

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

TÜRKÇE - TÜRKMENÇE BİLGİSAYARLI ÇEVİRİ SİSTEMİ

YÜKSEK LİSANS TEZİ

Guychmyrat AMANMYRADOV

**HAZİRAN 2006
TRABZON**

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİMDALI

TÜRKÇE - TÜRKMENÇE BİLGİSAYARLI ÇEVİRİ SİSTEMİ

Guychmyrat AMANMYRADOV

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
“Yüksek Lisans (Bilgisayar Mühendisi)”
Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

**Tezin Enstitüye Verildiği Tarih : 09.06.2006
Tezin Savunma Tarihi : 14.07.2006**

**Tez Danışmanı : Yrd. Doç. Dr. Cemal KÖSE
Jüri Üyesi : Prof. Dr. Rıfat YAZICI
Jüri Üyesi : Yrd. Doç. Dr. Ali GANGAL**

Enstitü Müdürü : Prof. Dr. Emin Zeki BAŞKENT

Trabzon 2006

ÖNSÖZ

Bu tez, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği Yüksek Lisans Programı'nda hazırlanmıştır.

1991 yılında SSCB'nin dağılması ile Türk halkları arasındaki diyalog güçlenmiştir. Böylece, asırlarca siyasi ve coğrafi sınırlarla ayrılmış Türk toplulukları arasında dil farklılığı yüze çıkmıştır. Ne yazık ki Türk toplulukları liderleri bir araya geldiğinde Rusça konuşur hale gelmişlerdir! Bu sıkıntıları azaltmak, diller arasındaki uzaklığı yakınlaştırmak amacı ile bir başlangıç olarak Türkçe Türkmençe sohbet (chat) programı geliştirilmiştir. İnternet'te sohbet odasına bağlanarak Türkmençe bilmeyen bir Türk, Türkçe bilmeyen bir Türkmen'le sohbet ederek anlaşabilmektedir.

Yüksek lisans tez danışmanlığımı üstlenerek, çalışmalarım boyunca yardımlarını esirgemeyen değerli hocam Yrd. Doç. Dr. Cemal KÖSE'ye, çalışmalarım program kodlama sürecinde ve bilgi vermede her zaman desteğini yanımda gördüğüm Ahmet Afşın AKIN'na, Türkmençe ve Türkçe kaynak bulmama yardımcı olan Doç. Dr. Berdi SARI'ya, çalışmalarımda yön gösteren Prof. Dr. Vasif V. NABIYEV'e teşekkür etmeyi bir borç bilirim.

Guychmyrat AMANMYRADOV

Trabzon, 2006

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	II
İÇİNDEKİLER.....	III
ÖZET.....	V
SUMMARY.....	VI
ŞEKİLLER DİZİNİ.....	VII
TABLolar DİZİNİ.....	VIII
SEMBOLLER DİZİNİ.....	IX
1. GENEL BİLGİLER.....	1
1.1. Giriş.....	1
1.2. Sohbet (Chat) Programları.....	2
1.3. Bilgisayarlı Çeviri (BÇ).....	3
1.4. Yakın Diller Arası Bilgisayarlı Çeviri Sistemi.....	6
1.5. Türk Dillerinin Genel Özellikleri.....	8
1.6. Türk ve Türkmen Dilleri Gramerleri.....	10
1.6.1. Fonem ve Alfabe.....	10
1.6.2. Kelime.....	11
1.6.3. Ekler.....	13
1.7. Türkçe ve Türkmençe Gramerlerinin Karşılaştırılması.....	16
1.7.1. Alfabe.....	16
1.7.2. İsim Ekleri.....	18
1.7.3. Fiil Ekleri.....	19
1.7.3.1. Basit Kipler.....	19
1.7.3.2. Birleşik Kipler.....	22
1.7.3.3. Edilgen ve Ettirgen Ekleri.....	22
1.7.3.4. Soru Cümleleri.....	23
1.7.3.5. Olumsuz Cümleler.....	24
1.8. İmla Hatalarının Düzeltilmesi.....	25
2. YAPILAN ÇALIŞMALAR.....	26

2.1.	Giriş.....	26
2.2.	Sohbet Motorunun Tasarımı.....	26
2.3.	Türkçe ve Türkmençe Morfolojik Analiz Modülü.....	32
2.4.	Türkçe ve Türkmençe ve Üretim Çeviri Modülü.....	38
2.5.	İmla Hatası Düzeltme Modülü.....	40
3.	BULGULAR.....	41
3.1.	Giriş.....	41
3.2.	Ekler.....	41
3.3.	Kökler.....	45
3.4.	Çeviri İstatistikleri.....	47
4.	SONUÇLAR VE ÖNERİLER.....	52
5.	KAYNAKLAR.....	53
	ÖZGEÇMİŞ.....	56

ÖZET

Bu yüksek lisans tez çalışmasında, yapıları birbirine yakın olan Türkiye Türkçe'si ile Türkmenistan Türkmençe'si arasında, bilgisayarlı çeviri sistemini geliştirilmiştir. Geliştirilen bu sistem Türkçe ve Türkmençe konuşan insanların birbirleri ile Internet aracılığıyla iletişim kurabilmeleri olanağını da sunmaktadır. Türkçe ve Türkmençe ortak kökenli, yani aynı dil ailesine aittir. Türkiye Türkçe'si ile Türkmenistan Türkmençe'si ortak kökenli diller olmalarına rağmen, morfolojik olarak bazı farklılıkları vardır. Türk dilleri dilbilgisel düzenlerine göre bitişimli (eklemlili) dil grubuna girdiğinden, bir ad yada eylem kökünden, değişik yapım ekleriyle çeşitli kavramları karşılayan yeni sözcükler türetilir. Ayrıca, çekim ekleri bir ad veya eylem köküne eklenerek de kelimenin farklı durumlarını belirten yeni sözcüklerde türetilir. Yakın diller arasında çeviri sistemi geliştirmek, birbirine yakın olmayan dillere nazaran kolaydır. Yapısal olarak birbirine yakın diller için geliştirilen çeviri sistemleri, çoğu zaman anlamsal ve sözdizimi analizi yapmadan, bazı çeviri kuralları ve morfolojik analiz yöntemleri kullanarak yeterli düzeyde çeviri yapabilmektedir. Böylece, yakın diller arasında çeviri sistemlerini geliştirmek için daha az çaba göstermek yeterli olacaktır.

Geliştirilen sistem ilk olarak Türkçe veya Türkmençe cümleyi alır ve bütün kelimeleri morfolojik olarak ayrıştırır. Böylece kelimeler morfolojik olarak eklerle ve köklere ayrılmış olur. Köklere ve eklerle göre, hedef dildeki sözlükten uygun kök ve ekler seçilerek kelime kelime çeviri gerçekleştirilir. Son olarak, hedef dile çevrilen kökleri ve ekleri birleştirerek süreç tamamlanır.

Anahtar Kelimeler : Doğal Dil İşleme, Bilgisayarlı Çeviri, Eklemlili Diller, Türkçe, Türkmençe, Sohbet Sistemi.

