır. Belli birtakım kurallar çerçevesinde, “A” ise “a” ya da “e” olarak değişir. Burada iki düzeyli kural tanımlayabilmek için Uygurca ünlü ve ünsüz harfler alt gruplara ayrılmıştır, daha ayrıntılı bilgi için [114]’e bakılabilir. Ayrıca tezin Ek bölümünde de önemli kurallar verilmiştir.

- Son hecesinde BACKV (Dil arka ünlü harf) içerirse, "lar" eklenir.

"01- 'A' ise 'a' olarak görünür"

A:a =>[[:CONS]*[:BACKV][CONS]*]+(%+:0)[[:CONS |CONS:]]* _;

- Son hecesinde FRONTV (Dil arka ünlü harf) içerirse, "ler" eklenir

"02- 'A' ise 'e' olarak görünür "

A:e =>[[:CONS]*[:FRONTV][CONS]*]+(%+:0)[[:CONS |CONS:]]* _;

Bu iki kurala göre isimlere eklenecek çokluk ekleri, yapısal düzeyde, “a”, ”e” sadece “A” yazılabilir. Böylece çokluk eki yapısal düzeyde “lAr” şekline dönüşmüş olur. “A” ise ekleneceği sözcüğün son hecesinde yer alan ünlü harflere göre, “a” ya da, “e” olarak değişir.

Örnek:

kitap + lAr = kitaplar

kalem + lAr = kalemler

qoy + lAr = qoylar

baG + lAr = baGlar

kOz + lAr = kOzler

Sözcüklerin son hecesinde yer alan ünlü harfler (E, i) dil arası ünlü harfi olursa, durum çok karmaşık olur. Çünkü dil arası ünlü harfler, beraber kullanıldıkları ünsüz harflere göre bazen, dil altı ünlü harfler olarak kabul edilir, “ler” eklenir, bazen, dil arkası ünlü harfler olarak kabul edilir ve “lar” eklenir.

Örnek:

tilsim + lar

kil + ler

mis + lar

gEzit + ler

iyEq + lar

Bundan dolayı, son hecesinde dil arası ünlü harfler gözüktüğünde, o sözcüklerde geçen ünsüz harflere de bakılması zorunludur. Dil arası ünlü harfler belli bir takım ünsüz harflerle geldiğinde dil altı ünlü harf olarak, belli bir ünsüz harfler ile geldiğinde dil arkası ünlü harf olarak kabul edilmektedir [42], [110-113].

Dil arası ünlü harflerin bu durumunu göz önüne alarak isimlerin çoğul eki alma kuralı aşağıdaki gibi yazılabilir.

- Son hecesinde MIDV (Dil arası ünlü harf) ve CONSBV içerirse, "ler" eklenir

"03- 'A' ise 'a' olarak görünür "

A:a=>[[CONSBV+][:MIDV][:CONSBV]*]+(%+:0)[:CONSBV
CONSBV:]* _;

- Son hecesinde MIDV (Dil arası ünlü harf) ve CONSFV içerirse, "lar" eklenir

"04- 'A' ise 'e' olarak görünür "

A:e => [[[CONSFV][:MIDV][CONS]*]+][[:MIDV][CONS]*]*(%+:0)
[:CONS|CONS:]* _;

Yukarıda tanımlanan kurallarda CONSBV ve CONSFV kümesi ise “dil arası” ünlü harfleri “dil altı” ve “dil arka” ünlü harfler yapan ünsüz harfler kümesidir [114]. Oluşturulmuş bu kurallara göre, daha önce verilmiş örnekler şöyle değiştirilebilir.

Örnek:

tilsim + lAr = telsimler

kir + lAr = kirler

mis + lAr = mislar

gEzit + lAr = gEzitler

iyEq + lAr = iyEqlar

Böylece, Uygurcada çokluk ekleri, yapısal düzeyinde “lAr” ile gösterilmiş olur. Görünen kısımda ise duruma göre, ”lar” ya da “ler” olarak görünür. İki düzenli biçimbilimsel yönteminde, tüm kurallar birbirinden bağımsız olarak paralel çalışır [115]. Dolayısıyla, yukarıda tanımlanan kuralları iki kısım altında toparlanabilir.

- "03- 'A' ise 'a' olarak görünür "

A:a => [[:CONS]*[:BACKV][CONS]*]+(%+:0)[[:CONS |CONS:]*_ ;
 [[CONSBV+][:MIDV][CONSBV]*]+(%+:0)[[:CONSBV|
 CONSBV:]*_ ;

- "04- 'A' ise 'e' olarak görünür "

A:e => [[:CONS]*[:FRONTV][CONS]*]+(%+:0)[[:CONS |CONS:]*_ ;
 [[[CONSFV][:MIDV][CONS]*]+][[:MIDV][CONS]*]*(%+:0)[[:CONS|
 CONS:]*_ ;

3.3.2 İyelik ekleri

Uygurcada kullanılan iyelik ekleri Çizelge 3.1’de verilmiştir. İyelik ekleri genelde 3. şahıs eklerini içermektedir. 2.şahıs için, saygın anlamına gelen bir ek daha bulunmaktadır. 3. şahısın tekil ve çoğul şekli aynıdır. Çizelge 3.1’de gösterildiği gibi, tüm eklerin sonunda,”m”, “miz”, “Nlar” gelir ve sabittir. Ama bu eklerin önünde, i, u, U ünlü harfleri gelir ve bazen gelmez.

Örnek:

kOz+Um = kOzUm

bala+m = balam

kOz+UNlar = kOzUNlar

Çizelge 3.1 : Uygurcanın iyelik ekleri.

Şahıs		Ekler	Örnekler
1- Şahıs	Tekil	m, im, um, Um	balam, kOzUm
	Çoğul	miz, imiz	balamiz, kOzUmiz
2-Şahıs	Normal, tekil	N, iN, uN, UN	balaN, kOzUN
	Normal, çoğul	Nlar, iNlar, uNlar, UNlar	balaNlar, kOzUNlar
	Saygın, tekil	Niz, iNiz	balaNiz, kOzUNiz
	Saygın, çoğul	liri	baliliri, kOzliri
3-Şahıs		si, i	balisi, kOzi

İyelik eklerinin ekleneceği sözcükler, ünlü harf ile sonlanırsa, ekler doğrudan eklenir.

Eğer, sözcükler, ünsüz harf ile sonlanırsa, eklerden önce kesin olarak bir ünlü harf olması gerekir.

Sonuç olarak, burada bir ünlü harf bazen eklenir, bazen eklenmez, eklenirken de, “i”, “u”, “U” şeklinde farklı biçimlerde değişmektedir. Değişim kuralı ise “yuvarlak dudaklı” ünlü harfler ve “yuvarlak dudaklı” olmayan ünlü harflere göre belirlenmektedir. Bu değişen ünlü harf için yapısal düzeyde “H” harfi kullanılmıştır. Yani burada tanımlanan “H” harfi belli bir takım kurallar çerçevesinde görünen düzeyde, “i”, “u” ya da “U” olarak değişir. İki düzeyli biçimbilimsel çözümleyici için aşağıda gösterildiği gibi kurallar tanımlanmıştır.

- "05- 'H' ise '0' olarak görünür"

H:0 => [:CONS]*[:VOWEL] (%+:0) _ ;

Kullanılan programda, yapısal düzeyde “0” görününce, görünen düzeyde yazıda boş bırakılır. Bu kuralda ünlü harflerden sonra “H” gelirse, “H” yerine “0” ile değişir. Çünkü Uygurcada iki ünlü harf peş peşe gelemmez.

- "04- 'H' ise 'u' olarak görünür"

H:u => [:CONS]*[:Lou][:CONS]+ (%+:0) _ ;

- "05- H ise U olarak görünür"

H:U => [:CONS]*[:LOU][:CONS]+ (%+:0) _ ;

Bu kurallar tanımlandıktan sonra, yapısal düzeyde, Uygurcanın iyelik ekleri “H” harfine göre değişir.

Örnek:

kOz + Hm = kOzUm (H:U)

bala + Hm = balam (H:0)

kOz + HNlar = kOzUNlar (H:U)

kOz + HNlar = kOzUNlar=kOzUNlAr=kOzUNlar (H:U, A:a)

bala + HNiz = balaNiz (H:0)

Çizelge 3.1’de gösterildiği gibi, 3-şahsa “si”, eki eklenirken, “a” harfi ise “i” ile değişmektedir. Bu durum için aşağıdaki kural tanımlanmıştır.

- "06- 'a' ise 'i' olarak görünür"

a:i <=> [CONS]*[VOWEL][CONS] _ s [H:i] ;

Bu kural tanımlandıktan sonra, isimlere 3. şahıs iyelik ekleri eklenirken yapısal düzeyde farklı yazılır.

Örnek:

bala + sH = balisi

taGa + sH = taGisi

ana + sH = anisi

kala + sH = kalisi

3.3.3 Durum ekleri

İsmin bir başka sözcük ile ilgi kurmak üzere ekli veya eksiz bulunduğu durumlara ismin halleri denir. Uygurcada tüm isimler hal ekleri alabilmektedir. İsimlere ekler eklenirken o sözcüklerde geçen harflere göre farklı ekler gelmektedir.

Örnek:

qol + Ga

kOl + ge

baG+qa

kUc + ke

İsimlerin yönelme hal ekleri "Ga", "qa", "ge", "ke", belli bir kuralara göre eklenmektedir. Yani sözcüklerde geçen ünlü ve ünsüz harfler tarafından belirlenmektedir. "A" harfi ise "e", "a" ile değişebilir Burada, yapısal düzeyde "Y" harfi tanımlanmakta ve ünlü-ünsüz uyum sağlama kuralları çevresinde "Y" harfi "G", "g", "k", "q" ler ile değişmektedir.

Örnek:

qol + YA = Ga (Y:G, A:a)

kOl + YA = kOlge(Y:g, A:e)

baG + YA = baGqa (Y:q, A:a)

kUc + YA = kUcke (Y:k, A:e)

Dolayısıyla yapısal düzeyde Uygurcanın yönelme hal ekleri “YA” dır.

Bu durum için tanımlanan iki düzenli kural:

- "07- ‘Y’ ise ‘q’ olarak görünür "

Y:q => [:CONS]*[: [o | u | a]][CONSUV0]+ (%+:0) _ ;

- "08- ‘Y’ ise ‘k’ olarak görünür "

Y:k => [:CONS]*[: [e | i | E | U | O]][CONSUV0] (%+:0) _ ;

- "09- ‘Y’ ise ‘G’ olarak görünür "

Y:G => [:CONS]*[: [o | u | a]][CONSVOaz]+ (%+:0) _ ;

- "10- ‘Y’ ise ‘g’ olarak görünür "

Y:g => [:CONS]*[: [e | i | E | U | O]][CONSVOaz]+ (%+:0) _ ;

Çizelge 3.2 : Uygurcanın durum ekleri.

Hal isimleri	Ekler	Örnekler
Yalın hali	0 (eksiz)	meydan, kitap, Oy, ders
İlgi hali	niN	meydanniN, kitapniN, OyniN, dersniN
Yönelme hali	Ga, qa, ge, ke	meydanGa, kitapqa, Oyge, derske
Yükleme hali	Ni	meydanni, kitapni, Oyni, dersni
Çıkma hali	din, tin	meydandin, kitaptin, Oydin, derstin
Bulunma hali	da, de, ta, te	meydanda, kitapta, Oyde, derste
Sınırlama hali	Gice, gice, qice, kice	meydanGice, kitapqice, Oygice, derskice
Eşitlik hali	çilik, ce	meydancilik, kitapçilik, Oycilik, dersçilik
Benzerlik hali	dek, tek	meydandek, kitaptek, Oydek, derstek
Bulunma miktar hali	diki, tiki	meydandiki, kitaptiki, Oydeki, derstiki

Benzeri yöntemler ile yapısal düzeyde iken, çıkma hal eki “DHn”, bulunma hal eki “DA”, sınırlama hal eki “YicA”, benzerlik hal eki “Dak”, benzerlik hal eki “Diki” olarak kabul edilmiş ve Uygurcanın uyumluluğuna göre kurallar yazılmıştır (EK.A). Daha önce belirtildiği gibi, Uygurcada isimler aynı anda birden fazla ek alabilirler. Bu ekler eklenirken, kesin olarak ünlü ve ünsüz harflerin uyum sağlaması

gerekmektedir. Bu tez kapsamında geliştirilmiş kurallar ile harflere ekler eklenirken ses uyumu sağlanmıştır.

Örnek:

kitap + lAr + Hm + DA= kitaplarımDa

kitap + Hm + DA = kitapımDa

meydan+ lAr + YA = meydanlarGa

ana+Hm+YA = anamGa

Tanımlanmış bu kurallar aracılığıyla, Uygurca isimlerin biçimbilimsel yapısı çıkartılmıştır.

Uygurca isimler yukarıda belirtilen ekler dışında, başka farklı ekler de alabilmektedir. Bu ekler eklenirken, isimler bazen eyleme, bazen sıfatlara, bazen de zarflara dönüşmektedir ve yeni anlam taşımaktadır.

Örnek:

iS -> iS+ci-> iS+ci+liq

iS+can-> iS+can+liq

iS+siz-> iS+siz+liq

iS+le -> iS+le+mci->iS+le+mci+lik

iS+le +t

iS+le +n

iS+hana

iS (isim)

iSle (eylem, çalışmak)

iSci (isim, iş yapan biri)

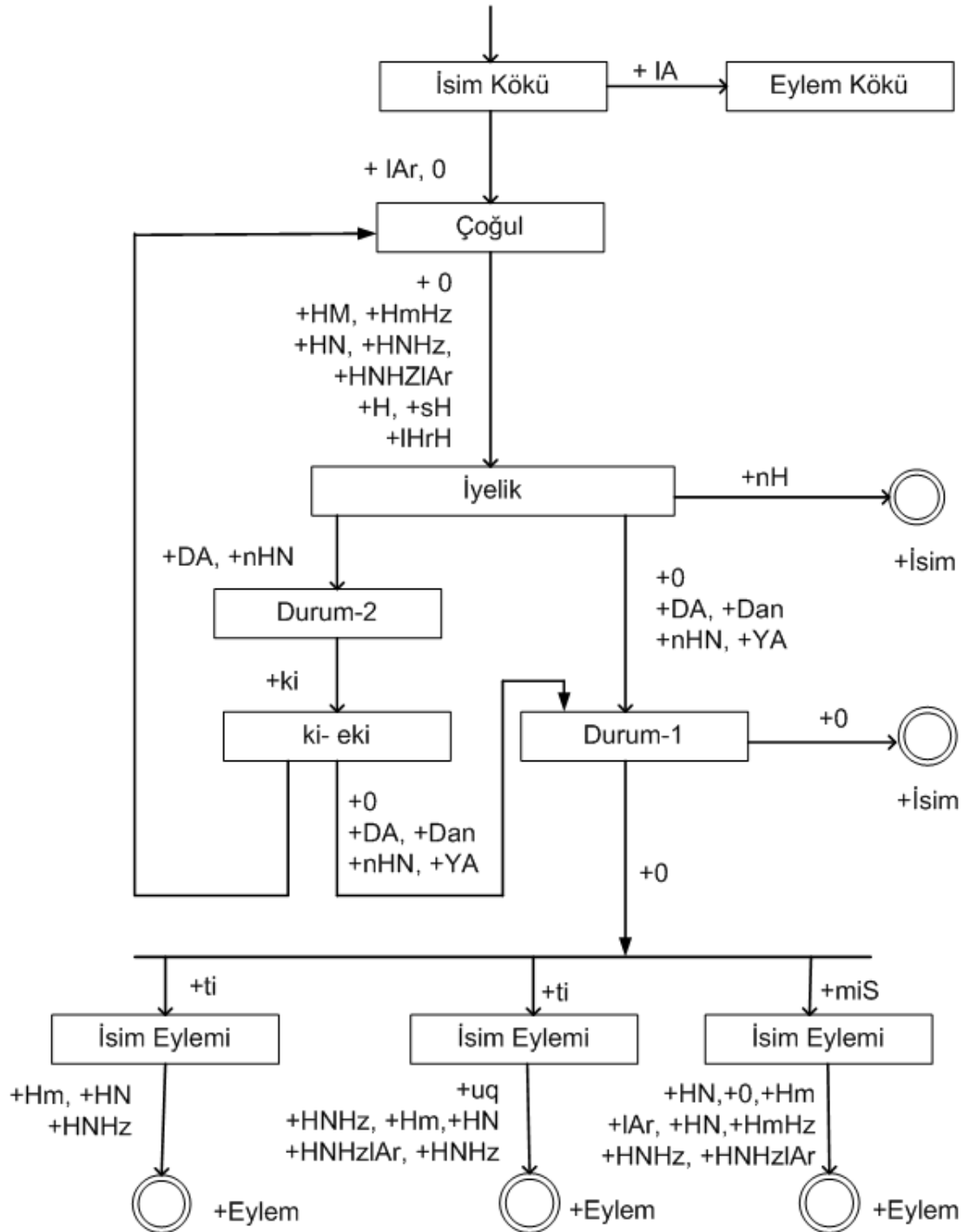
iScan (sifat, çalışan)

iSsiz (sifat, işsiz, güçsüz)

iSlet (eylem emir, 3-şahısa karşın)

iSlen (eylem, edilen, pasif)

Bu ekler sadece bazı isimlere eklenebilirler. Dil yapısı yönünden Uygurca çok karışık olduğundan dolayı, özellikle genel kurallar üzerinde çalışmaya özen gösterilmiştir. Uygurca isimlerin sonlu durumlu makineler ile yapılmış biçimbilimsel çözümleyicisi Şekil 3.7’ de verilmiştir.



Şekil 3.7 : İsimlerin sonlu durumlu makinler ile çözümlenmesi.

Şekil 3.7’de gösterilen çözümleyicide, kareler bir ara durumu göstermektedir, yani son durumlar değildir. Bu durumlardayken geçerli sözcük olarak kabul edilemez. Çift yuvarlak çizgiler ise son durumlardır. Yani başlangıç durumundan, ekler eklenerek, ya da ek almadan geçerek (yani boş geçerek “0”) son duruma gelebiliyorsa, bu bir geçerli giriş katarı, yani sözcük olarak kabul edilir. Örneğin yapısal düzeye “kitapnH” sözcüğü gönderilirse, işlemler şöyle gerçekleşir: İlk önce çoğul eki alıp almadığına bakılır. Bu sözcükte çoğul eki “lAr” olmaması dolayısıyla bir sonraki duruma “0” eki, yani “boş” ek ile geçer. Bir sonraki adımda ise “nH” eki aranır. “nH” eki ile alırken, çift yuvarlak çizgiye, yani “son” duruma gelir. Bundan dolayı “kitapnH” sözcüğünü geçerli bir sözcük olarak kabul eder ve biçimbilimsel bilgilerini geçiş yaptığı durumlardaki eklere göre yorumlar.

3.4 Uygurca Sıfatların Biçimbilimsel Çözümlemesi

Uygurcada sıfatlar çok yaygın olarak kullanılır. Eklerin eklenmesi ile sıfatlar oluşturulabilir. Aynı anda asıl sıfatlar bulunmaktadır.

Örnek: uzun, qısqa, yahSi, yaman, cirayliq, qizil, sEriq

Sıfatlara farklı ekler eklenerek sıfatların dereceleri değiştirilebilir. Dolayısıyla Uygurcada sıfatların üç farklı dereceleri vardır [42].

- Kuvvetlendirmeli derecelendirme

Sıfatın kuvvetlendirme şekli, önce sıfatın birinci hecesinin sonuna “p” ünsüz harfini eklemek ve sonra sıfatı eklemek suretiyle oluşturulur.

Örnek:

qizil qipqizil

sUzUk sUpsUzUk

yESil yapyESil

yENi yEpyENi

Sıfatların kuvvetlendirilme derecesi, bu tezde tasarlanan kurallarda ayrı bir sıfat şeklinde verilmiştir. Yani “kiqizil” ve “qipqizil” ise tam birbirinden bağımsız, ilişkisiz iki sözcük olarak kabul edecektir.

- Küçültmeli derecelendirme

Sıfatların küçültme derece şekli sıfatta ifade edilen vasfın asıl dereceye göre biraz küçüldüğünü bildirir.

Örnek:

sEriq sarGuS (sarımsı)

kOk kOkUS (mavimsi)

Sıfatların küçültme derecesi pek fazla olamayan ve renk ifade eden asıl derece sıfatlarının sonuna “S”, ”C”, c” –mtol” eklerinin eklenmesiyle oluşur.

Örnek:

aq aq + iS (beyazımsı)

qara qara + mtol (siyahımsı)

Bazı sıfatların küçültme derecesi ise, ”ac”, “sus” gibi eksikliği bildiren sıfatların eklemek suretiyle oluşur.

Örnek:

seriq acsEriq

qizil acqizil

kOk suskOk

qara susqara

Bu yöntem ile oluşan sıfatlarda, “ac”, “sus” gibi sıfatlar, ön ek olarak kabul edilmiştir. Sıfatların küçültme derecesinin yeni bir çeşidi, sıfat sayılan sözcük köklerine “Gina”, “kina”, “qina”, “kine” eklerinin eklenmesiyle oluşur.

Örnek:

salqin salqinGina

omaq omaqqina

tüzük tüzükkine

epcil epcilGine

kicik kicikkine

Bu tarzda elde edilecek sıfatlar için, “YinA” eki kural olarak tasarlanmıştır. Dolayısıyla bir önceki örnekler leksikonda farklı görünür.

Örnek:

salqin	salqin+YinA	salkinGina
omac	omac+YinA	omacqina
tüzük	tüzük+YinA	tüzükkine
epcil	epcil+YinA	epcilGine
kicik	kicik+YinA	kicikkine

Sıfatların küçültmeli derecesi, genel bir kuralla oluşturulmadığından ayrı bir sıfat şeklinde verilmiştir. Yani “qara” ve “qaramotol” ise tam birbirinden bağımsız, ilişkisiz iki sözcük (sıfat) olarak kabul edilir.

Uygurcada sıfat yapabilen çok sayıda ek bulunmaktadır.

- İsimlerden sıfat yapan ekler: “-liq”, ”-lik”, ”-luq”, ”-IUK”

Örnek:

tuz+luq
saqal+liq
qoral+liq
payda+liq

Yapısal düzeyde, ”liY” eki kabul edilmiştir,

- Sıfattan sonra gelen ekler: “siz”

Örnek:

qoral+siz
payda+siz

- Sıfatlardan önce gelen ekler, “na”, “bet”

Örnek:

na+mert
bet+qiliq

- Zarflardan sıfat yapan ekler “+ki”, “+qi”

Örnek:

hazir+qi

tUnUgUn+ki

Bu iki ek “YH” ile belirtilmiştir.

Örnek:

hazir+ YH hazirqi

tUnUgUn + YH tUnUgUnki

Bu tarzdaki ekler sıfatlara genel eklenecek ekler kümesine eklenmiştir.

3.5 Uygurca Zarfların Biçimbilimsel Çözümlemesi

Uygurcada zarflar pek çok ek almaz. Genel olarak eklenen ekler ise:

- “-raq”, ” rek”

Örnek:

baya + raq

neri + raq

Ete + rek

Bu ekler ise , “rAY” eki ile belirtilmiştir.

Örnek:

baya + rAY bayaraq

ete + rAY eterek

- “-ki”, “ce”

Örnek:

bUgun + ki

ete + ki

bUgUn + ce

Başka sözcüklere de eklerin eklenmesiyle zarflar oluşmaktadır.

- “+up, +Up”

Örnek:

oltur+up

kOr+Up

Bu ekler için “Hp” eki kabul edilmiştir ve buna göre kurallar tasarlanmıştır.

- “+Gice”, “+gice”, “+ qice”, “kice”

Örnek:

yazGice

yatqice

kelgice

ketskice

Bu ekler için “YHce” eki kabul edilmiştir ve buna göre kurallar tasarlanmıştır.

- “+Gili”, “+gili”, “+ qili”, “kili”

Örnek:

yazGili

yatqili

kelgili

ketskice

Bu ekler için, “YHli” eki kabul edilmiştir ve buna göre kurallar tasarlanmıştır.

- “+Gacqa”, “+gecke”, “+ qacqa”, “kecke”

Örnek:

yazGacqa

yatqacqa

kelgecke

ketkecke

Bu ekler için, “YacYA” eki kabul edilmiştir ve buna göre kurallar tasarlanmıştır.

- “+Gac”, “+gec”, “+ qac”, “kec”

Örnek:

yazGac

yatqac

kelgec

ketkec

Bu ekler için, “YAc” eki kabul edilmiştir ve buna göre kurallar tasarlanmıştır.

- “+Ganseri”, “+genseri”, “+ qanseri”, “kenseri”

Örnek:

yazGanseri

yatqanseri

kelgenseri

ketkenseri

Bu ekler için, “YAnseri” eki kabul edilmiştir ve buna göre kurallar tasarlanmıştır.

3.6 Uygurca Eylemlerin Biçimbilimsel Çözümlemesi

Uygurcada eylemler, tıpkı diğer bitişken dillerde olduğu gibi çok karmaşık bir yapıya sahiptir. Eylemlere çeşitli eklerin eklenmesi ile yeniden isim, sıfat, zarf gibi sözcükler oluşur. Eylemlerde olduğu gibi diğer sözcüklere farklı yapım eklerinin eklenmesiyle yeniden eylemler oluşur. Eylemler Uygurcada zengin ve aynı zamanda karmaşık bir yapıya sahiptir [37-39].

3.6.1 Uygurca eylemlerin çeşitleri

Eylemler yapısına göre asıl eylem, türetilmiş eylem ve bitişken eylem diye üç türe ayrılır.

- **Asıl Eylem:** Eylemlerin öz ekidir, yani bu eylemlerin hiçbir ek almadığı durumudur.

Örnek:

maN (yürü)

tur (bekle)

kUl (gölmek)

- **Türetilmiş eylemler:** Farklı eklerin eklenmesiyle diğér sözcüklerden yeniden oluşan eylemler.

Örnek:

teqdir (isim) (takdir)

teqdir + le (ödüllendir) ->(eylem)

muz (isim) (buz)

muz + la (üşümek) -> (eylem)

- Bitişken eylemler. Birden fazla eylemlerin birleşmesinden oluşan eylemlerdir.

Örnek:

qarSi (karşı) + al (al) -> karşılamak

qarSi (karşı) + tur (dur) -> direnmek

tarmar (dağılım)+ bol (ol) -> dağılmak

tarmar (dağılım)+ qil (yap) -> dağıtmak

care (çare) + kOr (gör) -> çözümle

3.6.1.1 Geçişli eylem ve geçişsiz eylem

Uygurcada eylemlerin bazıları nesne alır, bazıları almaz. Eylemler bu özelliklerine göre geçişli eylemler ve geçişli olmayan eylemler diye iki türe ayrılırlar. Ancak bu çeşit eylemlerin yapısında hiçbir değişiklik görülmez. Sadece yüklemlere göre fark görülür. Bundan dolayı biçimbilimsel kapsamında çözümlenemez.

3.6.1.2 Hareket eylemi ve durum eylemi

Eylemlerin anlamına göre hareket eylemi ve durum eylemi diye iki türe ayrılır ve anlamına bakılarak ayırt edilir. Yani eylemlerin etkiledikleri nesnenin durumunu belirtir. Burada herhangi bir ek etkileşimli değildir. Ancak eylemin anlamına bakılarak karar verilir. Dolayısıyla biçimbilimsel kapsamda çözümlenemez.

Anlamına göre belirsizlik giderme kendi alanında ayrıca bir araştırma konusudur ve bu tez kapsamı içerisinde değildir.

3.6.1.3 Bağımsız eylem ve yardımcı eylem

Bir bütün tümce de kendi başına anlamlı bir parça oluşturabilen eylemlere bağımsız eylem denir. Uygurcada eylemlerin çoğu bağımsız eylemlerdir. Bazı eylemler diğer sözcükler ile beraber kullanılarak anlam belirtir. Örneğin: idim, iken, ketti. Bu türdeki eylemlere yardımcı eylemler denir. Yardımcı eylemler, kendi içerisinde bağlaç ve durum eylemleri diye iki türe ayrılır. Bağlaçlar farklı ekler ile gösterildiğinden burada ayrı ayrı olarak gösterilecektir.

- Dolaysız hüküm bağlayıcısı

Dolaysız hüküm bağlayıcıları ise bir eylemin doğrudan, yani dolaysız yapıldığını belirtir. Çizelge 3.3'te eylemlerin iyelik ekleri ve zamana göre eklenmeleri gösterilmiştir.

Zaman ve şahıs yönünden fark edilirler ve iyelik ekleri alırlar.

Örnek:

Ötken yili bu yerge kelgen idiN. (Geçen sene buraya geldin).

Çizelge 3.3 : Dolaysız hüküm bağlayıcısı.

Şahıs ve zaman		Şimdiki zaman	Geçmiş zaman
1- Şahıs	Tekil	men	idim
	Çoğul	miz	iduq
2- Şahıs	Tekil	sen	idiN
	Tekil, nezaket	siz	idinNiz
	Tekil, hürmet	la	idile
	Çoğul	siler	idiNlar
3- Şahıs		dur // tur	idi

- Dolaylı hüküm bağlayıcısı (+ken)

Dolaylı hüküm bağlayıcıları bir eylemin doğrudan değil, yani dolaylı yapıldığını belirtir. Çizelge 3.4' te eylemlerin iyelik eklerine göre eklenmeleri gösterilmiştir.

Çizelge 3.4 : Dolaylı hüküm bağlayıcısı.

1-şahıs		2-şahıs				3-şahıs
Tekil	Çoğul	Tekil	Nezaket	Hürmet	çoğul	İken
iken+men	İken+miz	İken+sen	İken+siz	iken+la	iken+siler	

Dolaylı hüküm bağlayıcıları sadece şahıs yönünden fark etmektedir. Çünkü dolaylı yönünden ifade ettiğinden hep geçmiş zaman anlamında gelmektedir. Eylem köklerine “ken” bağlayıcısı ve iyelik eklerinin eklenmeleri ile oluşmaktadır.

Örnek:

u bu mektepte beS yıl iSligen iken. (o bu okulda beş sene çalışmış).

- Anlatma hüküm bağlayıcısı

Bir olayı anlatma, aynı hikaye tarzında anlattığını bildirir. Dolayısıyla zaman ve şahıs yönünden ve aldığı eklerde farklar olabilir. Çizelge 3.5’de zaman ve şahsa göre aldığı ekler gösterilmiştir. Eylem köklerine “dek” bağlacının ve iyelik eklerin eklenmesi ile oluşur.

Örnek:

ularniN sOzlerige qariGanda, bUguN kelgUdek.
(Onların konuştuklarına göre, bugün geleceklermiş).

Çizelge 3.5 : Anlatma hüküm bağlayıcısı.

Şahıs & Zaman		Şimdiki Zaman	Geçmiş Zaman
1-Şahıs	Tekil	dekmen (imiSmen)	ikenmiSmen (ikenmenmiS)
	Çoğul	dekmiz (imiSmiz)	ikenmmiSmiz (ikenmizmiS)
2-Şahıs	Tekil	deksen (imiSsen)	ikenmiSsen (ikensenmiS)
	Tekil nezaket	deksiz (imiSsiz)	ikenmiSsiz (ikensizmiS)
	Tek hürmet	dekla (imiSla)	ikenmiSla (ikenlamıS)
	Çoğul	deksiler (imiSsiler)	ikenmiSsiler (ikensilermiS)
3-Şahıs		dek (imiS)	ikenmiS

- Sübjektif tahmini hüküm bağlayıcısı

Bir olay ile ilgili yapılacağı ya da yapılmayacağına göre tahminlere olumsuz yönden yaklaşımı bildirir. Bu çeşit bağlaçlar zaman ve şahıs yönünden fark edilmektedirler. Genelde şimdiki ve geçmiş zamanda kullanılır. Dolayısıyla geçmiş zamandaki bir olaya karşın tahminlerini bildirmiş olur.

Örnek:

O vakıtta belkim sen memlikette bolGiydiN.

(O zaman belki sen memlekette olabilirdin).

Çizelge 3.6’da şahıs ve zamana göre aldığı ekler gösterilmiştir.

Çizelge 3.6 : Sübjektif tahmini hüküm bağlayıcısı.

Şahıs & Zaman		Şimdiki Zaman	Geçmiş Zaman
1- Şahıs	Tekil	-durmen (-dimen)	bolGidim (idimGu deymen)
	Çoğul	-durmuz (dimiz)	bolGiduq (iduqqu deymen)
2- Şahıs	Tekil	-dursen	bolGidiN (idiNGu deymen)
	Tekil nezaket	-dursiz	bolGidiNiz (idiNizGu deymen)
	Tekil hürmet	-durla (-dila)	bolGidile (idileGu deymen)
	Çoğul	-dursiler	bolGidiNlar (idiNlarGu deymen)
3- Şahıs		-dur (-du)	bolGidi (idiGu deymen)

- Objektif Tahmini hüküm bağlayıcısı

Bir olay ilgili yapılacağı ya da yapılmayacağına göre tahminlere olumlu yönden yaklaşımı bildirir. Bu çeşit bağlayıcılar zaman ve şahıs yönünden fark edilirler. Genelde şimdiki ve geçmiş zamanda kullanılır. Dolayısıyla geçmiş zamandaki bir olaya karşın tahminlerini bildirmiş olur.

Örnek:

O vakıtta belki sen memlikte bolGiydiN.

(O zaman belki sen memleketteydin).

Çizelge 3.7 : Objektif tahmini hüküm bağlayıcısı.

Şahıs Türleri		Şekilleri
1- Şahıs	Tekil	oxSaymen (bolsam kErek, cEGimba)
	Çoğul	oxSaymiz (bolsaq kErek, cEGimizva)
2- Şahıs	Tekil	oxSaysen (bolsaN kErek, cEGiNva)
	Tekil nezaket	oxSaysiz (bolsiNiz kErek, cEGiNizva)
	Tekil hürmet	oxSayla (bolsila kErek , cEGliriva)
	Çoğul	oxSaysiler (bolsaNlar kErek, cEGiNlarva)
3-Şahıs		oxSaydu (bolsa kErek, cEGi)

Çizelge 3.7’de şahıs ve zamana göre aldığı ekler gösterilmiştir.

Eylemler, eklendiği ekler ve tümcelerde üstelendiği görevlere göre üç ana çeşide ayrılırlar.

1. Eylemlerin haber şekilleri
2. Eylemlerin sabit sözcük şekilleri
3. Eylemlerin öz ek şeklidir

Burada eylemlerin sabit söz şekli ve haber şekilleri eylemlerin öz ek şekline çeşitli eklerin eklenmesi ile oluşur. Bunlar hariç eylemlerin haber şekli ise çeşitli yönelme ve kip şekillerini içerir.

Bundan sonraki bölümlerde, tüm özellikler detaylı şekillerde gösterilecektir.

3.6.2 Eylemlerin sabit sözcük şekilleri

Eylemlere bazı yapım eklerinin eklenmesi ile farklı anlam içeren ve buna karşılık olarak da tümcede farklı görev üstelenen sözcükler oluşur. Uygurcada genel olarak dört çeşidi içerir.

- Eylemlerden oluşan isimler
- Eylemden oluşan zarflar
- Eylemden oluşan sıfatlar
- Eylemden oluşan amaç eylemleri

3.6.2.1 Eylemlerden oluşan isimler

Bazı eylemlere eklerin eklenmesi ile eylemler isimlere dönüşecektir ve isimler gibi kullanılabilir. İsimler için tanımlanan tüm kurallar ve çözümlemeler geçerli olur.

1- maq, -mek (mAY)

yaz + maq	yaz + mAY
kel + mek	kel + mAY
tur + maq	tur + mAY

“maq” ve “mek” ekleri ise yapısal düzeyde “mAY” ile temsil edilmiştir ve tanımlanan kurallara göre görünen düzeyde “maq/mek” olarak görünmektedir.

2- S,-iS,-uS,-US (IS)

sOzle + S	sOzle + IS
tur + uS	tur + IS
yaz + IS	yEz + IS (a:E) dönüşümü yapıyor

3- Gu,-qu,-gU,- kU (YI)

yaz + Gu	yaz + YI
yat + qu	yat + YI
ket + kU	ket + kU
sez + gU	sez + gU

4-gan,-qan-,gen- ken (YAn)

yaz + Gan	yaz + YAn
kel + gen	kel + YAn

5- dek, -dak,-tak,-tek (DAk)

yaz + gan + dek	yaz + YAn + DAk
kel + gen + dek	kel + YAn + DAk

3.6.2.2 Eylemlerden oluşan zarflar

Bazı eylemlere eklerin eklenmesi ile eylemler zarflara dönüşecektir ve zarflar gibi kullanılabilir. Zarflar için tanımlanan tüm kurallar ve çözümlemeler geçerli olur.

1- p, -ip,-up, -Up (Hp)

sOzle+p	sOzle+Hp
yaz+ip	yaz+ Hp
kOr+Up	kOr+Hp
yul + up	yul + Hp

2- Gice, -gice, -kice,-qice (YHce)

yaz + Gice	yaz+YHce
yat + qice	yat+ YHce
kel + gice	kel + YHce
ket + kice	ket + YHce

3-Gili, -qili,-gili, -kili (YHli)

yaz + Gili	yaz + YHli
yat + qili	yat + YHli
kel + gili	kel + YHli
ket + kili	ket + YHli

4-Gacqa, -qacqa,-gecke,-kecke (YAcYA)

yaz + Gac +qa	yaz + YAc + YA
yat + qac + qa	yat + YAc + YA

kel + gec + ke	kel + YAc + YA
ket + kec + ke	ket + YAc + YA

3.6.2.3 Eylemlerden oluşan sıfatlar

Bazı eylemlere eklerin eklenmesi ile eylemler sıfatlara dönüşecektir ve sıfatlar gibi kullanılabilir. Sıfatlar için tanımlanan tüm kurallar ve çözümler geçerli olur.