SUMMARY

A Turkish -Turkmen Machine Translation System

This paper describes the details of a machine translation system for Turkish-Turkmen chat. The machine translation system is developed to provide a web based chat medium for Turkish and Turkmen speaking people. Turkish and Turkmen languages are closely related but they have rather complex morphotactics. Machine translation between closely related languages is easier than translation between languages that are not closely related. Thus, less effort needed to develop a translation system between closely related languages because many parts of related languages such as grammars and vocabularies are common. Hence, syntactic and semantic analysis may not be necessary since the translation system makes morphological analysis most of the time. So, simple translation rules and context dependent bilingual dictionaries may be sufficient for the translation. Turkish and Turkmen are agglutinative languages with respect to word structures formed by productive affixations of derivational and inflectional suffixes to root words. Therefore, different words can be constructed from roots and morphemes to describe objects and concepts. Basically, the translation system takes a Turkish or Turkmen sentence and analyses all the words morphologically. Then, it translates the root words and morphemes, and finally generates the Turkish or Turkmen sentence morphologically. Most of the time the input sentence is translated to at least one or more true output sentence. Finally, a context dependent decision mechanism is employed to determine the true output sentences.

Key Words: Natural Language Processing, Machine Translation, Agglutinative Languages, Turkish, Turkmen, Chat System.

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Vauquois üçgeni.....	7
Şekil 2. Ural Altay Dil Ailesi.....	9
Şekil 3. Eklerin diziliş sırası.....	14
Şekil 4. Türkçe ve Türkmençe'nin karşılıklı harfleri.....	17
Şekil 5. Garagol rumuzlu Türkmençe bilen istemcinin penceresi.....	30
Şekil 6. Leylek rumuzlu Türkçe bilen istemcinin penceresi.....	31
Şekil 7. Sonsuz döngüde istemci bekleyen server.....	31
Şekil 8. Ağaçtan köke ulaşma.....	33
Şekil 9. Köklerin değişmiş halleri ile Ağaç sözlüğüne eklenmesi.....	34
Şekil 10. Eşsesli köklerin Ağaç sözlüğüne eklenmesi.....	34
Şekil 11. Türkçe Çözümleyici.....	37
Şekil 12. Türkmençe Çözümleyici.....	37
Şekil 13. Türkmençe girilmiş bir cümlenin Türkçe çevirisi.....	39
Şekil 14. Türkmençe girilmiş bir cümlenin Türkçe çevirisi.....	39
Şekil 15. Türkçe'den ve Türkmençe'ye Çeviri İstatistiği.....	50
Şekil 16. Türkmençe'den ve Türkçe'ye Çeviri İstatistiği.....	51

TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Ünlü Sesler.....	11
Tablo 2. Türkçe ve Türkmençe kelimeyi oluşturan unsurlar.....	12
Tablo 3. Türkçe ve Türkmençe kelimelerin söz içindeki işleyişleri.....	12
Tablo 4. Türkçe ve Türkmençe Alfabe.....	16
Tablo 5. Türkçe ve Türkmençe isimin yaklaşma hali ekleri.....	18
Tablo 6. Türkçe ve Türkmençe isimin ilgi hali ekleri.....	19
Tablo 7. Türkçe ve Türkmençe geniş zaman ekleri.....	20
Tablo 8. Türkçe ve Türkmençe şimdiki zaman ekleri.....	20
Tablo 9. Türkçe ve Türkmençe bilinen geçmiş zaman ekleri.....	21
Tablo 10. Türkçe ve Türkmençe öğrenilen geçmiş zaman ekleri.....	21
Tablo 11. Türkçe ve Türkmençe gelecek zaman ekleri.....	22
Tablo 12. Türkçe ve Türkmençe birleşik kipler ekleri.....	22
Tablo 13. Türkçe ve Türkmençe edilgen ekleri.....	23
Tablo 14. Türkçe ve Türkmençe zamanların soru şekilleri.....	23
Tablo 15. Türkçe ve Türkmençe gelecek zaman soru şekli.....	24
Tablo 16. Türkçe ve Türkmençe bilinen geçmiş zaman olumsuzluk eki.....	24
Tablo 17. Sözlükte tutulan kelime tipleri.....	32
Tablo 18. Levenshtein Algoritması.....	40
Tablo 19. Türkçe ve Türkmençe eklerin benzerlik istatistikleri.....	42
Tablo 20. Türkçe ve Türkmençe isim çekim ekleri.....	42
Tablo 21. Türkçe ve Türkmençe fiil Çekim Ekleri.....	43
Tablo 22. Türkçe ve Türkmençe aynı anlamda yazılan ve kullanılan kelimeler.....	45
Tablo 23. Türkçe ve Türkmençe yakın anlamda yazılan ve kullanılan kelimeler.....	46
Tablo 24. Türkçe ve Türkmençe birbirinden farklı yazılan kelimeler.....	46
Tablo 25. Türkçe ve Türkmençe zıt anlamda yazılan ve kullanılan kelimeler.....	46
Tablo 26. Türkçe ve Türkmençe köklerinin benzerlik istatistiği.....	47

SEMBOLLER DİZİNİ

()	:	son harfin duruma göre gelebilirde gelmeye de bilir
A	:	Türkçe için a,e. Türkmençe için a,e, ä
BÇ	:	Bilgisayarlı Çeviri
Chat	:	Sohbet
d	:	Türkçe için d,t. Türkmençe için d
DDİ	:	Doğal Dil İşleme
H	:	Türkçe için ı,i,u,ü. Türkmençe için y,i,u,ü
IRC	:	Internet Relay Chat
LU	:	Levenshtein Uzaklığı
OSR	:	Optical Character Recognition

1. GENEL BİLGİLER

1.1. Giriş

Dilin birçok tanımı yapılmıştır ve bu tanımlarından biride N. Chomsky'nin tanımıdır: "Dil, sözcük ve cümle birimleri aracılığıyla düşünceyi konuşma ile ilişkilendiren çok seviyeli bir sistemdir". İnsan anlatım ve bildirişim için hareket eder veya da ses çıkarır yahut belirli işaretler çizer. Konuşma dili, yazı dili, hareket dili, (insan dili) nin üç ayrı görüntüsüdür. Dil, toplumsal bir olgudur ve başkaları için var olan pratik bilinçtir. Dil bilinçle birlikte var olmuş ve onunla karşılıklı etkileşerek gelişmiştir. Dilsiz hiç bir düşünce var olamaz, insan kendi kendine düşündüğü zaman bile ancak sözcüklerle eş değer dille düşünebilir.

Doğal Dil İşleme (DDİ) üzerinde elli yıldan fazla süredir araştırma yapılmaktadır. İlk başlarda DDİ (çeviri) üzerine kamu alanlarında çalışma başlatılmıştır. Ama çok fazla başarılı olamamıştır. Şirketler de her gün sayısı artan dokümanları, insan gücü ile çeviri yapmaya çalışmakta çok yetersiz kalmaktadır. Artık ticari şirketlerde bilgisayar ile çevirinin önemini anlamışlardır. Bu yüzden, son yıllarda şirketler bilgisayarlı çeviri üzerine çok sayıda yoğun araştırma ve geliştirme çalışmaları başlatmışlardır. Çeviri yapan kişiler dokümanları anlarlar ve anlama göre hedef dilde o anlamda cümle üretirler. Ama insan sınırlı sayıda çeviri yapabilir ve istenilen kalitede çevirmen her zaman bulunamayabilir. Son yıllarda DDİ üzerinde ticari firmalarında çalışmaya başlamasından, DDİ'de araştırmalar yoğunluk kazanmıştır.

Ticari şirketlerin yoğun çalışmalarına rağmen şu anda, tek tıklama ile bir dilden diğer dile tam anlamlı çeviri yapan bir sistem yoktur. Yakın tarihte de tek tıklama ile çeviri yapan sistemin mümkün olmayacağı görülmektedir. Bilgisayarlı çevirisinin tam anmalıyla başarılı olmamasının sebebi olarak, her dilin kendine has karmaşıklığını ve belirsizliğini gösterebiliriz.

Daha çok teknolojiye sahip olan gelişmiş ülkelerde bilgisayarlı çeviri üzerinde araştırmalar yapılmaktadır. En çokta İngilizce, Fransızca ve Almanca gibi batı dilleri üzerine çalışmalar yapılmıştır. Diğer diller üzerine çalışmalar, daha çok o dil ile İngilizce arasında olmuştur [1], [2] , [3], [4]. Öte yandan, yapıları birbirine yakın diller arasında fazla

çalışma yapılmamıştır. Özellikle Türk dilleri üzerine yok denilebilecek kadar az çalışma yapılmamıştır. Türk dilleri arasında şu ana kadar yapılan bilgisayarlı çeviri çalışması Bilkent Üniversitesi'nde İlyas Çiçekli tarafında Türkçe – Kırım Tatarca'sı arasında tek taraflı Türkçe'den Kırım Tatarcasına çeviri sistemidir. Diğer bir çalışmada İstanbul Üniversitesi doktora öğrencisi Ahmet Cüneyd TANTUĞ tarafından “Türkmence Biçimbirimsel Çözümleyici” üzerinde yapılmaktadır.

Bu çalışmada, yakın diller olarak Türk dilleri baz alınıp bilgisayarlı çeviri sistemi tanıtılmaktadır. Türkiye Türkçe'si ile Türkmenistan Türkmence'si arasında geliştirilen bir çeviri sistemin Türkçe veya Türkmence ile herhangi diğer Türk dilleri arasında kolayca uygulanabileceğini düşünülmektedir.

Türk ve Türkmen dilleri Oğuz dil ailesindedir. Bu iki dil aynı dil ailesinden olduğundan gramerleri, kelimeleri ve morfemleri benzerlik gösterirler. Genellikle bir kök bazı değişiklik ile veya hiç değişikliğe uğramadan her iki dilde de kullanılmaktadır. İki dilin arasında en çok morfem düzeyinde değişiklik görülür. Türkçe ve Türkmence için geliştirilmiş bu modelin diğer Türk dillerine de kolayca uygulanabileceği öngörülmektedir. Daha önce Çek-Slovak ve İspanyol–Katalan dilleri yakın diller arasında geliştirilmiş bilgisayarlı çeviri yapan sistemlerin başarılı bir şekilde kendi dil ailesindeki yakın dillere uygulanabileceği görülmüştür [5], [6], [7], [8].

Bu çalışmada, Türkçe ve Türkmence için bilgisayarlı çeviri yapan sohbet sistemini tanıtılacak. Bu çeviri sistemi, Türkçe girilmiş olan bir cümleyi Türkmence'ye, Türkmence cümleyi ise Türkçe'ye çevirir. Genel bilgiler bölümünde, bilgisayarlı çeviri, sohbet sistemi hakkında gelen bilgi verilecek. Türk ve Türkmen dillerinin gramer yapıları incelenecek. Yapılan çalışmalar bölümünde, Türkçe ve Türkmence için bilgisayarlı çeviri için neler yapıldığını ve çeviri sisteminin nasıl çalıştığı tanıtılacak. Bulgular bölümünde, Türkçe ve Türkmence çeviri sisteminin sonuçlarını üzerinde durulacak. Öneriler bölümünde ise, daha iyi bir sistem için neler yapılacağı üzerinde durulacak.

1.2 Sohbet (Chat) Programları

Sohbet programları, bilgisayar arkalı uzaktaki kişilerin birbirleriyle haberleşmelerini sağlarlar. İnternet'in ilk kullanımından bu yana sohbet programları vardır. İlk başlarda, sadece IRC(Internet Relay Chat) kanallarından haberleşseler de, günümüzde istemci/kullanıcı daha gelişmiş sohbet programları vardır. Gün geçtikçe sohbet

programları İnternet’te artmaktadır. Günde milyonlarca kiři sohbet programları arkalı haberleşmektedirler. En popüler olanları ise ICQ, Yahoo Messenger, Google Talk ve MSN Messenger. İnternet’e bağlanabilen cep telefonlarının artmasıyla, SMS’den daha maliyeti düşük olduğundan kullanışlılığı artırmaktadır.

MSN Messenger, ICQ gibi programlar, çevrimiçi olan iki kiřinin birbirleriyle çabuk haberleşebilmelerini sağlamaktadır. Programların fonksiyonları geliştirilerek, birden fazla kiři ile aynı anda haberleşebilmek için özel kanallar açabilmene izin veriyorlar. Programların kurulumları çok kolaydır ve sağladıkları gelişmiş özellikleri ile kullanıcıya birçok kolaylıklar sağlıyor. Her programı kullanabilmen için, o programın servisine üye olmak zorundasın.

Sohbet programları ilk başlarda bilimsel çalışmalar için üniversiteler arası haberleşe için kullanılsa da, son yıllarda bazı şirketler ticari amaçlı sohbet programları geliştirmektedir. Son 20 yıldır çok büyük bir evrim geçirmese de, rekabetin artmasıyla şirketler programlarına sürekli yeni özellikler kazandırıyorlar.

Bu popülerleşmiş programlara görüntü ve ses özellikleri eklenmiştir. Genel olarak görüntü yetersiz kalitede olsa da ses kalitesini, yeterince kaliteli yapan programlar vardır. Ses kalitesi en yüksek olan programlar Skype ve Google Talk programlarıdır.

1.3. Bilgisayarlı Çeviri

Bilgisayarlı Çeviri (BÇ) terimi, bilgisayarların insan yardımıyla veya yardımsız olarak girilen metni çeviri yapmasıdır. Genel olarak, insan yardımı ile bilgisayarlı çeviri ve insan yardımı olmadan bilgisayarlı çeviri olmak üzere ayrıştırılır. Her ne kadar, insan yardımı olmadan çok başarılı bilgisayarlı çeviri beklentisi olsa da, pratikte buna daha ulaşılammıştır. Bilgisayara verilen bir metnin daha önceden verilen kurallara göre çevirisi yapılsa da, dilin kendisine has olan belirsizliklerden dolayı, istenilen mükemmel çeviri gerçekleşmemektedir. Çevirisi yapılmış bir metni yeniden bir kontrol edici mekanizma ile tekrar çevirisi yapılarak başarı oranı artırılabilir ama günümüzde mükemmel çeviri mümkün değildir. Bu yüzden, çok kaliteli çeviriye ihtiyaç duyulduğunda bir tercümanın çeviriyi tekrardan gözden geçirmesi gerekmektedir. Eğer çeviriyi, metni kabaca anlamak için kullanılıyorsa, genel olarak günümüzdeki çeviri sistemleri yeterli olabilmektedirler.

Günümüzde okuma, yazma, anlama ve konuşma gibi insan yeteneğine yakın modeller geliştirilmesinden uzak olsa da, DDİ üzerinde belli bir ilerlemeye gelinmiştir. Ticari amaçlı birçok programlar geliştirilmiştir. Doğal dil üzerine daha çok Doğal Dil İşleme, Dilbilim Anlama, Konuşmayı Tanıma ve Sentez etme alanlarında çalışma yapılmaktadır.