1-Gan,-qan,-gen,-ken (YAn)

yaz + Gan	yaz + YAn
kel + gen	kel + YAn
qaC + qan	qaC + YAn
kUt + ken	kUt + YAn

2- wat,-iwatat,-uwat,-Uwat (iwatat)

yaz + i+ vat +qan	yaz+Ivat+ qan
kUt+i + vat + qan	kUt+Ivat+ qan

3- idi + Gan , +ydi, + Gan (+idi+YAn)

qara + ydi + Gan	qara + idi + YAn
yaz + idi + Gan	yaz + idi + Gan
kOr + idi + Gan	kOr + idi + Gan
dey +ydi + Gan	dey + idi + Gan

4- ar, -er (Ar)

qara + r	qara + Ar
yaz + ar	qara + Ar
kOr + er	kOr + Ar
Eyt + ar	Eyt + Ar

3.6.2.4 Eylemlerden oluşan amaç kipi

Eylemlere “maq, mek” eklerinin eklenmesi ile oluşan eylem isimlerine “ci” eklerinin eklenmesi ile oluşur.

-maq+ci, -mek+ci	(mAY)+ci
yaz + maq + ci	yaz + mAY + ci
kel + mek + ci	kel + mAY + ci
tur + maq + ci	tur + mAY + ci
sOzle+ mek + ci	sOzle + mAY + ci

3.6.3 Eylemlerin haber şekilleri

Uygurcada eylemlerin olayları anlatma yönünden üç sınıfa ayrılırlar.

- 1) Şahıs sınıfı
- 2) Çekim sınıfı
- 3) Zaman sınıfı

3.6.3.1 Eylemlerin şahıs sınıfı

Uygurcada eylemler genelde iyelik ekleri ile gelebilirler. İyelik eklerinin eklenme yöntemleri ise isimlerin farklı şahıs eklerinin eklenmesiyle oluşurlar. Yani eklenirken ünlü ve ünsüz uyuşma kurallarına uyarlar. İsimlere iyelik eklerinin eklenme yöntemleri bu bölümün birinci kısmında anlatılmıştır.

3.6.3.2 Eylemlerin çekim (mood) sınıfı

Uygurcada farklı kipler bulunmaktadır. Bunların birçoğu farklı ekler ile fark edilirken, bazıları ise tümcelerde nasıl bir görev üstelendiğine bakılarak fark edilmektedir.

- Doğrudan beyan çekimi
yazimen (yaz +i) + Şahıs ekleri
yazisen
yazimiz
yazisiler
yazidu

Örnek:

ahmet anasiGa mektup yazidu.

men muellimge xet yazimen.

- Dolaylı beyan çekimi
yEziptimen
yEziptimiz
yEzipsiler
yEzipsiz(yaz +ip) + iyelik eki
yEziptu

Örnek :

ahmet anasiGa mektup yeziptu.

men muellimge het yEziptimen.

- Anlatma beyan çekimi

yEziptudek (yaz +iptu+ DAK)

yEziptudek (yaz +iptu+ DAK)

yEziptumiş (yaz + iptu+ miş)

yEziptumişmen (yaz + iptu + miş +men)

yEziptumişsen (yaz + iptu + miş +sen)

yEziptumişmiz (yaz + iptu + miş +miz)

Örnek:

bu makaleyi ahmet yEziptudek (yEziptimiş).

- Sübjektif değerlendirme çekimi

“Gu” ekinin eklenmesi ile oluşur

Örnek :

Ahmet maqale yazdiGu deymen.

siz tamaka cikisizGu deymen.

- Objektif değerlendirme çekimi

Eylemlerin belli bir kanıt ile yapıldığını belirtir. Genelde “ohSaydu, kErek” gibi sözcükler ile beraber gelirler.

Örnek:

ehmetmu maqale yazGan oxaydu.

ular hazir yiGin ECivatsa kErek.

sen bir yerge barmaqci oxaysen.

men sEnin puliNni alGan oxSaymen.

- Emir – Talep çekimi

Emir ve Talep çekimlerinin aldığı ekler aynıdır, ancak kullandıkları ortama göre ayırt edilir. Çizelge 3.8’de bu çekim için kullanılan ekler verilmiştir.

- Kaygı çekimi

Eylemlere “Gay” eklerinin eklenmesi ile oluşur.

Örnek:

saat beş oldu, Ahmet kEçikip qalmıGay.

Çizelge 3.8 : Eylemlerin emir-talep çekim ekleri.

Şahıs Türleri	Ekler	Örnekler
1- Şahıs	Tekil	Y, ay, ey
	Çoğul	Yli, ayli, eyli
2- Şahıs	Tekil	Eylemin Özeki, Gin, qin, gin, Kin
	Tekil nezaket	N, iN, uN, UN
	Tekil hürmet	sila, sile,
	Çoğul	Nlar, iNlar uNlar, Unlar
3-Şahıs	sun	qara+sun, kel+sun, tur+sun, otur+sun

- Varsayım- itiraz çekimleri.

Varsayım ve itiraz çekimlerinin aldığı ekler aynıdır, ancak kullanıldıkları ortama göre ayırt edilir. Çizelge 3.9’da çekim için kullanılan ekler verilmiştir.

Çizelge 3.9 : Eylemlerin varsayım-itiraz çekim ekleri.

Şahıs Türleri	Ekler	Örnekler
1- Şahıs	Tekil	sam, sem
	Çoğul	saq, sek
2- Şahıs	Tekil	saN, seN
	Tekil nezaket	siNiz
	Tekil hürmet	sila, sile,
	Çoğul	saN+lar, seN+lar
3-Şahıs	sa, se	qara+sa, kel+se, tur+sa, otur+sa

Varsayım ve itiraz çekimlerinin aldığı ekler aynıdır, ancak kullanıldıkları ortama göre ayırt edilir. Çizelge 3.9’da çekim için kullanılan ekler verilmiştir.

- Ümit - teklif çekimi

Bu kip varsayım – itiraz çekimi ile yapı olarak aynıdır. Anlamına bakılarak fark edilir.

- Pişmanlık çekimi

Bu kip varsayım – itiraz çekimi ile gelen fiillere, “boptiken” ya da, “bolGanilen” sözcüklerinin beraber gelmesi ile oluşur. Bazen “-cu” ekinin eklenmesi ile de oluşur.

Örnek:

tunUgUnki dokilatni menmu aNlasam boptiken.

yazGandikin, yaxSiraq yazsaN boptiken.

- Yalvarma çekimi

Emir çekiminde gelen fiillere “-cu” ekinin eklenmesi ile oluşur.

Örnek:

baray + cu

yazay + cu

kelsi + cu

sOzleN + cu

- Zaruriyet çekimi

Varsayım - itiraz çekiminde gelen fiillere “bolidu” sözcüğünün eklenmesi ile oluşur.

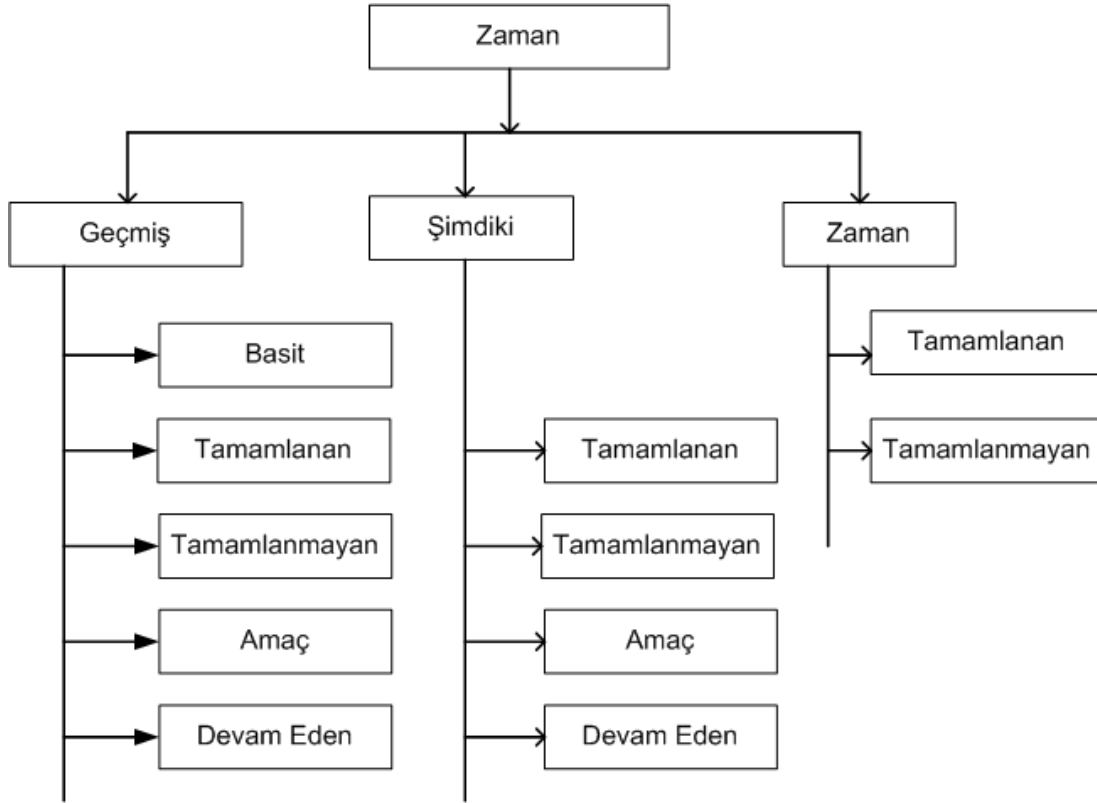
Örnek:

aniNiz bek ansireydiken, pat-pat het yEzip tursiNiz bolidu.

3.6.4 Eylemlerin hal-zaman sınıfı

Uygurcada eylemler genel olarak üç farklı zamanda gelir. Eylem zamanları dolaylı beyan, dolaysız beyan, anlatma beyan, sübjektif beyan, objektif beyan kipleri hal-zaman yönünden birbirine belirgin olarak fark eder. Şekil 3.8’de Uygurcada eylemlerin zaman sınıfları gösterilmiştir [110].

Uygurca eylemlerin zaman sınıflarına bakıldığında, bazı sınıflar, örneğin, “şimdiki tamamlanmayan zaman” şekil ve yapı yönünden, “eylem sıfat” ile aynıdır. Böylece aynı yapı aynı anda sıfat ve eylem olarak yorumlanmaktadır. Aslında Uygurcada eylemler temel olarak, şimdiki zaman, gelecek zaman ve geçmiş zamanlardan oluşur. Şekil 3.8’de gösterildiği sınıflar ise İngiliz diline karşılık olarak sınıflara ayrılmıştır.



Şekil 3.8 : Uygurcada eylemlerin zaman kategorisi.

Bundan dolayı birkaç sınıf, aldığı ek yönünden tamamı ile benzer. Bunlar arasındaki fark sadece anlamsal farktır. Bu anlamsal farkları çıkartabilmek için, tüm tümcenin, tümce bazında çözümlenmesi gerekmektedir. Bazı durumlarda ise tümce bazında değil, sayfa bazında çözümlenmesi gerekmektedir. Bundan dolayı Uygurca eylemlerin sonlu durumlu biçimbilimsel çözümleyicisi tasarlanırken en güncel kaynakların önerdiği kümeleme yöntemleri kullanılmıştır [42], [113]. Şekil 3.9 ve 3.10’da Uygurca eylemlerin sonlu durum geçişleri verilmiştir. Şekillerde gösterildiği gibi Uygurca eylemler, eylem köküne farklı eklerin eklenmesiyle farklı sözcükler ve yapılar oluşur. Eylemlerin olumsuz şekli ise kök sözcüklere “ma /me” eklerin eklenmesiyle oluşur. Şekilde, kareler ara durumlardır. Eylemler ara durumdayken, başka eklerin eklenmesiyle geçişler devam edecek ve son duruma ulaşacaktır. Son durumlar ise çift yuvarlaklar ile belirtilmiştir. Ek almadan geçişler ise “0” ile belirtilmiştir. “0” ise yapısal düzeyde “boş” geçiş anlamını verir. Görünen biçimde hiçbir şey yazdırmaz. Şekilde, zarf, isim ve sıfatlara ulaşan durumlar kare ile gösterilmiştir. İsimlere geçtikten sonra, isimler için geliştirilen çözümleyicilere göre çalışacaktır.

3.6.5 Eylemlerin özel şekilleri

Uygurcada eylemlerin tümcede geçen özneler ve çatı eklerine göre beş şekli bulunmaktadır.

1. Etken eylem
2. Edilgen eylem
3. Dönüslü eylem
4. İşteş Eylemi
5. Ettirgen Eylemler.

- Genelde, etken eylemlerde tümcenin nesnesi bellidir. Yani eylemi gerçekleştiren özne kesindir.

Örnek:

Uy: Ahmet xet yazdi.

Tr: Ahmet mektup yazdı.

- Edilgen eylemlerde ise tümcenin öznesi belli değildir. Yani meçhuldür. Bu tarzdaki eylemlere “meçhul eylem” de denir. Meçhul şekilde gelen eylemlere, çatı ekleri olarak, eylemlere, “il, in, l, n” eklerin eklenmesi ile oluşur.

Örnek:

Uy: yemekler yEpildi.

Tr: yemekler yapıldı.

- Dönüslü eylemde, öznenin yaptığı işin etkisi kendine dönüyorsa bu eylemlere dönüslü eylemler denir. Dönüslü eylemlerde özne genellikle canlı varlıklardır. Uygurcada dönüslü eylem diğer eylemlere, il, ul, Ul, l, n, in, un, Un eklerinin eklenmesi ile oluşur.

Örnek:

Uy: u her zaman mahtinidu.

Tr: O her zaman övünür.

- İşteş eylemi: Bir işin birden çok özne tarafından yapıldığını gösteren eylemlere işteş eylemi denir. İşteş eylemi yapmak için eylemin kök ya da gövdesine, ”S, iS, uS, US” eklerinin eklenmesi ile oluşur.

Örnek:

Uy: biz ular bilen paraNlaStuq.

Tr: biz onlar ile konuştuk.

- Bir işin başkalarına yaptırıldığını belirten eylemlere ettirgen eylemler denir. Genellikle, özne işi yapandır. Ancak ettirgen eylemlerle kurulan tümcelerde özne işi yaptırandır. Ettirgen eylem, eylem kök ya da gövdelerine, “dur, tur, t, Guz, quz, guz, kuz” çatı eklerinin birinin ya da ikisinin üst üste getirilmesiyle kurulur.

Örnek:

Uy: muellim oquGucilarGa kitap oqutti.

Tr: Öğretmen öğrencilere kitap okuttu.

Uy: u ukisini oynatti.

Tr: O kardeşini gezdirdi.

Uygurcada eylemler aldığı çatı ekleri ile bir eylemden ikinci bir eylem oluşur, ya da o çeşit eylem bir daha oluşur.

Örnek:

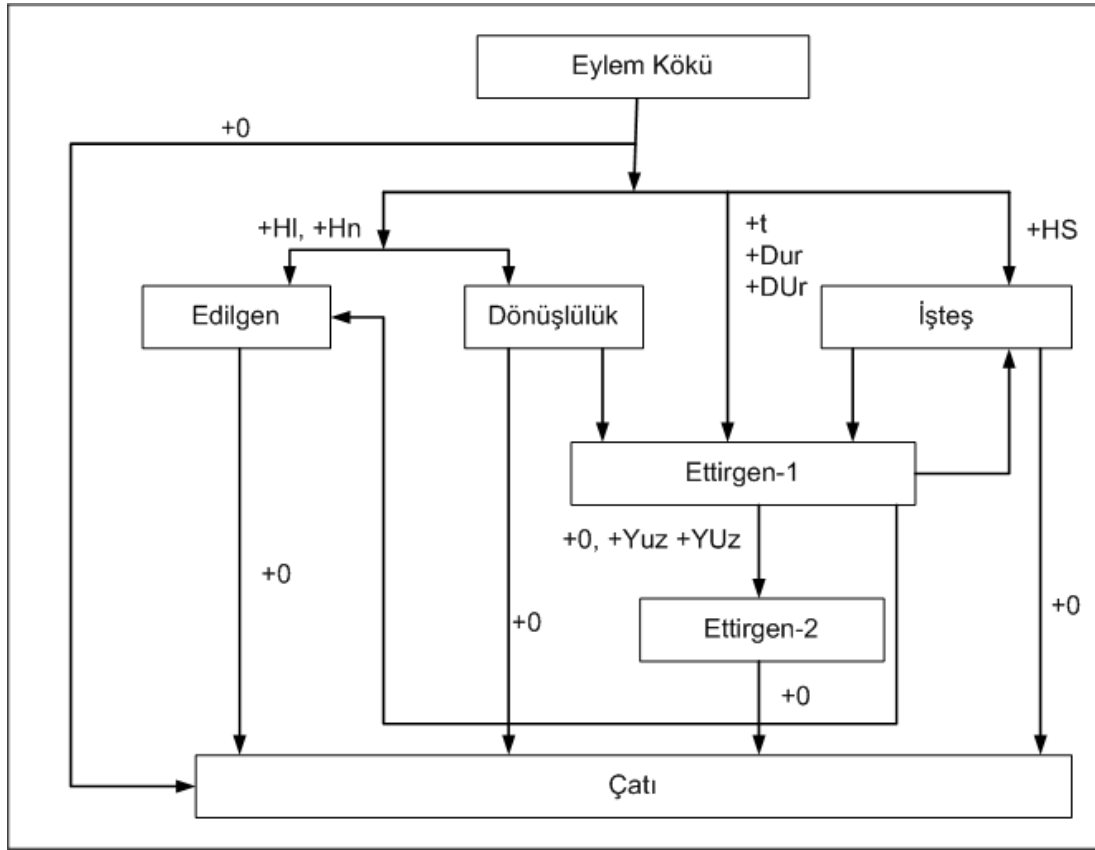
Oyna+t+quz.

Bu örnekte, ”onaya “ kendisi emir kipinde bir eylemdir. “t” ekinin eklenmesiyle, ettirgen eylem oluşur. Sonra “quz” ettiren ekinin bir daha eklenmesiyle, bir daha ettirgen eylem oluşur. Bazen bir çeşit eylem başka bir eyleme dönüşür.

Örnek: kOrUStUrduq. Türkçe anlamı “görüşürdük”. Burada bir işteş eylemi ise ettirgen eyleme dönüşmüş oldu. Uygurca eylemlerin kendi aralarındaki dönüşüm ise Şekil 3.12’de gösterilmiştir.

Uygurca eylemler için geliştirilen ve tasarlanmış biçimbilimsel çözümleyici örnekleri [116]’da verilmiştir.

Sayılar, adılar (zamirler) ve bağlaçların yapısı çok basit olduğundan, bunların çözümleme yöntemi bu tezde anlatılmamıştır. Ama bunlar için geliştirilen çözümler ve kurallar ise EK. B’de verilmiştir.



Şekil 3.11 : Eylem özeklerinin sonlu durum geçişleri.

4. UYGURCA İLE TÜRKÇENİN KARŞILAŞTIRILMASI VE ÇEVİRİ KURALLARI

Uygurca ile Türkçe akraba diller olduğundan söz dizimi birbirine benzemektedir. Ama tümcelerde geçen kök sözcüklere eklenen ekler farklı sözcüklere eklenmektedir ya da aynı anlamı verebilmek için değişmektedir. Örneğin Uygurcada özneye eklenen bir ek, Türkçe karşılığında eylemlere ya da nesnelere eklenmesi gerekmektedir. Yani Uygurcadan Türkçeye çeviri yaparken, genel anlamı ile tümcelerdeki kök sözcükler aynı sırada gelir ama kök sözcüklere eklenmesi gereken ekler aynı sırada gelmez [117]. Bundan dolayı eklerin doğru yere ve doğru sözcüklere eklenmesi gerekmektedir.

Örnek:

Uy: mendin soridi.

Tr: bana sordu.

Bu örnekte, Uygurca cümleinin öznesi olan kök sözcük “men” dır ve “çıkış” eki olan “din” eki eklenmiştir. Ama Türkçe karşılığında ise , “men” sözcüğünün karşılığı olan “ben ” sözcüğüne, “yönlenme” eki olan “ a” eki eklenmiştir.

Bu bölümün alt bölümlerinde, Uygurca ile Türkçe tümce bazında karşılaştırılarak daha ayrıntılı açıklanmıştır ve çeviri kuralları tanımlanmıştır.

4.1 Uygurca ile Türkçenin Söz Dizimi

Genel olarak, Uygurca ile Türkçenin söz dizimi aynıdır. Yani kural olarak aşağıdaki sırayı izlerler. Ama her zaman bu öğelerin bulunması şart değildir. Bazen bu öğeler tekrar edebilir.

Özne -> Belirteç -> İlgeç -> Tümleç -> Yüklem.

Örnek:

Uy: serpil tUnUgUn poyiz bilen mekteptin keldi.

Tr: Serpil dün tren ile okuldan geldi.

Uy: men bazarGa kettim.

Tr: Ben pazara gittim.

Uy: sen su ictiN mu?.

Tr: Sen su içtin mi?

Uy: qoylar taGqa ketti.

Tr: koyunlar dağa gitti.

Uy: men kitap okuwatimen.

Tr: Ben kitap okuyorum.

Yukarıda verilmiş örneklerde, Uygurca tümcelerdeki sözcükler ile Türkçe sözcükler birbirine denk gelmektedir. Bu nedenle doğrudan doğruya sözcükten sözcüğe çeviri yapılabilmektedir.

Örnek:

Uy: men uniNdin bir sual soridim.

Tr: Ben ona bir soru sordum.

Bu örnekte Uygurca tümce de bulunan “uniNdin” sözcüğü, “çıkış” durumunda gelmiştir. Ama onun karşılığı olan Türkçe çevirisinde ise “ona”, yani “yönelme ” durumunda gelir.

Uygurca ile Türkçenin sözdizimi aynı olduğundan, çeviri işlemi yaparken, kaynak tümce yani Uygurca tümce ile Türkçe tümcedeki sözcüklerin sırasını korumak yeterlidir. Bundan dolayı, Uygurca ile Türkçe arsında çeviri işlemi gerçekleştirirken sözcüklerin doğru aktarılması ve sözcükleri çevirirken oluşabilecek belirsizliğin giderilmesi ya da en aza indirilmesi gerekmektedir.

4.2 Uygurca ile Türkçenin Zaman ve Kip Yönünden Karşılaştırılması

Uygurca ile Türkçe, söz dizimlerinin benzemesine karşın, içerdiği ekler yönünden çok farklıdır. Bu nedenle zaman ve kip yönünden birbirinden farklı olmaktadır. Çizelge 4.1’de ise Uygurcanın biçimbilimsel çözümleyicisi ile çözümlenen yapılar ve onların Türkçe karşılığı örnekler ile gösterilmiştir [113].

Çizelge 4.1’de gösterildiği gibi, Uygurcada şimdiki zaman, Türkçede 3 farklı zamana denk gelir.

Uygurcada, şimdiki zaman, anlam yönünden, aynı yapıyı kullanarak, şimdiki zaman, geniş zaman ve gelecek zaman anlamlarında kullanılır. Ama Türkçede geniş zaman, gelecek zaman ve şimdiki zaman tamamen farklı ekler alır.

Örneğin, “bilmek” eylemi, Türkçede de, “bilmek” anlamını verir.

Örnek:

Uy: men Türkçe bilimen

Tr: ben Türkçe bilirim

Uy: ben Türkçe bileceğim

Tr: ben Türkçe biliyorum.

Örnekte verilen Uygurca cümlelerin Türkçede çevirisi ise, yukarıda gösterildiği gibi üç farklı şekilde yazılabilir. Bu durum genelde eylemlerde görülür. Tam olarak hangi zamanda geldiği ise, o an konuşulan duruma göre belirtilmektedir.

Uygurca ile Türkçe arasındaki bir başka belirgin fark ise, Uygurcada gereklilik kipi genel itibarıyla iki sözcükten oluşur. Yani “eylem+ lazım” ya da “eylem+kirek” sözcüklerin birikmesinden oluşur. Oysa Türkçede “eylem+malıyım” şeklinde oluşur.

Örnek:

Uy: men ete mektepke bEriSim kerek.

Tr: Ben yarın okula gitmeliyim.

Uy: biz u yerge barماسلىقىمىز kerek.

Tr: Biz oraya gitmemeliyiz.

Uygurcada bazı kipler tek sözcük ile ifade edilirken, Türkçe karşılığında birden fazla sözcükle ifade edilmek zorundadır. Örneğin Uygurcanın “gelecek zaman rivayeti” Türkçeye çevirilirken, iki sözcük ile çevrilmektedir.

Örnek:

Uy: men taziliq qilidikenmen.

Tr: Ben temizlik yapmak istiyordum.

Çizelge 4.1 : Uygurca ile Türkçenin zaman ve kiplerinin karşılaştırılması.

Zaman-kip	Uygurca	Türkçe
Görülen geçmiş zaman	yaz+di+m yaz+Gan iSli+gen	Yazdım Yazdım(Yazmışım) İşledim(iSlemişim)
Duyulan geçmiş zaman	yaz+Gan+men iSli+gen+men yaz+hp+(ti)+men	Yazmışım Çalışmışım Yazmışım
Şimdiki zaman	yaz+i+vati+men yaz+maq+ta+men	Yazıyorum Yazmaktayım
Geniş zaman (Şimdiki bitmemiş zaman)	yaz+ar+men iSle+r+men	Yazarım Çalışırım
Gelecek zaman	yaz+i+men kUt+i+men yaz+maq+ci	Yazacağım Bekleyeceğim Yazacağım
Şart	yaz+sa+m kUt+se+m	Yazsam Beklesem
İstek	yaz+Gay+men kut+kUm yoq yaz+Gu+m var	Yazayım Bekliyesim yok Yazasım var
Gerekliklik	yaz+sa+m kerek kUt+se+m kerek	Yazmalıydım Beklemeliydim
Emir	yaz+ma+y kUt+ey iSle+y	Yazmayayım Bekleyeyim Çalışayım
Duyulan geçmiş zaman (hikaye)	qil+i+vidim	Yapmış idim
Şimdiki zaman (hikaye)	qili+i+vatatti+m qil+maqta+idim	Yapıyordum Yapmaktaydım
Geniş zaman (hikaye)	qil+ar idim qil+lat+tim	Yapar idim Yapardım
Gelecek zaman (hikaye)	qil+maqci idim qil+idiGan idim	Yapacaktım Yapardım,yapıyordum
Şart (hikaye)	qil+sa+m Edim	Yapsa idim
İstek (hikaye)	qil+Gay idim	Yapaydım
Duyulan geçmiş zaman (rivayet)	qil+Gan iken+men qil+i+ptiken+men	Yapmışmışım Yapmışmışım
Şimdiki zaman (rivayet)	qil+i+vaticiden+men qil+maq+ta ikenmen	Yapıyormuşum Yapıyormuşum
Gelecek zaman (rivayet)	qil+idi+kenmen qil+maqci qil+Gi+dek +miS+men	Yapacaktım Yapmak istiyormuşum Yapıyormuşum, yapacaktım
Şart çekimi (rivayet)	qil+sa+m iken	Yapsaymışım
İstek çekimi (rivayet)	qil+Gay+iken men	Yapamışım
Gerekliklik çekimi (rivayet)	qil+sa+m kerek iken	
Geniş zaman (rivayet)	qil+ar ikenmen qil+mas ikenmen	Yaparmışım Yapmazmışım
Duyulan geçmiş zaman (şart)	qil+Gan bolsam	Yapmış isem
Geniş zaman (şart)	qil+ar bolsam	Yapar isem
Gelecek zaman (şart)	qil+maq+ci bolsam qil+idi+Gan bolsam	Yapacak isem Yaparsam, yapacaksam

Biçimbilimsel çözümleyici sözcük bazında çalıştığından, birden fazla sözcüklerin aynı anda çözümlemesi oldukça zordur.

Bu nedenle, Uygurca ve Türkçe arasında çeviri sistemi tasarlamadan önce, ekler üzerinde araştırmaya ağırlık verilmiştir. Çizelge 4.1’de verilmiş olan örnekler [113]’de verildiği gibi değişik sözcükler ile gösterilmiştir.

4.3 Uygurcadan Türkçeye Çeviri Kuralları

Çeviri kuralları tasarlariken, XFST'nin sunduğu koşullu, koşulsuz ve sıralı değiştirme yöntemleri kullanılmıştır. Bu yöntemler yazmış olduğumuz düzenli ifadelerle göre çalışmaktadır. Bu kurallar daha önce tasarlamış olduğumuz iki seviyeli kurallardan tamamen farklıdır. Bazen birden fazla sözcük ile ilgili biçimbilimsel çözümlemeler ile değiştirmek, yorumlar yapmak zorunda kalınmaktadır. Böyle durumlarda birkaç işlemin aynı anda yapılması gerekmektedir. XFST farklı değiştirme işlemleri yapabilmektedir.

- Koşulsuz değiştirme kuralı: Bu kural ile metinde bir bilgiyi, örneğin biçimi (morphem) değiştirmek istenirse, hiç koşul aranmaz. Örneğin: Uygurcada ikinci şahıs etiketlerinden “A1ss” etiketini Türkçede doğrudan bir karşılığı bulunmamaktadır. Dolayısıyla bu etiket, Türkçede “A1sg” etiketi ile gösterilebilir. Bu nedenle XFST ile “A1sg <- A1ss” bir kural yazılabilir. Böylece, hiçbir kural aranmadan herhangi bir yerde “A1ss” etiketi doğrudan “A2sg” ile değişecektir.
- Koşullu değiştirme kuralı: Bu kural ile belli bir koşul sağlanırsa, sonra değiştirme işlemi gerçekleşecektir. Örneğin Uygurcada “gerekliklik” kipi, eylemlere şahıs ekleri eklendikten sonra, “kirek” ve “lazım” eklerinin eklenmesiyle oluşur. Eğer bu “lazım” ya da “kirek” eki gelmezse, eylemler doğrudan çevrilebilir, yani eylemlerin sadece kendisi ile ilgili biçimbilimsel bilgileri göz önüne alınabilir.
- Paralel değiştirme kuralı: Belli bir koşul sağlandığında birden fazla işlemin aynı anda gerçekleştirilmesidir. Bu kurallar ile ilgili örnekler ise, sonraki bölümlerde daha detaylı olarak gösterilecektir.

Uygurcadan Türkçeye çeviri yaparken, Uygurcanın ve Türkçenin biçimbilimsel çözümleyicilerinde sözcükleri etiketlemek için aynı etiketler kullanılmıştır [30],

[114], [116], [118]. Yani Uygurca için geliştirilen biçimbilimsel çözümleyicilerde, Türkçe için geliştirilen biçimbilimsel çözümleyicilerde kullanılan takıların kullanılmasına özen gösterilmiştir.

Örneğin: Uygurcada birinci şahıs tekil etiketi: “P1sg”, birinci şahıs çoğul etiketi “P2pl”, ikinci şahıs tekil etiketi, “P2sg”, ikinci şahıs çoğul etiketi, “P2pl”, üçüncü şahıs tekil etiketi “P3sg”, üçüncü şahıs çoğul etiketi, “P3pl” ile belirtilmiştir. Diğer etiketlerin anlamı ise EK.C’ de¹⁵ verilmiştir.

Uygurcada bulunan ama Türkçede bulunmayan özel durumlar için, özel etiketler tanımlanmıştır. Uygurcadan Türkçeye çeviri sistemi tasarlarırken, çeviri kuralları tanımlanarak, Uygurca için tanımlanan bu özel etiketlerin Türkçe karşılığı bulunmaya çalışılmıştır.

Böylece, Türkçe için geliştirilen biçimbilimsel çözümleyicinin eski yapısı korunmuş ve aynı anda Uygurca için özel etiketler tanımlanabilmiştir.

Örneğin: Uygurcada, ikinci şahıs için, saygı ile hitap edilirken, “siz” sözcüğü kullanılır ve her zaman tekil anlamı vardır. Ama Türkçede ise, “siz” çoğul anlamı içerir, ayrıca ikinci şahsa karşı saygı anlamı da içerir.

Uygurcada bu “siz” sözcüğü “A2sp” ile belirtilmiştir. Bu sözcüğün çoğul şekli ise, “sizler” ve “A2pp” ile belirtilmiştir.

Ayrıca Uygurcada ,”senler” şeklinde bir sözcük bulunuyor ve yaygın kullanılıyor. Genelde ikinci şahsa, çoğul anlamında kullanılır ve çok aşağılayıcı, eleştirici ve sert bir konuşma tarzı içerir.

Çeviri yaparken, kaynak dilde kullanılan sözcüklerin gerçek anlamının hedef dilde tam anlamı ile aktarılması çoğu zaman eksik kalmaktadır. Bu da anlam kaybına ve anlamın değişmesine sebep olmaktadır. Bu durum sadece bilgisayarlı çeviride değil, insan aracılığıyla yapılan çevirilerde bile görülen genel olaydır.

Bu bölümün alt bölümlerinde Uygurcadan Türkçeye çeviri yapmak için gereken genel çeviri kuralları verilmiştir.

¹⁵ EK. C’de kullanılan etiketler, Türkçe ve Türkmençe için geliştirilen biçimbilimsel çözümleyicilerde kullanılan etiketlerin kullanılmasına özen gösterilmiştir ve bu ekler çizelgesi [32]’den alınmıştır. Uygurcada bulunan özel durumlar için yeni etiketler tanımlanmıştır.

Bilindiği gibi, biçimbilimsel çözümleyiciler bir sözcük ile ilgili olası tüm çözümleri üretmektedir. Bu nedenle bir sözcüğe karşın birden fazla çözümleme söz konusudur. Özellikle bitişken dillerde bir sözcüğe denk gelen çözüm sayısı fazla olmaktadır. Örneğin, Türkçede üçte bir sözcük belirsizlik içermektedir [119]. Bu da belirsizliğe neden olmaktadır. Çeviri kurallarını tasarlarken, bir sözcük ile ilgili birden fazla biçimbilimsel çözümlemeden hangisinin kullanılacağı çok önemlidir.

Çünkü farklı biçimbilimsel çözümlemeler farklı anlamlara neden olmaktadır [86]. Bu nedenle bir sözcük ile ilgili biçimbilimsel çözümlemelerdeki belirsizliğin giderilmesi gerekmektedir.

Uygur dilinin biçimbilimsel belirsizliği ile ilgili çalışmalar bir sonraki bölümde ayrıntılı bir şekilde anlatılmıştır.

Bu bölümde çeviri kuralı tasarlarken, Uygurcanın biçimbilimsel belirsizliliğinin giderildiği ve o sözcük ile en uygun çözümün bulunduğu varsayılmıştır.

4.3.1 Kök sözcüklerin çevirilmesi

Uygurca için geliştirilmiş biçimbilimsel çözümleyici sözcükleri çözümlerken, Türkçe için geliştirilmiş biçimbilimsel çözümleyicide kullanılan etiketi kullanmıştır [30], [114], [116].

Örnek:

Uy: mektep: mektep+Noun+A3sg+Pnon+Nom

Tr: okul: okul+Noun+A3sg+Pnon+Nom

Çözümlemede, “Noun” etiketinin anlamı, çözümlenen sözcüğün “isim” olduğu, “A3sg” etiketinin anlamı, “Tekil 3-Şahıs Uyum Özelliği”, “Pnon” etiketinin anlamı, “Belirsiz İyelik” ve “Nom” etiketinin anlamı, “Yalın Durum” yani her hangi bir durum eki almadığı anlamını verir. Geliştirilmiş biçimbilimsel çözümleyicide, eğer isimler “çoğul” eki almazsa “A3sg” etiketi ile eğer “çoğul” eki alırsa “A3pl” etiketi ile etiketlenmiştir.

Çözümlemede gösterilen “mektep” ve “okul” ise, “kök” sözcüklerdir. Yani bu sözcükler ile biçimbilimsel bilgiler “+” ile gösterilmiştir.

Genelde kök sözcükler farklı biçimbilimsel bilgiler içerebilir ve çeviri, bu biçimbilimsel bilgilere göre yapılır.

Örnek:

Uy: mekteplerimizni: mektep+Noun+A3pl+P1pl+Acc

Tr: okullarımızı: okul+Noun+A3pl+P1pl+Acc

Bu çözümlemede, Uygurca sözcük “mekteplerimizni” nin kök sözcüğü “mektep” ve bu sözcük “isim” dir ve “Noun” etiketi ile etiketlenmiştir. Ayrıca bu sözcük “A3pl”, “P1pl” ve “Acc” etiketleri ile etiketlenmiştir. Bu etiketlerin anlamı:

- A3pl: 3.şahıs çoğul uyum özelliği
- P1pl: 1.şahıs çoğul İyelik eki
- Acc: Belirtme durumu eki

Türkçe sözcük, “okullarımızı” ise, kök sözcük “okul”, “Noun”, “A3pl”, “P1pl” ve “Acc” etiketleri ile çözümlenmiştir.

Bundan dolayı, genel olarak sadece kök sözcükler üzerinde çeviri yaparken, isim, sıfat ve sayı gibi, çevrilecek kök sözcük üzerinden doğrudan çeviri yapılabilir.

Örnek:

[“okul” <- “mektep”]

Bu kuralın anlamı hiçbir koşul aramadan, “mektep” sözcüğünü , “okul” sözcüğü ile değiştir (replace) anlamına gelmektedir.

Bu değiştirme sonucu, sadece kök sözcük değişecektir ve bu sözcükler ile ilgili biçimbilimsel bilgiler üzerinde hiçbir değişiklik yapılmayacaktır. Bu nedenle “mektep” sözcüğüne uygun olarak Türkçeye “okul”, “mekteplerimizni” sözcüğü “okullarımızı” olarak çevrilecektir.

Uygurca ile Türkçede birçok sözcük aynen benzemektedir ve aynı anlamı taşımaktadır.

Örnek:

mektep, kitap, su, ana, yaz, at.

Gerçi bu sözcüklerin kök sözcüğü benzese bile, ekleneceği eklerden dolayı farklı anlamlar oluşturmaktadır.

Bu nedenle sözcük kökü benzese bile bu çeşit sözlüklerin çözümlenmesi ve çeviri kurallarının uygulanması yapılmaktadır.

4.3.2 Sadece biçimbilimsel bilgilerinin değişmesi

Bu durum Uygurcadan kaynaklanan gruplamadan, etiketlemeden oluşur. Yani Uygurcada bulunan birden fazla biçimbilimsel bilginin, Türkçedeki tek bir bilgiye denk gelmesinden kaynaklanır.

Uygurcada bir eylemin tamamıyla yapılmadığını (Imperfect) belirtmek için “ImPer” etiketi kullanılmıştır. Bu etiket de, Türkçenin biçimbilimsel çözümleyicisinde bulunan, geniş zaman etiketi olan “Aor” etiketi ile koşulsuz olarak değiştirilmiştir.

Örnek:

["+Narr"] <- ["+ImPer"]

Uygurcada “yazamen” sözcüğü şöyle çözümlenir

yaz+Verb+Pos+ImPer+A1sg

Bu çözümlemede, sözcük kökünün “yaz” olduğu, sözcük türünün “eylem ” olduğu, bu eylemin, “olumlu” olduğu, ancak bu eylemin kesin olarak “bitmediği” ve bu eylemin , “1.şahıs tekil uyum özelliği” taşıdığı anlaşılmaktadır.

Bu çözümlemede geçen “ImPer” etiketi, Türkçede bulunan “Aor” etiketi ile değiştikten sonra:

yaz+Verb+Pos+Aor+A1sg

Şeklinde değişir.

Bu çözümleme ise, Türkçede bulunan “yazarım” sözcüğü ile aynı anlamı taşımaktadır.

Sonuç olarak, Uygurcada biçimbilimsel çözümleyicide bulunan “ImPer” etiketinin hepsi, Türkçede bulunan “Aor” etiketi ile koşulsuz değiştirilmiştir.

Bu değişim kuralı, sözcük köklerinden bağımsız olarak değişmektedir.

4.3.3 Biçimbilimsel bilgilerin ve sözcüklerin aynı anda değişmesi

Çözümlemelerdeki biçimbilimsel bilgilerin ve sözcüklerin aynı anda değişme durumları. Uygurca sözcükler çözümlenirken köklerinin farklı olması ve bu bilgilere karşılıklı gelen ya da gelmesi gereken, Türkçede bulunan biçimbilimsel belirticisinin de farklı bilgi oluşturmasından kaynaklanır. Örneğin: ”A2sr”, “A2pr” gibi etiketler. Bu etiketlerin Türkçede tam karşılığı yoktur. Uygurcada ikinci şahıs için “saygı”

belirtmek için tekil durumlarda “A2sr” etiketi kullanılırken, ikinci şahıs için “nezaketini” belirtmek için de “A2pr” etiketi kullanılmıştır.

Ancak Türkçede, ikinci şahıs için sadece iki etiket bulunmaktadır. Yani ikinci şahıs tekil uyumu için “A2sg” kullanılırken, ikinci şahıs çoğul uyumu için ise, “A2pl” etiketi kullanılmıştır. Ancak, Türkçede ikinci şahıs için saygı ve nezaketi göstermek içinse, ikinci şahıs çoğul uyum etiketi olan, “A2pl” kullanılır.

Sonuç olarak, Türkçede ikinci şahıs için ayrıca nezaket etiketi içermediğinden, Uygurcada bulunan nezaket etiketleri, doğrudan Türkçede bulunan ikinci şahıs çoğul uyum etiketleri ile değiştirmek üzere aşağıdaki çevir kuralları tanımlanmıştır.

[“A2pl” <- “A2sr”]

[“A2pl” <- “A2pr”]

Örnek:

Uy: sili: sili+Pron+Pers+A2sr+Pnon+Nom

Tr: siz: siz+Pron+Pers+A2pl+Pnon+Nom

Yani, Uygurcada bulunan ve ikinci tekil şahıs için “saygıyla” söylenen “sili” zarfı şöyle çözümlenmiştir.

Sözcük kökü “sili” dir. Bu sözcüğün türü “Zamir” (Pron) dir ve bu zamir kişisel zamir (Pers) kümesindedir. Bu zamir ikinci şahıs için “tekil saygı” uyumu sağlar (A2sr), iyelik eki taşımamaktadır yani belirsiz iyelik (Pnon) dir ve durum eki taşımamaktadır, yalın durum (Nom) dur.

Bu çözümlmeyi Türkçeye çevirmek için, sözcük kökü “sili” ise “siz” e çevirilirken, Uygurcada bulunan, ikinci şahıs tekil saygı etiketi “A2sr” ise, Türkçede bulunan ikinci şahıs çoğul uyum eki ile değiştirilmiştir.

4.3.4 Eylemlerden dolayı bir önceki sözcüğün farklı ekler ile değişmesi

Uygurcada bazı eylemleri içeren tümceler, Türkçeye doğrudan çevirildiğinde yanlış sonuç ortaya çıkmaktadır. Örneğin, Uygurcada bir kişiden bir şey sorulduğunda şöyle yazılmaktadır.

oquGucilardin soridi.

Bu sözcükler şöyle çözümlenir.

oquGucilardin: oquGuci + Noun+A3pl+Pnon+Abl

soridi: sora+Verb +Pos+Past +A3sg

Çözümlemelerin anlamı ise; “oquGucilardin” sözcüğünün, kökü, “oquGuci”, “İsim (Noun)” dır. Şu özelliklerdedir: 3-şahıs çoğul uyum durumu (A3pl), belirsiz iyelik (Pnon), ve çıkma durumu(Abl) da gelen.

İkinci sözcük “soridi” sözcüğün kök sözcüğü, “sora”, bir eylem(Verb) dir. Bu eylem olumlu (Pos), geçmişte yapılmış (Past) ve 3-şahıs tekil uyum durumundadır.

Bu biçimbilimsel bilgilere göre bu iki sözcükten oluşmuş tümce Türkçeye çevrilirken yalnızca kök sözcükler çevrilirse aşağıdaki gibi yapı oluşmaktadır.

Öğrencilerden: öğrenci+Noun+A3pl+Pnon+Abl

Sordu: sor+Verb+Pos+Past+A3sg

Sonuç olarak, Türkçede “öğrencilerden sordu” şeklinde bir sonuç elde edilir. Ancak Türkçenin dilbilgisine göre bu yanlış tümcedir. Doğrusu, “öğrencilere sordu” şeklinde olmalıdır. Türkçenin dil bilgisine göre yazılan bu tümce şöyle çözümlenir.

Öğrencilere: öğrenci + Noun +A3pl +Pnon +Dat

Sordu: sor+Verb +Pos +Past +A3sg

Uygurca tümcede bulunan “oquGucilardan” sözcüğündeki, “Çıkma” durum etiketi (Abl), Türkçede bulunan “Yönelme” durum etiketinin “Dat” ile değişmesi gerekmektedir. Bu değişimin gerçekleşmesinin koşulu bu sözcüğün kesin olarak “sora” eyleminden önce gelmesidir.

[Dat <- Abl || _ [%]+ ["sora"] %+ Verb]

Bu koşul oluştuğunda “sora” eylemi “Çıkma” durumu içeren herhangi bir sözcükten sonra gelirse “Çıkma” durumu bilgisi “Yönelme” durumuna dönüştürülecektir.

4.3.5 Tümce anlamına göre biçimbilimsel bilgilerin değişmesi

Uygurca tümcelerde ifade edilen anlamı aynen Türkçeye aktarmak için bazı durumlarda peş peşe gelen her iki sözcüğün biçimbilimsel yapısının değiştirilmesi gerekir.

Örneğin, Uygurcada birinin bir yerden geldiğini başkalarına anlatırken veya Ankara’dan geldiğini anlatırken, altta verilmiş olduğu gibi bir yapı kullanılır.

kelgen yerim Ankara.

Bu tmcede yer alan szckler Uygurcanın bimbilimsel mleyicisi ile ldğnde altta verildiğ gibi sonular elde edilir (Burada bimbilimsel belirsizliğn giderildiğ kabul edilmiřtir.)

kelgen: kel+Verb+Pos^DB+Adj+PastPart+Pnon

yErim: yer+Noun+A3sg+P1sg+Nom

Ankara: Ankara+Noun+Prop+A3sg+Pnon+Nom

Bu szckler ile ilgili bimbilimsel mlemelerde, daha nceki bimbilimsel mlemelerden farklı olarak, “^DB” etiketi bulunmaktadır.

“^DB” etiketi, bir řit szckten, eklerin eklenmesiyle farklı řit szcğn tretildiğni belirtir. rneğn “kelgen” szcğnde geen bimbilimsel mlemelerin verdiğ anlam ise: kelgen szcğnn “kk” szcğ, “kel” ise, bu szck eylem (Verb)ise, olumlu (Pos) dur. Bu szcğe “eylemden” sıfat yapma eki “gen”in eklenmesiyle, sıfat tretilmiřtir “Adj”. Burada aslında “eylem ” olan bir szckten, bir sıfat “Adj” tretilmiřtir. Bu nedenle “^DB” etiketi kullanılmıřtır.

Burada oluřan sıfat (Adj) ise anlam olarak, gemiřte tamamlanmıř olan bir olayı belirttiğnden dolayı “gemiř zaman orta” etiketi olan “PastPart” ile belirtilmiřtir. Ayrıca bu eylem her hangi bir řahsı belirtmediğnden, belirsiz iyelik etiketi “Pnon” ile belirtilmiřtir, belirsiz iyelik ve yalın durumdadır.

Tmcede yer alan ikinci szck, “yErim” ile ilgili bimbilimsel mlemelerin anlamı ise: “yErim” szcğnn “kk” szcğ ise “yer”, bu szck “İsim” (Noun), 3-řahıs tekil uyumu (A3sg), 1-řahıs iyelik eki (P1sg) ve yalın durum (Nom)dadır.

Tmcede yer alan nc szck “Ankara” ile ilgili bimbilimsel mlemelerin anlamı ise, “Ankara” szcğnn “kk” szcğ “Ankara” dır. Bu bir “zel” (Prop) isimdir, 3-řahıs tekil uyumundadır (A3sg). Bu szck belirsiz “iyelik (Pnon) ve “yalın” durumdadır (Nom).

Eğr bu Uygurca Tmceyi mlendiğ bimbilimsel bilgisi ile doğrudan Trkeye evirilirse, ařağdaki gibi bir eviri szcğ oluřur.

Gelen: gel+Verb+Pos^DB+Adj+PresPart

Yerim: yer+Noun+A3sg+P1sg+Nom

Ankara: Ankara+Noun+Prop+A3sg+Pnon+Nom

Bu çözümlemeden, “gelen yerim Ankara” şeklinde bir Türkçe tümce oluşturulur. Ancak böyle bir tümce, Türkçe dilbilgisine aykırıdır.

Türkçe dilbilgisine göre olması gereken tümce ise: “geldiğim yer Ankara” şeklindedir.

“geldiğim yer Ankara” tümcesini oluşturan sözcüklerin biçimbilimsel çözümlenmesi aşağıda verilmiştir.

Geldiğim: gel+Verb+Pos^DB+Adj+PastPart+P1sg

Yer: yer+Noun+A3sg+Pnon+Nom

Ankara: Ankara+Noun+Prop+A3sg+Pnon+Nom

Uygurca tümcede yer alan, “yerim” sözcüğündeki, birinci şahıs iyelik eki (P1sg) ise, belirtisiz iyelik eki “Pnon” ile değişecektir. Aynı anda bu birinci şahıs iyelik eki (P1sg) kendinden bir önceki “gelen” sözcüğünde yer alan geniş zaman bağlayıcı “PresPart” etiketi ile değişmelidir. Birde bu etiketten sonra belirsiz iyelik eki “Pnon” eklenmelidir. Yani burada paralel değişim söz konusudur ve koşullar sağlanmalıdır. Buradaki koşul örneğin, “yerim“ sözcüğünde geçen “iyelik” uyum ekidir ve hangi şahıs ise, o şahıs uygun olarak bir önceki sözcüğe o şahsın iyelik eki aktarılmalıdır.

Bu kural ile ilgili tanımlanan paralel kural aşağıda verilmiştir.

[PastPart %+ P1sg <- PresPart | | DB %+Adj %+ _ [%]+ [?+ %+ Noun
%+ +? %+P1sg] , ,

Pnon <-P1sg | | DB %+Adj %+PresPart [%]+ ?+ %+Noun %+ [Inf %+]*
[A3sg | A3pl] %+ _]

Bu kuralda iki tane değişme işlemi aynı anda paralel olarak gerçekleşmektedir ve gerçekleşen işlemler:

PastPart %+ P1sg <- PresPart

Pnon <-P1sg

“PresPart” etiketi “PastPart+P1sg ” etiketi ile değişirken, bir sonraki sözcükteki “P1sg” etiketi ise, “Pnon” etiketi ile değişmiştir. Bu kuralda verilen “%” karakteri XFST de tanımlanan özel karakterleri ayırt etmek için kullanılmıştır. Çünkü XFST’de “+” karakterinin anlamı bir karakterin bir ya da birden fazla tekrar ettiği

anlamı verir. Ancak bu tanımlanan kuralda “+” sadece bir karakter anlamı taşımaktadır.

Bu iki değiştirme işlem aynı anda paralel olarak gerçekleştirildiğinden “ , , ” işareti ile birleştirilmiştir.

Bu kural sadece birinci şahıs tekil iyelik uyum durumu için tanımlanmıştır. Birinci şahıs çoğul iyelik uyum durumu için aşağıdaki gibi tanımlanır.

[PastPart %+ P1pl <- PresPart || DB %+Adj %+ _ [%] + [?+ %+ Noun %+] + ? %+ P1pl , ,

Pnon <-P1pl | | DB %+Adj %+PresPart [%]+ ?+ %+ Noun %+ [Inf %+]* [A3sg | A3pl] %+ _]

Bu kuralda da iki tane değişme işlemi aynı anda paralel olarak gerçekleşmektedir ve gerçekleşen işlemler:

PastPart %+ P1pl <- PresPart

Pnon <-P1pl

“PresPart” etiketi “PastPart+P1pl ” etiketi ile değişirken, bir sonraki sözcükteki “P1pl” etiketi ise, “Pnon” etiketi ile değişmiştir.

Diğer şahıs iyelik uyum ekleri için tanımlanan çeviri kuralları, EK. D’de verilmiştir.

4.3.6 Birden fazla sözcüğün bir sözcük ile değişmesi

Uygurcanın dil yapısı ve kullanılan sözcük, deyim yönünden, bazen birden fazla sözcük ile ifade edilen bir durum, eylem ya da olay, Türkçede sadece bir sözcük ile ifade edilebilmektedir.

Örneğin, Uygurcada Türkçede olduğu gibi tek sözcükten oluşan gereklilik kipi yoktur. Gereklilik kipi genelde eylemlerin sonuna “gerek” veya “lazım” gibi zarfların eklenmesi ile yapılır. Ancak Türkçede eylemlerin sonuna sadece “malı” ekinin eklenmesi ile yapılır.

Türkçede “yazmalıyım” sözcüğünü Uygurcada gösterebilmek için, “yEziSim kirek” gibi iki sözcükten oluşan bir tümce oluşturulur.

Bu tümcede yer alan sözcükler üzerinde biçimbilimsel çözümleme yaparken aşağıdaki gibi sonuçlar elde edilir.

yEziSim: yaz+Verb^DB+Verb+Recip+Pos+Neces+A1sg

kirek: kirek+Adverb

“yEziSim” sözcüğünün kök sözcüğü ise, “yaz” dır. Diğer etiketlerin anlamı ise:

- Verb: belirtilen “kök” sözcük ise, eylemdir.
- ^DB+Verb+Recip: kök sözcüğe “iS” ekini eklenmesiyle bir başka “eylemin” oluştuğunu belirtiyor. Buradan oluşan yeni eylem ise, “Dönüştürme” eylemidir.
- Pos: Eylemin olumlu olduğunu belirtir
- Neces: Gerekliklik kipi
- A1sg: 1-tekil şahıs özelliği

“kirek” sözcüğü ile ilgili biçimbilimsel çözümleme ise:

kirek: kirek +Adverb

Bu biçimbilimsel çözümlemenin anlamı “kirek” tir, “kök” sözcüğü “kirek” tir, ve bu sözcük zarftır (Adverb).

Ancak, bu Uygurca tümcenin karşılığı olan Türkçe sözcük “yazmalıyım” ın biçimbilimsel çözümlemesi:

yazmalıyım: yaz+Verb+ Pos+ Neces+ A1sg

Bu çözümlemenin anlamı ise, “yazmalıyım” sözcüğünün “kök” sözcüğü ise, “yaz” dır ve başka etiketlerin anlamı ise:

- Verb: kök sözcük ise, eylemdir.
- Pos: kök sözcük, eylem, olumludur
- Neces: Gerekliklik kipi
- A1sg: 1-tekil şahıs özelliği

Uygurca tümcede yer alan “kirek” sözcüğü ile “yaz” sözcüğünün çözümlemesinde yer alan “^DB+Verb+Recip” etiketleri, Türkçe karşılığı olan “yazmalıyım” sözcüğünün biçimbilimsel çözümlemesinde yer almaz.

Yani Uygurca tümcelerde gereklik kipi, Türkçeye çevrilirken, Uygurca gereklik kipi oluşturan “kirek” ve “lazim” zarflarının silinmesi ve aynı anda, “kirek”, ve

“lazım” zarfında bir önce gelen eylemlerde geçen fazla biçimbilimsel bilgilerin silinmesi gerekir.

Uygurcandan Türkçeye birinci şahıs ile ilgili gereklilik kipini çevirmek için aşağıdaki çeviri kuralı tanımlanmıştır.

```
[[ ]<- ^DB %+Verb %+Recip | | %+Verb _ %+Pos %+Neces %+A1sg  
WhiteSpace [kirek |lazım] %+Adverb , ,  
[ ] <- [kirek | lazım] %+Adverb | | %+Verb^DB %+ ?+ %+Neces %+A1sg  
WhiteSpace _ ]
```

Burada yapılan işlem:

```
[ ]<- ^DB %+Verb %+Recip  
[ ] <- [kirek | lazım]
```

Burada yapılan işlem, “^DB %+Verb %+Recip” etiketinin ve “[kirek | lazım]” sözcüklerin aynı anda silinmesidir.

Bu kural koşullu çalışır. “^DB %+Verb %+Recip” silinmesi için “WhiteSpace” ile ayırt edilen bir sonraki sözcük “kirek” ya da “lazım” olmalıdır. Eğer bu koşul sağlanmazsa, silme işlemi gerçekleşmez.

Eğer koşul sağlanırsa “^DB %+Verb %+Recip” silinecektir ve aynı anda, “kirek” ya da “lazım” sözcükleri ve bu sözcükler ile ilgili biçimbilimsel bilgisi “Adverb” silinecektir.

Burada sadece birinci şahıs ile ilgili gereklilik kipinin çevirilmesiyle ilgili kural verilmiştir. Diğer şahıslar ile ilgili çeviri kuralları ise, EK. D’de verilmiştir.

4.3.7 Bir sözcüğün birden fazla sözcük ile değişmesi

Uygurcada bir sözcüğe birden fazla sözcüğün denk geldiği durumlar bulunmaktadır. Bazen Uygurcada bir sözcüğü Türkçede birden fazla sözcüğün birleşmesi ile göstermek zorunludur.

Uygurcada soru eki sadece “mu” ekini bir önceki sözcüğe eklenmesiyle oluşur. Ancak, Türkçede soru eki dört tanedir,”mı, mi, mu, mü” ve soru sözcüklerden ayrı yazılmaktadır. Aynı anda bir önceki sözcük ile kalın ince ses uyumunun sağlanması zorunludur.

Biçimbilimsel çözümlemelerde, ekler ile ilgili, kalın ya da ince, bilgilerin edinilmesi imkânsızdır. Bundan dolayı, önce Türkçe soru sözcüğü elde edilecektir ve daha sonra Türkçeye çevirildikten sonra, ses uyumunu sağlamak için, oluşan Türkçe tümcede düzenleme yapılır.

Örneğin, Uygurca soru sözcüğü “yazdimu” sözcüğünün çözülmesi ise:

yazdimu: yaz+Verb+Pos+Past+A3sg+QuesMu

Bu biçimbilimsel çözümleme, “yazdimu” sözcüğünün “kök” sözcüğünün “yaz” olduğunu belirtir. Başka etiketlerin anlamı ise:

- Verb: “kök” sözcük ise eylemdir
- Pos: eylem olumludur
- Past: Belirli Geçmiş zaman
- A3sg: 3-şahıs ekil uyumu
- QuesMu: “mu” ekinin eklenmesi ile oluşan soru zamiri

Uygurca “yazdimu” sözcüğünün Türkçe karşılığı ise, “yazdı mı”, iki ayrı sözcük ile belirtilir ve bu iki sözcük, “yazdı mı”, ile biçimbilimsel çözümlemeler aşağıda verilmiştir.

yazdı: yaz+Verb+Pos+Past+A3sg

Bu biçimbilimsel çözümlemenin anlamı ise: “yazdı” sözcüğünün “kök” sözcüğü, “yaz” dır. Diğer etiketlerin anlamları ise:

- Verb: “kök” sözcük ise eylemdir.
- Pos: Eylem ise olumludur.
- Past: Belirli geçmiş zaman.
- A3sg: 3. şahıs tekil uyumu.

mı: mı+Ques+Pres+A3sg

Bu biçimbilimsel çözümlemenin anlamı ise: “mı” sözcüğünün “kök” sözcüğü, “mı” dır. Diğer etiketlerin anlamları ise:

- Ques: Soru zamiridir.
- Pres: Şimdiki zaman.

- A3sg: 3. şahış tekil uyumu.

Bu nedenle Uyurca soru sözcüğü “yazdimu” ile ilgili yapıyı Türkçeye çevirmek için, “yazdimu” sözcüğünün biçimi için bilimsel çözümlemesi ile ilgili bilgiler üzerinde parçalama ve ayırma işlemi yapılması gerekmektedir.

Bundan dolayı Uyurca soruların Türkçeye çevrilmesi için bir koşulsuz çeviri kuralı tanımlanmıştır.

WhiteSpace "mi+Ques+Pers+A3sg" <- "+QuesMu"

Bu kuralın anlamı, Uyurca sözcükler ile ilgili biçimbilimsel çözümlemelerde, “QuesMu” etikti, doğrudan ve hiç koşul aramadan:

" WhiteSpace mi+Ques+Pers+A3sg" etiketleri ile değiştirilecektir.

“WhiteSpace” anlamı ise bir boşluktur.

Burada "mi+Ques+Pers+A3sg" etiketleri, Türkçede bulunan “mi” soru ekini oluşturmak içindir. Aslında Türkçede, “mı”, “mu”, “mi” ve “mü” gibi dört farklı soru eki vardır. Bu ekler genelde bir önce gelen sözcükler ile ses uyumu sağlayarak eklenirler.

Örnek:

- yazdın mı?
- geldin mi?
- gördün mü?
- okudun mu?

Uyurca biçimbilimsel çözümleme aşamasında, Uyurca “kök” sözcüğün Türkçe karşılığı henüz çevrilmediğinden ve Türkçe sözcüğün görünen şekli elde edilmediğinden, bu dört farklı ekin hangi birinin kullanılmasına karar verilemez. Bu nedenle doğrudan “mi” oluşturan biçimbilimsel çözümleme varsayılan çözüm olarak kabul edilmiştir.

Bu kural sonucu, Uyurca soru sözcüğü “yazdimu” ile ilgili biçimbilimsel çözümleme aşağıdaki gibi iki biçimbilimsel çözümlemeye ayrılır.

yazdimu: yaz+Verb+Pos+Past+A3sg+QuesMu

yazdimu: yaz+Verb+Pos+Past+A3sg WhiteSpace mi+Ques+Pers+A3sg

Bu biçimbilimsel bilgilere dayanarak Türkçe çözüm elde edilirken, “yazdı mı” şeklinde bir çeviri elde edilir. Burada soru eki “mı” ile bir önceki sözcük “yazdı” arasında ses uyumu sağlanmamıştır.

Türkçe dil bilgisine göre bu bir yanlış çeviridir. Ancak, böyle bir çeviri yapıldıktan sonra, Türkçe soru cümlelerinde ses uyumu kontrolü yapılmıştır ve “mı” soru eki, “mi” eki ile değiştirilmiştir. Bu nedenle Uygurcadan Türkçe çeviri elde edildikten sonra ses uyumu kontrolü yapılmıştır. Bunun için, karma yöntemde önerilen yüzeysel biçim düzeyinde tanımlanan tümce geneli kurallar doğrudan kullanılmıştır [32].

4.3.8 “İdi” yardımcı eyleminin çevirilmesi

Uygurcada birçok yardımcı eylem bulunmaktadır. Bu yardımcı eylemle yalnız kullanıldığında herhangi bir anlam taşımazken, başka eylemler ile beraber kullanılarak, o eylemlere ek anlam ekleme görevi üstlenir.

Örnek:

Uy: kelgen idi

Tr: gelmiş

Uy: yazGan idi

Tr: yazmış

Örneklerde gösterildiği gibi, “idi” yardımcı eylemleri, “kelgen”, ve “yazGan” eylemlerine ek anlamı ekleme görevi üstleniyor. Türkçe karşılığı ise “gelmiş” ve “yazmış” gibi tek sözcükler olur.

“Kelgen idi” sözcüğünün biçimbilimsel çözümü altta verilmiştir.

kelgen: kel+Verb+Pos^DB+Adj+PresPart

Bu biçimbilimsel çözümlemenin anlamı, kelgen sözcüğünün , “kök” sözcüğü ise “kel” dir. Başka etiketlerin anlamı ise:

- Verb: kök sözcük ise eylemdir.
- Pos. Eylem ise olumludur.
- ^DB+Adj: ise Türetilmiş sıfat.
- PresPart: Geniş zaman ortacı.

İdi: i+Verb+Pos+Past+A3sg

Bu biçimbilimsel bilgilerin anlamı, sözcük “kök” ise, “i” dir ve başka etiketlerin anlamı ise:

- Verb: Kök sözcük ise eylemdir.
- Pos: Eylem ise olumludur.
- Past: Belirli Geçmiş zaman.
- A3sg: 3. şahıs tekil uyum durum.

Bu sözcüklerin Türkçe karşılığı olan “gelmiş” sözcüğünün biçimbilimsel çözümlenmesi:

Gelmiş: gel+Verb+Pos+Narr+A3sg

Bu çözümlemelerin anlamı, “Gelmiş” sözcüğünün “kök” sözcüğü, “gel” dir. Başka etiketlerin anlamı ise:

- Verb: Sözcük kökü ise eylemdir.
- Pos: Bu eylem ise olumludur.
- Narr: Belirsiz geçmiş zaman.
- A3sg: 3-şahıs tekil uyum durum.

Bu nedenle, Uygurca sözcüklerin biçimbilimsel çözümlemesinde geçen “i” eylemi ve bu eylemin biçimbilimsel bilgileri silinecektir. Aynı anda, “i” yardımcı eyleminde bir önce gelen eylemde bulunan “^DB+Adj+PresPart” etiketleri “Narr+A3sg” ile değişecektir.

Uy: kel+Verb+Pos^DB+Adj+PresPart WhiteSpace

i+Verb+Pos+Past+A3sg

Tr: gel+Verb+Pos+Narr+A3sg

Bu değiştirme işlemini gerçekleştirmek üzere aşağıda bulunan koşullu çeviri kuralı tanımlanmıştır.

```
" +Narr+A3sg" <- % ^ DB % + Adj % + PresPart || % + Verb % + Pos _  
WhiteSpace + i % + Verb % + Pos % + Past % + A3sg ,,
```

[%]<- i %+ Verb %+ Pos %+ Past %+ A3sg || %+ Verb %+ Pos %^ DB %+
Adj %+ PresPart [%]+ _]

Burada yapılan paralel deęiřtirme:

+Narr+A3sg <- ^DB + Adj + PresPart

[] <- i + Verb + Pos + Past + A3sg

Bu bir kořullu kuraldır ve kořul ise, “i” yardımcı ekinin olmasıdır. Bu kuralda sadece birinci řahıs ile ilgili çeviri kuralı verilmiřtir. Bařka řahıslar iin verilen kurallar ise EK. D’de gsterilmiřtir.

5. BELİRSİZLİKLERİN GİDERİLMESİ

Belirsizlik giderilmesiyle ilgili çalışmalar, doğal dil işlemenin en temel çalışmalarını oluşturur. Belirsizlik çalışmaları farklı düzeylerde yapılmaktadır. Örneğin anlamsal belirsizlik ve biçimbilimsel belirsizlik vs.

Bu bölümde geliştirilecek yeni sistemin temelini oluşturan biçimbilimsel belirsizlik giderme ve anlamsal belirsizlik giderme çalışmaları örnekler ile ayrıntılı biçimde anlatılmıştır. Ardından bu çalışmalar ile ilgili sonuçların sayısal değerleri ile verilmiştir.

5.1 Biçimbilimsel Belirsizliğin Giderilmesi

Çeviri sisteminde, harf düzeltme işlemi tamamlandıktan sonra, Türkçeye çevirilecek Uygurca tümceler, tümce bazında sisteme gönderilecektir. Sisteme gönderilen Uygurca sözcükler üzerinde biçimbilimsel çözümlemeleri üretilecektir.

Genelde biçimbilimsel çözümleyiciler, bir sözcük ile ilgili olarak olası tüm çözümlemeleri üretir. Bu nedenle çeviri işlemleri için, üretilmiş çözümlemelerden hangisinin kullanılacağına karar verilmesi gerekir. Üretilen çözümlemelerden hangisinin kullanılacağına karar verilmesi ise biçimbilimsel belirsizliğini giderme olarak adlandırılır.

Biçimbilimsel belirsizlikler ile ilgili çalışmalar doğal dil çalışmalarının sonraki aşamaları için önemlidir. Çünkü bir sözcük ile ilgili ne kadar ayrıntılı bilgi edinebilirsek o sözcük ile doğru işlem yapabiliriz. Aynı anda bilgisayarlı çeviri işlemleri için, üretilmiş tüm çözümlemeler arasından, kaynak tümce sözcükleri hedef tümceye en uygun şekilde aktarabilecek bir çözümün olması hedeflenir. Bu da bir sözcük ile ilgili biçimbilimsel çözümlemenin, tümcede yer alan diğer sözcüklerin biçimbilimsel çözümlemeleri ile uyum sağlaması anlamına geliyor. Tümcede yer alan sözcük sayısı fazla olunca, özellikle tümcelerde yer alan sözcükler ile ilgili biçimbilimsel çözümleme sayısı fazla olunca, biçimbilimsel belirsizlik çalışmaları daha karmaşık bir duruma dönüşür. Bu nedenle, biçimbilimsel belirsizlik

özmlerleriyle ilgili eřitli alıřmalar yapılmıřtır. Trkiye Trkesi iinde ok nemli alıřmalar yapılmıřtır [53], [86-88], [120].

Bu tez alıřması ncesinde Uygurcanın biimbilimsel belirsizliklerinin giderilmesi ile ilgili herhangi bir alıřmaya rastlanmamıřtır. Bu nedenle Uygurcanın biimbilimsel belirsizliėi ile ilgili alıřma yapılmıřtır ve Uygurcanın belirsizliėi ile ilgili olarak, rastgele 144 szck zerinde test yapıldıėında, 1 szcėe karřın 1.4 oranda zm rettiėi hesaplanmıřtır [121]. Bu demektir ki, ortalama her 2 szcėe karřın 3 biimbilimsel zmlemesi denk gelmektedir.

Bu tez alıřmasında Uygurcanın biimbilimsel belirsizliėini gidermek iin kural tabanlı bir belirsizlik giderici geliřtirilmiřtir.

Kural tabanlı biimbilimsel belirsizlik gidericiler ile ilgili kořullar, genel olarak, tmcede geen szck ile ilgili zmlerinin, tmcede belli bir yapısal dzeydeki forma ya da yapısal dzeydeki yer zelliklerine gre kořullandırmasıyla oluřuyor [86]:

- Yapısal dzeydeki kesin ya da greli yerine (tmcenin bařlangıcı ya da sonuna gre veya o anki szckten bir sonraki szcėe gre).
- Kk szcėe ve en son ekim grubuna gre.
- Tretimsel trne gre.
- Durum ekleri, řahısların iyelik ekleri ve belli anlamsal formlara gre.
- Eylemlerin uyumu, eylemlerin kiplerine uyumuna gre.

Biimbilimsel belirsizlik alıřmaları doėal dil alıřmalarının karmařık konularından biridir. nk bu konu sadece mhendislik bilgisi deėil, aynı zamanda, alıřılacak dil ile ilgili derin dilbilgisi gerektirir. zellikle, kural tabanlı biimbilimsel belirsizlik gidericilerde dil bilgisine olan talep n plana ıkmaktadır.

Kural tabanlı Uygurca biimbilimsel belirsizlik giderici ile ilgili kořullar, Uygurca dil bilgilerine gre tanımlanmıřtır [42], [110-113]. Geliřtirilen biimbilimsel belirsizlik gidericinin iřlevi, alt blmlerde rnekler ile anlatılmıřtır.

5.1.1 Eylemlerde belirsizliėin emir kipine gre giderilmesi

Uygurcada bazı eylemler zmlenirken, aynı anda, isim, sıfat, eylem tipinde zmleniyor. rneėin, at, yaz, bar gibi eylemler.

Uygurcada, genel olarak eylemler tümcenin sonunda gelir, özellikle “emir kipinde” gelen eylemler her zaman tümcenin sonunda gelir.

Uygurcanın bu özelliğinden yararlanarak, eylemlerin emir kipine göre şöyle kurallar vardır: Eğer tümcenin sonuna gelen sözcüğün biçimbilimsel çözümlemesinde, eylem varsa ve bu eylem emir kipinde gelirse, bu sözcük ile ilgili başka biçimbilimsel çözümlemeler silinecektir. Eğer tümcede gelen sözcüklerde eylem varsa ve bu eylem emir kipinde gelip, son sözcük değilse, o zaman bu “emir kipi” ile ilgili çözüm silinecektir.

Bu nedenle eylemlerin “emir kipi” çözümü üreten sözcüğün tümcedeki sözcük sırasına göre, “eylem” ile ilgili çözümlemenin kullanılıp kullanılmayacağına karar verilebilir.

Örneğin, “sen xet yaz” (sen mektup yaz) tümcesi çözümlenirken aşağıdaki çözümlemeler üretilir.

sen:sen+Pron+Pers+A2sg+Pnon+Nom

xet:xet+Noun+A3sg+Pnon+Nom

yaz:yaz+Noun+A3sg+Pnon+Nom

yaz:yaz+Verb+Pos+Imp+A2sg

“sen” ve “xet” sözcükleri tek çözüm ürettiğinden, bu çözümler doğrudan geçerli çözümlemeler olarak kabul edilir. “yaz” sözcüğü çözümlenirken iki farklı çözüm üretilir.

1: yaz+Noun+A3sg+Pnon+Nom

2: yaz+Verb+Pos+Imp+A2sg

- 1. çözümde ise, “yaz” sözcüğü isim (Noun) olarak çözümlenmiş ve bu sözcük, üçüncü şahıs tekil uyumu (A3sg), belirsiz iyelik (Pnon) ve yalın durumdadır.
- 2. çözümde ise, “yaz” sözcüğü eylem (Verb) olarak çözümlenmiş ve bu eylem, olumlu (Pos), emir kipinde (Imp) ve ikinci şahıs tekil uyumundadır (A2sg) .