Doğal Dil İşlemede görülen en büyük problemlerden biride belirsizliklerdir. Eğer bir giriş için birden fazla dilbilimsel çıkış varsa, belirsizlik vardır deriz. Mesela Türkçe “O okula gitti” cümlesini Rusça’ya çevirmek istediğimizde

-Он пошел в школу. (Bay okula yürüyerek gitti)

-Она пошла в школу. (Bayan okula yürüyerek gitti)

-Он ездил в школу. (Bay okula araçla gitti)

-Она ездила в школу. (Bayan okula araçla gitti)

Belirsizlikten dolayı dört farklı cümle ortaya çıkmıştır. Türkçe’deki o zamirine karşılık Rusça’da он (bay o) ve она (bayan o) zamirleri karşılık gelmektedir. Türkçe’de gitti dediğimizde yayamı yoksa bir araçla gittiğini belirtmeyiz. Ama Rusça’da bunlar belirtilir. Bu belirsizlikler yüzünden yukarıda dört farklı cümle ortaya çıkmıştır.

Doğal dil üzerindeki ilk çalışmalar arasında Alan Turing’in çalışması geliyor. Alan Turing çalışmasında, Turing testinin nasıl olması gerektiğini tanıttı. Tezinde bir test oyunu önerdi, oyunda bilgisayarın dili kullanımı bilgisayarın düşünmediğine karar verebilecek. Oyunu eğer bilgisayar kazanırsa, bilgisayarın akıllı olduğu düşünülecek.

Turing’in oyununda 3 katılımcı olacak: İki insan ve bir bilgisayar. Bir tarafta sürekli sorgulayıcı insan olacak, diğer tarafta ise insan ve bilgisayar olacak. Sorgulayıcı kişi, cevaplardan bilgisayarın cevabı ile insanın cevabını ayırt etmeye çalışacak. Bilgisayarın görevi, sorgulayıcı kişiye kendinin bir insan olduğuna inandırmaya çalışmak. Diğer katılımcı insanın görevi ise, sorgulayıcı kişiye kendisinin insan olduğunu, diğer katılımcının bilgisayar olduğunu inandırmaya çalışmak. Eğer sorgulayıcı kişi insanın cevapları ile bilgisayarın cevaplarını ayırt edemezse, bilgisayar kazanmış, aksi halde kaybetmiş sayılıyor. Aşağıda yapılan Turing testi ile ilgili küçük bir örnek verilmiştir [9].

S: Dışarıda hava nasıl?

C: Bu gün dışarıya pek çıkmadım. Ben kendime çok düşkünüm.

S: Kaç yaşındasın?

C: Yaklaşık olarak 30.

S: Yaklaşık olarak?

C: Ben Star Trek için konuşabilirim.

20. y.y. sonunda, 10 gigabyte belleğe sahip bir bilgisayarın 5 dakika sonra sorgulayıcı kişiyi kandırma şansının %30 olduğu gözetlenmiştir. Çoğu zaman bilgisayar birilerini kandırabiliyor, ama kesin bir standardı yoktur.

Turning testinden sonra birçok çalışma yüze çıkmıştır. ELIZA programı ise kullanıcı ile belli bir konuşma yeteneğine sahip olan programdı. ELIZA programı insanlara psikolojik destek veren bir program olarak yüze çıkmıştı. Turning testi ile ELIZA programının arasındaki fark, ELIZA programının kullanıcıyı ve problemlerini gerçekten anladığına inanılmaktadır. ELIZA programı ilk 1966 yılında Weiznebaum tarafından oluşturulsa da daha sonraları geliştirilmiştir.

Günümüzde konuşma ve dil üzerinde çalışmalar popülerleşmiştir. Birçok kullanıcı yardımcısı uygulamalar, web üzerinden uygulamalar geliştiriliyor. Aşağıda birkaç uygulama günümüzde yaygın kullanılanlardan birkaçıdır.

- Kanada'da geliştirilmiş bir bilgisayar programı günlük hava durumları bilgilerini alır ve İngilizce ve Fransızca raporlar sunar.
- Altavista sitesinde *Babel Fish* çeviri programı üzerinden günde 1.000.000 üzerinde çeviri gerçekleştirilmektedir.
- Cambridge ve Massachusetts üniversitelerinde geliştirilen bir programda, kullanıcı günlük konuşma dili ile restoranları sorar ve bilgisayar veritabanından en uygun restoranlar hakkında bilgi verir.
- Bir program, yüzlerce öğrencinin makalelerini okur ve onlara not verir.
- Rus çocukları için geliştirilen bir programda, çocuklara masal, hikâye okuyarak, çocukların telaffuzunu geliştirmesine yardımcı olur.

Kendi kendimize neden bilgisayarlı çeviri sistemi önemli sorusunu sorabiliriz. İlk ve belki de en önemli sebebi, gitgide artan çevrilmesi gereken dokümanların insan gücü ile yetiştirilememesi. İkinci olarak ta, teknik çevirilerin insanlar için sıkıcı olması, bu yüzden teknik çevirilerde bilgisayarlardan medet umuyorlar. Üçüncü olarak, bazı büyük şirketlerde belli bir terimler dizisi aynı yöntemle sürekli olarak çeviri yapılması gerekebiliyor. Böyle

bir durumda insanlar aynı yöntemden gitmek istemezler, farkı yöntemler araştırmasına gidiyorlar, bu yüzden bilgisayarlı çeviri daha yararlıdır. Dördüncü olarak, bilgisayarlı çeviri ile yapılan çeviri çok fazla dokümanı çok kısa zamanda çeviri edebilmektedir. Beşinci olarak, bazı şirketlerde bilgisayarlı çeviri yeterli olmaktadır. Mükemmel bir tercümana ihtiyaç olmayabiliyor. Altıncı olarak, bir dilden başka bir dili öğrenen birisine basit kalpların çevirisinde yardımcı olarak kullanılabilir. Sonuç olarak, mükemmel çeviri gerektirmeyen yerlerde bilgisayarlı çeviri geniş olarak kullanılmaktadır. Son zamanlarda, şirketler bilgisayarlı çevirisine çok önem vermekteler ve çeviri maliyetini azaltmaya uğraşmaktalar. Aslında, yukarıdaki sebeplerden herhangi biride, bilgisayarlı çeviri sistemi için yeterli derecede önemlidir.

1.4. Yakın Diller Arası Bilgisayarlı Çeviri Sistemi

Bir dilden diğer bir dile çeviri yapmak zor bir iştir. Farklı kültürle kaynaşmanın etkisi, dile ve gramerlere de yansımaktadır [10], [11], [1], [12]. Kaynak dilinde kullanılan bir kelime veya kavram hedef dilde olmayabilmektedir. Bu durumlarda çeviri problemleri yüze çıkıyor. Örneğin, kaynak dildeki bir kelimenin hedef dilde birden fazla anlamı varsa, doğru kelimeyi seçmek kolay bir iş değildir. Doğru seçilemezse, cümlede belirsizlik veya yanlış cümle olacaktır.