Bu tmcede, 2.zmn seilmesinin koulu ise, 2.zmn kk szcg olan “yaz” eyleminin, tmcenin sonuna gelmesidir. Kural uygulandıktan sonra, aağıdaki zmler seilmiş olur.

sen:sen+Pron+Pers+A2sg+Pnon+Nom

xet:xet+Noun+A3sg+Pnon+Nom

yaz:yaz+Verb+Pos+Imp+A2sg

Eęer “yaz” szcg tmcenin sonunda deęil, tmcenin ortasında geliyorsa, bu sefer “emir” kipi ieren zm silinir. rnek:

“bu yaz issiq” (bu yaz sıcak) tmcesi zmlenirken:

bu:bu+Pron+Demons+A3sg+Pnon+Nom

yaz:yaz+Noun+A3sg+Pnon+Nom

yaz:yaz+Verb+Pos+Imp+A2sg

issiq:issiq+Adj

Uygulanan kural sonucu “emir kipi” (Imp) etiketi¹⁶ ieren zm silinecektir ve aağıdaki zmler seilmiş olur:

bu:bu+Pron+Demons+A3sg+Pnon+Nom

yaz:yaz+Noun+A3sg+Pnon+Nom

issiq:issiq+Adj

Bu kural sadece eylemlerin “emir” kipi etiketine gre alıtıęından baka zmlerleri etkilemez. Geliştirilmiş biimbilimsel zmleyicide, bu kurallar tasarlanırken genel, yani dz yazı dili dlmtr. iir dizesi ieren tmcede bu kural yanlı alıacaktır. Ayrıca bu kural genel bir durum olduęundan, biimbilimsel zmleyici de ilk sırada alıtırılmıtır.

5.1.2 Gelecek zaman etiketine gre belirsizlik giderme-1

Uygurcanın biimbilimsel zmleyicisi, eylemlere “maęci\mekci” ekleri eklendięinde, “ama” (Inten) kipi ve “gelecek” (Fut) zaman etiketi ile etiketlenir. Bu

¹⁶ Sadece kurallar iin gereken etiketin anlamları verilmitir. Tm etiketlerin anlamları ise EK. C’de verilmitir.

nedenle bir sözcük ile ilgili çözümlemede, aynı anda amaç kip etiketi ve gelecek zaman etiketi varsa, gelecek zaman etiketi içeren çözüm seçilir.

Örnek: “u xey yazmaqci” (o mektup yazacak) tümcesi çözümlendiğinde altaki çözümlemeler¹⁷ üretilir.

u: u+Pron+Pers+A3sg+Pnon+Nom

xet: xet+Noun+A3sg+Pnon+Nom

yazmaqci: yaz+Verb+Pos+Fut+A3sg*

yazmaqci: yaz+Verb+Pos^DB+Noun+Inf1+A3sg+Pnon+Nom^DB+Adj +Agt

yazmaqci: yaz+Verb+Pos^DB+Noun+Inf1+A3sg+Pnon+Nom^DB+Noun

+Agt+A3sg+Pnon+Nom

yazmaqci: yaz+Verb+Pos+Inten+A3sg*

Üretilen çözümlemede, “yazmaqci” sözcüğü ile ilgili dört farklı çözüm üretilmiştir. Tasarlanmış kural ile amaç kipi içeren çözümleme silinecektir ve geriye aşağıdaki çözümlemeler kalır.

u: u+Pron+Pers+A3sg+Pnon+Nom

xet: xet+Noun+A3sg+Pnon+Nom

yazmaqci: yaz+Verb+Pos+Fut+A3sg

yazmaqci: yaz+Verb+Pos^DB+Noun+Inf1+A3sg+Pnon

+Nom^DB+Adj+Agt

yazmaqci: yaz+Verb+Pos^DB+Noun+Inf1+A3sg+Pnon

+Nom^DB+Noun+Agt+A3sg+Pnon+Nom

Bu kural çalışırken, sözcüklerin tümcedeki sırası ile hiçbir bağlantısı göz önüne alınmaz, sadece etiketlere göre çalışır.

¹⁷ Sadece üzerinde işlem yapılacak ve bu kuralın uygulanacağı çözümlemeler “*” işareti ile belirtilmiştir.

5.1.3 Gelecek zaman etiketine göre belirsizlik giderme-2

Uygurcada “maqci/mekci” eki, gelecek zaman eylemlerini etiketlemek için kullanılır. Aynı anda, biçimbilimsel çözümleyici, yapım eki (Agt) olan “ci” ekine göre çözüm üretir. Örneğin, bir önceki bölümde verilen çözümlemede, “yazmaqci” sözcüğü ile ilgili çözümlemeler.

u: u+Pron+Pers+A3sg+Pnon+Nom

xet: xet+Noun+A3sg+Pnon+Nom

yazmaqci: yaz+Verb+Pos+Fut+A3sg*

yazmaqci: yaz+Verb+Pos^DB+Noun+Inf1+A3sg+Pnon

+Nom^DB+Adj+Agt*

yazmaqci: yaz+Verb+Pos^DB+Noun+Inf1+A3sg+Pnon+Nom^DB+Noun

+Agt+A3sg+Pnon+Nom*

Eğer çözümlemelerde, aynı anda gelecek zaman etiketi ve “ci” eki ile ve ilgili yapım eki (Agt) bulunuyorsa, gelecek zaman etiketi içermeyen tüm etiketler silinebilir.

Bu kural aslında bir önceki kuralın tamamlayıcı kuralıdır. Bir önceki kural ve yeni tanımlanmış kuralın uygulanması sonucu, “yazmaqci” sözcüğü ile ilgili belirsizlik giderilir ve aşağıdaki çözüm elde edilir.

u: u+Pron+Pers+A3sg+Pnon+Nom

xet: xet+Noun+A3sg+Pnon+Nom

yazmaqci: yaz+Verb+Pos+Fut+A3sg

Eğer çözümlemede sadece “ci” yapım ekinin eklenmesiyle oluşan “Agt” etiketi varsa ve gelecek zaman etiketi yoksa o zaman doğrudan “Agt” etiketi içeren çözüm seçilir.

Bu kural sadece etikete göre çalışır ve komşu sözcükler ile etkileşimi olmaz. Örneğin, “mEniN dadam kitapci” (benim babam kitapçı) tümcesi aşağıdaki çözümlemeleri üretir.

mEniN: men+Pron+Pers+A1sg+Pnon+Gen

dadam: dada+Noun+A3sg+P1sg+Nom

kitapci: kitap+Noun+A3sg+Pnon+Nom^DB+Adj+Agt*

kitapci: kitap+Noun+A3sg+Pnon+Nom^DB+Noun+Agt+A3sg +Pnon
+ Nom**

Çözümlemede gösterildiği gibi, “kitapçı” sözcüğü ile ilgili olarak iki çözüm üretilmiştir. Bu iki çözümden herhangi biri seçilebilir.

Geliştirilen kuralda, eğer bir sözcük ile ilgili yapım eki “Agt” içeren birden fazla çözüm olursa, “DB+Noun+Agt” etiketi içeren çözümleme silinir. Bu kuralın uygulanması sonucu, tümce ile ilgili aşağıdaki çözümleme üretilir.

mEniN: men+Pron+Pers+A1sg+Pnon+Gen

dadam: dada+Noun+A3sg+P1sg+Nom

kitapci: kitap+Noun+A3sg+Pnon+Nom^DB+Noun+Agt +A3sg+Pnon+Nom*

5.1.4 Sıfatlara göre zarfların silinmesi

Uygurcada, Türkçede olduğu gibi birçok sözcük aynı anda sıfat ve zarf olarak çözümlenir. Örneğin, yoq (yok), yalGuz (yalnız), awwal (evvel), deslep (ilk) gibi. Tümcelerde, bir sözcük ile aynı anda sıfat (Adj) ve zarf (Adverb) içeren çözümleme oluştuğunda, zarfı içeren çözümleme kaldırılmıştır.

Örnek: “uning dadasi yoq” (onun babası yok) cümlesi çözümlendiğinde, aşağıdaki çözümler üretilir.

uniN: u+Pron+Pers+A3sg+Pnon+Gen

dadasi: dada+Noun+A3sg+P3sg+Nom

yoq: yoq+Adj*

yoq: yoq+Adverb**

Burada zarfın kaldırılmasının nedeni ise, Uygurca tümcelerinde sıfatların kullanılma olasılığının zarflardan daha yüksek olmasıdır [110].

Başka sözcüklerde sıfat ya da zarf geçmediğinden hiçbir değişiklik olmaz. Tümcenin belirsizlik giderme kuralı uygulandıktan sonraki biçimbilimsel çözümlenmesi ise:

uniN:u+Pron+Pers+A3sg+Pnon+Gen

dadasi: dada+Noun+A3sg+P3sg+Nom

yoq: yoq+Adj

Bu kural sadece bir sözcük aynı anda zarf ve sıfat olarak çözümlendiğinde çalışır. Burada silme şartı ise, silinecek sıfat ile zarfların aynı “kök” olmasıdır. Yani sıfatların yalın durumlarıdır. Sıfatlar ek aldığıında, zarfla ilgili çözüm üretmez ve böyle bir çözüm oluşmaz.

5.1.5 Zarfların seçilmesi

Uygurcada bazı sözcükler çözümlenirken, eklendiği eklerden dolayı, zarf ,sıfat, isim gibi sözcük çeşitleri ile çözümlenir. Ancak bunların üretilme şeklinde bir çok yönden fark vardır. Örneğin, bir sözcük zarf olarak çözümlenirken, bu sözcük hiçbir ek almaz. Ama bazı sözcüklere eklerin eklenmesiyle türetilmiş sözcükler oluşur. Bunlar biçimbilimsel çözümlemeye, çekim grubu belirticisi (^DB) ile belirtilir.

Örneğin: “u belki yaxshi bala” (o belki iyi çocuk) tümcesi çözümlendiğinde, tümcede yer alan sözcükler, aşağıdaki çözümlemeleri üretir.

u: u+Pron+Pers+A3sg+Pnon+Nom

belki: belki+Adverb*

belki: belki+Adverb+Indicate**

belki: bel+Noun+A3sg+Pnon+Nom^DB+Adj+Rel***

belki: bel+Noun+A3sg+Pnon+Nom^DB+Pron+Rel+ A3sg+Pnon+Nom****

yaxSi: yaxSi+Adj

bala: bala+Noun+A3sg+Pnon+Nom

Burada “belki” sözcüğü dört farklı sözcük üretir. Hiçbir ek almadan “zarf” etiketi üreten bir çözüm bulunduğundan, bu etiket hariç silinecektir. Bu kural çalışırken başka tümcede bulunan başka sözcük çözümlemeleri ile bağlantısı bulunmaz ya da tümcedeki bulunduğu yer ile bağlantısı bulunmaz.

Bu kural sonucu tümce ile ilgili çözümleme aşağıdaki gibi olur.

u: u+Pron+Pers+A3sg+Pnon+Nom

belki: belki+Adverb*

yaxSi: yaxSi+Adj

bala: bala+Noun+A3sg+Pnon+Nom

5.1.6 Eylemlere göre zarfların seçilmesi

Uygurcada zarflar çoğu zaman eylemlerden önce gelir. Bu nedenle, eğer bir sözcük ile çözümlemeye, aynı anda sıfat ve zarf ile ilgili çözümleme varsa ve bu sözcük bir eylem özelliği içeren bir sözcük içeriyorsa, o zaman , o sözcüğün sıfat etiketi içeren çözümlemesi silinir. Aksi durumda , zarf özelliği taşıyan çözümleme silinir.

Örneğin: “men yalGuz keldim” (ben yalnız geldim) sözcüğü aşağıdaki gibi çözümlenir.

men: men+Pron+Pers+A1sg+Pnon+Nom

yalGuz: yalGuz+Adj*

yalGuz: yalGuz+Adverb**

keldim: kel+Verb+Pos+Past+A1sg

Burada “yalGuz” sözcüğü ile ilgili iki çeşit, zarf ve sıfat, çözümlemesi bulunmaktadır. Aynı anda burada “yalGuz” sözcüğü eylem özelliği taşıyan “keldim” sözcüğünden önce geldiğinden, zarf içeren ikinci çözüm seçilir ve tümcenin çözümü aşağıdaki gibi olur.

ben: men+Pron+Pers+A1sg+Pnon+Nom

yalGuz: yalGuz+Adverb**

keldim: kel+Verb+Pos+Past+A1sg

Farklı tümce “men yalGuz bala” (ben yalnız çocuk) tümcesi üzerinde ile ilgili çözümleme yapılırken aşağıdaki çözümlemeler üretilir.

men: men+Pron+Pers+A1sg+Pnon+Nom

yalGuz: yalGuz+Adj*

yalGuz: yalGuz+Adverb**

bala: bala+Noun+A3sg+Pnon+Nom

Bu tümcede yer alan “yalGuz” sözcüğü, bir isimden önce geldiğinden, “yalGuz” sözcüğünün sıfat içeren çözümlenmesi seçilecektir. Bu tümce ile ilgili biçimbilimsel çözümlemesi:

men: men+Pron+Pers+A1sg+Pnon+Nom

yalGuz: yalGuz+Adj*

bala: bala+Noun+A3sg+Pnon+Nom olur.

Bu kural, kendisinden bir sonra gelen sözcüğün türüne göre çalışır.

5.1.7 Tümce sonunda eylemlerin seçilmesi

Uygurcada tümcenin sonunda gelen sözcük birden fazla çözüm ürettiğinde ve bu çözümlemeler içinde eylem içeren çözümler varsa, bu eylem içeren çözümleme seçilir. Çünkü Uygurcada tümcenin sonunda, eylemlerin gelme olasılığı başka sözcüklerin gelme olasılığından daha yüksektir.

Örneğin, “men kitap aldım” (ben kitap aldım) sözcüğü çözümlendiğinde aşağıdaki çözümlemeler üretilir.

men: men+Pron+Pers+A1sg+Pnon+Nom

kitap: kitap+Noun+A3sg+Pnon+Nom

aldım: aldi+Noun+A3sg+P1sg+Nom*

aldım: al+Verb+Pos+Past+A1sg**

Burada “aldım ” sözcüğü iki farklı çözümleme üretir. Tek yıldızlı çözüm isim olarak çözümlenirken, Türkçe anlam olarak (önüm, ön tarafım) çevrilir ve tümce anlamına hiç uyum sağlamaz. İki yıldızlı çözüm ise, eylem olarak çözümlenir. Bu nedenle bu kural sonucu iki yıldızlı çözüm seçilir ve bu tümce ile ilgili biçimbilimsel çözümlemeler aşağıdaki gibi olur.

men: men+Pron+Pers+A1sg+Pnon+Nom

kitap: kitap+Noun+A3sg+Pnon+Nom

aldım: al+Verb+Pos+Past+A1sg**

Bu kuralın çalışma koşulu ise sözcüğün tümce sonunda gelmesi ve eylemin, emir kipi olmamasıdır.

5.1.8 Tümce başında gelen eylemlerin silinmesi

Uygurcada, çoğu durumlarda tümcenin başına eylemler gelmez. Bu nedenle tümce başlangıcında sözcükler ile eylem içeren çözümleme tespit edildiğinde, eylem içeren çözümlemeler silinecektir.

Örneğin: “alma aldım” (elma aldım), Tümcesi çözümlendiğinde aşağıdaki çözümlemeler üretilir.

alma: alma+Noun+A3sg+Pnon+Nom*

alma: al+Verb+Neg+Imp+A2sg**

aldım: aldi+Noun+A3sg+P1sg+Nom

aldım: al+Verb+Pos+Past+A1sg

Burada “alma” sözcüğü iki farklı çözüm üretir. Biri meyve anlamına gelen “elma” iken o ikincisi ise, “al” anlamına gelen “al” eyleminin olumsuz şeklidir. Bu kural sonucu isim anlamına gelen çözümleme seçilecektir. Bir sonraki çözümleme ise, bölüm 5.2.7’ de tanımlanmış kural ile çözümlenir ve eylem olan çözümleme seçilir.

Kurallar sonucu aşağıdaki çözümleme elde edilir.

alma: alma+Noun+A3sg+Pnon+Nom

aldım: al+Verb+Pos+Past+A1sg

Bu kural tümcede geçen sözcük yerine göre çalışır yani tümcenin başında ve tümcenin ortasında geçerlidir.

5.1.9 Şahıs eklerine göre belirsizliğin giderilmesi

Uygurcada tümce oluşturulurken, tümcede geçen bazı sözcüklerin cümlede geçen şahıs ve özneler ile uyum sağlaması gerekir. Bu nedenle tümce içinde geçen sözcüklerin aldığı ekler ile cümlemin öznesinin eklerinde uyum sağlayıp sağlamadığı aranmalıdır.

Örneğin: “men bir qalem aldım” (ben bir kalem aldım) tümcesi çözümlenirken aşağıdaki çözümlemeler üretilir.

men: men+Pron+Pers+A1sg+Pnon+Nom

bir: bir+Num+Card

qelem: qelem+Noun+A3sg+Pnon+Nom*

qelem: qele+Noun+A3sg+P1sg+Nom**

aldim: aldi+Noun+A3sg+P1sg+Nom

aldim: al+Verb+Pos+Past+A1sg

Burada “qelem” sözcüğü iki farklı çözüm üretir. Tek yıldızlı çözümde, “qelem” sözcüğü isim, belirsiz iyelik ve yalın durumda olarak çözümlenmiştir. Uygurca dilbilgisine göre, tümce öznesi olan “ben” sözcüğü ile uyum sağlanır. Çünkü “ben” sözcüğü belirsiz iyelik ve yalın durumda olduğundan, tümcede anlam uyuşması sağlanır.

“qelem” ile ilgili iki yıldızlı çözümlemelerde ise, “qelem” sözcüğü ise, 1.tekil şahıs iyelik eki içerir. Bu çözümlemenin seçilebilmesi için, tümcenin öznesinin, bitirme durumu (Acc) eki alması gerekir. Ancak cümle öznesinde bitiminde durum eki olmadığından bu çözümleme seçilmemiştir. Bu kuralın uygulanması sonucu tümce ile son çözüm ise aşağıdaki gibi olur:

men: men+Pron+Pers+A1sg+Pnon+Nom

bir: bir+Num+Card

qelem: qelem+Noun+A3sg+Pnon+Nom

aldim: al+Verb+Pos+Past+A1sg

Bu kuralın çalışma koşulu ise tümcede geçen eklerin tümcede geçen şahıs eklerine uyma koşuludur.

5.1.10 Öznesiz tümcelerde belirsizliğin giderilmesi

Uygurcada bazı tümcelerin öznesi gizli olarak gelir. Bu durumlarda birden fazla çözüm içeren sözcük çözümlerinin bu sözcüklerden sonra gelen sözcükler ile uyum içinde olması aranır.

Örneğin: "bir qelem aldim" (bir kalem aldım). Bu tümce çözümlenirken aşağıdaki çözümlemeler üretilir.

bir: bir+Num+Card

qelem: qelem+Noun+A3sg+Pnon+Nom*

qelem: qele+Noun+A3sg+P1sg+Nom**

aldim: aldi+Noun+A3sg+P1sg+Nom

aldim: al+Verb+Pos+Past+A1sg

Burada “aldim” sözcüğü ile ilgili çözümlerden “eylem “ içeren çözümlemenin seçildiği kabul edilmiştir. Çünkü bu kural “eylem” ile belirsizlik giderme işlemlerinden sonra çalışır. Uygurca dilbilgisine göre, “aldim” sözcüğünde yer alan 1.şahıs tekil uyum özelliği (A1sg) ise “qelem” sözcükleri ile ilgili çözümlemelerde, belirsiz iyelik (Pnon) içeren sözcük ile uyum sağlaması gerekir. Bu nedenle tek yıldız içeren çözümleme seçilir.

Bu kural uygulama sonucu, tümce ile ilgili sonuçlar aşağıdaki gibi olur.

bir: bir+Num+Card

qelem: qelem+Noun+A3sg+Pnon+Nom

aldim: al+Verb+Pos+Past+A1sg

Bu kural tümcede yer alan başka sözcükler ile özellikle şahıs eklerine uyum sağlama koşulu ile çalışır.

5.1.11 Biçimbilimsel belirsizlik giderici ile ilgili sonuçlar

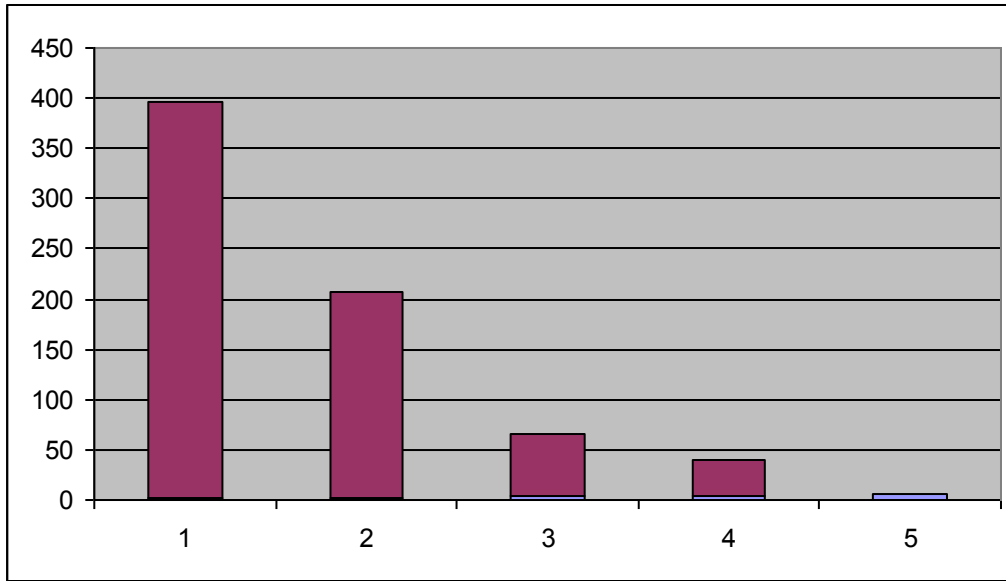
Uygurcanın biçimbilimsel belirsizlik gidericisi için, Uygurcanın genel dil bilgisinden yararlanarak bir belirsizlik giderici geliştirilmiştir.



Şekil 5.1 : Biçimbilimsel belirsizlik giderici.

Bu belirsizlik giderici çıkışında, tümcede geçen her sözcük için tek bir biçimbilimsel çözüm seçilmiş olur. Eğer biçimbilimsel çözümleyici için tanımlanan kurallar yetersiz kalırsa ve bu nedenle birden fazla çözümlemeler tek çözüme indirgenemezse, o zaman ilk çözüm seçilir.

Biçimbilimsel belirsizlik gidericinin genel çalışma yapısı Şekil 5.1’ de verilmiştir. Belirsizlik gidericiyi sınamak için öncelik ile 527 adet sözcük içeren 105 adet tümcenin üzerinde biçimbilimsel çözümle işlemi yapılmıştır ve Şekil 5.2’de gösterilen sonuç elde edilmiştir. Şekil 5.2’de yatay eksen de gösterilen 1,2,3,4 ve 5 sayıları ise, ilgili tümcelerde geçen sözcüklerin üretebileceği çözüm sayısını göstermektedir. Dikey eksen de gösterilen sayılar ise, tümcede toplam olarak, 1, 2, 3, 4 ve 5 çözüm üreten sözcük sayısını verir. Görüldüğü gibi 1 adet çözüm üreten sözcük sayısı yaklaşık 400 iken, 5 adet çözüm üreten sözcük sayısı ise 0 olarak gerçekleşmiştir.



Şekil 5.2 : Tümcelerin biçimbilimsel çözümlemesi.

Geliştirilmiş biçimbilimsel belirsizlik giderici ile Şekil 5.2’de üretilen biçimbilimsel çözümlemeler üzerinde belirsizlik giderme işlemi yapılırken Çizelge 5.1’de gösterilen sonuçlar elde edilmiştir.

Çizelge 5.1’de farklı yapı içeren 105 tümce kullanılmıştır. Tümcede toplam 527 adet sözcük kullanılmıştır. Sözcükler üzerinde biçimbilimsel belirsizlik çözümleme yapıldığında toplam 700 adet çözüm elde edilmiştir. Bir sözcüğe gelen belirsizlik oranı ise 1.33 olarak gerçekleşmiştir.

Çizelge 5.1 : Biçimbilimsel belirsizlik giderici sınama sonuçları.

Tümce sayısı	Sözcük sayısı	Biçimbilimsel çözümleyici ile üretilen çözüm	Kurallarla indirgenemeyen çözümleme sayısı	Doğru çözümlenmeyen sayısı	Hata oranı
105	527	700	13	22	3.14%

Geliştirilmiş biçimbilimsel belirsizlik giderici ile, bu çözümler üzerinde belirsizlik giderme işlemi uygulandığında, 540 adet çözümleme elde edilmiştir. Yani biçimbilimsel giderici asıl tümcede geçen sözcük sayısından fazla çözüm vermektedir. Dolayısıyla 13 adet çözümleme üzerindeki belirsizlik giderilememiştir. Bu nedenle belirsizlik gidericinin başarı oranı 96.86% olarak gerçekleşmiştir. Belirsizliği çözümlenemeyen 13 çözümleme arasından her zaman ilki seçilmiştir.

5.2 Sözcük Anlamsal Belirsizliğin Giderilmesi

Bir tümcede geçen bir sözcük, eğer birden fazla anlam içeriyorsa, bu sözcüğün hangi anlamda geldiğinin tespit etme çalışmaları işlemi bilgisayar destekli doğal dil çalışmalarında henüz tamamıyla çözümlenmeyen açık bir sorundur. Bilgisayarlı dil bilimlerinde, bu işlemler için yapılan çalışmalar, sözcük anlamındaki belirsizliği giderme çalışmaları (WSD) olarak adlandırılır.

Anlamsal belirsizlik giderme çalışmalarında en temel bileşenler ise, bir sözcüğün tüm anlamlarını içeren bir sözlük, ikinci bileşen ise, o dil için geliştirilen sözcüklerin anlamını ayırt edebilecek bir derlemenin olmasıdır.

Anlamsal belirsizlik giderme ile ilgili çalışmalar 1940 yıllarında, bilgisayarlı çeviri işlemlerindeki en temel konu olarak tanımlanmıştır. Warren Weaver 1949 yılında yayınladığı bildirisinde, anlamsal belirsizlik ile ilgili sorunu hesaplamalı metinlerde kullanmıştır [122]. İlk zamanlarda, araştırmacılar anlamsal belirsizliğin karmaşık bir problem olduğunu, ve sadece elektronik bilgisayarlar ile çözülemeyeceğini ve dil bilimleri ile genel kavramlarda gerektiğinde hem fikir olmuştur. 1970'li yıllarda anlamsal belirsizlik ile ilgili çalışmalar yoğun olarak el ile yazılan kural tabanlı çözümlerden oluşmuştur. 1980 yılında geniş kapsamlı sözcük kaynaklarının erişilebilmesinden sonra, örneğin, Oxford Advanced Learner's Dictionary of Current English (OALD) , sözlüğünün elektronik ortama taşınması ile sözcük anlamları ile

ilgili yazılan kurallar yerine, bu sözlükler kaynak olarak kullanılmıştır. Anlamsal belirsizlik giderme yöntemi sözlük tabanlı ya da bilgi tabanlı yöntem olarak devam etmektedir. 1990 yıllarında kullanılmaya başlayan istatistiksel yöntemler bilgisayarlı dil bilimlerinde bir çığır (devrim) açmıştır. Bununla beraber anlamsal belirsizliği istatistiksel yöntemler ile, sözlük kullanılarak çözümlene yapılmıştır. Denetimsel çözümlene ana konu olmuştur. 2000’li yıllarda istatistiksel yöntemlerle denetimli teknikler, anlamsal belirsizliğin giderilmesi için en yüksek başarılı çözümü üretmiştir. Bu nedenle farklı yöntemler arayışına girmiş; örneğin, donanımlara uyum sağlayabilen, yarı yada hiç denetimli olmayan yöntemler, derlem tabanlı yöntemler ve farklı yöntemleri içeren karma yöntemler araştırılmıştır. Ancak, başarıyı en yüksek sonuçlar denetimli (supervised) sistemlerde görülmüştür [123].

Sözcük temelinde anlamsal belirsizlik gidermede izlenen yöntemler genel olarak aşağıdaki gibi sıralanabilir.

- Sözlük ve Bilgi tabanlı yöntemler.
- Denetimli (supervised) yöntemler.
- Kısmen denetimli (semi-supervised) yöntemler.
- Denetimsiz (unsupervised) yöntemler.

Aşağıda bu yöntemler hakkında kısaca açıklamalar yapılmıştır.

5.2.1 Sözlük ve bilgi tabanlı yöntemler

Lesk algoritması sözlük kullanarak anlamsal belirsizlik gidermek için kullanılan ilk yöntemdir. Bu yöntem, bir tümcede kullanılan sözcüklerin hepsi birbiri ile ilişkilidir ve sözcüklerin arasındaki bu ilişkiler, o sözcüklerin tanımı ve anlamı için kullanılan sözcüklerden elde edilir, ilkesine dayanır. Örneğin “at” sözcüğünün bir “hayvan” ya da “fırlat” olarak çözümlenebilmesi için, tümcede, “at bindim” ya da, “taş attım” gibi ilgili sözcükler ile beraber gelmesi gerekir. Böylece “at” sözcüğünün “hayvan” olarak tanımı da geçen sözcükle “bindim” sözcüğünün tanımında geçen sözcüklerin uyumuna göre, “at” sözcüğünün tümcede nasıl bir anlamda geldiğine karar verilir.

Birden fazla anlama sahip herhangi bir sözcüğün tümce içerisindeki yeri önemli olmaksızın, hangi anlamda olduğu ile ilgili genellikle bir belirsizlik mevcuttur [124]. Bu nedenle çeviri yaparken tümcede hangi anlam geçerse o anlamı verecek şekilde

hedef dile aktarılmalıdır. Bir başka sözcük ile anlatıldığında, hedef dile çevrilen sözcük, kaynak cümledeki anlamı vermesi gerekmektedir.

Lesk algoritmasının çalışma yöntemi ise, tümcede geçen her sözcük için, BOS’de geçen tanımı aranır. Yani tümcede geçen tüm sözcükler ile, belirsizliği giderilecek sözcükleri tanımlamak için kullanılan sözcükler üzerinde karşılaştırma yapılır. Eğer hangi anlamda karşılaşılan sözcük sayısı fazla ise, o anlam seçilir.

Örneğin: İngilizcede “ash”¹⁸ sözcüğünün, iki farklı anlamı vardır Çizelge 5.2’de olduğu gibi tanımlanmıştır.

Çizelge 5.2 : İngilizcede “ash” sözcüğünün tanımı.

Anlamı (Sense)	Tanımı (Definition)
A1: Tree	A tree of the olive family
A2: Burned stuff	The solid residue left when combustible material is burned.

“ash” sözcüğünün “A1” anlamı ise bir çeşit zeytin ağacı olarak tanımlanırken, “A2” anlamında ise, “kül” olarak tanımlanmıştır. Lesk algoritması ile, içinde “ash” sözcüğü geçen tümce üzerinde hesaplama yaparken, anlamların topladığı puana göre hangi anlamanın seçileceğine karar verilir.

Örneğin: Çizelge 5.3’de, ilk Tümcede “A1” anlamı hesaplanırken, yani “tree” sözcüğünün tanımı için, tümcedeki tüm sözcükler, “A1” nin tanımı için geçen sözcükler ile karşılaştırılır. Ancak, hiç bir sözcük eşleşmesi bulunmamaktadır. Bu nedenle “A1” anlamı için toplanan puan ise “0” olur.

“A2” anlamı için tanımlanan sözcükler üzerinde karşılaştırma yapıldığında, “burn” sözcüğü ile eşleşir. Bu nedenle “A2” anlamının topladığı puan “1” olur. “A2 “ anlamının topladığı puanın yüksek olması nedeniyle, tümcede geçen “ash” sözcüğünün, “A2” anlamı taşıdığı kanaatine varılır. Yani tümcede “ash” sözcüğü “kül” anlamında gelmiştir.

Çizelge 5.3’de yer alan ikinci tümce üzerinde hesaplama yapılırken, tümcede geçen “tree” sözcüğü, “ash” sözcüğünün “A1” anlamını tanımlamada kullanılan

¹⁸ Bu örnekler ve çizelgeler [125]’ dan alınmıştır.

Çizelge 5.3 : Sözcük anlamının hesaplanması.

Puan		Tümce
A1	A2	
0	1	This sigar burns slowly and creates a stiff ash
1	0	The ash is one of the last trees to come into leaf

sözcüklerden “tree” sözcüğü ile eşleşiyordu. Bu nedenle “A1” sözcüğü “1” puan toplar. “A2” anlamını tanımlamak için herhangi bir sözcük eşleşmediğinden, “0” puan toplayacaktır. Bu nedenle, ikinci tümcede, “ash” sözcüğünün bir “tree” (ağaç) anlamında geldiğinin kanaatine varılır.

```
for every word w[i] in the phrase
  let BEST_SCORE = 0
  let BEST_SENSE = null
  for every sense sense[j] of w[i]
    let SCORE = 0
    for every other word w[k] in the phrase, k != i
      for every sense sense[l] of w[k]
        SCORE = SCORE + number of words that occur
          in the gloss of both sense[j] and sense[l]
      end for
    end for
    if SCORE > BEST_SCORE
      BEST_SCORE = SCORE
      BEST_SENSE = w[i]
    end if
  end for
  if BEST_SCORE > 0
    output BEST_SENSE
  else
    output " Could not disambiguate w[i] "
  end if
end for
```

Şekil 5.3 : Asıl Lesk algoritmasının sözde kodu [127].

Eğer her iki anlam eşit puan alırlarsa, rasgele biri seçilir. Bu nedenle, Lesk algoritması sözcük tanımından etkilenmesi çok belirgindir. Lesk algoritmasının başarımının İngilizce için yüzdesinin 50-70% arasında olduğu hesaplanmıştır [126].

Yukarıda anlatılan Lesk algoritması ise, basitleştirilmiş algoritmanın çalışma yöntemidir.

Asıl algoritmada ise, tümcede geçen bir sözcüğün anlamı aranırken, tümcede geçen sözcük ile doğrudan karşılaştırılma yerine, o sözcüğün sözlükteki tanımı ile karşılaştırılarak yapılır. Eğer tümcede karşılaştırılacak sözcüğün birden fazla anlamı

varsa, tüm anlamlar ile karşılaştırılır. Yani bir anlamın tanımında geçen sözcükleri, Tümcedeki tüm sözcüklerin olası anlamının tanımında kullanılan sözcükler ile karşılaştırılır. Burada amaç ise, karşılaştırma uzayının genişletilmesidir. Şekil 5.3’de asıl Lesk algoritmasının sözde kodu verilmiştir

5.2.2 Denetimli yöntemler

Denetimli yöntemler, bir tümce, bu tümcede geçen belirsiz sözcükleri çözümleyebilecek bilgileri içerir ilkesine dayanır. Bu nedenle makine öğrenmesi algoritmalarının örneğin, özellik seçme, parametre optimize etme, tip öğrenmesi gibi yöntemleri kullanıldığı görülmüştür. Makine öğrenmesi algoritmaları içinde, vektör destekli öğrenim (Support vector machines) ve bellek tabanlı öğrenim (memory-based learning) algoritmalarının başarımının yüksek olduğu görülmüştür. Yüksek başarımların nedenleri ise yüksek boyutlardaki özellikler kullanarak hesaplama yapmalarıdır. Ancak bu yöntemlerin olumsuz yönü ise, çok büyük eğitim verilerinin hazırlanmasıdır. Yani etiketlenmiş derlemelerin oluşturulmuş olması, hazır bulunmasıdır. Bu da yüksek maliyet içeren ve zaman isteyen bir ön çalışmadır. Örneğin bir tümcede geçen sözcükler ile ilgili tüm anlamsal belirsizliklerin giderilmesi için, tüm sözcüklerin, tüm anlamlarını içeren derlemin bulunması gerekmektedir. Bunun için yoğun kullanılan derlem ise, WordNet tarzında insanlar tarafından etiketlenen, Brown derleminin bir parçası olan SemCor derlemidir. Bu derlemde 234,000 sözcük etiketlenmiştir [128-129].

5.2.3 Kısmen denetimli yöntemler

Eğitim verilerinin yeterli olmaması nedeniyle bir çok anlam belirsizlik giderme algoritmaları etiketlenmiş ve etiketlenmemiş veriler üzerinde işlem yapabilen kısmen denetimli yöntemler kullanır. Bu yöntem için Yarowsky algoritması örnek olarak gösterilebilir [130]. Bu yöntemin çalışma ilkesi ise, insanlar konuşmasında, “bir tümcede, bir sözcük konu ile ilgili sadece bir anlam içerir ”ve“ bir konudan bahsederken o konu ile ilgili anlam içeren sözcük kullanır” ilkesine dayanır. Araştırmalara göre, bir konudan bahsederken, o tümcede kullanılan sözcüklerin, o konu ile ilgili sadece bir anlam verdiği görülmüştür. Bu nedenle ilk önce o kullanabilen bilgi kullanarak, tüm sözcükler ile belirsizlik giderme işlemi gerçekleştirilmektedir. Denetimli yöntemde olduğu gibi, tüm sözcükler tüm anlamları ile bilgilerin bulunması gerekmez. Yarowsky algoritması da bu amaç ile

çalışmaktadır. Yani ilk önce az sayıda etiketlenmiş sözcük anlamları ile, çok sayıda etiketlenmeyen bir derlemde işlem yapmaya başlar. Algoritma bu az sayıdaki etiketlenmiş çekirdek anlamları kullanarak bir karar-listesi oluşturur. Sonra bu karar listesini kullanarak etiketsiz derlemi etiketler. Sonra algoritma derlemde bulunan, ve en uygun olan örnekleri seçer ve eğitim kümesine ekler. Bu örnekleri kullanarak yeniden karar-listesi oluşturur ve böyle devam eder. Sonunda derlem tümüyle (ya da belli bir eşik değerine kadar) etiketlenmiş olur.

5.2.4 Denetimsiz yöntemler

Denetimsiz yöntem, sözcük anlamsal belirsizlik giderme araştırmacılarının ilgilendiği önemli bir konudur. Bu yöntemde ön görülen çözüm ise, “ bir sözcüğün anlamı, aynı yapıdaki bir başka tümcede de aynı anlam veriyor ilkesidir. Bu nedenle ilgili anlamı bulmak, aynı yapıdaki tümcelerden ilgili sözcüğün görülmesi ile hesaplanır” [131]. Sonra, yeniden bulunan sözcükler, ilgili anlam kümesine dahil edilir. Bu yöntemin başarımı başka yöntemlere göre düşük olduğu görülmüştür . Ancak bu yöntem her hangi bir insan emeği ile hazırlanan bilgi kaynaklarına bağlı olmadığından, bilgi erişim kısıtlaması yönünden avatantajlı olduğu söylenebilir.

5.3 Uyurcada Anlamsal Belirsizliğin Giderilmesi

Bu tez çalışması yapılırken, Uyurca için geliştirilen herhangi bir derlem yada bir sözcüğün farklı anlamlarının açıklamalarını içeren elektronik sözlük bulunamamıştır. Bu nedenle Lesk alogortması kullanılarak bir belirsizlik giderici çalışması yapılmıştır.

Tümcede geçen sözcükler üzerinde biçimbilimsel belirsizlik işlemleri yapıldıktan sonra, her sözcük ile tek çözüm elde edilir. Geliştirilecek çeviri sisteminde, biçimbilimsel çözümlemelerden sonra, Uyurca sözcük kökleri Türkçeye aktarmaktadır. Aktarma işlemleri için, Uyurcadan Türkçeye çeviri sözcüğü kullanılmaktadır [84].

Tıpkı başka dillerde olduğu gibi, Uyurcada da bazı sözcükler, aynı anda birden fazla anlam içerir. Lesk algoritması ile geliştirilen yöntemlerde, basitleştirilmiş Lesk algoritmasının başarısı, asıl Lesk algoritmasına göre yüksek olduğu tespit edilmiştir. Uyurca bitişken bir dil olduğundan ve eklendiği eklere göre anlamın değişmesi yüzünden, geliştirilen anlamsal belirsizlik gidericide, sadece basitleştirilmiş Lesk

algoritması kullanılmadan, ayrıca asıl Lesk algoritması da kullanılmıştır. Ayrıca, sözcükler üzerine, tümcede geçen sözcüklerin “kök” sözcükleri, bu “kök” sözcüklerin anlamın tanımlamak için kullanılan sözcüklerin “kök” sözcükleri üzerinde uyuşma sayısı hesaplanmıştır.

Aşağıda Lesk algoritmasını Uygurcanın sözcük anlamsal belirsizliğini gidermek için kullanılması ayrıntılı biçimde anlatılmıştır.

Uygurca için BOS bulunmadığından öncelikle bir BOS sözcüğü oluşturulmuştur [85]. Oluşturulan sözcükler Uygurcada en çok kullanılan sözcüklerden seçilmiştir ve en çok kullanılan anlamları yer almıştır.

Geliştirilen anlamsal belirsizlik gidericinin çalışma yöntemi ise Şekil 5.4’de gösterilmiştir.

Geliştirilen anlamsal belirsizlik giderici, biçimbilimsel belirsizliği çözümlenmiş sözcükleri “giriş” kısmından okuma yapmaktadır.

Örnek :

Uy: men taghda at mindim

Tr: ben dağda at bindim

Uygurca sözcük, biçimbilimsel belirsizlik giderici tarafından, belirsizliği giderildikten sonra, aşağıdaki çözümlemeler seçilmiş oluyor.

men+Pron+Pers+A1sg+Pnon+Nom

taG+Noun+A3sg+Pnon+Loc

at+Noun+A3sg+Pnon+Nom

min+Verb+Pos+Past+A1sg

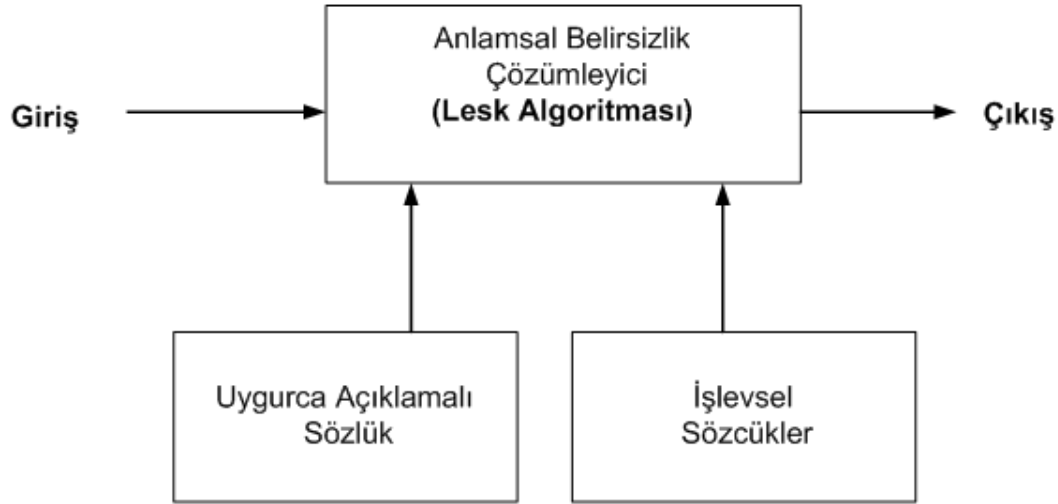
Tümcede geçen bu sözcükler ve tanımlamalar ise, BOS’de aşağıdaki gibi tanımlanmıştır.

*men

#men(0): birinji Sehis qoSumcisi, azliq, sOzligiciniN Ozi, sOzligici, pEqir. Ozem.

*taG

#taG(0): taS,qoram say bilen oralGan haNlar, taS,igiz yerler. bu taGda qaraGay, arca var. keklik. bu taGnin havasi bek yaxSi. Bu taGda gUzel yaylaqlar bar.



Şekil 5.4 : Anlamsal belirsizlik giderici.

*at

#at(0): adem we nerilerge birilgen isim. u balaGa at qoydi.atamaq. buni dep ataymiz.

#at(1): at azsa tayGa egiSir.at mingen.sUt emgiCi haywan.men at mindim. bu at tiz yUgUreydu.

#at(2): bir nesini atmaq, cOrimek, taSlamaq. miltiq atti. oq atti. taS atti.top atti, qaS atti.

*al

#al(0): nersini ucumlap , siqimlap yaki birer eswap bilen tutup qolGa kirgUzmek. cilekten su almaq. yancuqtin pul almaq. telefon almaq.

#al(1): boSatmaq, bikar qilmaq, cUSurmek,wezipidin almaq,xizmettin almaq.

#al(2): nerse kirek sEtiwalmaq, bu ayGice necce pul aldiNiz, bu kitapni necce pulGa aldiNiz?. men Oy aldim.

#al(3): igellemek, iSGal etmek, orun almaq, nam almaq, Seher almaq.

*min

#min(0): bir nersige iki putini icip oturmaq, at minmek, eSek minmek,welispit minmek. boyniGa minmek.

*qoy

#qoy(0): sUt emgici haywan, yaylaq, gOS, su, ot

#qoy(1):qolidiki nersini birer jayGa qoymaq. qoydi. at qoydi.isim qoydi.

Örnekte gösterildiği gibi, “men”, “taG” ve “min” sözcüklerin “tek” anlamı bulunmaktadır. Bu nedenle bu sözcükler üzerinde anlamsal belirsizlik çalışması yapılmasına ihtiyaç bulunmamaktadır.

Ancak , “at” sözcüğünün 3 farklı anlamı¹⁹ bulunmaktadır. Tümcenin anlamına göre, hayvan anlamına gelen “at(1)” sözcüğü seçilmelidir.

Burada tanımlanan “at(0)”, “at(1)”, “at(2)” sözcüklerine üretilmiş yapay sözcükler denir. Yapay sözcükler aracılığı ile, görünen biçimde tek şekil ile yazılan “at” sözcüğünden üç farklı sözcük ile yazılmış oluyor.

Tanımlanan kök aktarma sözcüklerinde ise, bu yapay sözcüklerin taşıdığı anlama göre, hedef dilde uygun sözcükler ile çevrilecektir.

“at(0)” sözcüğü ile ilk çözüm, “varsayılan ” çözüm ise doğrudan Türkçeye “ad” olarak çevrilir.

İkinci anlam taşıyan yapay sözcük “at(1)”ise, hayvan anlamına gelen “at” olarak çevrilirken, üçüncü anlam taşıyan yapay sözcük “at(2)” ise, “fırlat” olarak çevrilir.

Sadece “kök “ sözcükler üzerinde aktarma yapılır. Sözcükler ile ilgili biçimbilimsel bilgiler üzerinde herhangi bir değişiklik yapılmaz.

“ad” <- “at(0)”

“at” <- “at(1)”

“fırlat” <- “at(2)”

Böylece Uygurcada bulunan “at” sözcüğü, tümcede geçen farklı anlamlarına göre, Türkçedeki farklı anlamlarına göre çeviri sağlanmış olur. Uygurcada bulunan “al” sözcüğü ile ilgili dört farklı anlam bulunmaktadır. Bir önceki yöntemler ile 4 farklı anlamı için yapay sözcükler tanımlanmıştır ve yapay sözcükleri içeren anlamına göre Türkçe sözcüklere çevrilmiştir.

¹⁹ Sözcüğün kendisi “*” ile belirtilmiştir. Bu sözcükler ile ilgili anlamlar “#” ile belirtilmiştir. Anlaşılmasının kolay olması için, farklı anlamları ise sözcüğün sonuna rakamlar eklenmesi ile belirtilmiştir.

Yapay sözcükler içinden, uygun olanı seçebilmek için, o sözcüğün tümcede geçen sözcükler ile ne kadar uyum sağladığına bakılması gerekiyor. Bu nedenle Lesk algoritması kullanılmıştır.

Lesk algoritması, çözümlerde ilk sözcük olan “men” sözcüğü ile ilgili çözümlerden, “kök” sözcük olan “men” i , okuyor. Ancak, bu çalışma için geliştirilen sözlükte “men” ile ilgili tek anlam verir. Bu nedenle bu sözcük ile ilgili olarak anlamsal belirsizlik giderme çalışması yapmaya gerek görülmez. Bir sonraki adımda ikinci sözcük olan “ taG” ile ilgili kaç anlam olduğunu sorguluyor. Bu sözcüğünde “tek” anlamı vardır. Bu tek anlam doğrudan seçilmiş anlam olarak kabul edilir.

Üçüncü adımda ise tümcede yer alan “at” sözcüğü ile ilgili sorgulama yapılır. “at” sözcüğünün üç farklı anlamı bulunmaktadır. Bu nedenle tümcede yer alan tüm sözcüklerin (kendisi hariç) tamamında geçen sözcükler ile, “at” sözcüğün her bir yapay sözcüğü tanımlamak için kullanılan sözcükler üzerine karşılaştırma yapılır.

Bu karşılaştırma, “at” sözcüğü ile ilgili 3 farklı yapay sözcüğü tanımlamak için kullanılan tüm sözcükler bitene kadar devam eder. En sonunda hangi “yapay” sözcük ile ilgili uyuşan sözcük sayısı fazla ise, o yapay sözcük ile belirtilir, tümcedeki gerçek anlam kabul edilir ve bu yapay sözcük, Türkçeye çevirilecek sözcük olarak seçilir.

Geliştirilen sistemde, “at” sözcüğünün farklı 3 yapay sözcük ile uyum sağlayan sözcük sayıları ise aşağıda verilmiştir.

“at(0)” varsayılan anlamı “0” puan toplarken, “at(1)” anlam ise “3” puan, “at(2)” anlamı da “0” puan toplamış. Bu sonuçlara göre, “at(1)” anlamı taşıyan, yani “hayvan” anlamına gelen yapay sözcük “at(1)” seçilir.

Anlamlar üzerinde hesaplama yaparken, asıl Lesk algoritması ve basitleştirilmiş Lesk algoritması test edilmiştir. Ayrıca Uygurca bitişken dil olduğundan, tümcede geçen sözcükler ile, sözlüklerde geçen sözcükler yüzeysel biçimleri ile birebir karşılaştırılmıştır. Sonradan, tümcede geçen “kök” sözcükler ile, sözlükte geçen sözcüklerin “kök” sözcükleri karşılaştırılmıştır.

Ayrıca Uygurcada, bir çok yardımcı sözcük bulunmaktadır ve bunların tek anlamı olduğundan, algoritmaya dahil edilerek ve edilmeden hesaplamalar yapılmıştır.

Hesaplama sonucu, yardımcı sözcüklerin herhangi bir etkisinin olmadığı gözlemlenmiştir.

Test için kullanılan parametreleri ve belirsizlik gidericinin başarımını doğrudan etkileyen etkenleri göstermek için test sonuçları farklı çizelgelerde gösterilmiştir. Belirsizlik gidericinin testi için, özellikle birden fazla anlam içeren sözcüklerden oluşan tümceler seçilmiştir ve sonuçları ilgili çizelgeler ile gösterilmiştir.

Örnek 1:

Uy: u at qoydi

Tr: u ad koydu.

Bu örnek tümcede geçen sözcükler için, biçimbilimsel belirsizlik giderici sonucu aşağıdaki biçimbilimsel çözümlemeleri seçilmiştir.

u: u+Pron+Pers+A3sg+Pnon+Nom

at: at+Noun+A3sg+Pnon+Nom

qoydi: qoy+Verb+Pos+Past+A3sg

Biçimbilimsel çözümlemelerde geçen kök sözcükler ile ilgili tanımlamalar BOS'de aşağıdaki gibi tanımlanmıştır.

*u

#u(0): Ucuncu Sahis. birlik.

*at

#at(0): adem we nerilerge birilgen isim. u balaGa at qoydi. atamaq. buni dep ataymiz.

#at(1): at azsa tayGa egiSir.at mingen.sUt emgiCi haywan.men at mindim. bu at tiz yUgUreydu.

#at(2): bir nesini atmaq, cOrimek, taSlamaq. miltiq atti. oq atti. taS atti.top atti, qaS atti.

*qoy

#qoy(0): sUt emgici haywan, yaylaq, gOS, su, ot

#qoy(1):qolidiki nersini bir jayGa qoymaq.qoydi.at qoydi.isim qoydi.

Örnek 1 ile ilgili test sonucu ise Çizelge 5.4’de verilmiştir. Sonuca göre basitleştirilmiş Lesk algoritması tüm koşullarda doğru çalışırken, asıl Lesk algortması, “qoy” sözcüğünü doğru çözümlerken, “at” sözcüğünü yanlış çözümlemiştir. Yani “isim” olarak çözümlemesi gereken, “hayvan” olarak çözümlemiştir. Ayrıca yardımcı sözcüklerin farklı bir etki göstermemiştir. Belirsizlik gidericinin başarısı test edilirken, test için kullanılan tümcelere göre başarı oranı değişmektedir.

Çizelge 5.4 : Örnek 1 ile ilgili sınama sonuçları.

		Asıl Algoritma		Basitleştirilmiş Algoritma	
Yardımcı sözcükler içeriyor	Test sözcükleri	at	qoy	at	qoy
	Kök sözcüklerle uyuşması	Doğru	Doğru	Doğru	Doğru
	Tümce sözcüklerle uyuşması	Yanlış	Doğru	Doğru	Doğru
Yardımcı sözcükler içermiyor	Kök sözcüklerle Uyuşması	Doğru	Doğru	Doğru	Doğru
	Tümce sözcüklerle uyuşması	Yanlış	Doğru	Doğru	Doğru

Lesk algoritmasının temel çalışma mantığında sözcüklerin uyuşma sayısına göre karar verildiğinden, sözcüklerin farklı anlamlarının tanımı için kullanılan sözcük sayısından belirgin bir şekilde etkilenmektedir [125].

Ayrıca tümcede kullanılan sözcük sayısı da etkili olmaktadır.

Örnek 2:

Uy: men taghda at mindim.

Tr: ben dağda at bindim.

Bu örnek ile ilgili biçimbilimsel çözümlemeler ve sözcük tanımları daha önceki sayfalarda²⁰ anlatıldığından burada tekrar verilmemiştir. Bu örnekte ise “at” sözcüğü hayvan anlamında gelmiştir.

Bu örnek ile yapılan test sonucu Çizelge 5.5’de verilmiştir.

Çizelge 5.5 : Örnek 2 ile ilgili sınama sonuçları.

		Asıl algoritma	Basitleştirilmiş algoritma
Yardımcı sözcükler içeriyor	Test sözcükleri	at	at
	Kök sözcüklerle uyuşması	Yanlış	Yanlış
	Tümce sözcüklerle uyuşması	Doğru	Doğru
Yardımcı sözcükler içermiyor	Kök sözcüklerle uyuşması	Yanlış	Yanlış
	Tümce sözcüklerle uyuşması	Doğru	Doğru

Bu örnekte, bir önceki örneğe karşın sözcük sayısı farklı olan , yani sözcük sayısı fazla olan bir tümce kullanılmıştır. Ayrıca birden fazla anlam içeren tek bir sözcük kullanılmıştır.

Test sonucu, Asıl Lesk algoritması doğru sonuç göstermiştir.

Örnek 3:

Uy: men deryagha tash attim.

Tr: ben ırmağa taş attım.

Bu örnekte “at” sözcüğü eylem anlamında gelmiştir ve bu örnek ile test sonuçları ise Çizelge 5.6’da verilmiştir.

²⁰ Sayfa 115-117 sayfalara bakılması tavsiye edilir.

Çizelge 5.6 : Örnek 3 ile ilgili sınama sonuçları.

		Original algoritma	Basitleştirilmiş algoritma
Yardımcı sözcükler içeriyor	Test sözcükleri	at	at
	Kök sözcüklerle uyuşması	Yanlış	Doğru
	Tümce sözcüklerle uyuşması	Yanlış	Yanlış
Yardımcı sözcükler içermiyor	Kök sözcüklerle uyuşması	Doğru	Doğru
	Tümce sözcüklerle uyuşması	Doğru	Yanlış

Bu örnekte geçen “taş” (taş) ve “derya” (nehir) ile ilgili sözcüklerin tanımı ise aşağıdaki gibi verilmiştir. Başka sözcüklerin tanımı²¹ daha önceden verildiğinden burada tekrar gösterilmemiştir.

*taş

#taş(0): qattiq nerse. taGlarda deryalarda kOp. taş Oy. taş kOrUk.taş yol.

*derya

#derya(0): sularniN aqtiGi yer. bu suni deryadin ekeldim . deryaGa suGa cUSkili kettuq.coN Eriq . deryaGa kOrUk salduq.

Çizelge 5.6’da gösterildiği gibi “kök” sözcükler üzerinden uyuşma arandığında basitleştirilmiş Lesk algoritması doğru sonuç üretmektedir. Tümcede geçen sözcükler üzerinden karşılaştırma yapıldığında, yardımcı sözcükler hesaba alındığında Asıl Lesk algoritması yanlış sonuç üretirken, yardımcı sözcükler hesaba alındığında ise, doğru sonuç üretmiştir.

²¹ Sayfa 115-117 bakılması tavsiye edilir.

Uygurca için anlamsal belirsizlik için tercih edilebilecek parametreleri seçmek amacıyla birden fazla anlam içeren ve çeşitli sözcük sayısından oluşan tümceler ile test edilmiştir ve Çizelge 5.7’de gösterildiği sonuçlar elde edilmiştir.

Çizelge 5.7 : Uygurca anlamsal belirsizlik giderici sınaama sonuçları.

	Kullanılan algoritma ve sözcük karşılaştırma yöntemi	Asıl algoritma	Basitleştirilmiş algoritma
Yardımcı sözcükler içerir	Kök sözcüklerle uyumu	56%	72%
	Tümce sözcüklerle uyumu	56%	50%
Yardımcı sözcükler içermez	Kök sözcüklerle uyumu	67%	83%
	Tümce sözcüklerle uyumu	72%	56%

Çizelge 5.7’de gösterilen verilerin anlamı ise, test sonucu veren doğru çözümleme sayısını vermektedir.

Bu çizelgede 80 adet sözcükten oluşan 18 adet tümce kullanılmıştır. Tümcelerde kullanılan sözcüklerden bazıları tek anlam içerirken, bazıları ikiden fazla anlam içermektedir.

Yani basitleştirilmiş Lesk algoritması kullanarak, kök sözcükler üzerinde hesaplama yaparken, doğruluk oranı yüzde 83 iken, asıl Lesk algoritması kullanarak kök sözcükler üzerinden hesaplandığında ise, başarı oranı yüzde 67 olmaktadır.

Bu nedenle Uygurcanın sözcüksel anlam belirsizliği üzerinde hesaplama yaparken, basitleştirilmiş Lesk algoritması kullanılması, kök sözcükler üzerinden karşılaştırma yapılması ve yardımcı sözcüklerin hesaba alınmaması kanaatine varılmıştır.

6. UYGURCADAN TÜRKÇEYE BİLGİSAYARLI ÇEVİRİ YÖNTEMİ

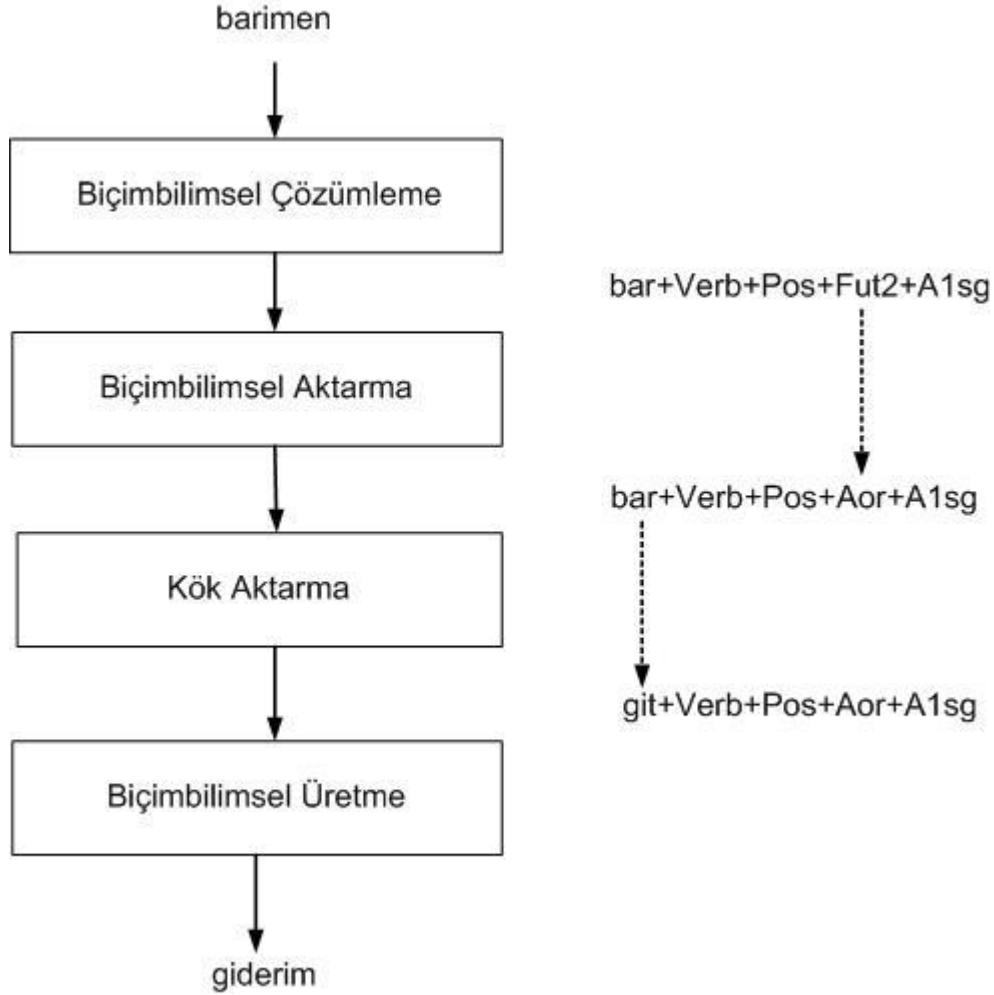
Uygurca'dan Türkçe'ye çeviri sistemi için, bitişken diller, özellikle akraba diller için önerilen kural tabanlı çeviri yöntemlerinden biri olan doğrudan aktarma yöntemi tercih edilmiştir. Doğrudan aktarma yönteminin kullanma nedenleri ise kısaca şöyle özetlenebilir:

1. Uygurca ile Türkçenin sözdiziminin benzer olması nedeniyle, Uygurcanın sözdizimi için ayrıca araştırma yapmaya gerek yoktur.
2. İstatistiksel ve örnek tabanlı çalışan yöntemleri kullanabilmek için Uygurca ile ilgili olarak büyük eğitim verileri içeren derlemelerin olması gerekir. Ancak bu tez çalışılırken, Uygurca ile herhangi bir derlem bulunamamıştır.

Uygurca ile Türkçe bitişken dil olduğundan, doğrudan aktarma yöntemi kullanılmadan önce bu diller için biçimbilimsel çalışmalarının yapılması gerekmektedir. Bu yöntem olduğu gibi doğrudan kullanılamaz. Dolayısıyla Uygurca ile ilgili biçimbilimsel çözümleyici çalışmalar, ilk defa bu tez kapsamında gerçekleştirilmiştir. Uygurcadan Türkçeye doğrudan aktarma yöntemine, biçimbilimsel çözümleyiciler dahil edilmesi Şekil 6-1'deki yapı üzerinde görülmektedir. Uygurca bitişken dil olduğundan dolayı, Uygurca'da tümedeki bir sözcüğün karşılığı sözlükte olduğu gibi bulunamaz. Bu nedenle önce Uygurca sözcükler üzerinde biçimbilimsel çözümleme yapılması gerekmektedir. Biçimbilimsel çözümleme sonucunda, çözümleme yapılan sözcük ile ilgili kök sözcük ve o sözcük ile ilgili biçimbilimsel yapısı elde edilmiş oluyor sonra daha önceden oluşturulmuş Uygurcadan Türkçeye kök aktarma sözcüğü ile Uygurca sözcüklerin Türkçe çevirisi elde edilir. Böylece kök aktarma işlemi gerçekleşir. Son olarak Türkçenin biçimbilimsel yapılarına göre çözümleyicisi, Uygurcadan aktarılmış kök sözcük ve biçimbilimsel yapılarına göre Türkçe sözcükler üretir.

Önerilen doğrudan aktarma yönteminde, biçimbilimsel çözümleme ve sözcük kökü aktarımında belirsizlikler ortaya çıkıyor [121].

Bu tez çalışmasında belirsizliklerin giderilmesi için özel olarak kural tabanlı belirsizlik giderme yöntemleri geliştirilmiştir ve bu yöntem Türkmençe ile Türkçe arasında geliştirilen karma yöntem ile karşılaştırılmıştır [32].



Şekil 6.1 : Uygurca sözcüğün Türkçe karşılığının doğrudan oluşturulması ([32]'den uyarlanmıştır).

Aşağıdaki iki bölümde, bu iki yöntemin Uygurcadan Türkçeye bilgisayarlı çeviri sisteminde nasıl uygulandığı anlatılmıştır.

6.1 Karma Yöntemin Kullanılması²²

Akraba ve bitişken diller arasında bilgisayarlı çeviri için önerilen Karma Model yöntemi ilk olarak Türkmenceden Türkçeye çeviri sisteminde kullanılmıştır [32].

²² Bu bölümde anlatılan çalışmaların tümü Cüney Tantuğa aittir [32]. Tezin bütünlüğünü sağlamak amacıyla özet şeklinde anlatılmıştır.

Çeviri yapılacak diller bitişken olması durumundan dolayı aktarım fonksiyonu ve istatistiksel dil modeli bileşeni, kullanılmıştır.

6.1.1 Aktarma modelleri

Türkmenceden Türkçeye çeviri sisteminde üç farklı aktarma modeli kullanılmıştır.

- Temel Model: Bu modelde Türkmençe sözcükler üzerinde biçimbilimsel çözümleme yapıldıktan sonra, doğrudan kök sözcükler Türkçeye çevirilmiştir. Sonra bu kök sözcükler ile Türkmençe biçimbilimsel çözümleyicisinin çözümlediği biçimbilimsel yapılar alınmıştır ve Türkçenin biçimbilimsel çözümleyecisinden sentezlenerek Türkçeye çevirilmiş sözcükler üretilmiştir. Türkmençe ile Türkçenin söz diziminin aynı olduğu göz önünde bulundurularak herhangi bir değişiklik yapılmamıştır.
- Aktarma Modeli I: Bu model de temel model oluşan bazı sorunları gidermek için geliştirilmiştir. Temel model, sözcük bazında çalışır. Dolayısıyla aktarma yaparken bir sözcüğe karşın bir sözcük aktarma yapabilir. Ya da bir sözcüğe karşın birden fazla sözcük aktarılabilir. Bu sorunu gidermek için Temel Modele çoktan-çoka aktarma kuralları eklenmiştir. Bu kurallar ile Türkmençe’de çoklu sözcük grupları belirlendikten sonra Türkçeye aktarılmıştır
- Aktarma Modeli II: Bu modelde Türkçeye aktarılma yapıldıktan sonra, Türkçenin genel dilbilgisine göre biçimbilimsel yapılarında değişiklik yapılan kuralların tanımlanmasıdır. Çünkü diller yakın olmasına karşın, kök sözcüklerin sırası değişmezse bile, köklere ya da sözcüklere eklenecek eklerin yerleri değişebilir.

6.1.2 İDM oluşturulması

Türkmenceden Türkçeye aktarılma yapıldığında, Türkmencenin oluşturduğu tüm biçimbilimsel çözümlemeler aktarılmıştır. Bu nedenle biçimbilimsel ve anlamsal belirsizlik söz konusudur. Bu sorunu gidermek için Türkçe için bir eğitim derlemi kullanılmıştır. Derlemde eğitilen söz dizimdeki sözcük, olasılığı en yüksek olan söz dizimi seçilmiştir. Ancak Türkçe ve Türkmençe bitişken diller olduğundan, eğitim verisi için görünen biçim yerine yapısal olarak biçim verileri üzerinden hesaplama yapılmıştır. Bu sistem için geliştirilen İDM türleri kısaca anlatılmıştır.

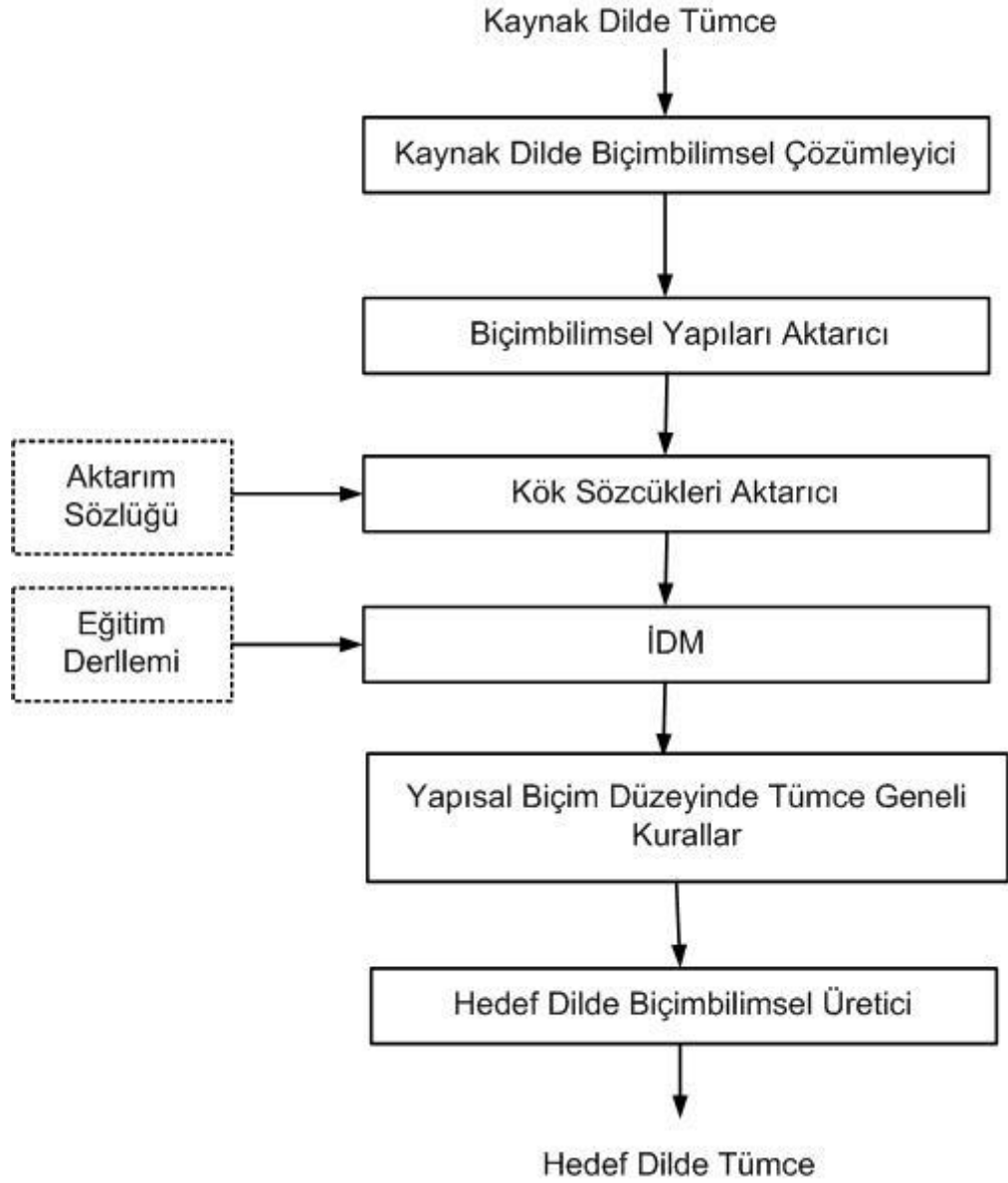
- Tip-I : Bu model sözcüklerin n-gram olasılıklarından oluşmuştur. Girdi verisi için sadece sözcük kökleri kullanılır.
- Tip-II : Sadece sözcük türlerinden oluşan İDM tipidir. Türetilmiş sözcükler için son çekim grubunun sözcük türü kullanılmıştır.
- Tip-III: Tip-II'de kullanılan modelde, tümce ile ilgili modelle karşın, aynı sözcük türüne sahip çözümlemeleri ayıkamaz. Bunları ayıklamak için son çekim grubundaki sözcük türüne o sözcük ile ilgili biçimbilimsel özelliklerde eklenir.
- Tip-IV: Sözcük kökü ile son CÇ'da bulunana sözcük türü ile oluşturulur.
- Tip- V: Sözcük kökü ile son ÇG dan oluşur.
- Tip-VI: Tüm biçimbilimsel çözümlemeyi içerir.

Karma yöntemde ile geliştirilen sistemde, sözcüklerin aktarılmasında model-II kullanılmıştır. Sözcükler kök aktarma işlemlerinde sonra İDM modeline gönderilmiştir ve değişik İMD modelleri ile eğitilmiştir. İşlem sonucu sistemi oluşturduğu yüksek olasılıklı tümce söz dizimi yapısal düzeyde elde edilmiştir. Bu yapısal düzeydeki tümceleri üzerinde Türkçenin genel tümce kuralları uygulanmıştır. Sonra tüm yapı Türkçenin biçimbilimsel çözümleyicisi ile sentezlenerek görünen biçimde tümce elde edilmiştir.

Karma yönteminde kullanılan yapı ile Uygurca ile Türkçe arasındaki öngörülen çeviri sisteminin yapısı Şekil 6.2'de verilmiştir. Şekil 6.2 nin Uygurca için uygulanan yapısında kaynak dil olarak Uygur dili kullanılmıştır ve Uygurca için geliştirilen biçimbilimsel çözümleyici kullanılmıştır [114], [116]. Hedef dil olan Türkçe kelimelerinin üretimi için, daha önceden geliştirilmiş olan biçimbilimsel çözümleyicisi kullanılmıştır [30].