Dünya dilleri coğrafi veya kültür dağılımına göre sınıflandırılabilir. Tarihsel olarak bakıldığında aynı kökten gelen halkların dilleri de birbirine yakın olur. Kaynak bakımından birbirine yakın olan diller bir aile teşkil ederler. Dünya dilleri bu şekilde çeşitli dil ailelerine ayrılırlar. Bir dil ailesi tarihin bilinmeyen devirlerinde bir ana dilden çıkan dillerin oluşturduğu topluluktur. Bu diller arasındaki benzerlikler böyle bir varsayım kuvvetlendirmektedir. Bir ana dilin yazılı belgeleri olmadığı halde birçok özelliklerini kendisinden türemiş bulunan ailedeki dilleri karşılaştırarak tespit etmek mümkün olabilmektedir.

Yakın diller arası çeviri kuralları ile morfolojik kurallar uygulanarak, anlamsal ve sözdizimsel analizi hiç yapılmasa bile kabul edilebilir bir sistem elde edilmiş olur [2]. Bu yüzden yakın diller arası bilgisayarlı çeviri sisteminde yakın olmaya diller arasındaki çeviriye göre daha az kural yeterli olacaktır. Kaynak dilde biçimbirimsel analiz yapılacak. Biçimbirimsel analiz olmuş kök ve ekler hedef dildeki karşılıkları alınacak. Sonra, hedef

dile biçimbirimsel üretici kullanılarak aktarılacak. Burada sözdizimsel analiz yapılmayacağı için kelimenin cümledeki durumu bilgisine ihtiyaç olmayacaktır.

Günümüze kadar bilgisayarlı çeviri üzerine yapılan çalışmalar daha çok birbirine yakın olmayan diller arasında gerçekleşmiştir. Yakın diller arasında örneğin, Türk dilleri arasında bilgisayarlı çeviri üzerine yok denilebilecek kadar az çalışma yapılmıştır. Yakın diller yapı bakımından birbirine yakındırlar. Bu bakımdan, yakın olmayan dillere göre birçok özellik iptal edilse bile anlaşılabilir bir çeviri elde edilebilir. Örneğin: Türk dillerinde cümlede kelimenin sırası belli bir kalıp ile sınırlandırılmamıştır. Bu yüzden Türkçe Türkmence arasında bilgisayarlı çeviri işleminde kelimelerin sırasını göz ardı edilebilir. Ama bu, Türkçe ve İngilizce için göz ardı edilemez. Böylece, yakın diller arası bilgisayarlı çeviride, kelime sırası göz ardı edilerek başka özelliklerine daha çok dikkat edilebilir.

Her ne kadar yakın diller gramer ve sözdizimsel olarak birbirlerine yakın olsalar da kendine has zorlukları vardır. Bazı bileşik kelimeleri veya o dile has kelimeleri karşı dile çevirmek istenirse karmaşıklık doğabiliyor. Bu karmaşıklığı giderebilmek için belli kalıpları sözlüğe eklenebilir.

Aşağıda Kalam Oflazer'in yakın diller için önerdiği Vauquois üçgeni verilmiştir.



Şekil 1. Vauquois üçgeni

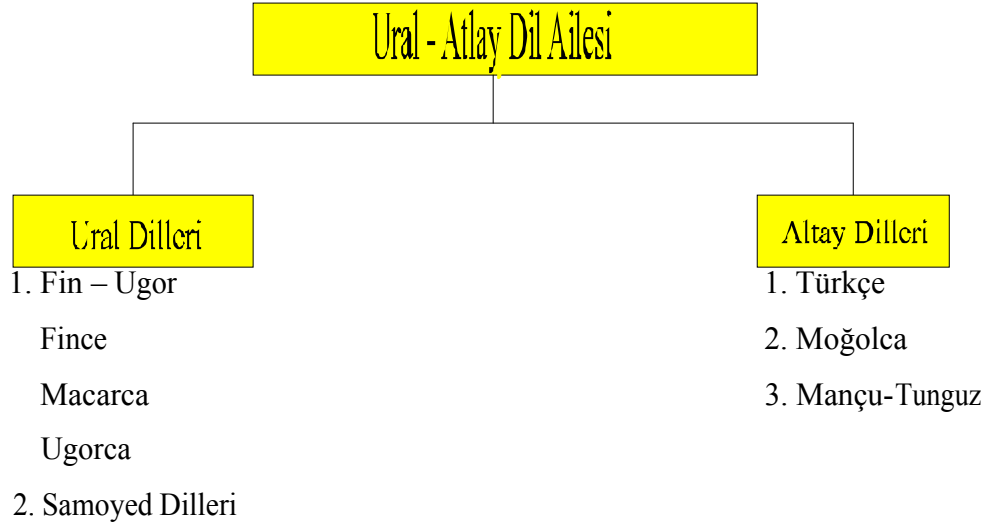
Yakın diller arası çeviri için ilk girişim 1985 yılında Prag'daki Charles Üniversitesinde başladı. Projenin adı RUSLAN-dı ve Çekçe – Rusça çeviri üzerine geliştirilmişti. 1985 yılında başlamasına rağmen çalışır bir sistemi 1990 yılında elde edebildiler. Çekçe girilmiş bir metni biçimbirimsel ve sözdizimsel olarak analiz eder ve Rusça metni biçimbirimsel transfer ederek, Rusça cümle üretir. Kabaca %40-ı doğru çeviri yapabiliyor, %40-ı editör tarafından gözden geçirilerek çeviri yapabiliyor, %20-si ise yeniden çeviri gerektiriyordu.

1.5. Türk Dillerinin Genel Özellikleri

Yeryüzünde ortalama 4000 dil konuşulmaktadır. Bu diller arasında en çok ve en yaygın konuşulan ilk yedi dil Çince, Hintçe, İspanyolca, İngilizce, Fransızca, Rusça ve Türkçe'dir. Bu sıralama, Türkçe'nin dünya dilleri arasında ne kadar büyük bir yeri olduğunu göstermesi bakımından son derece önemlidir.

Ural – Altay dilleri dünyada en çok kullanılan dil ailesinden biridir. Türk dilleri Ural – Altay dil grubundandır. Türk dilleri birbirine fonetik, morfoloji, sentaks ve gramer bakımından birbirlerine benzerdirler. Ural – Altay dillerinin aynı kaynaktan gelip gelmedikleri tartışma konusudur. Ural – Altay gruptaki dillerin yakınlık ve benzerlikleri Hint – Avrupa dil ailesindeki dillerin yakınlık ve benzerliği kadar net olarak ortaya konulamamıştır. Ancak bu diller yapı bakımından birbirlerine benzer. Ural–Altay dillerinin hepsi eklemeli dillerdir. Hepsinde belli bir derecede ünlü uyumu vardır. Sözdiziminde sözcüklerin sıralanışı aynıdır.

Ural – Altay dil grubu iki ana gruba ayrılır. Bu gruplarda içlerinde alt gruplar oluşturur[13].



Şekil 2. Ural Altay Dil Ailesi

Dil bilimcilerine göre, Türk dilleri tarihi ve coğrafi bakımdan ele alındığında altı dala ayırmak mümkün. İlk dalı, Oğuz veya güneybatı dalıdır, Azeri, Türk, Türkmen ve Horasan Türk dillerini içerir. İkinci dalı ise, Kıpçak veya kuzeybatı dalıdır, Kazak, Karakalpak, Nogay, Kırgız, Tatar, Başkir, Batı Sibiryaya diyalekti, Kırım Tatarca, Kumık, Karaçay – Balkar ve Karaite dillerini içerir. Üçüncü dalı, Uygur – Karatay veya güneydoğu dalıdır, Özbek, Uygur, Sarı Uygur ve Salar dillerini içerir. Dördüncü dalı Sibiryaya veya kuzeydoğu dalı, Yakut, Dolgan, Altay, Kakas, Şor, Tuvan, ve Tofa dillerini içerir. Beşinci dalı, Çuvaş dalı ve Volga kenarlarında çok kullanılır. Altıncı dalı da, Kalaj dalıdır, merkezi İran civarlarında çok kullanılan dildir.

Türk dillerinin birkaç önemli özelliği vardır. Bunlardan biri, “ses uyumu” dur. Türk dillerinde iki çeşit ses uyumu vardır. Türk dillerinde ünlüler ince ve kalın olmak üzere ikiye ayrılırlar. İnce ünlüler (Türkçe: *e, i, ö, ü* ve Türkmençe: *ä, e, i, ö, ü*) ağzın ön kısmında üretilir ve kalın ünlüler (Türkçe: *a, ı, o, u* ve Türkmençe: *a, y, o, u*) ağzın arka kısmında üretilir. Türk dillerinde kelimenin son ünlü harfine göre ince veya kalın ünlü sesli uyumuna tabii tutulur [14], [13], [15]. Örneğin: Türkçe *yıldız* kelimesine, *yıldız* k elimesinin son seslisi kalın ünlü olduğundan ekleyeceğimiz ekler kalın ek olmak zorundadır: *yıldız + lar*, *yıldız + lar + ı*. İnce sesli ile biten *çiçek* kelimesine ise ince ünlü uyumlu ekler eklenir: *çiçek + ler*, *çiçek + ler + i*. Türk dillerinin ikinci önemli özelliği morfolojisidir. Türk

dilleri eklemli diller olduğundan, köke ekler ekleyerek yeni kelimeler üretilir. Örneğin: *kitaplarım* kelimesinde *kitap*(kök), *lar*(çoğul eki), *ım*(birinci şahıs eki) olur.

1.6. Türk ve Türkmen Dilleri Gramerleri

Türkiye Türk dilinin İstanbul diyalektinin içermek üzere birkaç diyalekti vardır. Bu çalışma, İstanbul diyalekti Türkiye Türkçe'sinin resmi dili olduğundan, İstanbul diyalektine göre yapılmıştır. Türkmenistan Türkmençe'sinin de birkaç diyalekti vardır. Genel yaygın diyalektleri şunlardır: Teke, Yomut, Arsarı, Sarık, Salır, Göklen, ve Çovdur. Türkmenistan Türkmençe'si için, Teke diyalekti Türkmenistan'ın resmi dili olduğundan Teke diyalekti göz önünde bulundurulmuştur. Bu tezde, Türkmençe terimi "Türkmenistan'ın Türkmençe literatür dili" ve Türkçe terimi "Türkiye'deki Türkçe literatür dili" anlamlarını vermektedir.

Türk dillerinin Oğuz dalının gramerleri yani, Azerice Türkçe ve Türkmençe dilleri birbirine çok yakındırlar. Bu dillerin morfoloji bakımından genel olarak ek uyumları az değişerek veya hiç değişmeden uyumluluk gösterirler. Genel olarak Türk dilleri arasında kelime kelime bilgisayarlı çeviri yapılabilir [16]. Her iki dildeki belirsizlikler genel olarak aynıdır.

1.6.1. Fonem ve Alfabe

Ses dilin en küçük parçasıdır. En küçüğünden en büyüğüne kadar bütün dil birlikleri seslerden yapılır. Heceler (seslemler), ekler, kökler, kelimeler, kelime gruplar ve cümleler birtakım seslerin birleşmesinden meydana gelirler. Bu dil birliklerini unsurlarına ayırdığımız zaman en sonunda karşımıza bir unsur çıkar ki artık onu parçalayamayız. Fonem, ses parçalanamayan ve küçük dil birliğidir.

Ses sözlü, harf yazılıdır. Ses kulağa, harf göze hitap eder. Sesler söylenilir, ağızdan çıkan, işitilen, yani konuşulan küçük dil birlikleridir. Dili yazıya geçirmek için bu küçük unsurlar birtakım işaretlerle karşılanır. Bu işaretlere harf adı verilir. Harf, seslerin yazdaki işaretleridir.

Türkçe ve Türkmençe'de sesler ünlü ve ünsüz olmak üzere ikiye ayrılırlar. Akciğerden gelen havanın ağız boşluğunda herhangi bir engele uğramadan dil ve dudakların oluşturduğu seslere ünlü sesler denir. Ünlüler ağızdan çıkarken bazılarının

söyleniş seresi kısa, bazılarının ise uzun olur. Söylenişi kısa olan ünlülere kısa ünlüler, söylenişi uzun olan ünlülere ise, uzun ünlüler denir. Türkçe’de genellikle kısa ünlüler bulunur. Ancak, bölge ağızlarında ve Türkçe’de bulunan bazı yabancı sözcüklerde uzun ünlüler bulunur. Türkmençe’de kısa ve uzun ünlüler bulunur. Aşağıdaki tabloda Türkçe ve Türkmençe ünlülerin özellikleri gösterilmiştir.

Tablo 1. Ünlü Sesler

Türkçe					Türkmençe				
	Düz		Yuvarlak			Düz		Yuvarlak	
	Geniş	Dar	Geniş	Dar		Geniş	Dar	Geniş	Dar
Kalın	a	ı	o	u	Kalın	a	y	o	u
İnce	e	i	ö	ü	İnce	e, ä	i	ö	ü

Akciğerden gelen havanın ağız boşluğunda bir engele çarpması ile oluşan seslere ünsüz sesliler denir. Ünsüzler çıkarken ses yolunda açılma, kapanma, daralma ve gırtlakta bulunan ses tellerinde titremeler olur.

Türkçe’de toplam 21 ünsüz harf vardır. Onlar:

b, c, d, f, g, ğ, h, j, k, l, m, n, p, r, s, ş, t, v, y, z.

Türkmençe’de 21 ünsüz harf vardır. Onlar:

b, ç, d, f, g, j, f, k, l, m, n, ñ, r, s, ş, t, w, ýz.

1.6.2. Kelime

Kelime, anlam veya görevi bulunan ve tek başına kullanılabilen ses veya sesler topluluğudur. Kelimeler anlamlı veya görevli dil birlikleridir. Kelimelerin genellikle anlamları vardır. Dışarıda bir varlığı, bir nesneyi bir hareketi karşılarlar. Kelimeyi okuduğumuz veya duyduğumuz zaman o varlık veya hareket gözümüzün önünde canlanır: kitap, daktilo, koyun, kuzu, cetvel, koşmak, yazmak...

Ancak, anlam bulunmayan kelimeler de vardır: gibi, ile, ve, için, fakat, ama, kadar vb. Bunların anlamlar yoktur ve hiçbir varlığı veya hareketi karşılamazlar. Cümlede

anlamly kelimelerle birlikte kullanılır. Onların manalarına yeni ifadeler katarlar: aslan gibi, onun için gelmiřtim, sabaha kadar ađladı örneklerinde olduđu gibi.

Kelimededen küçük ses birlikleri olan sesler, heceler, ekler ve baz kökler tek başlarına kullanılmadıkları halde kelimeler tek başlarına kullanılır.

Ařađıdaki tabloda kelimeyi oluřturan unsurları Türkmençe ve Türkçe gösterilmiřtir.