6.2 Belirsizlik Giderme Yönteminin Kullanılması

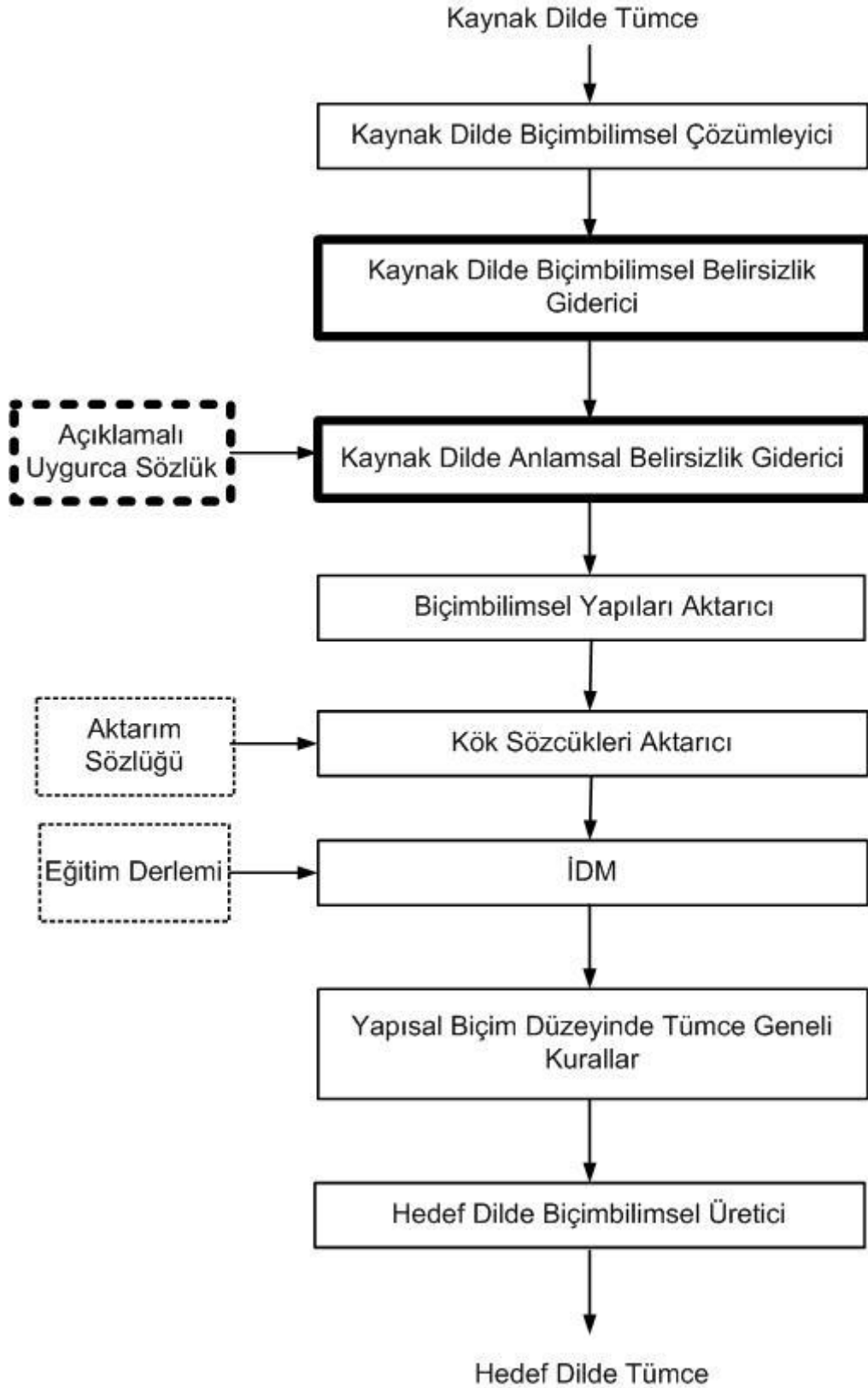
Karma yönteme alternatif olarak geliştirilen yeni sistemde, istatistiksel dil modelleri ve aktarma yöntemleri çalıştırmadan önce, belirsizlik giderme yöntemi çalıştırılmıştır. Belirsizlik yöntemi ayrı ayrı olarak biçimbilimsel çözümlemeler üzerinde ve sözcük anlamı üzerinde yapılmıştır. Karma yöntem üzerine geliştirilen yeni modüller Şekil 6.3'de kalın çizgiler ile gösterilmiştir.



Şekil 6.2 : Karma yöntemi ([32]' den alınmıştır).

Biçimbilimsel belirsizliği giderici kullanılarak bir sözcük ile ilgili biçimbilimsel belirsizlik ise tek çözüme indirgenmiştir. Uygurca için geliştirilen biçimbilimsel belirsizlik gidercinin çalışma yöntemleri ve başarımı Bölüm 5'te anlatılmıştır. Bu nedenle bu bölümde tekrar anlatılmamıştır.

Biçimbilimsel belirsizliği giderilen çözümlemelerin, aktarılacak kök sözcük üzerinde kök aktarma işlemleri yapılmadan önce, anlamsal belirsizliği giderme işlemleri yapılacaktır. Böylece bir Uygurca sözcük Türkçeye çevrilirken, birden fazla sözcük üretme sorunu çözülmüştür. Yani karma yönteminde eğitim derlemi kullanarak üretilen birden fazla çözümlerden, en çok olasılığa göre seçme işlemi ile eşit değer



Şekil 6.3 : Belirsizlik giderme yöntemi([32]'den uyarlanmıştır).

işlem burada yapılmış oluyor. Uyurca için geliştirilen anlamsal belirsizlik giderici ve çalışma yöntemi Bölüm 6'da anlatıldığından bu bölümde tekrar anlatılmamıştır.

Anlamsal belirsizlik giderici sonucu elde edilen sözcüklere kök aktarma işlemi uygulanır. Bundan sonraki işlem ise karma yöntemde kullanılan işlemler ile aynıdır. Geliştirilen yeni yöntem ile sistemin çalışma aşaması Şekil 6.3’de gösterilmiştir.

Burada kök aktarma işleminden sonra İDM’in yaptığı işlem ise Türkçede kullanılan eş anlamlı sözcüklerden hangisinin kullanılacağına karar vermek amacıyla yapılan işlemidir. Yani biçimbilimsel ve anlamsal belirsizlik giderme amacıyla işlem yapmayacaktır. Örneğin Uyгурcada “yıl” sözcüğüne karşılık olarak Türkçede “sene” ve “yıl” gibi iki eş anlamı bulunmaktadır. Çeviri yapıldıktan sonraki Türkçe tümcede “yıl” sözcüğümü kullanılacak ya da “sene” sözcüğü mü kullanılacak, buna ancak İDM işlemi aracılığı ile karar verilecektir.

Bu nedenle geliştirilen yeni sistemde İDM kaldırılmamıştır ve eş anlamlı sözcükler arasından seçim yapmak için kullanılmıştır.

Uyғurcadan Türkçeye bilgisayarlı çeviri için önerilen yöntemlerin uygulanması bir sonraki bölümde detaylı olarak örnekler ile anlatılmıştır.

7. ÇEVİRİ SİSTEMİNİN UYGULANMASI

Uygurcadan Türkçeye bilgisayarlı çeviri sistemin uygulanması için, öncelikle Tantuğ'un önerdiği karma yöntem kullanılmıştır. Daha sonra, bu tez kapsamında geliştirilen yeni yöntem olarak, biçimbilimsel belirsizlik giderme ve sözcük anlamsal belirsizlik giderme yöntemi kullanarak çeviri sistemi geliştirilmiştir. Geliştirilen yeni yöntemde de aynı çerçeve kullanılmıştır. Bölümün son kısmında ise bu iki yöntemin sonuçları karşılaştırılmıştır.

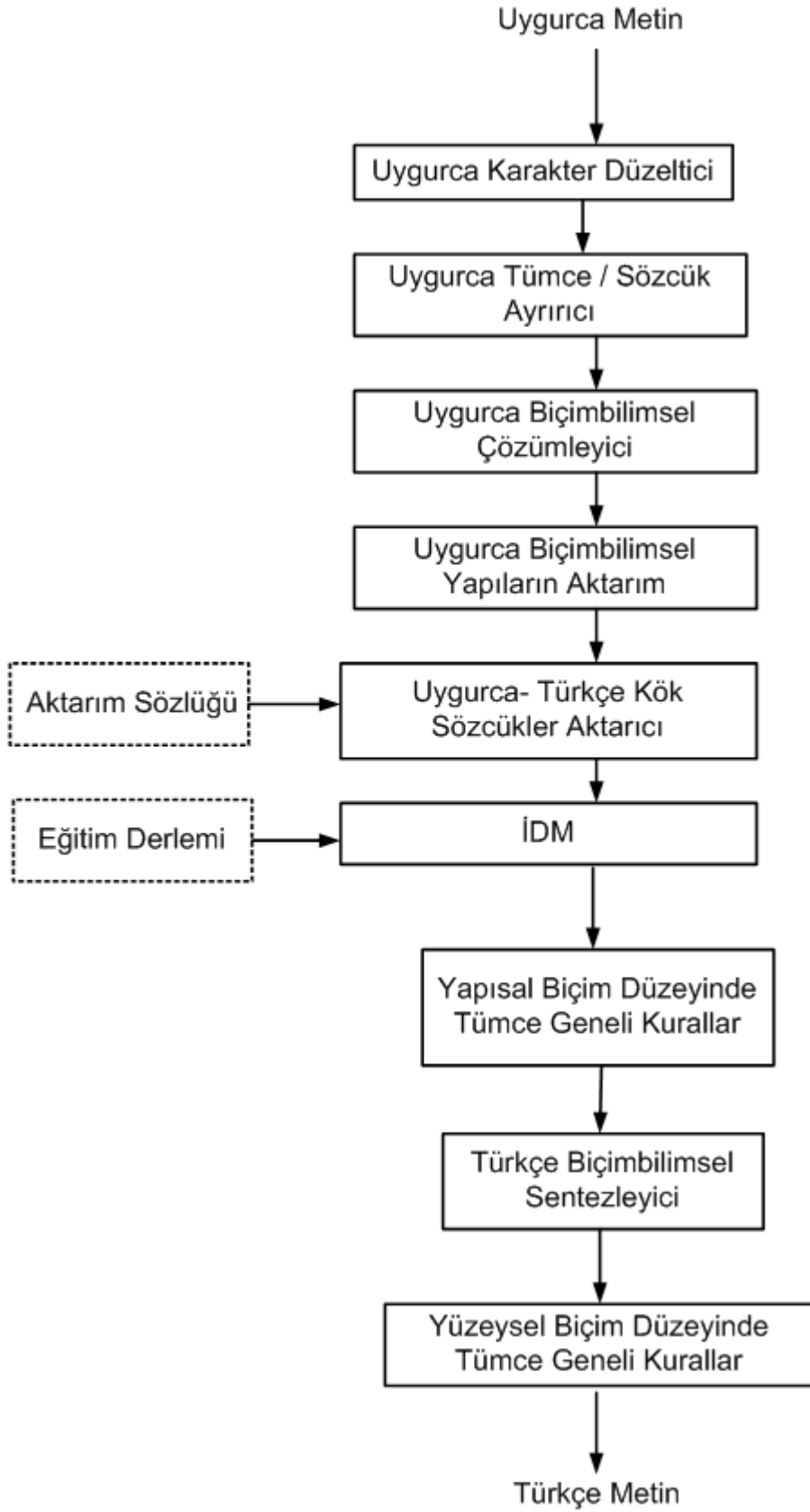
7.1 Karma Model Tabanlı Çeviri Sisteminin Uygulanması

Tantuğ'un geliştirdiği akraba ve bitişken diller arasında bilgisayarlı çeviri için geliştirdiği karma model [32], tüm Türk dillerinden Türkiye Türkçesine çeviri yapmak için bir ortam oluşturmaktadır. Çünkü bu model de Türkçe için İDM oluşturulmuştur ve başka dillerden Türkiye Türkçesine çeviri yapmak için doğrudan kullanma imkânı sunmaktadır.

Tantuğ'un geliştirdiği yöntemde, karakter düzeltici, sözcük ayrıştırıcı, biçimbilimsel çözümleyici, aktarma kuralları ve çeviri kuralları gibi bileşenler SDD kullanılarak geliştirilmiştir. Sonunda bu modellerin bir arada çalışmasını sağlayan çerçeveye geliştirilmiştir. Bu nedenle çeviri yapılacak diller ile ilgili çalışmalar SDD ile yapıldığında, bu çerçeve doğrudan kullanma ortamı sunmaktadır. Tantuğ geliştirdiği modeli test etmek için Türkmenceden Türkçeye bilgisayarlı çeviri sistemi geliştirmiştir [32].

Bu tez çalışmasında, Uygurcadan Türkçeye bilgisayarlı çeviri sistemi geliştirileceğinden, Tantuğ'un geliştirdiği çerçeve doğrudan kullanılmıştır. Türkmence için geliştirilen SDD ise Uygurca için geliştirilen SDD ler ile değiştirilmiştir.

Uygurcadan Türkçeye karma model kullanarak geliştirilen çeviri sisteminin yapısı Şekil 7.1'de gösterilmiştir.



Şekil 7.1 : Karma model ile Uygurcadan Türkçeye çeviri sistemi ([32] ‘den uyarlanmıştır).

Sistemin çalışması, Şekil 7.1’de gösterildiği gibi, ilk önce çeviri yapılacak Uyгурca metinler üzerinde karakter düzeltme işlemleri gerçekleştirilir. UKIJ’in önerdiği çift karakterler ve ASCII olmayan karakterlerin hepsi ASCII karakterlerine dönüştürülür.

Örnek:

u bala taghda at mindi.

İşlem sonrası bu tümce ise:

“u bala taGda at mindi”.

olarak değişir. Yani tümcede bulunan “gh” karakteri ise “G” ile değiştirilmiştir.

Bir sonraki adımda ise, karakter düzeltilmesi yapılan tümce üzerinde sözcük ayrıştırıcı yapılır.

“u”, “bala”, “taGda”, “at”, “mindi”, “.” gibi altı ayrı parçaya ayrıştırılır.

Sonraki adımda, bu sözcükler üzerinde biçimbilimsel çözümleme yapılıyor ve aşağıdaki sonuçlar elde edilir.

u+Pron+Pers+A3sg+Pnon+Nom

bala+Noun+A3sg+Pnon+Nom

taG+Noun+A3sg+Pnon+Loc

at+Noun+A3sg+Pnon+Nom

at+Verb+Pos+Imp+A2sg

min+Verb+Pos+Past+A3sg

+.Punc

Bu adımdan sonra, Türkçeye aktarma yapmak için aktarma kuralları uygulanır. Bu örnekte verilen tümce ise Türkçeye olduğu gibi aktarıla bilindiğinden dolayı biçimbilimsel çözümlemelerde herhangi bir değişiklik olmaz. Aktarma sözlüğüne bakılarak, sözcüklerin karşılığı aranır.

o <- u+Pron

çocuk <-bala +Noun

delikanlı <-bala +Noun

genç <-bala +Noun

dağ <- taG + Noun

at <- at + Noun

fırlat <- at + Verb

bin <- min + Verb

.+Punc <- .+Punc

Görüldüğü gibi “bala” sözcüğüne karşılık olarak, Türkçede, “çocuk”, ”delikanlı”, genç” gibi üç farklı sözcük denk gelmektedir.

Ayrıca “at” sözcüğü de iki farklı anlam ile çözümlenmiştir. Bir anlamı “at” anlamında gelen “hayvan” olarak çözümlenirken, bir başka anlamı ise “fırlat” yani bir eylem olarak çözümlenmiştir.

Bu sözcüklerden, “genç”, “delikanlı” ya da “çocuk” birini seçmek için İDM modeli kullanarak, bu tümcede kullanma olasılığı en yüksek olan sözcük seçiliyor [32]. Aynı yöntem “at” ve “fırlat” sözcüğü içinde kullanılır. Yani bir söz ile ilgili birden fazla biçimbilimsel çözüm oluştuğunda birinin seçilmesi için İDM kullanılır.

Çalışma sonucu “çocuk” ve “at” (isim) sözcükleri seçilir.

İDM işleminde sonra, tümce ile ilgili sonuçlar ise aşağıdaki gibi olur.

o+Pron+Pers+A3sg+Pnon+Nom

çocuk+Noun+A3sg+Pnon+Nom

dağ+Noun+A3sg+Pnon+Loc

at+Noun+A3sg+Pnon+Nom

bin+Verb+Pos+Past+A3sg

.+Punc

Görüldüğü gibi Uygurca kök sözcükler ile Türkçe kök sözcükler değiştirilmiştir.

Bundan sonraki adımda ise, Türkçenin biçimbilimsel çözümleyicisi kullanarak sentezleme, yani sözcük üretme işlemi gerçekleştirilir. Biçimbilimsel çözümleyici SDD ile geliştirildiğinden, sözcükler üzerinde çözümleme yapılabilirdi. Aynı anda bir sözcük ile ilgili biçimbilimsel bilgiler gönderildiğinde, o sözcük üretilebilirdi. Bu nedenle yukarıda elde edilen biçimbilimsel çözümlere biçimbilimsel belirsizlik gidericisi ile sentezleme yapıldıktan sonra aşağıdaki sözcükler elde edilir.

“o”, “çocuk”, ”dağda”, “at”, “bindi”, “.” .

Uygurca ile Türkçenin sözdizimi aynı olduğundan bu sözcükler kendi sıraları ile birleştirilebilir ve Türkçe tümce elde edilir.

“o çocuk dağda at bindi.” .

Bir sonraki adım ise, yüzeysel biçimde tümce genel kurallarının uygulanmasıdır. Uygurcada tek bir çeşit soru eki vardır ve bir önceki sözcük ile bitişik yazılır. Dolayısıyla Uygurcada soru eki için ses uyumu söz konusu değildir.

Ancak Türkçede, soru sözcüğünün , “mi, mı, mu, mü” gibi dört farklı şekli vardır ve bir önceki sözcük ile ses uyumu sağlamak zorundadır. Bu tarz durumların düzeltilmesi için yüzeysel biçim tümce genel kuralları uygulanır.

En son sistemin ürettiği çeviri sözcüğü ise:

O çocuk dağda at bindi.

Şeklinde gözükecektir.

Bu modelde, İDM modelinden başlayarak Türkçe ile ilgili hesaplama yapılır. Yani artık Uygurca ile bir ilgisi kalmaz. Bu nedenle bu çerçeve başka dillerden Türkçeye çeviri yaparken, Türkçe ile ilgili çalışmalardan kurtarmış olur.

7.2 Belirsizlik Giderme Yöntemi ile Çeviri Sisteminin Uygulanması

Geliştirilen bu sistemde, karma yönteminde kullanılan İDM in işlevi sadece eş anlamlı sözcüklerin seçme işlemi için gerçekleştirilecektir. Yani İDM’in yapması gereken belirsizlik giderme ile ilgili çalışmalar, yeni sistemde geliştirilmiş belirsizlik giderici çalışmalar ile giderilmiş olacaktır.

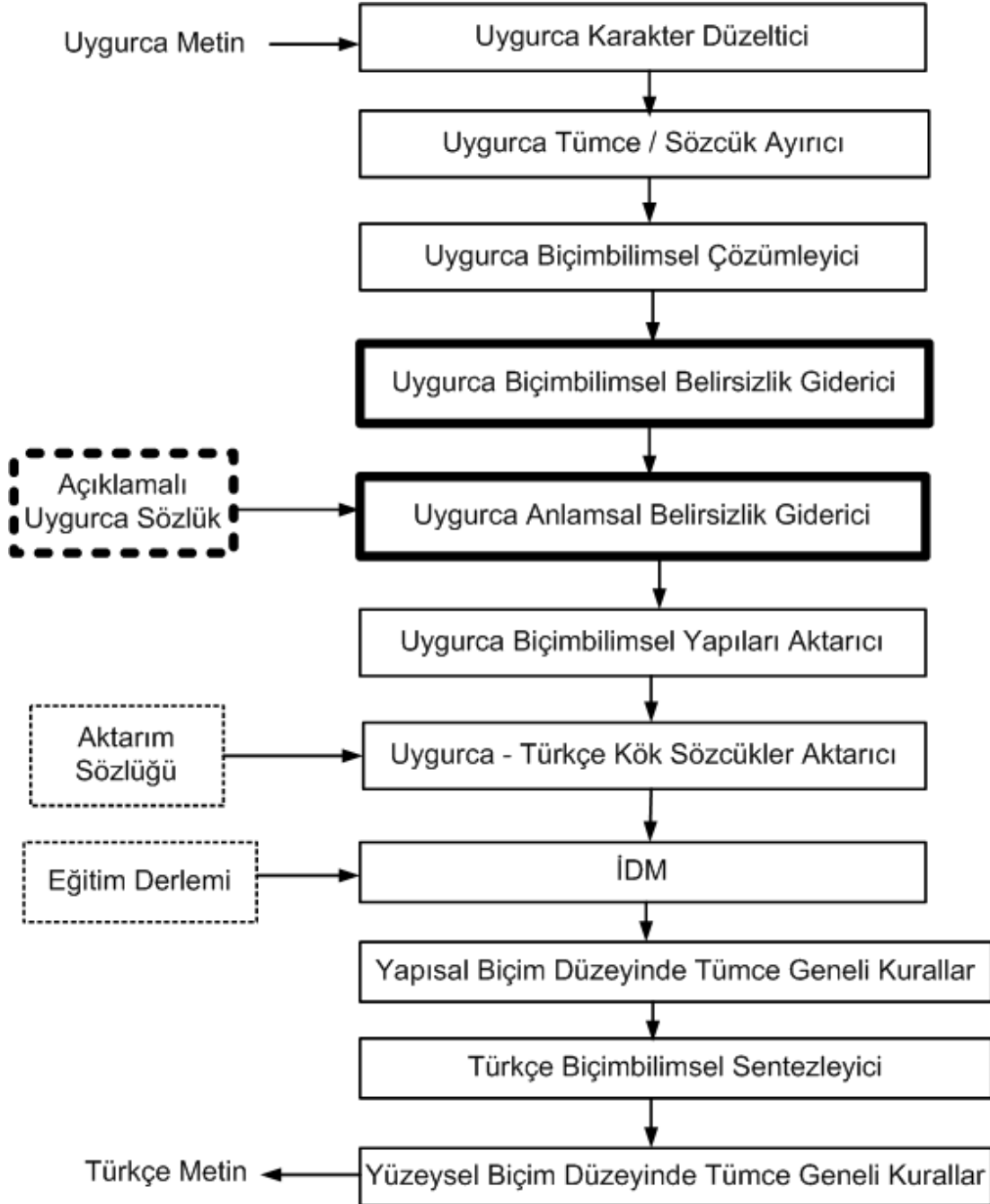
Burada karma yöntemine yapılan bir katkı ise, İDM de yapılan işlemlerin yükü azaltılmış olmasıdır. Geliştirilen yeni sistemde, sistemin çalışması karma yöntemde olduğu gibi biçimbilimsel çözümleyicinin çalışmasına kadar olan süreçler aynıdır. Bu nedenle bu bölümde sadece karma yöntemler göre farklı olan süreçler anlatılacaktır. Sistemin yapısı Şekil 7.2’de verilmiştir ve karma yönteme eklenen modüller kalın çizgiler ile gösterilmiştir. Bir tümcede karakter düzeltme, ayrıştırma ve biçimbilimsel çalışmaları yapıldıktan sonra, sözcükler ile ilgili çözümlenen biçimbilimsel çözümlemeler üzerinde belirsizlik giderme işlemleri yapılacaktır.

Biçimbilimsel belirsizlik giderme sonucu, tümce ile ilgili çözümlemeler aşağıdaki gibi olur:

u+Pron+Pers+A3sg+Pnon+Nom

bala+Noun+A3sg+Pnon+Nom

taG+Noun+A3sg+Pnon+Loc



Şekil 7.2 : Belirsizlik giderme yöntemi ile çeviri sistemi.

at+Noun+A3sg+Pnon+Nom

min+Verb+Pos+Past+A3sg

.+Punc

Yani “at” sözcüğü ile ilgili olarak “isim” ve “eylem” olarak çözümlenen çözümlemelerden, “isim” ile ilgili çözümme seçilmiş oluyor.

Başka sözcükler ile ilgili sadece tek bir çözümlemeler olduğundan, geçerli çözüm olarak kabul edilir.

Biçimbilimsel belirsizlik giderme işlemlerinden sonra, sözcükler üzerinde anlamsal belirsizlik giderme çalışmaları başlanacaktır.

Bu çalışmada, bir önceki adımda elde edilen sözcükler üzerine, Uygurcanın açıklamalı sözcüğüne göre arama yapılır ve sözcükler ile ilgili anlamları üzerine, tümceye göre en uygun anlamı bulmaya çalışılır. Tüm sözcüklerin farklı anlamları ise “0” den başlayarak numaralandırılmıştır ve önerirken ilgili olarak sözcüklerin ilk yani “0” numaralı anlamı seçilmiştir.

bu(0)+Pron+Demos+A3sg+Pnon+Nom

bala(0)+Noun+A3sg+Pnon+Nom

taG(0)+Noun+A3sg+Pnon+Loc

at(0)+Noun+A3sg+Pnon+Nom

min(0)+Verb+Pos+Past+A3sg

Görüldüğü gibi, “bu”, ”bala”, “taG”, “min” sözcükleri ile ilgili doğru anlam seçilmiştir.

Ancak “at” sözcüğü ile ilgili olarak “0” ncı anlamı seçilmiştir, yani “isim”, “ad” anlamı içeren anlamı seçilmiştir. Yani yanlış anlamı seçilmiştir.

Bir sonraki adımda biçimbilimsel yapısı aktarma işlemi gerçekleştiriliyor ancak, bu örnekte, sözcüklerin biçimbilimsel yapısında herhangi bir değişiklik yapılmaz ve kök aktarma işlemleri gerçekleştirilir. Kök aktarma işlemi ise kök aktarma sözlüğüne bakılarak aşağıdaki gibi gerçekleştirilir.

bu <- bu(0)+Pron
delikanlı <- bala(0)+Noun
yiğit <- bala(0)+Noun
çocuk <- bala(0)+Noun
dağ <- taG(0)+Noun
at <- at(0)+Noun
bin <- min(0)+ Verb

Görüldüğü gibi “at” sözcüğü, seçilen anlamı “at(0)” na göre Türkçeye “at” olarak aktarılırken, “bala” ile ilgili üç eş anlamlı sözcük , “yiğit”, “delikanlı”, “çocuk” hepsi aktarılır. Bu üç eş anlamlı sözcüğün hangisini seçileceği ise İDM ile gerçekleşir. İDM çalışma sonucu, “çocuk” sözcüğü seçilmiştir. İDM işlemi sonucu tümce de geçen tüm sözcükler ile ilgili çözümlemeleri ise aşağıdaki gibi olur.

İDM işleminde sonra, tümce ile ilgili sonuçlar ise aşağıdaki gibi olur.

o+Pron+Pers+A3sg+Pnon+Nom
çocuk+Noun+A3sg+Pnon+Nom
dağ+Noun+A3sg+Pnon+Loc
ad+Noun+A3sg+Pnon+Nom
bin+Verb+Pos+Past+A3sg
.+Punc

Bundan sonraki çalışma adımları ise, karma yöntemde yapılan adımlar ile bire bir benzer. Yani bu çözümlemeler Türkçenin biçimbilimsel sentezleyicisinden geçirilir ve Türkçe sözcükler üretilir sonra tümce genelinde yüzeysel kurallar uygulandıktan sonra tümce ile ilgili son çeviri sözcüğü ise aşağıdaki gibi üretilecektir.

“o çocuk dağda at bindi .”

Sistemin çalışma sonucu, Uygurca tümcede bulunan “at” sözcüğü, Türkçeye “at” olarak çevirilmiştir. Bu yanlışlık ise sözcük anlam belirsizliğinden kaynaklanmaktadır.

8. SONUÇLAR

8.1 Sonuçlar

Sistemin genel başarımını test etmek ve karma yöntem ile karşılaştırmak için 240 örnek tümce kullanılmıştır. Bu tümceler göre sistemin ürettiği sonuçlar ile referans olarak verilen çeviri tümcelerine göre BLEU sonucu hesaplanmıştır.

Karma yöntem, yani Türk dillerinden Türkçeye çeviri sisteminin çerçevesinin çekirdek kısmını oluşturan İDM'nin eğitilmesi için bir eğitim derlemesine gerek duyulmaktadır. Bu tez çalışmada Tantug'un geliştirdiği çerçeve kullanıldığından, Tantug'un geliştirdiği eğitim derlemi doğrudan kullanılmıştır [32]. Burada tezin bütünlüğünü sağlamak amacıyla eğitim derlemi ile ilgili bilgiler verilmiştir ²³. İDM'in oluşturulması için 1 milyon sözcükten ve 500 bin tümceden oluşan bir derlem kullanılmıştır ve derlem ile ilgili istatistikler Çizelge 8.1'de verilmiştir. Uygurca için yaygın olarak kullanılan 1250 kök sözcüğünden oluşan bir kök aktarma sözcüğü kullanılmıştır.

Çizelge 8.1 : Eğitim derlemi ile ilgili istatistikler [32].

Toplam Kelime Adedi	948.403
Toplam Tümce Adedi	50.674
Farklı Sözcük Kökleri Adedi	25.897
1 Defa Geçen Sözcük Kökleri	11.447
2 Defa Geçen Sözcük Kökleri	3505
2'den Fazla Geçen Sözcük Kökleri	10.945

²³ İDM ile ilgili olarak istatistiksel veriler verilmiştir. Verilerin nasıl oluşturulduğuna ilgili ayrıntılı bilgiler için [32] kaynağa bakılması tavsiye edilir.

Bu kök aktarma sözlüğü aynı anda eş anlamlı sözcükleri de içermektedir. Sözcük anlamsal belirsizliğini gidermek için 578 adet kök sözcük içeren, 650 farklı anlam için açıklama içeren bir elektronik sözlük oluşturulmuştur. Lesk algoritmasında sözcük anlamı için içeren sözcük sayısı, anlamsal belirsizlik gidermek için çok belirgin etki gösterdiğinden, tüm anlam için eşit sözcükler ile açıklama yapılmasına özen gösterilmiştir.

Sistemin başarımını sınamak için, karma yöntem ve belirsizlik giderme yöntemleri üzerinde aynı tümceler üzerinde çeviri işlemleri gerçekleştirilmiştir. Bu amaçla 240 adet tümceden oluşan 1220 sözcüklü bir sınama kümesi oluşturulmuştur. Sınama yapılırken önerilen farklı aktarma yöntemler içersinden, başarımı en yüksek olan aktarma yöntemi II tercih edilmiştir.

Sistemde bulunan İDM modeli, Karma yöntem ve belirsizlik giderme yöntemlerin her ikisinde kullanılmaktadır. İDM'nin başarımı ise girilen sınama verilerinin tiplerine göre değiştiğinden, yüksek olasılığa sahip olan tümce yapısını bulmak için değişik tipler göre sınama verileri ile test edilmiş ve sistemin ürettiği çeviri sonuçlar BLEU puanları üzerinden karşılaştırılmıştır. İDM'nin farklı derecelerine göre etkisinin görülebilmesi amacı ile $n=1,3$ ve 5 seçilerek deneyler tekrarlanmıştır. Ancak İDM Tip-VI'nin başarımı çok düşük olacağı Tantuğ'un çalışmasında tespit edilmiştir [32]. Bu nedenle bu tez kapsamında Tip-VI ile ilgili test etme ve karşılaştırma işlemi yapılmamıştır. Sonuçlar yorumlanırken, her iki sistemin İDM modülleri ve derecelirine göre, sonuçlar aynı çizelgede verilmiştir.

Çizelge 8.2'de İDM modellerinden Tip-I kullanılarak hesaplanan BLEU puanı verilmiştir. Görüldüğü gibi İDM derecesi 1 den 3 yükseldiğinde sistemin başarımı her iki yöntemde de artışı gözlemlenmiştir.

Çizelge 8.2 : Karma yöntem ve yeni yöntemin İDM Tip- I başarımı.

Dil Modeli Derecesi(n)	Karma Yöntem (BLEU)	Belirsizlik Giderme (BLEU)
n=1	35.80	35.81
n=3	35.73	35.90
n=5	35.73	35.90

Ancak İDM derecesi 3 den 5 yükseltildiğinde bir fark gözlenmemiştir. Çizelge 8.2'de verilen iki sistemin başarımı karşılaştırıldığında geliştirilen yeni sistemin başarımını

yüksek olduğu görülmüştür. Bunun temel nedenleri ise, yeni sistem ile belirsizlik giderme sonucu İDM için gönderilen sözcüklerin belirgin olduğundandır İDM.

Tip-II modeline göre sistemlerin başarımlarını değerlendirme sonuçları Çizelge 8.3’de verilmiştir:

Çizelge 8.3 : Karma yöntem ve yeni yöntemin İDM Tip- II başarımları.

Dil Modeli Derecesi(n)	Karma Yöntem	Belirsizlik Giderme
n=1	33.00	35.64
n=3	33.18	35.64
n=5	33.61	35.64

Tip-II’de edilen sonuçlar Tip-I’e göre düşük olmasındır. Bunun temel nedeni ise, sadece sözcük türüne göre hesaplama yaparken yetersiz kalmasıdır. Geliştirilen yeni sistemin başarımları da Tip-I’e göre düşük olmuştur ancak karma yöntemine göre yüksek olduğu gözlemlenmiştir.

Çizelge 8.4’de İDM Tip-III’ün sonuçları verilmiştir.

Çizelge 8.4 : Karma yöntem ve yeni yöntemin İDM Tip- III başarımları.

Dil Modeli Derecesi(n)	Karma Yöntem	Belirsizlik Giderme
n=1	34.18	35.64
n=3	34.09	35.64
n=5	34.12	35.64

Tip-III nin başarımları Tip-II göre yüksek olmuştur karma yöntemde. Bunun temel nedeni ise Tip-II ye göre daha fazla biçimbilimsel verisi kullanmasıdır. Ancak yeni sistemin başarımları Tip-II’e göre herhangi bir değişiklik olmamıştır. Ama yeni sistemin başarımları ise karma yöntem sonucundan yüksek başarı vermiştir.

Tip-IV İDM’e göre çalışan sistemin başarımlarını Çizelge 8.5’de verilmiştir. Sınama verilerine “kök” bilgisinin eklenmesi ile n=1 için sistem başarımları Tip-II ve Tip-III’ de her iki yönteme göre artış gözlemlenmiştir. Ayrıca n=3 ve n=5 dereceleri de karma yönteminde Tip-I nin yakaladığı başarımlara yaklaşmıştır. Ancak karma yöntem başarımları geliştirilen yeni yöntem başarımlarından düşük olmaktadır.

Çizelge 8.5 : Karma yöntem ve yeni yöntemin İDM Tip- IV başarımı.

Dil Modeli Derecesi(n)	Karma Yöntem	Belirsizlik Giderme
n=1	35.48	35.81
n=3	35.74	35.81
n=5	35.74	35.81

Tip-V İDM ‘e göre yapılan sınamaların sonuçları Çizelge 8.6’da verilmiştir.

Çizelge 8.6 : Karma yöntem ve yeni yöntemin İDM Tip -V başarımı.

Dil Modeli Derecesi(n)	Karma Yöntem	Belirsizlik Giderme
n=1	36.07	35.64
n=3	36.17	35.74
n=5	36.17	35.74

Bu tipte biçimbilimsel çözümlemelerin kökleri ile son çekim grubu sınama verisi olarak aktarılır. Bu Tip’in başarımı daha önce yapılan başka modüller ile karşılaştırıldığında, karma yöntem n=1 olduğunda en yüksek başarımla elde edildiği görülmüştür ve yeni sistemin başarımında da yüksek olmuştur. Ayrıca dil model derecesinin 1 den 3’e yükselmesiyle başarımlarında artış gözlemlenirken, ancak dil model derecesinin 3 den 5’e yükseldiğinde hiçbir fark gözlemlenmemiştir.

Sistemlerin başarımı çizelgelerden elde edilen verilere göre şöyle yorumlanabilir. Karma yöntem İDM modellerinde kullanılan sınama verilerine göre başarımında belirgin olarak farklılık göstermektedir ve Tip –II’ göre çeviri yaparken, diğer Tip’lere göre en düşük sonuç üretmektedir. En yüksek sonuç ise Tip-V de elde edilmiştir.

Yeni sistem, yani belirsizlik giderme yöntemi, İDM aktarma modellerinden etkilenmesi Karma yöntemi kadar belirgin olmadığı gözlemlenmiştir. Yeni sistem daha çok İDM dericilerine göre farklılık göstermektedir. Yeni sistemin Karma yönteminde olduğu gibi İDM modellerinden doğrudan etkilenmemesinin nedeni ise, sözcükler ile belirsizliklerin kurallarla giderilmesi, bu nedenle İDM için tek verinin gönderilmesidir. Belirsizlik giderme yöntemi ise, sadece “kök” sözcüklere göre çalışan Tip-I ve İDM derecesi 3 olduğunda en yüksek sonuç üretmiştir.

Geliştirilen belirsizlik giderme yönteminin doğrudan karma yöntemine göre en belirgin olumsuz etkeni ise, çalışma süresinin uzun olmasıdır. Bunun temel nedeni ise, Lesk algoritması çalışırken tüm sözcükler üzerinde anlam belirsizliği gidermek için tarama yapmasıdır.

8.2 Çeviri Örnekleri

Bu bölümde bazı çeviri örnekleri verilerek, gerçekleştirilen sistemin başarısı ve bu başarı ile ilgili BLEU puanlaması verilecektir. Örneklerde kaynak tümce, karma yöntem ve yeniden geliştirilen sistemin çıktıları, aday tümce ve BLEU puanlamasında kullanılan iki adet referans tümce alt alta verilmiştir. Çizelge 8.7’de gösterilen örnekte, eş anlamlı sözcüklere göre referans oluşturulmuştur ve geliştirilen sistemin ürettiği aday çeviriler üzerindeki BLEU puanlaması karşılaştırılmıştır.

Çizelge 8.7 : Çeviri örnekleri-I.

Kaynak	men bu mektepte 5 yıl okudum.	BLEU
Aday Çeviri (Karma)	ben bu okulda 5 sene okudum.	100
Aday Çeviri (Yeni)	ben bu okulda 5 sene okudum.	100
Referans 0	ben bu okulda 5 sene okudum	
Referans 1	ben bu okulda 5 yıl okudum	

Görüldüğü gibi her iki sistem aynı oranda başarı göstermiştir ve istenen çeviri tam anlamıyla elde edilmiştir. Çünkü aday çevirilerde referans tümcelerinde geçen tüm sözcükler geçmektedir.