Tablo 2. Türkçe ve Türkmençe kelimeyi oluřturan unsurlar

	Türkçe	Türkmençe
1	Kök	Kök (Asyl Söz)
2	Yalın kelime	Sada Söz
3	Birleřik Kelime	Gořma Söz
4	Kısaltma	Gysgaldylan Gořma Söz
5	Ek	Gořulma
6	Yapım Eki	Söz ýasaýyjy gořulma
7	Çekim Eki	Söz üýtgediji gořulma
8	Kelime Türleri	Söz Toparlary

Ařađıdaki tabloda da Türkmençe ve Türkçe kelimelerin söz içindeki işleyiřleri bakımından gösterilmiřtir.

Tablo 3. Türkçe ve Türkmençe kelimelerin söz içindeki işleyiřleri

	Türkçe	Türkmençe
1	İsim	At
2	Sıfat	Sypat
3	Sayı	San
4	Zamir	Çalyřma
5	Fil	İřlik
6	Zarf	Hal
7	var, yok, mümkün, evet, hayır	Modal Sözler
8	Yansımalar	Yaňsılı äheňler ve Őekiller

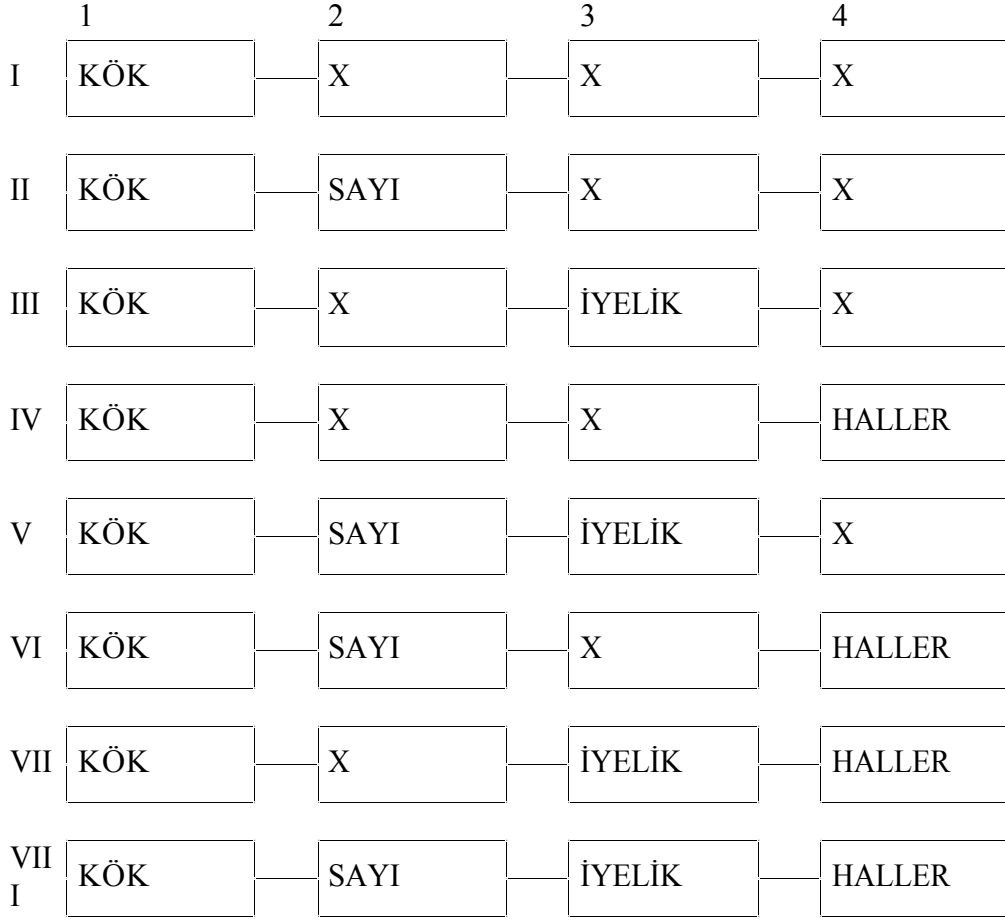
Tablo 3'ün devamı

9	Ünlemler	Ümlükler
10	Bağlaçlar	Bağlayjylar
11	Edatlar	Posleloglar
12	-da, a, -ä, hä...	Ownuk Bölekler

1.6.3. Ekler

Ekler, kendi başlarına anlam ifade etmezler, bu nedenle onlar tek başlarına kullanılamazlar. Ekler kelimelere eklenerek, ek adını alırlar. Ekler, kelimelerde zincir gibi görev üstlenirler. Ne kadar değerli olduğunu görmek için kısa bir örnek verelim. Farz edelim ki dildeki ekler bir gün çalışmayı reddetsinler, sadece köklerle kendimizi ifade etmenin ne kadar zor olduğunu görürsünüz. Ekler özellikle eklemli diller için çok önemlidirler. Kelimelerdeki her bir eke baktığımızda, bir zincirin oluşumunu hatırlatıyor. Ekler her ne kadar demirden olmasalar da, eklerin görevleri demirinkinden kadar güçlüdürler.

Türk dillerinin kendine has ek sırası vardır. Kökler kelimedede kendisine uyulan, ekler ise köke uyan unsurlardır, ekler zincirinin başladığı nokta köktür. Genel Türk dillerinin isimlere eklerinin sırasını inceleyelim.



Şekil 3. Eklerin diziliş sırası

Örneğin kitap kelimesini ele alalım.

1. *Kitap+*

2. *Kitap+lar+*

3. *Kitap+lar+ım+*

4. *Kitap+lar+ım+dan.*

Türk dillerinde eklerin sırası çok önemlidirler. Eğer bunu değiştirmeye çalışırsak anlamsız kelime elde ederiz. Yukarıdaki eklerin sırasını değiştirerek deneyelim:

kitap + ım + dan + lar

Örnekte görüldüğü üzere denememiz başarısız oldu. Çünkü zincirlerin sırasını değiştirdik. Eklerin zincirdeki sıraları çok önemlidirler.

Türk ve Türkmen dillerinde genellikle kelimelere ekler eklenerek yeni kök veya kelime türetilir. Köke eklenin ekler iki çeşittirler: yapım ve çekim ekleri. Yapım ekleri köke eklendiğinde, yeni bir kök türetilir. Örneğin: Türkçe *çay* + *cı* (çaycı) kelimesinde, *çay* köküne *cı* yapım eki eklenerek yeni kök türetilmiştir. Türkmençe *myhman* + *hana* (misafirhane) kelimesi *myhman* kelimesine *hane* eki eklenerek yeni bir kök türetmiştir. Çekim ekleri ise o kökün durumu, hali, şartı, sayısı, kişisi ve zamanı ile ilgili bilgi verir. Türkçe *kitap* + *lar* (kitaplar) kelimesinde, *kitap* köküne *lar* çekim eki eklenerek, kitapların çok olduğunu belirtir.