Örnek II de, iki sistemin farklı çeviri ürettiği durum ilgili örnek verilmiştir. Bu örnek de karma yöntemin başarısı Çizelge 8.8’de gösterildiği gibi 100 olarak verirken, geliştirilen yeni sistemin başarısı 93.06 olarak hesaplanmıştır.

Gerçekleştirilen yeni sistemin ürettiği aday çeviri, referans tümcesi ile benziyor olsa da, BLEU hesabı yapılırken “n” gramlara göre hesaplama yapılır ve bu sonuçlar genel BLEU puanlamasına yansıyor. Yani $n=4$ olduğunda, hesaplanan BLEU ise 75 olarak gerçekleşir ve bu da genel BLEU puanını düşmesine neden olmaktadır.

Çizelge 8.8 : Çeviri örnekleri-II.

Kaynak	Men bazardin 2000 dane yéngi kitap aldım.	BLEU
Aday Çeviri (Karma)	ben pazardan 2000 tane yeni kitap aldım.	100
Aday Çeviri (Yeni)	ben pazardan 2000 tane yeni kitap satın aldım.	93.06
Referans 0	ben pazardan 2000 tane yeni kitap aldım.	
Referans 1	ben pazardan 2000 adet yeni kitap satın aldım	

Örnek III de sistemlerin yanlış sonuç ürettiği ile ilgili örnek verilmiştir. Bu örnekte yanlışlık oluşmasının nedeni anlamsal belirsizlik gidericiden kaynaklanmaktadır.

Çizelge 8.9 : Çeviri örnekleri-III.

Kaynak	u bala taghda at mindi	BLEU
Aday Çeviri (Karma)	O çocuk dağda at bindi.	84.09
Aday Çeviri (Yeni)	O çocuk dağda at bindi.	0.0
Referans 0	O çocuk dağda at bindi.	
Referans 1	O delikanlı dağda at bindi.	

Bu nedenle geliştirilen yeni sistemin ürettiği aday çeviri ise geçersiz olarak nitelendirilmiştir (Çizelge 8.9). Bu çeviri için BLEU sonucunun 0 olmasının nedeni ise, tek tümce üzerinde çeviri yapıldığındandır. Bu tez çalışmasında kullanılan BLEU algoritması ise birden fazla tümce üzerine çeviri yapan sistemin Başarısını değerlendirmek için tasarlanmıştır.

Örnek IV’de yardımcı sözcükler içeren tümceler ile ilgili örnek verilmiştir (Çizelge 8.10). Geliştirilen yeni sistemde “manga” sözcüğü anlamsal çözümleme yapılırken yanlış anlam ile çözümlendiğinden “yürümek” anlamı ile çözümlenmiştir. Bu nedenle geliştirilen yeni sistemin çeviri başarısı karma yönteme göre düşük olmuştur.

BLEU puanlarına göre Türkçe gibi bitişen diller ile ilgili bilgisayarlı çevirilerin başarımını ölçmek, yetersiz kalıyor [32]. Ancak çeviri başarımını otomatik hesaplama yapmak için tercih edilmiştir.

Çizelge 8.10 : Çeviri örnekleri-IV.

Kaynak	men sanga qayta qayta xet yazdim, amma sen manga hich jewap yazmading.	BLEU
Aday Çeviri (Karma)	ben sana tekrar tekrar mektup yazdım ancak sen bana hiç cevap yazmadın.	100
Aday Çeviri (Yeni)	ben sana tekrar tekrar mektup yazdım ancak sen yürü hiç cevap yazmadın.	73.49
Referans 0	ben sana tekrar tekrar mektup yazdım ancak sen bana hiç cevap yazmadın.	
Referans 1	ben sana defalarca mektup gönderdim, ama sen bana hiç cevap yazmadın.	

8.3 Değerlendirmeler

Bu tez çalışmasında Türk dilleri ailesinde bulunan dillerden Uygurca ile Türkçe arasında bir bilgisayarlı çeviri sistemi geliştirilmiştir. Uygurca ile Türkçenin söz dizimi birbirine benzediğinden yakın diller arasında bilgisayarlı çeviri yapmada tercih edilen doğrudan aktarma yöntemi tercih edilmiştir. Söz konusu diller bitişken diller olduğundan temel aktarma yöntemlerine biçimbilimsel çözümleyici ve belirsizlik giderici çalışmaları ve çeviri kuralları eklenmiştir. Uygurca ile Türkçenin söz dizimi benzediğinden tez ile yapılan çalışmalarda, Uygurca ve Türkçenin söz dizimleri üzerinde ayrıca araştırma yapılmamıştır.

Bu tez kapsamında yapılan çalışmalar yoğun olarak Uygurca üzerinde yapılmıştır. Ancak Uygurca ile ilgili doğal dil çalışmaları ve bilgisayarlı çeviri çalışmaları için kaynak olabilecek kaynaklar bulunamamıştır. Türk dili ise Türk dilleri arasında doğal dil ve bilgisayarlı çalışmalar için en çok araştırılan dildir. Bu tezde çeviri yapılacak hedef dil Türkçe olduğundan, Türkçe ile ilgili yapılmış çalışmalardan yararlanılmıştır. Bu nedenle bu tezi çalışmasını yapabilmek için, Türkçe için yapılan daha önceki çalışmalardan örnek alarak bunları Uygurca için nasıl uygulanabileceği üzerinde araştırmalar yapılmıştır.

Araştırma sonucu, Türk dilleri arasına bilgisayarlı çeviri için temel taşı olabilecek çalışmanın, “akraba ve bitişken diller arasında bilgisayarlı çeviri için kavram model” adında çalışmaya rastlanmıştır [32]. Bu çalışma kapsamında, Türkmenceden Türkçeye bilgisayarlı bir çeviri sistemi geliştirilmiştir. Çeviri sistemi, herhangi bir

Türkî dilden Türkçeye çeviri sistemi gerçekleştirilecekse, sadece kaynak dilden hedef dile aktarma kuralları²⁴ tanımlandıktan sonra, Türkçe kısmı ile çalışmaya bir daha ihtiyaç olmayacak şekilde tasarlanmıştır. Bu nedenle yapılacak çalışma ise, sadece kaynak dil üzerinde yapılması gerekmektedir. Olası belirsizlikleri gidermek için İDM modeli kullanılmıştır. Herhangi bir bitişken dili Türkçeye çevire bilen bir çevirici çerçevesi geliştirilmiştir. Bu çerçeve hedef dil için, herhangi bir dilden bağımsız çalışabilmektedir. Çerçevede kullanılan Türkçe biçimbilimsel sentezleyici SDD kullanılarak geliştirildiğinden, yeni diller için de aynı yöntemin kullanılması gerekmektedir.

Bu nedenle, Uygurcadan Türkçeye çeviri sistemi geliştirirken, Karma yöntem için geliştirilen altyapı yani çerçeve kullanılmıştır, Karma yönteme değişiklik olarak, kaynak dil için biçimbilimsel giderici ve anlamsal belirsizlik giderici geliştirilmiştir. Ancak Uygurca ile ilgili herhangi bir derlem bulunamadığından, kural tabanlı belirsizlik giderici geliştirilmiştir. Karma modelde kullanılan İDM ise sadece eş anlamalı sözcükleri arasında seçme işlemi yapmak amacı için kullanılmıştır.

Bu geliştirilen yeni sistemin karma yöntem için yaptığı katkı ise, İDM in gönderilecek bilgiler üzerinde kesin bilgi gönderme imkânı sunmasıdır. Ancak bu yeni sistemin olumsuz yönü ise, İDM'e gelecek bilgi tek bilgi olduğundan, başka farklı bilgi seçme yeteneği kısıtlanmış oluyor. Bölüm 8'de verilen çeviri örneklerinden III böyle durum söz konusudur. Karma yöntemi ile “at” sözcüğü doğru bulunurken, belirsizlik giderme yöntemi nedeniyle “ad” yani “isim” anlamında çözümlenmiştir.

Tez sonunda, yapılan başarı değerlendirme sonucu, Karma yöntemi ile geliştirilen yeni sistemin başarıyı ufak bir farkla aynı sonuç vermiştir. Bu tez kapsamında, biçimbilimsel belirsizlik giderici için tanımlanan genel kurallar tanımlandığı, ayrıca anlamsal belirsizlik gidericini başarı oranının yüzde 50 ile 70 arasında olduğu göz önüne alındığında, geliştirilen yeni sistemin iyi bir sonuç verdiğinden söz edebiliriz.

Yani biçimbilimsel belirsizlik giderici için daha ayrıntılı kurallar tanımlandığında ve anlamsal belirsizliğin başarıyı yükseltildiğinde, sistemin genel başarısı daha da yükselebilir.

²⁴ Burada biçimbilimsel çözümlemelerde aynı etiket kullanılması gerekmektedir.

Sonuç olarak bu tezin bilimsel katkısı aşağıdaki gibi özetlenebilir.

1. Bu tezde Uyurca ile ilgili olarak ilk defa doğal dil çalışmaları yapılmıştır. Doğal dil çalışmalarını en temel taşları olan biçimbilimsel çözümleyici geliştirilmiştir. Geliştirilen bu çözümleyici ile Uyurca sözcükler üzerinde biçimbilimsel belirsizlik analizi yapılarak Uyurca sözcüklerde bulunan belirsizlik oranı somut veriler ile gösterilmiştir.
2. Uyurca ile ilgili herhangi bir derlem bulunamadığından, kural tanımlama yöntemleri kullanılarak, biçimbilimsel belirsizlik giderici geliştirilmiştir.
3. Geliştirilen sistemde, Uyurcadan Türkçeye sözcük aktarma yaparken, oluşabilecek anlamsal belirsizleri gidermek için kural çalışmaları yapılmıştır. Uyurca bitişken dil olduğundan, anlam belirsizlik giderici için bilinen yöntemler doğrudan kullanıldığında başarıyı düşük olduğundan, Uyurcaya özgü yöntem geliştirilmiştir. Yani anlamsal belirsizlik giderme çalışması yaparken, sözcüklerin görünen biçimine göre değil, sözcük köküne göre hesaplandığında daha doğru sonucun alınabileceği kanıtlanmıştır.
4. Derlemi bulunmayan bitişken diller üzerinde, anlamsal belirsizlik giderme çalışması yaparken, Lesk algoritmasının kök sözcükler üzerinde çalıştırması önerilmiştir. Önerilen bu yöntem Uyurca için yapılan sına verileri ile kanıtlanmıştır.
5. Türkçe ile ilgili daha önce yapılan ilgili çalışmalardan yararlanarak Uyurcadan Türkçeye kural tabanlı çeviri sistemi gerçekleştirilmiştir. Bu çeviri sisteminde, bir çeviri sisteminde bulunması gereken, belirsizlik giderici, aktarma kuralları gibi bileşenleri içermektedir. Geliştirilen çeviri sistem, daha önceden yapılan alt yapılarda da çalışabilmektedir.

Bu tez çalışmasında, yapılan Uyurca ile yapılan çalışmaların tümü ilk defa yapılmıştır. Bunun temel nedeni ise, bu dil ile ilgili daha önceden yapılan doğal dil çalışmalarının çok az olmasıdır. Ancak Uyurcanın Türkçe benzemesi yönünden yararlanarak, Türkçe için yapılan doğal dil çalışmalarının Uyurca için nasıl uygulanabileceği üzerinde araştırmalar yapılmıştır.

Tez çalışma sonucu, eksik görülen yöntemler ve öneriler şöyle sıralanabilir:

1. Bu tez kapsamında, daha önceden Türk dilleri için yapılan bilgisayarlı çeviri yöntemlerine göre en belirgin yönü ise belirsizlik giderme yönteminin kullanılmasıdır. Ancak, bu tezde belirsizlik giderme de tanımlanan kurallar genel durumlar için tasarlanmıştır. Bu nedenle eğer tanımlanan durum haricinde bir durum ile karşılaşıldığında, sistem yanlış sonuç verecektir. Dolayısıyla özel durumlar için araştırmalar yapılması gerekir.
2. Anlamsal belirsizlik gidermek için kullanılan BOS, gündelik yaşamda en çok kullanılan sözcüklere göre oluşturulmuştur ve bu sözcüklerin anlamı basit sözcükler ile oluşturulmuştur. Bu nedenle daha karmaşık sözcükler içeren tümce üzerinde çalıştırıldığında yanlış çözümlere neden olabilir. Bu çeşit durumların giderilmesi için, BOS' da geçen her sözcük adil bir şekilde, örnekler ile açıklanmalıdır. Birde açıklamak için kullanılan sözcük sayısının eşit olmasına çaba harcanmalıdır. Çünkü Lesk algoritması, açıklaması daha fazla sözcük içeren tümcelerde çalıştığında yanlış yapma olasılığı daha yüksektir.
3. Geliştirilen sistem, Akra diller arasında çeviri yapmak için tasarlanan çerçeve üzerinde çalıştırılmıştır. Ancak bu çerçevenin bir İDM modeli içermesi, ayrıca belirsizlik giderme yöntemlerinde kök bulma işleminin de zaman alması nedeni ile sistemin çalışması yavaş olmaktadır. İDM yeni sistemde sadece eş anlamlı sözcükleri seçme görevi görmektedir. Bu nedenle İDM modeli belirsizlik giderme yapısından çıkartılabilir.
4. Bu tezin en önemli kısmını oluşturan çalışma ise, biçimbilimsel çözümleyicidir. Bu çözümleyici, Türk dilleri için kullanılan etiketleri aynen kullanarak, SDD üzerinden geliştirilmiştir. Ancak bu çözümleyici sadece tek yönlü çalışabilmektedir ve kök sözcük kümesinde kullanılan sözcük sayısı kısıtlıdır. Bu nedenle sözcük kümesinde bulunmayan sözcükler ile ilgili çözümleme yapamaz. Biçimbilimsel çözümleyicisinin bir başka eksik yönü ise, tek yönlü çalışmasıdır. Bundan dolayı başka dillerden Uygurcaya bilgisayarlı çeviri çalışmalarında sentezleyici olarak kullanılamaz. Bu olumsuzlukları gidermek için, kök sözcük kümesinin genişletilmesi ve ters yönde çalışabilecek şekilde kuralların tanımlanması gerekir.

KAYNAKLAR

- [1] **Weaver, W.**, 1955. Translation. In Locke , W. N. ve Boothe, A. D., Machine Translation of Language, MIT Press, Cambridge, MA, pp. 15-23.
- [2] **Nagao, M., Tsujii, J., Yada, K., Kakimoto, T.**, 1982. An English Japanese Machine Translation System Of the Titles Of Scientific And Engineering Papers. International Conference On Computational Linguistics, Porceedings of the 9th conference on Computational Linguistics- Vol.1, Prague, Czechoslovakia, pp. 245- 225.
- [3] **Chandioux, J.**, 1976. Météo: Un système opérationnel pour la traduction automatique des bulletins météorologiques destinés au grand public. Meta, **21**, pp. 127-133.
- [4] **Url-1** <http://en.wikipedia.org/wiki/METEO_System#cite_note-0>, alındığı tarih 15.09.2009.
- [5] **Url-2** <http://en.wikipedia.org/wiki/Environment_Canada>, alındığı tarih 15.09.2009.
- [6] **Url-3** <http://www.chandioux.com/press_meteo20_eng.html >, alındığı tarih 25.09.2009.
- [7] **Hutchins, J. W.**, 2001. Machine translation over fifty years. Histoire, Epistemologie, Langage, Tome XXII, fasc. **1**, pp. 7–31.
- [8] **Url-4** <<http://en.wikipedia.org/wiki/SYSTRAN>>, alındığı tarih 25.10.2009.
- [9] **Toma P. P.**, 1977. Overcoming the language barrier, Systran As A Multilingual Machine Translation System. World Translation Center, (München: Verlag Dokumentation, 1977) 569, **Vol.1**.
- [10] **Seite. B., Bachut, D., Maret, D. ve Roudaud S. B.**, Presentation Of The Eurolang Project 12 rue de Reims - 94700 Maisons-Alfort.
- [11] **Hutchins, W. J.**, 1994. ‘Research methods and system designs in machine translation: a tenyear review, 1984-1994.’, Machine Translation, Ten Years On, 12-14 November 1994, Cranfield University.
- [12] **Hutchins, W. J.**, 1997. ‘From first conception to first demonstration: the nascent years of machine translation, 1947-1954. A chronology.’ Machine Translation **12** (3), 195-252. Hutchins, W.J. 1998. ‘The origins of the translator’s workstation.’ Machine Translation, **13** (4), pp. 287-307.
- [13] **Hutchins, W. J.**, 1999. ‘The development and use of machine translation systems and computerbased translation tools.’ International Conference on Machine Translation & Computer Language Information Processing. Proceedings of the conference, 26-28 June 1999, Beijing, China, ed. Chen Zhaoxiong, pp. 1-16.

- [14] **Url-5** <<http://www.cs.cmu.edu/~abberger/mt.html>>, alındığı tarih 20.09.2009.
- [15] **Url-6** <<http://spraakbanken.gu.se/pedant/parabank/node6.html>>, alındığı tarih 15.10.2009.
- [16] **Berger, A. B., Brown, P. F., Pietra, S. A. D., Pietra, V. J. D., Gillett, J. R., Lafferty, J. D., Mercer, R. L., Printz, H. Ve Ures, L.**, 1994. The Candide System for Machine Translation, Proceedings of the workshop on Human Language Technology, Plainsboro, NJ.
- [17] **Yamron, J., Baker, J., Bamberg, P., Chevalier, H., Dietzel, T., Elder, J., Kampmann, F., Mandel, M., Manganaro, L., Margolis, T ve Steele E.**, 1993. LINGSTAT: an interactive, machine-aided translation system, Proceedings of the workshop on Human Language Technology, Princeton, New Jersey, pp. 191-195.
- [18] **Alshawi, H.**, 1996. Head Automata and Bilingual Tiling: Translation with Minimal Representations, Proceedings of the 34th annual meeting on Association for Computational Linguistics, Santa Cruz, California, pp. 167- 176.
- [19] **Knight, K.**, 1993. Building A Large Ontology For Machine Translation. Proceedings of the workshop on Human Language Technology, Princeton, New Jersey, pp. 185-190.
- [20] **Hovy, E.**, 1994. Pangloss: Knowledge-Based Machine Translation, Proceedings of the workshop on Human Language Technology, Princeton, New Jersey, pp. 478-478.
- [21] **Dekai, W. ve Hkust, H. W.**, 1998. Machine Translation with a Stochastic Grammatical Channel, Proceedings of the 17th international conference on Computational linguistics- Vol.2, Montreal, Quebec, Canada, pp. 1408- 1415.
- [22] **Nagao, M.**, 1984. A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle, *Artificial and Human Intelligence*, Elithorn, A. and Banerji, R. (eds), Elsever Science Publishers.
- [23] **Uchida, H. ve Sugiyama, K.**, 1980. A Machine Translation System From Japanese Into English Based On Conceptual Structure, Proceedings of the 8th conference on Computational linguistics, Tokyo, Japan, pp. 455- 462.
- [24] **Hanneman, G., Huber, E., Agarwal, A., Ambati, V. Parlikar, A. Peterson, E. ve Lavie, A.**, 2008. A Statistical Transfer Systems for French–English and German–English Machine Translation, Proceedings of the Third Workshop on Statistical Machine Translation, Columbus, Ohio, pp. 163- 166.
- [25] **Hajič, J.**, 1987. RUSLAN - An MT System Between Closely Related Languages, in *Third Conference of the European Chapter of the Association for Computational Linguistics (EACL'87)* Copenhagen, Denmark.
- [26] **Hajič, J., Hric, J., ve Kuboň, V.**, 2000: Machine translation of very close languages, in *Proceedings of the sixth conference on Applied natural language processing*, Morgan Kaufmann Publishers Inc, pp. 7-12.

- [27] **Dvořák, B., Homola, P., ve Kuboň, V.**, 2006. Exploiting similarity in the MT into a minority language, Fifth International Conference on Language Resources and Evaluation, 5th SALT MIL Workshop on Minority Languages: "Strategies for developing machine translation for minority languages" Genoa, Italy.
- [28] **Hamzaoglu, İ.**, 1993. Machine translation from Turkish to other Turkic languages and an implementation for the Azeri languages, in *Institute for Graduate Studies in Science and Engineering*. MSc Thesis İstanbul: Bogazici University.
- [29] **Altıntaş, K.**, 2000. Turkish to Crimean Tatar Machine Translation System, in *Bilgisayar Mühendisliği Bölümü*. MSc Ankara: Bilkent Üniversitesi.
- [30] **Oflazer, K.**, 1995. Two-level Description of Turkish Morphology, *Literary and Linguistic Computing*, **Vol. 9**, pp. 137-148.
- [31] **Altıntaş, K. ve Çiçekli, İ.**, 2001. A Morphological Analyser for Crimean Tatar, in *Proceedings of the 10th Turkish Symposium on Artificial Intelligence and Neural Networks*, TAINN North Cyprus, pp. 180-189.
- [32] **Tantuğ, A. C.**, 2007. Akraba ve Bitişken Diller Arasında Bilgisayarlı Çeviri İçin Karma Bir Model. . *Bilgisayar Mühendisliği Bölümü*. Doktora Tezi. İstanbul: İstanbul Teknik Üniversitesi.
- [33] **Tantuğ, A. C., Adalı E., ve Oflazer K.**, 2006. Computer Analysis of The Turkmen Language Morphology, *FinTAL Lecture Notes in Computer Science*, Vol.4139, Springer, pp. 186-193.
- [34] **Tantuğ, A. C., Adalı, E., ve Oflazer, K.**, 2007. Machine Translation between Turkic Languages. *Proceedings of the ACL 2007 Demo and Poster Sessions*, Prague, pp. 189–192.
- [35] **Duval, J. R. ve Janbaz, W. A.**, 2006. An Introduction to Latin-Script Uyghur. *Middle East & Central Asia Politics, Economics, and Society Conference*. Sept 7-9, University of Utah, Salt Lake City, USA.
- [36] **Saleh, I. ve Janbaz, W.A.**, 2006. Web Development Considerations for Unicode-based Text Processing in Uyghur Language. *The 30th Internationalization and Unicode Conference*, Washington, DC USA.
- [37] **Belikiz**, 2007. The 3253 different word forms Uyghur Verb "qıl". *Corpus Linguistics and Corpus Based Reseach. Dept of Linguistics*, College of Anthropology Xinjiang Normal University.
- [38] **Belikiz, Ablajan ve Dilmurat**, 2007. The 2107 different words forms of Uyghur verb "bol". *Corpus Linguistics and Corpus Based Reseach, Dept of Linguistics*, College of Anthropology Xinjiang Normal University.
- [39] **Eziz, G.**, 2007. Resistance to Borrowing of Uyghur Verbs. *Annual Conference*, University of Washington.
- [40] **Aibaidulla, Y. ve Kim-Teng, L.**, 2003. The development fo Tagged Uyghur Corpus, *Proceedings of PACLIC17*, 1-3 Sentosa, Singapore, pp. 228-234.

- [41] **Abaidula, Y., Rezwangul ve Sali, A.**, 2005. The Research and Development of Computer Aided Contemporary Uighur Language Tagging System. *Journal of Chinese Language and Computing* 15(4), pp. 203-210.
- [42] **Kaşgarlı, S. M.**, 1992. “*Modern Uygur Türkçesi Grameri*” İstanbul Orkun Yayınevi.
- [43] **Mahsut, M., Ogawa, Y., Sugino, K., Tuyama, K., ve Inagaki, Y.**, 2004. "An Experiment on Japanese- Uighur Machine Translation and Its Evaluation". *AMTA 2004, LNAI 3265*, pp. 208-216.
- [44] **Köksal, A.**, 1976. Türkçenin Özdevimli Biçimbilim Çözümlemesi, Doktora tezi, Hacettepe Üniversitesi, Ankara.
- [45] **Tantuğ, A. C., Adalı, E., ve Oflazer, K.**, 2006. Lexical Ambiguity Resolution for Turkish in Direct Transfer Machine Translation Models, The 21st International Symposium on Computer and Information Sciences, ISCIS 2006, İstanbul, Turkey, Lecture Notes in Computer Science, Springer, **Vol. 4263**, pp. 230-238.
- [46] **Sagay, Z.**, 1981. A computer translation from English to Turkish. Yüksek Lisans Tezi, Bilgisayar Mühendisliği Bölümü, ODTÜ. Ankara.
- [47] **Tantuğ, A. C., Cebiroğlu, G., Adalı, E., ve Erenler, Y.**, 2003. Synthetic Turkish Word Root Generation, Proceedings of the Turkish Artificial Intelligence and Neural Networks, TAINN 2003, Canakkale, Turkey
- [48] **Hankamer, J.**, 1986. Finite state morphology and left-to-right morphology, West Coast Conference on Formal Linguistics, Bildiri Kitabı.
- [49] **Güvenir, H. A.**, 1991. Drill and Practice for Turkish Grammar, Intelligent Tutoring Systems for Foreign Language Learning. The Bridge to International Communication, NATO ASI Serisi, Springer-Verlag, Berlin.
- [50] **Eryiğit, G., Adalı, E. ve Oflazer, K.**, 2006. Türkçe Cümlelerin Kural Tabanlı Bağlılık Analizi, In: Proceedings of the 15th Turkish Symposium on Artificial Intelligence and Neural Networks, Muğla, pp. 17-24.
- [51] **Stoop, A.**, Transit in the world of Machine Translation, Towards automatic translator for Dutch and Turkish. 3. Türkçe Dilbilimi Konferansı Bildiri Kitabı Tilburg, Hollanda.
- [52] **Tantuğ, A. C. ve Eryiğit, G.**, 2004. Probabilistic Turkish Word Root Generation, Proceedings of the 3rd Asia Pacific International Symposium on Information Technology, Istanbul, Turkey, pp. 443-447.
- [53] **Hakkani-Tür, D. Z., Oflazer, K. ve Tür, G.**, 2002. Statistical Morphological Disambiguation for Agglutinative Languages, Computers and the Humanities, **Vol. 36**, No.4.
- [54] **Solak, A. ve Oflazer, K.**, 1993. Design and Implementation of a Spelling Checker for Turkish, Literary and Linguistic Computing, **Vol. 8**, No:3, Oxford University Press.
- [55] **Tantuğ, A. C., Oflazer, K. ve El-Kahlout, İ. D.**, 2008, BLEU+: A Tool for Fine-Grained BLEU computation, Proceedings of LREC, Morocco.

- [56] **Url-7** <<http://ddi.ce.itu.edu.tr>>, alındığı tarih 15.09.2009.
- [57] **Url-8** <<http://www.tdk.gov.tr>>, alındığı tarih 15.09.2009.
- [58] **Url-9** <<http://www.tubitak.gov.tr>>, alındığı tarih 15.09.2009.
- [59] **Url-10** <<http://ddi.ce.itu.edu.tr/projects>>, alındığı tarih 15.09.2009.
- [60] **D. Jurafsky ve J. H. Martin**, 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*: Prentice Hall.
- [61] **Hutchins, W. J. ve Somers, H. L.**, 1992. *An Introduction to Machine Translation* Academic Press, London.
- [62] **Wilks, Y.**, 1992. *Systran: It Obviously Works, But How Much Can It Be Improved? Computers in Translation: A Practical Appraisal*, Routledge, London, pp. 166-188.
- [63] **Maas H. D.**, 1987. The MT System SUSY Machine Translation Today: the State of the Art, *Proceedings of the Third Lugano Tutorial*, Edinburgh University Press, Edinburgh, pp. 209-246.
- [64] **Vauquois B. ve Boitet C.**, 1985. *Automated Translation at Grenoble University, Special Issue on machine translation*, MIT Press, Cambridge, MA, USA, pp. 28-36.
- [65] **Appleby. S. ve Prol, M. P.**, 2000. Multilingual World Wide, Web, *BT Technology Journal*, Millenium Edition, Vol. **18**, No. 1.
- [66] **Wahlster, W.**, 1993. *Verbmobil Translation of Face-To-Face Dialogs*, Proceedings of MT Summit IV, Kobe, Japan.
- [67] **Dorna, M. ve Martin C. E.**, 1996. Semantic-based Transfer, Proceedings of the 16th conference on Computational linguistics, Vol (**1**), Copenhagen, Denmark, pp. 316- 321.
- [68] **Landsbergen, J.**, 1987. Isomorphic Grancakrs and Their Use in the Rosetta Translation System Machine Translation Today, 1987: the State of the Art, *Proceedings of the Third Lugano Tutorial*, Edinburgh University Press, Edinburgh, pp. 351-377.
- [69] **Goodman K. ve Nirenburg S.**, 1991. *The KBMT Project: A Case Study in Knowledge- Based Machine Translation*, Press, Morgan Kaufmann, San Mateo, California.
- [70] **Mitkov, R.**, 2003. *The Oxford Handbook of Computational Linguistics*, Oxford University Press, New York.
- [71] **Hakkani-Tur, D.**, 2000. *Statistical Modelling of Agglutinative Languages*, PhD Thesis, Department of Computer Engineering, Bilkent University.
- [72] **Çicekli, İ. ve Guvenir, H. A.**, 2001. Learning Translation Templates from Bilingual Translation Examples, *Applied Intelligence*, Vol. **15**, No. 1, pp. 57-76.
- [73] **Brown, R. D.**, 1996. Example-Based Machine Translation in the Pangloss System, in *The 16th International Conference on Computational Linguistics (COLING-96) Copenhagen, Denmark*. Vol.**1**, pp. 169-174.

- [74] **Güvenir, H. A., ve Çiçekli, İ.**, 1998. "Learning Translation Templates from Examples," *Information Systems*, Vol. **23**, pp. 353-363.
- [75] **Somers, H.**, 1999. Review Article: Example-based Machine Translation. Vol. **14**: Kluwer Academic Publishers, pp. 113-157.
- [76] **Sproad, R.**, 1992. *Morphology and Computation*, MIT press.
- [77] **Cebiroğlu, G., Tantı, A. C., Adalı, E. ve Erenler, Y.**, 2003. Sentetik Türkçe Sözcük Kökleri Üretimi International XII. Turkish Symposium on Artificial Intelligence and Neural Networks - TAINN.
- [78] **Oflazer, K. ve Güzey, C.**, 1994. Spelling Correction in Agglutinative Languages, *Proceedings of the 4th ACL Conference on Applied Natural Language Processing*, Stuttgart, Germany.
- [79] **Do-Gil, L., Sang-Zoo, L. ve Hae-Chang R.**, An Efficient Method for Korean Noun Extraction Using Noun Occurrence Characteristics.
- [80] **Vural, E., Erdogan, H., Oflazer, K. ve Yanikoglu, B.**, 2005. An Online Handwriting Recognition System For Turkish, *Proceedings of SPIE Vol. 5676 Electronic Imaging*, San Jose.
- [81] **Yamashita, T., ve Matsumoto, Y.**, 2000. Graduate School of Information Science Nara Institute of Science and Technology , *Applied Natural Language Conferences archive, Proceedings of the sixth conference on Applied natural language processing table of contents* Seattle, Washington, pp. 232 – 238.
- [82] **Cercone, N.**, 1978. *Morphological Analysis and Lexicon Design for Natural-Language Processing Computers and the Humanities*, Pergamon Press, Vol. **11**, pp. 235-258.
- [83] **Eryiğit, G. ve Adalı, E.**, 2004. An Affix Stripping Morphological Analyzer For Turkish. *Proceedings of the IASTED International Conference Artificial Intelligence and applications*. Innsbruck, Austria.
- [84] **Necip, E. N.**, Yeni Uygur Türkçesi Sözlüğü.. Atatürk Kültür, Dil ve Tarih Yüksek Kurumu. Türk Dil Kurumu Yayınları: **615**. Sözlük Bilim ve Uygulama Kolu Yayınları, 2.
- [85] **Yaqub, A., Gheyurani, Gh., Qadir, İ., Abduraxman, H., Emet, A., Nur, P., Abduqadir, E., Tahir, A. ve Rehimcan, A.**, 1999. *Uygur Tilining İzahliq Lughiti*. Shin Jiang Xeliq Neshiryatı.
- [86] **Oflazer, K. ve Kuruöz, İ.**, 1994. Tagging and Morphological Disambiguation of Turkish Text, *The Fourth ACL Conference on Applied Natural Language Processing*, Stuttgart, Germany, pp. 144-149.
- [87] **Oflazer, K. ve Tür, G.**, 1996. Combining Hand-crafted Rules and Unsupervised Learning in Constraint-based Morphological Disambiguation, *Proceedings of the ACL-SIGDAT Conference on Empirical Methods in Natural Language Processing*, Philadelphia, PA, USA.

- [88] **Yüret, D., ve Türe, F.**, 2006. Learning Morphological Disambiguation Rules for Turkish, in North American Chapter of the Association for Computational Linguistics Annual Meeting (HLT-NAACL 2006) New York City.
- [89] **Mohri, M.**, 1996. On Some Applications of Finite-State Automata Theory to Natural Language Processing. *Natural Language Engineering*, 2, pp. 1–20.
- [90] **Koskenniemi, K.**, 1983. Two-Level Morphology : A General Computational Model for Word Form Recognition and Production, Department of General Linguistics, University of Helsinki.
- [91] **Karttunen, L.**, 1983. KIMMO : A General Morphological Processor, in Texas Linguistic Forum, Texas, USA, 1983, pp. 163-186.
- [92] **Antworth, E. L.**, 1990. PC-KIMMO: A Two-Level Processor for Morphological Analysis, Occasional Publications in Academic Computing No.6, Summer Institute of Linguistics, Dallas, Texas, USA.
- [93] **Alam, Y. S.**, 1983. A Two-Level Morphological Analysis of Japanese. *Texas Linguistics.Forum*, 22, pp. 229-252.
- [94] **Karttunen, L. ve Wittenburg, K.**, 1983. A Two-Level Morphological Analysis of English. *Texas Linguistics Forum*, 22, pp. 217-228.
- [95] **Kim, DB., Lee, SJ., Key-Sun Choi Gil-Chang Kim**, 1994. Proceedings of the 15th conference on Computational linguistics – Vol.1 Kyoto, Japan, pp. 535 – 539.
- [96] **Lun, S.**, 1983. A Two-Level Morphological Analysis of French, in Texas Linguistic Forum, Texas, USA, 1983, pp. 271-278.
- [97] **R. Khan, R.**, 1983. A Two-Level Morphological Analysis of Rumanian, in Texas Linguistic Forum, Texas, USA, pp. 253-270.
- [98] **Koskenniemi, K.**, 1985. An Application of the Two-Level Model to Finnish, University of Helsinki Department of General Linguistics .
- [99] **Karttunen, L., Gaal, T. ve Kempe, A.**, 1983. Xerox Finite-State Tool, XEROX Research Centre, Europe, Technical Report 1997. Linguistic Forum, Texas, USA, pp. 253-270.
- [100] **Beesley, K. R. ve Karttunen, L.**, 2003. Finite State Morphology. Stanford: CSLI Publications.
- [101] **Papineni, K., Roukos, S., Ward, T. ve Wei-Jing Z.**, 2002. Bleu: a Method for Automatic Evaluation of Machine Translation, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, pp. 311-318.
- [102] **Doddington, G.**, 2002. Automatic Evaluation of MT Quality Using n-gram Co-occurrence Statistics. In Proceedings of Human Language Technology Conference, San Diego, CA, pp. 128–132.
- [103] **Turian, J. P., Shen, L. ve Melamed, I. D.**, 2003. Evaluation of Machine Translation and its Evaluation, MT Summit IX, New Orleans, LA.

- [104] **Banerjee, S. ve Lavie, A.**, 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments, in ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization Ann Arbor, MI, USA.
- [105] **Nießen, S., Och, F., Leusch, G., ve Ney, H.**, 2000. An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research. In Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC), Athens, Greece, pp. 39–45.
- [106] **Url-11** < http://en.wikipedia.org/wiki/Levenshtein_distance > alındığı tarih 05.09.2009.
- [107] **Levenshtein VI**, 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady **10**: 707–10.
- [108] **Url-12** < <http://www.ukij.org> >, alındığı tarih 06.06.2007.
- [109] **Url-13** < <http://unicode.org> >, alındığı tarih 06.06.2007.
- [110] **Tömür, H.**, 1997. Modern Uygur Grammar (Morphology). Yıldız Teknik Üniversitesi, Fen-Ed Fak. T.D.E Bölümü. İstanbul. (İngilizce baskı).
- [111] **Osmanof, M.**, 1997. Hazirqi Zaman Uyghur Edebiy Tilining İmla ve Teleppuz Lughiti. Shin Jiang Xeliq Neshiryatı. (Uygurca baskı).
- [112] **Shen, Y., Gao, X. ve Ju, X.**, 1998. "Uyghur Dili Grammatkası". Çin Halik Cumhuriyeti Merkezi Milletler Üniversitesi Baskısı. (Çince baskı).
- [113] **Öztürk, R.**, 1993. "Yeni Uygur Türkçesi Grameri". Türk dil kurumu yayınları: **593**.
- [114] **Orhun, M., Tantuğ, A. C. ve Adalı, E.**, 2008. Rule Based Analysis of the Uyghur Nouns, *Proceedings of the International Conference on Asian Language Processing (IALP)*. Chiang Mai, Thailand, pp. 8-14.
- [115] **Url-14** < <http://www.ling.helsinki.fi/~koskenni/esslli-2001-karttunen> >, alındığı tarih 25.09.2006.
- [116] **Orhun, M., Tantuğ, A. C. ve Adalı, E.**, 2009. Rule Based Tagging of the Uyghur Verbs, *Proceedings of the Fourth International Conference on Intelligent Computing and Information Systems*, Cairo, Egypt, pp. 811-816.
- [117] **Hengirmen, M.**, 2005. "Türkçe Dilbilgisi", ISBN: 975-7287-26-1, Engin Yayınevi, Selanik Caddesi 28/6. 06650 Kızılay- Ankara.
- [118] **Orhun, M., Tantuğ, A. C. ve Adalı, E.**, 2008. Uygur dili ve Bilgisayarlı Çeviri. *Akademik Bililşim 2008*. Çanakkale 18 Mart Üniversitesi, Çanakkale, Türkiye.
- [119] **Hakkani-Tur, D.**, 2000. Statistical Modelling of Agglutinative Languages, Ph.D Thesis, Department of Computer Engineering, Bilkent University.
- [120] **Sak, H., Güngör, T. ve Saraçlar, M.**, 2007. Morphological disambiguation of Turkish text with perceptron algorithm, *Proceedings of International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, Mexico City, pp. 107–118.

- [121] **Orhun, M., Tantuğ, A. C. ve Adalı, E.**, 2010. Uygucada Biçimbilimsel Belirsizlik, *Akademik Bilişim 2010*. Muğla Üniversitesi. Muğla, Türkiye.
- [122] **Booth, A.D, Locke, W. N.**, 1995. Machine Translation of Languages: Fourteen Essays, Cambridge, MA.
- [123] **Url-15** <http://en.wikipedia.org/wiki/Word_sense_disambiguation>, alındığı tarih 01.02.2010.
- [124] **Orhan, Z. ve Altan, Z.**, 2006. Anlam Belirsizliği İçeren Türkçe Sözcüklerin Hesaplamalı Dilbilim Uygulamalarıyla Belirginleştirilmesi, XX. Ulusal Dilbilim Kurultayı, İstanbul.
- [125] **Jurafsky, D. ve Martin, J.H.**, 2007. Speech and Language Processing, An introduction to natural language processing, computational linguistics, and speech recognition, Printice Hall.
- [126] **Lesk, M.**, 1986. Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Code From an Ice Cream Cone, Proceedings of the 5th Annual International Conference on Systems Documentation, ACM Press, pp. 24–26.
- [127] **Banerjee, S.**, 2002. Adapting the Lesk Algorithm for Word Sense Disambiguation to WordNet. Master of Science, Department of Computer Science University of Minnesota Duluth, Minnesota 55812, U.S.A. (Yüksek Lisans Tezi).
- [128] **Miller, G. A., Leacock, C., Teng, R., ve Bunker, R. T.**, 1993. A semantic concordance, Proceedings ARPA Workshop on Human Language Technology, ACL , pp. 303–308.
- [129] **Landes, S., Leacock, C. ve Teng, R. I.**, 1998. Building semantic concordances. In Fellbaum, C. (Ed.), WordNet: An Electronic Lexical Database, MIT Press, Cambridge, MA, pp. 199–216.
- [130] **Yarowsky, D.**, 1995. Unsupervised word sense disambiguation rivaling supervised methods. Proceedings of the 33rd annual meeting on Association for Computational Linguistics, Cambridge, Massachusetts, pp. 189-196.
- [131] **Schütze, H.**, 1998. Automatic word sense discrimination, Computational Linguistics, MIT Press, Cambridge, MA, 24(1), pp. 97-123.

EKLER

EK A: Uygurcanın Biçimbilimsel Kuralları

EK B: Uygurcanın Biçimbilimsel Çözümleme Örnekleri

EK C: Biçimbilimsel Etiketlerinin Açıklamaları

EK D: Uygurcadan Türkçeye Örnek Çeviri Kuralları

EK A: Uygurcanın Biçimbilimsel Kuralları

Alphabet

a e b p t j c x d r z Z s S G f q k g N l m n h o O u U w E i y %+:0 ;

Sets

CONS = b p t j c x d r z Z s S G f q k g N l m n h w y;

CONSVO = b j d r z Z G g N l m n h w y;

CONSVOaz = j r z Z G g N l m n w y;

CONSUVO = p t c x s S f q k ;

CONSFV = k g; !MIDV'leri FRONTV yapan CONS'lar.

CONSBV = b p t j c x d r z Z s S G f q N l m n h w y; !MIDV'leri BACKV yapan CONS'lar.

VOWEL = a e E i o O u U;

FRONTV = e O U ;

BACKV = a o u ;

MIDV = E i;

NLABV = a e E i ;

LABV = o O u U ;

İyelik ekleri için yeni kümelerin tanımlanması.

Lou = o u ;

LOU = O U ;

LeiE = e i E a;

La = a;

Rules

"01- 'A' ise 'a' olarak görünür "

A:a => [[:CONS]*[:BACKV][CONS]*]+(%+:0)[[:CONS |CONS:]* _ ;

[[CONSBV+][:MIDV][CONSBV]*]+(%+:0)[[:CONSBV |CONSBV:]* _ ;

"02- 'A' ise 'e' olarak görünür "

A:e => [[[CONSFV][:MIDV][CONS]*]+][[MIDV][CONS]*]*(%+:0)[:CONS|
CONS:]* _ ;

[:CONS]*[:FRONTV][CONS]*+(%+:0)[:CONS |CONS:]* _ ;

[[CONS]*[:MIDV][CONSFV]+]+(%+:0)[:CONS |CONS:]* _ ;

"03- 'H' ise 'i' olarak görünür "

H:i <=> [:CONS]*[:LeiE][:CONS]+ (%+:0) _ ;

[:CONS]*[:VOWEL][:CONS]* (%+:0) n _ N;

[:CONS]*[:VOWEL]*[:CONS]* (%+:0) n _ ;

[CONS]* D _ ;

"04- 'H' ise 'u' olarak görünür "

H:u => [:CONS]*[:Lou][:CONS]+ (%+:0) _ ;

"05- 'H' ise 'U' olarak görünür "

H:U => [:CONS]*[:LOU][:CONS]+ (%+:0) _ ;

"06- 'H' ise '0' olarak görünür "

H:0 => [:CONS]*[:VOWEL] (%+:0) _ ;

"07- 'H' ise 'u' olarak görünmez "

H:u /<= [:CONS]*[:VOWEL][:CONS]* (%+:0) n _ N;

[:CONS]*[:VOWEL][:CONS]* (%+:0) n _ ;

"08- 'H' ise 'U' olarak görünmez "

H:U /<= [:CONS]*[:VOWEL][:CONS]* (%+:0) n _ N;

[:CONS]*[:VOWEL][:CONS]* (%+:0) n _ ;

"09- 'D' ise t olarak görünür "

D:t <=> [:CONS]*[:VOWEL][CONSUV0]+ (%+:0) _ [H:i] ;

[:CONS]*[:VOWEL][CONSUV0]+ (%+:0) _ [A:a][A:e];

"10- 'D' ise 'd' olarak görünür "

D:d <=> [:CONS]*[:VOWEL][CONSV0az]+ (%+:0) _ [H:i];

[:CONS]*[:VOWEL][CONSV0az]* (%+:0) _ [A:a][A:e];

"11- 'Y' ise 'q' olarak görünür "

Y:q => [:CONS]*[: [o | u | a]][CONSUV0]+ (%+:0) _ ;

"12- 'Y' ise 'k' olarak görünür "

Y:k => [:CONS]*[: [e | i | E | U | O]][CONSUV0] (%+:0) _ ;

"13- 'Y' ise 'G' olarak görünür "

Y:G => [:CONS]*[: [o | u | a]][CONSV0az]+ (%+:0) _ ;

"14- 'Y' ise 'g' olarak görünür "

Y:g => [:CONS]*[: [e | i | E | U | O]][CONSV0az]+ (%+:0) _ ;

"15- 'a' ise 'E' olarak görünür "

a:E <=> .#. [:CONS]* _ [CONS] (%+:0) [[H:i] m [[H:i] N | [H:i] | [H:i] m [H:i] z
[[H:i] N [H:i] z] ;

"16- 'e' ise 'E' olarak görünür "

e:E <=> .#. [:CONS]* _ [CONS] (%+:0) [[H:i] m [[H:i] N | [H:i] | [H:i] m [H:i] z
[[H:i] N [H:i] z] ;

"17- 'a' ise 'i' olarak görünür "

a:i <=> [CONS]*[VaeE][CONS][[CONS]*[VaeE][CONS]]*[CONS]* _ [CONS]*
(%+:0)[[H:i] m [[H:i] N | [H:i] | [H:i] m [H:i] z [[H:i] N [H:i] z] ;
[CONS]*[VOWEL][CONS] _ s [H:i]

"18- 'e' ise 'i' olarak görünür "

e:i <=> [CONS]*[VaeE][CONS][[CONS]*[VaeE][CONS]]*[CONS]* _ [CONS]*
(%+:0)[[H:i] m [[H:i] N | [H:i] | [H:i] m [H:i] z [[H:i] N [H:i] z] ;

"19- 'u' ve 'i' karakterin silinmesi"

[u:0 | i:0] <=> [CONS]*[\$:0] _ [CONS][CONS]*[VOWEL]*(%+:0)H ;

"20- 'B' ise 'a' olarak görünür "

B:a <=> [CONS][Ba][CONS] _ y ;

"21- 'B' ise '0' olarak görünür "

B:0 <=> [CONS][VOWEL] _ y ;

"22- 'B' ise 'e' olarak görünür "

B:e <=> [CONS][Be][CONS] _ y ;

"23- 'F' ise 'y' olarak görünür "

F:y => [CONS]*[VOWEL] _ ;

"24- 'F' ise 'i' olarak görünür "

F:i => [CONS]*[VOWEL]*[CONS] _ ;

"25- 'I' ise '0' olarak görünür "

I:0 <=> [:CONS]*[:VOWEL] (%+:0) _ ;

"26- 'I' ise 'i' olarak görünür "

I:i <=> [:CONS]*[a | e | i | E][CONS] _ S ;

[:CONS]* [a | e | i | E][CONS] _ v a t [:CONS]*[:VOWEL] ;

"27- 'I' ise 'U' olarak görünür "

I:U <=> [:CONS]* [U | O][CONS] _ S ;

[:CONS]*[:VOWEL]*[:CONS]*[Y:g | Y:k] _ ;

[:CONS]* [U | O][CONS] _ v a t [:CONS]*[:VOWEL] ;

"28- 'I' ise 'u' olarak görünür "

I:u <=> [:CONS]*[u | o][CONS] _ S ;

[:CONS]*[:VOWEL]*[:CONS]*[Y:G | Y:q] _ ;

[:CONS]* [u | o][CONS] _ v a t [:CONS]*[:VOWEL] ;

EK B: Uygurcanın Biçimbilimsel Çözümleme Örnekleri

1. İsimler.

kelem: kelem+Noun+A3sg+Pnon+Nom

kelemler: kelem+Noun+A3pl+Pnon+Nom

kelemlerim: kelem+Noun+A3pl+P1sg+Nom

kelemlerimniN: kelem+Noun+A3pl+P1sg+Gen

insan: insan+Noun+A3sg+Pnon+Nom

insanlar: insan+Noun+A3pl+Pnon+Nom

insani: insan+Noun+A3sg+P3sg+Nom

insanimiz: insan+Noun+A3sg+P1pl+Nom

insanniN: insan+Noun+A3sg+Pnon+Gen

insanda: insan+Noun+A3sg+Pnon+Loc

insanGa: insan+Noun+A3sg+Pnon+Dat

insandin: insan+Noun+A3sg+Pnon+Abl

sanlarimiz: insan+Noun+A3pl+P1pl+Nom

insanlarimizda: insan+Noun+A3pl+P1pl+Loc

insanlarimizdaki: insan+Noun+A3pl+P1pl+Loc+Ki

2. Eylemler

2.1: Geçmiş zaman (Simple past)

yazdim: yaz+Verb+Pos+Past+A1sg

yazdiN: yaz+Verb+Pos+Past+A2sg

yEziptimen: yaz+Verb+Pos+Past+Ind+A1sg

yazGandimen: yaz+Verb+Pos+Past+Subj+A1sg

yEzipsen: yaz+Verb+Pos+Past+Ind+A2sg

yEziptimenmiS: yaz+Verb+Pos+Past+A1sg+Narr

yaz+Verb+Pos+Pres+Perfect+A1sg+Narr

yEziptimenmiS: yaz+Verb+Pos+Past+A1sg+Narr

yaz+Verb+Pos+Pres+Perfect+A1sg+Narr

yEziptidek: yaz+Verb+Pos+Past+Narr+A3sgl

yEzip: yaz+Verb+Pos+Past+Ind+A3sg

yaz+Verb+Pos^DB+Adv+Hp

yEziptimen: yaz+Verb+Pos+Past+Ind+A1sg

yEziptidek: yaz+Verb+Pos+Past+Narr+A3sg

2.2: Şimdiki Tamamlanmış zaman (Present Perfect)

yazGan: yaz+Verb+Pos+Pres+Perfect+A3sg

yaz+Verb+Pos^DB+Adj+Yan

yazGanken: yaz+Verb+Pos+Pres+Perfect+Ind+A3sg

yazGankenmen: yaz+Verb+Pos+Pres+Perfect+Ind+A1sg

yazGanmiS: yaz+Verb+Pos+Pres+Perfect+A3sg+Narr

yaz+Verb+Pos+Pres+Perfect+Narr+A3sg

yazGanmenmiS: yaz+Verb+Pos+Pres+Perfect+A1sg+Narr

2.3: Geçmişte tamamlanmış zaman (Past perfect)

yazGantim: yaz+Verb+Pos+Past+Perfect+A1sg

yazGantiN: yaz+Verb+Pos+Past+Perfect+A2sg

yazGankenmiSmen: yaz+Verb+Pos+Past+Perfect+Narr+A1sg

yazGankenmiSmen: yaz+Verb+Pos+Past+Perfect+Narr+A1sg

2.4: Şimdiki tamamlanmayan zaman (Present imperfect)

yazarmen: yaz+Verb+Pos+Pres+ImPer+A1sg

yazidikenmen: yaz+Verb+Pos+Past+Iperfect+Ind+A1sg

yaz+Verb+Pos+Pres+ImPer+Ind+A1sg

yazGidekmen: yaz+Verb+Pos+Pres+ImPer+Narr+A1sg

yazidiGandimen: yaz+Verb+Pos+Pres+ImPer+A1sg

2.5: Geçmişte tamamlanmayaz zaman (Past imperfect)

yazattim: yaz+Verb+Pos+Past+IPerfect+A1sg

yazattuq: yaz+Verb+Pos+Past+IPerfect+A1pl

yazidikenmen: yaz+Verb+Pos+Past+IPerfect+Ind+A1sg

yaz+Verb+Pos+Pres+ImPer+Ind+A1sg

yazidiGankenmen: yaz+Verb+Pos+Past+IPerfect+Ind+A1sg

yazidiGanken: yaz+Verb+Pos+Past+IPerfect+Ind+A3sg

2.6: Şimdiki devam eden zaman (Present continuous)

yEziwatimen: yaz+Verb+Pos+Pres+Prog+A1sg

yEziwatisen: yaz+Verb+Pos+Pres+Prog+A2sg

yEziwatidu: yaz+Verb+Pos+Pres+Prog+A3sg

yEziwitiptimen: yaz+Verb+Pos+Pres+Prog+Ind+A1sg

yEziwatqudekmen: yaz+Verb+Pos+Pres+Prog+Narr+A1sg

yEziwatarmiSmen: yaz+Verb+Pos+Pres+Prog+Narr+A1sg

yEziwatqandimen: yaz+Verb+Pos+Pres+Prog+A1sg

yazmawatimen: yaz+Verb+Neg+Pres+Prog+A1sg

2.7: Geçmişte devam eden zaman (Past continuous)

yEziwatattim: yaz+Verb+Pos+Past+PastProg+A1sg

yEziwatqankenmen: yaz+Verb+Pos+Past+PastProg+Ind+A1sg

yEziwatqankenmiSmen: yaz+Verb+Pos+Past+PastProg+Narr+A1sg

2.8: Şimdiki amaç zamanı (Present intention)

yazmaqci: yaz+Verb+Pos+Pres+Int+A3sg

yazmaqçimen: yaz+Verb+Pos+Pres+Int+A1sg

yazmaqçimiSmen: yaz+Verb+Pos+Pres+Int+Narr+A1sg

yazmaqçikenmen: yaz+Verb+Pos+Pres+Int+Ind+A1sg

2.9: Geçmişteki amaç zamanı (Past intention)

yazmaqçitim: yaz+Verb+Pos+Past+Int+A1sg

yazmaqçikenmenmiS: yaz+Verb+Pos+Past+Int+A1sg+Narr

yazmaqçikenmenmiS: yaz+Verb+Pos+Past+Int+A1sg+Narr

2.10: Özekler (Stem)

yEziS: yaz+Verb^{DB}+Verb+Recip+Pos+Imp+A2sg

yEziSmaçci: yaz+Verb^{DB}+Verb+Recip+Pos+Pres+Int+A3sg

yEzilmaqçikenmiS: yaz+Verb^{DB}+Verb+Pass+Pos+Past+Int+A3sg+Narr

yaz+Verb^{DB}+Verb+Pass+Pos+Past+Int+Narr+A3sg

yaz+Verb^{DB}+Verb+Reflex+Pos+Past+Int+A3sg+Narr

yaz+Verb^{DB}+Verb+Reflex+Pos+Past+Int+Narr+A3sg

yazGuz: yaz+Verb^{DB}+Verb+Cause+Pos+Imp+A2sg

yazGuzmaqçitim: yaz+Verb^{DB}+Verb+Cause+Pos+Past+Int+A1sg

yEzildi: yaz+Verb^{DB}+Verb+Pass+Pos+Past+A3sg

yaz+Verb^{DB}+Verb+Reflex+Pos+Past+A3sg

Örnek:

sen yUz miN qelem aldiN.

sen: sen+Pron+Pers+A2sg+Pnon+Nom

yUz: yUz+Noun+A3sg+Pnon+Nom

yUz+Num+Card

yUz+Verb+Pos+Imp+A2sg

miN: miN+Num+Card

qelem: qelem+Noun+A3sg+Pnon+Nom

qele+Noun+A3sg+P1sg+Nom

aldiN: al+Verb+Pos+Past+A2sg

EK C: Biçimbilimsel Etiketlerinin Açıklamaları

Etiket	Açıklama
+Noun	İsim
+Adj	Sıfat
+Adverb	Belirteç
+Det	Belirteç
+Dup	İkileme
+Interj	Ünlem
+Ques	Soru
+Verb	Eylem
+Postp	İlgeç
+Num	Sayı
+Pron	Zamir
+Punc	İşaretleme
+Card	Asıl Sayı
+Ord	Sırasal Sayı
+Percent	Yüzdesel Sayı
+Range	Aralıksal Sayı
+Real	Gerçel Sayı
+Ratio	Oransal Sayı
+Distrib	Üleştirme Sayı
+Time	Zaman Belirteci
+Inf	Master
+Prop	Özel İsim
+PastPart	Geçmiş Zaman Ortacı
+FutPart	Gelecek Zaman Ortacı
+PresPart	Geniş Zaman Ortacı
+Demos	İşaret Zamiri
+Ques	Soru Zamiri
+Reflex	Dönüşlü Zamir

+Pers	Kişi Zamiri
+Quan	Belgisiz Zamir
+A1sg	1. Tekil Şahıs Uyum Özelliği
+A2sg	2. Tekil Şahıs Uyum Özelliği
+A3sg	3. Tekil Şahıs Uyum Özelliği
+A1pl	1. Çoğul Şahıs Uyum Özelliği
+A2pl	2. Çoğul Şahıs Uyum Özelliği
+A3pl	3. Çoğul Şahıs Uyum Özelliği
+P1sg	1. Tekil Şahıs İyelik Eki
+P2sg	2. Tekil Şahıs İyelik Eki
+P3sg	3. Tekil Şahıs İyelik Eki
+P1pl	1. Çoğul Şahıs İyelik Eki
+P2pl	2. Çoğul Şahıs İyelik Eki
+P3pl	3. Çoğul Şahıs İyelik Eki
+Pnon	Belirsiz İyelik
+Nom	Yalın Durum
+Acc	Belirtme Durumu
+Dat	Yönelme Durumu
+Abl	Çıkma Durumu
+Loc	Kalma Durumu
+Gen	Tamlayan Durumu
+Ins	Aracılık Durumu
+Equ	Eşitlik Durumu
^DB+Verb+Pass	Edilgen Çatı
^DB+Verb+Caus	Ettirgen Çatı
^DB+Verb+Reflex	Yansımali Çatı
^DB+Verb+Recip	Dönüşlü Çatı
^DB+Verb+Able	Yetkinlik Çatı
^DB+Verb+Repeat	Sürerlik Yardımcı Eylemi (-dur eki)
^DB+Verb+Hastily	Tezlik Yardımcı Eylemi (-ver eki)
^DB+Verb+EverSince	Sürerlik Yardımcı Eylemi (-gel eki)
^DB+Verb+Almost	(-yaz eki)

^DB+Verb+Stay	(-kal eki)
^DB+Verb+Start	Sürerlik Yardımcı Eylemi (-koy eki)
+Pos	Olumlu
+Neg	Olumsuz
+Past	Belirli Geçmiş Zaman
+Narr	Belirsiz Geçmiş Zaman
+Fut	Gelecek Zaman
+Aor	Geniş Zaman
+Pres	Şimdiki Zaman
+Desr	İstek Kipi
+Cond	Şart Kipi
+Neces	Gereklilik Kipi
+Opt	Dilek Kipi
+Imp	Emir Kipi
+Prog1	Şimdiki Zaman (-yor)
+Prog2	Şimdiki Zaman (-makta)
+Cop	Kesinlik / kuşku anlamı (-dır eki)
+AfterDoingSo	Bağlama Ulacı (-ıp eki)
+SinceDoingSo	Zaman Ulacı (-alı eki)
+AsLongAs	Zaman Ulacı (-dıkça eki)
+When	Zaman Ulacı (-ınca eki)
+ByDoingSo	Durum Ulacı (-arak eki)
+While	Durum Ulacı (-ken eki)
+AsIf	Kıyaslama Ulacı (-arcasına eki)
+WithoutHavingDoneSo	Durum Ulacı (-maksızın eki)
+With	İsimden Sıfat Yapan Yapım Eki (-lı eki)
+Without	İsimden Sıfat Yapan Yapım Eki (-sız eki)
+FitFor	İsimden Sıfat Yapan Yapım Eki (-lık eki)
+Agt	Yapım Eki (-çı eki)
+Dim	Küçültme Sıfatı Eki (-cik eki)
+Ness	Sıfattan İsim Yapan Yapım Eki (-lık eki)
+Become	İsimden Eylem Yapan Yapım Eki (-laş eki)
+Acquire	İsimden Eylem Yapan Yapım Eki (-lan eki)

+Zero	Ek gelmeden türetim olduğunu gösterir
+PCAbI	İlgeç, çıkma durumunda bir isimden sonra gelir
+PCAcc	İlgeç, belirtme durumunda bir isimden sonra gelir
+PCDat	İlgeç, yönelme durumunda
+PCGen	İlgeç, tamlayan durumunda bir isimden sonra gelir
+PCIns	İlgeç, aracılık durumunda bir isimden sonra gelir
+PCNom	İlgeç, yalın durumda bir isimden sonra gelir
+QuesMu	“mu” eki ile oluşan soru etiketi
+RAq	Sıfatları artırma eki (-raq, rek)
+Endear	Sıfatları artırma eki, (-Gina, qina, gine, kine)
+DesrMu	İstek ekidir (-mu), ancak soru değil
+Inten	Amaç eylem etiketi
+ImPer	Tamamlanmayan eylem etiketi
+Perferct	Tamamlanan eylem etiketi
+DAk	Benzetme ekidir (bun-dağ, un-dağ)
+Cilik	İsimden isim yapma eki (kitap – cılık)
+Desr2	İstek eki(kel-gey, yaz-Gay)
+Fut2	Geniş zamanda gelen gelecek zaman etiketi
+A2sp	2. Tekil Şahıs, kibar Uyum Özelliği
+A2sr	2. Tekil Şahıs, saygı Uyum Özelliği
+A2pr	2. Çoğul Şahıs, saygı Uyum Özelliği
+A2pdr	2. Çoğul Şahıs, saygısız Uyum Özelliği

EK D: Uygurcadan Türkçeye Örnek Çeviri Kuralları

```
define A1SGNOM ["ben+Pron+Pers+A1sg+Pnon+Nom"];
define A2SGNOM ["sen+Pron+Pers+A2sg+Pnon+Nom"];
define A3SGNOM ["o+Pron+Pers+A3sg+Pnon+Nom"];
define A1PLNOM ["biz+Pron+Pers+A1pl+Pnon+Nom"];
define A2PLNOM ["siz+Pron+Pers+A2pl+Pnon+Nom"];
define A3PLNOM ["o+Pron+Pers+A3pl+Pnon+Nom"];
define BAGLAC ["ve+Conj"];
define WhiteSpace [[" " | "\n" | "\t"]];
define Karakter [? - WhiteSpace];

define sarteKi [% ] de %+ Conj <- %+ DesrMu || %+ Verb %+ [Pos|Neg] %+ Desr
    %+ [A3sg|A3pl] _ [% ] Karakter+;

define A1SGPnon ["ben+Pron+Pers+A1sg+Pnon"];
define A2SGPnon ["sen+Pron+Pers+A2sg+Pnon"];
define A3SGPnon ["o+Pron+Pers+A3sg+Pnon"];
define A1PLPnon ["biz+Pron+Pers+A1pl+Pnon"];
define A2PLPnon ["siz+Pron+Pers+A2pl+Pnon"];
define A3PLPnon ["o+Pron+Pers+A3pl+Pnon"];

define Sorular "+Dat" <- "+Abl" ∨ [.#. | %, | %; ][A1SGPnon|A2SGPnon|
    A3SGPnon| A1PLPnon| A2PLPnon|A3PLPnon] _ WhiteSpace
    "sor+Verb";

define Idi3sg ["+Narr+A3sg" <- %^ DB %+ Adj %+ PresPart || %+ Verb %+ Pos _
    WhiteSpace+ i %+ Verb %+ Pos %+ Past %+ A3sg ,, [% ] <- i %+
    Verb %+ Pos %+ Past %+ A3sg || %+ Verb %+ Pos %^ DB %+ Adj
    %+ PresPart [% ]+ _ ];
```

```

define Idi3pl ["+Narr+A3pl" <- %^ DB %+ Adj %+ PresPart || %+ Verb %+ Pos _
    WhiteSpace+ i %+ Verb %+ Pos %+ Past %+ A3pl ,, [% ] <- i %+
    Verb%+Pos %+ Past %+ A3pl || %+ Verb %+ Pos %^ DB %+ Adj
    %+PresPart[%]+ _ ];

define Idi1pl ["+Narr+A1pl" <- %^ DB %+ Adj %+ PresPart || %+ Verb %+ Pos _
    WhiteSpace+ i %+ Verb %+ Pos %+ Past %+ A1pl ,, [% ] <- i %+
    Verb%+Pos %+ Past %+ A1pl || %+ Verb %+ Pos %^ DB %+ Adj
    %+PresPart[% ]+ _ ];

define Idi1sg ["+Narr+A1sg" <- %^ DB %+ Adj %+ PresPart || %+ Verb %+ Pos _
    WhiteSpace+ i %+ Verb %+ Pos %+ Past %+ A1sg ,, [% ] <- i %+
    Verb%+Pos %+ Past %+ A1sg || %+ Verb %+ Pos %^ DB %+ Adj
    %+ PresPart [% ]+ _ ];

define Idi2sg ["+Narr+A2sg" <- %^ DB %+ Adj %+ PresPart || %+ Verb %+ Pos _
    WhiteSpace+ i %+ Verb %+ Pos %+ Past %+ A2sg ,, [% ] <- i %+
    Verb %+ Pos %+ Past %+ A2sg || %+ Verb %+ Pos %^ DB %+ Adj
    %+ PresPart [% ]+ _ ];

define Idi2pl ["+Narr+A2pl" <- %^ DB %+ Adj %+ PresPart || %+ Verb %+ Pos _
    WhiteSpace+ i %+ Verb %+ Pos %+ Past %+ A2pl ,, [% ] <- i %+
    Verb %+ Pos %+ Past %+ A2pl || %+ Verb %+ Pos %^ DB %+ Adj
    %+ PresPart [% ]+ _ ];

define IdiBirles Idi3sg .o. Idi3pl .o. Idi2sg .o. Idi2pl .o. Idi1sg .o. Idi1pl;

define gel1sg A1sg <- A3sg || A1SGNOM [WhiteSpace Karakter+]+ %+Fut %+ _ ;
define gel1pl A1pl <- A3sg || A1PLNOM [WhiteSpace Karakter+]+ %+Fut %+ _ ;
define gel2sg A2sg <- A3sg || A2SGNOM [WhiteSpace Karakter+]+ %+Fut %+ _ ;
define gel2pl A2pl <- A3sg || A2PLNOM [WhiteSpace Karakter+]+ %+Fut %+ _ ;

define gelecek gel1sg .o. gel1pl .o. gel2sg .o. gel2pl;

define sifat ["daha+Adverb" <- "iyi+Adj" || _ WhiteSpace "daha+Adverb" ,,
    "iyi+Adj" <- "daha+Adverb" || "iyi+Adj" WhiteSpace _ ];

define SoruParcila WhiteSpace "mi+Ques+Pers+A3sg" <- "+QuesMu" ;

```

```

define KL1sg [[]<- %^ DB %+ Verb %+ Recip || %+ Verb _ %+ Pos %+ Neces %+
    A1sg WhiteSpace [kirek |lazim] %+ Adverb ,, [] <- [kirek | lazim]
    %+ Adverb || %+ Verb %^ DB %+ ?+ %+ Neces %+ A1sg
    WhiteSpace _ ];

define KL1pl [[]<- %^ DB %+ Verb %+ Recip || %+ Verb _ %+ Pos %+ Neces %+
    A1pl WhiteSpace [kirek |lazim] %+ Adverb ,, [] <- [kirek | lazim] %+
    Adverb || %+ Verb %^ DB %+ %+ Neces %+ A1pl WhiteSpace _ ];

define KL2sg [[]<- %^ DB %+ Verb %+ Recip || %+ Verb _ %+ Pos %+ Neces %+
    A2sg WhiteSpace [kirek |lazim] %+ Adverb ,, [] <- [kirek | lazim]
    %+ Adverb || %+ Verb %^ DB %+ ?+ %+ Neces %+ A2sg
    WhiteSpace _ ];

define KL2pl [[]<- %^ DB %+ Verb %+ Recip || %+ Verb _ %+ Pos %+ Neces %+
    A2pl WhiteSpace [kirek |lazim] %+ Adverb ,, [] <- [kirek | lazim] %+
    Adverb || %+ Verb %^ DB %+ ?+ %+ Neces %+ A2pl WhiteSpace _
    ];

define KL3sg [[]<- %^ DB %+ Verb %+ Recip || %+ Verb _ %+ Pos %+ Neces %+
    A3sg WhiteSpace [kirek |lazim] %+ Adverb ,, [] <- [kirek | lazim]
    %+ Adverb || %+ Verb %^ DB %+ ?+ %+ Neces %+ A3sg
    WhiteSpace _ ];

define KL3pl [[]<- %^ DB %+ Verb %+ Recip || %+ Verb _ %+ Pos %+ Neces %+
    A3pl WhiteSpace [kirek |lazim] %+ Adverb ,, [] <- [kirek | lazim] %+
    Adverb || %+ Verb %^ DB %+ ?+ %+ Neces %+ A3pl WhiteSpace _
    ];

define kirekLazim KL1sg .o. KL1pl .o. KL2sg .o. KL2pl .o. KL3sg.o. KL3pl;

define P1sgtoPnon [PastPart %+ P1sg <- PresPart || DB %+ Adj %+ _ [% ]+ [ ?+ %+
    Noun %+ +? %+ P1sg ] ,, Pnon <-P1sg || DB %+ Adj %+
    PresPart [% ]+ ?+ %+ Noun %+ [Inf %+]* [A3sg | A3pl] %+ _ ];

define P2sgtoPnon ["PastPart+P2sg" <- PresPart || DB %+ Adj %+ _ [% ]+ [ ?+ %+
    Noun %+ ] +? %+ P2sg ,, Pnon <-P2sg || DB %+ Adj %+
    PresPart [% ]+ ?+ %+ Noun %+ [Inf %+]* [A3sg | A3pl] %+ _ ];

```

```

define P3sgtoPnon ["PastPart+P3sg" <- PresPart || DB %+ Adj %+ _ [% ]+ [?+ %+
Noun %+ ] +? %+ P3sg ,, Pnon <-P3sg || DB %+ Adj %+
PresPart [% ]+ ?+ %+ Noun %+ [Inf %+]* [A3sg | A3pl] %+ _ ];

define P1pltoPnon [PastPart %+ P1pl <- PresPart || DB %+ Adj %+ _ [% ]+ [?+ %+
Noun %+ ] +? %+ P1pl ,, Pnon <-P1pl || DB %+ Adj %+
PresPart [% ]+ ?+ %+ Noun %+ [Inf %+]* [A3sg | A3pl] %+ _ ];

define P2pltoPnon ["PastPart+P2pl" <- PresPart || DB %+ Adj %+ _ [% ]+ [?+ %+
Noun %+ ] +? %+ P2pl ,, Pnon <-P2pl || DB %+ Adj %+
PresPart [% ]+ ?+ %+ Noun %+ [Inf %+]* [A3sg | A3pl] %+ _ ];

define P3pltoPnon ["PastPart+P3pl" <- PresPart || DB %+ Adj %+ _ [% ]+ [?+ %+
Noun %+ ] +? %+ P3pl ,, Pnon <-P3pl || DB %+ Adj %+
PresPart [% ]+ ?+ %+ Noun %+ [Inf %+]* [A3sg | A3pl] %+ _ ];

define paralel P1sgtoPnon .o. P1pltoPnon .o. P2sgtoPnon .o. P2pltoPnon .o.
P3sgtoPnon .o. P3pltoPnon ;

regex paralel .o. IdiBirles .o. Sorular .o. gelecek .o. kirekLazim .o. sifat ;

save GlobalKurallarBicimBirimsel.fst

```


ÖZGEÇMİŞ

Ad Soyad:	Murat ORHUN
Doğum Yeri ve Tarihi:	Aktu, Doğu Türkistan, 22.07.1972.
Lisans Üniversitesi:	Sin Kiang Sanayi Üniversitesi, Elektrik Mühendisliği Bölümü, 1990-1995.
Yüksek Lisans:	Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, 1998-2001.
Doktora:	İstanbul Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, (2003 -----)

Yayın Listesi :

- **Orhun M., Tantuğ, A. C. ve Adalı. E.,** 2008. Machine Translation and Uyghur Language, *Akademik Bilişim 2008*, Çanakkale Onsekiz Mart Üniversitesi, 30 Ocak-1- Şubat, Çanakkale, Türkiye.
- **Orhun M., Tantuğ, A. C. ve Adalı. E.,** 2008. Rule Based Analysis of the Uyghur Nouns, *Proceedings of the International Conference on Asian Language Processing (IALP)*, November, 12-14, Chiang Mai, Thailand, pp. 8-24.
- **Orhun, M., ve Yavuz, A. G.,** 2009. Design and implementation of a web based system for the translation from Latin alphabet to Chinese characters, *Akademik Bilişim 2009*, Şubat 11-13, Harran Üniversitesi, Şanlıurfa, Türkiye.
- **Orhun M., Tantuğ, A. C. ve Adalı. E.,** 2009. Rule Based Tagging of the Uyghur Verbs. *Proceedings of the Fourth International Conference on Intelligent Computing and Information Systems*, Faculty of Computer & Information Science, Ain Shams University, March 19-22, Cairo, Egypt, pp. 811-816.
- **Orhun M., Tantuğ, A. C., Adalı. E. ve Sönmez A. C.,** 2009. Computational comparison of the Uyghur and Turkish Grammar. *Proceedings of the 2nd IEEE International Conference on Computer Science and Information Technology*, August 8-11, Beijing, China, Vol. 3, pp. 338-342.
- **Orhun M., Tantuğ, A. C. ve Adalı. E.,** 2009. Machine Translations Between Turkic Languages. *Naitonal Conference on Man-Machine Speech Communication (NCMMSC), International Symposium on Speech and Language Processing*, Kunlun Hotel, August 14-16, Urumqi, Xinjiang Uyghur Autonomous Region, China.
- **Orhun M., Tantuğ, A. C. ve Adalı. E.,** 2009. Rule Based Analysis of the Uyghur Nouns. *International Journal of Assian Language Processing*, **19** (1), pp. 33-43.

- **Orhun M., Tantuğ, A. C. ve Adalı. E.,** 2010. Uygurcada Biçimbilimsel Belirsizlik, *Akademik Bililşim 2010*, Şubat 10-12, Muğla Üniversitesi, Muğla, Türkiye.
- **Orhun M., Tantuğ, A. C. ve Adalı. E.,** 2010. Morphological Disambiguation Rules for Uyghur Language, *IEEE International Conference on Software Engineering and Service Science (ICSESS 2010)*, July 16-18, Bei Jing, China (Kabul edildi).