A. Yapım ekleri: Kökün veya kelimenin anlamında değişiklik yapan eklerdir. Bunlar eklendikleri kökten yeni bir kelime meydana getirirler. Mesela Türk kelimesinden –çe ekiyle Türkçe kelimesi yapılmıştır. Türk bir milletin ad, Türkçe o milletin dilinin adıdır. Dört çeşit yapım eki vardır:

1. **İsimden isim yapma ekleri:** İsim kök veya gövdelerinden yeni isimler yapan eklerdir: yıl-lık, köy-lü, balık-çı, kuzu-cuk, ev-cil vb.
2. **İsimden fiil yapma ekleri:** İsim kök veya gövdelerinden fiil türeten eklerdir: gece-le-mek, duru-l-mak, tür-e-mek, boz-ar-mak vb.
3. **Fiilden isim yapma ekleri:** Fiil kök veya gövdelerinden isim türeten eklerdir: aç-mak, gel-me, düş-üş, geç-im, düş-kün, sırt-kan vb.
4. **Fiilden fiil yapma ekleri:** Fiil kök veya gövdelerinden fiil yapan eklerdir: al-ınmak, gül-üş-mek, yat-ır-mak vb.

B. Çekim ekleri: Kökün veya kelimenin anlamında değişiklik yapmayıp kelimeleri kullanılış sahasına sokan eklerdir. Çekim ekleri anlama etki etmezler, köke ve kelimeye yeni bir anlam katmazlar. Fakat bir ifade katarlar. Bu ifade öteki kelimelerle münasebet ifadesidir. “Ev-den geldim.” Cümlesindeki –den eki münasebet ifadesidir ve evin anlamını değiştirmemektedir.

1. **Çoğul eki:** İsimlerin çokluk şekillerini ifade eder: -lar, -ler: kitap-lar, çocuk-lar, arabalar, ev-ler, gece-ler, sergi-ler vb.
2. **İyelik ekleri:** “İye” Türkçe’de “sahip” anlamına gelir. İyelik ekleri ismin karşıladığı nesnenin bir kişiye veya nesneye ait olduğunu gösteren çekim ekidir.
3. **Hal ekleri:** Bu ekler ismi bazen isme, bazen fiile bazen de edata bağlar
4. **Soru ekleri.**

1.7 Türkçe ve Türkmençe Gramerlerinin Karşılaştırılması

1.7.1. Alfabe

1920 yılına kadar Türkler Arap alfabesini kullanmışlardır. Türkiye resmi olarak 1928 yılında Latin alfabesine geçti. Günümüzde Arap alfabesi Çin’de, İran’da ve Arabistan’da yaşayan Türkler tarafından kullanılmaktadır. Türkmenlerde 1928 yılına kadar Arap alfabesini kullanmıştır. 1928 yılında diğer Türklerinle aynı zamanda Latin alfabesine geçmişlerdir. Eski SSCB’ye bağlı olduğundan Latin alfabeden Kiril alfabeğe geçmiştir. SSCB zamanında Kiril alfabe kullanmışlardır. Bağımsızlığını kazanmasıyla, tekrar 12 Nisan 1993 yılında Türkmenistan Latin alfabesine geçme kararı almıştır Şu anda Türkiye ile Türkmenistan’ın kullandığı Latin alfabe, birkaç harf dışında aynıdır. Türkmen alfabesinin Türk alfabesinden farklı olan harfleri şunlardır: ä, ñ, y, j, w, ý ve f. Bununla birlikte, bazı sesler aynı olmalarına rağmen yazılışları farklıdır. Türkçe’deki ı, c, v, y ve j harflerinin Türkmençe karşılığı sırasıyla y, j, w, ý ve f harfleridir. Aşağıdaki Tablo 4’te Türkçe ve Türkmençe alfabelerinin karşılaştırılması verilmiştir.

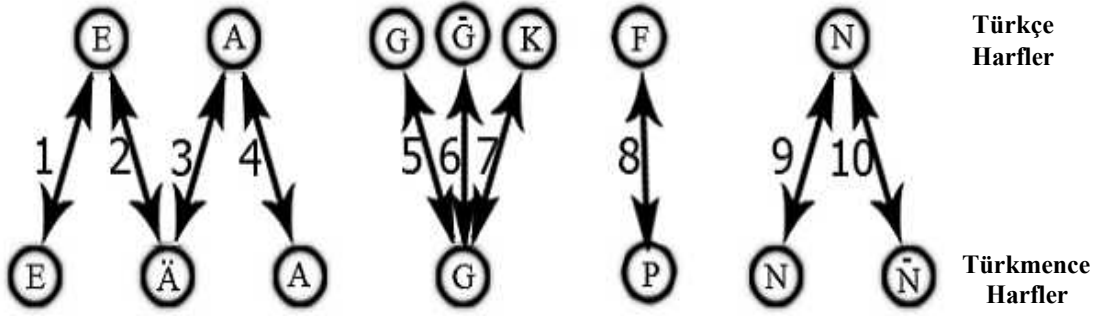
Tablo 4. Türkçe ve Türkmençe Alfabe

Sıra Numarası	1	2	3	4	5	6	7	8	9	10	11
Türkçe Karakterler	Aa	Bb	Cc	Çç	Dd	Ee	Ff	Gg	Ğğ	Hh	ıı
Türkmençe Karakterler	Aa	Bb	Jj	Çç	Dd	Ee	Ff	Gg	-	Hh	Yy
	12	13	14	15	16	17	18	19	20	21	22
	İi	Jj	Kk	Ll	Mm	Nn	Oo	Öö	Pp	Rr	Ss
	İi	žŽ	Kk	Ll	Mm	Nn	Oo	Öö	PP	Rr	Ss
	23	24	25	26	27	28	29	30	31		
	Şş	Tt	Uu	Üü	Vv	Yy	Zz	-	-		
	Şş	Tt	Uu	Üü	Ww	ý	Zz	Ää	Ññ		

Bazı Türkçe harflerin tam karşılığı Türkmençe alfabede, bazı Türkmençe harflerin ise tam karşılığı Türkçe alfabede yoktur. Türkçe alfabedeki ğ harfi yumuşak g harfi gibi söylenir. Örneğin: *kâğıt*(kagyzt), *bağ*(bag). Türkmençe alfabedeki ä harfi a ve e seslerinin

karışımı bir sestir ve bunun Türkçe alfabede karşılığı yoktur. Ä sesine lâle kelimesinde â sesi yakındır. Yine Türkmençe alfabede var olan ñ harfi Türkçe alfabede yoktur. Bu harfte, ng seslerinin birleşimi bir sestir. Örneğin: *deñiz* (deniz), *obañyz* (köyünüz). Bazı harfler birden fazla harfe karşılık gelebilir.

- Sesler benzer olmasına rağmen dillerin harf sayısı farklıdır. (Türkmençe ‘de 30, Türkçede ise 29 harf vardır.)
- Türkçedeki 29 harften 26’sına bire-bir izdüşümü vardır.
- Türkçedeki “E” harfi Türkmençedeki “E”, “Ä” harflerine; “A” harfi “A”, “Ä” harflerine; Türkçedeki “NG” harfleri ise Türkmençedeki “Ñ” harfine izdüşer.
- Türkmençedeki “G” Türkçedeki “G” ve “Ğ” harflerine izdüşer.



Şekil 4. Türkçe ve Türkmençe'nin karşılıklı harfleri

Yukarıdaki belirlenmiş harfler için gereken kuralları araştırılım.

1. “E” ünlüsünün “Ä” sesi gibi telaffuz olduğu durumlar.
beş -> baş, beri -> bəri, göre -> görä, er -> är, gedik->gädik, şehir->şäher, zehir->zäher, bebek->bäbek
2. “A” ünlüsünün “Ä” sesi gibi telaffuz olduğu durumlar.
çare -> çäre, şakirt->şägirt, adım->ädim, hakim->häkim (anlamları farklı)
lale->läle, alem->älem, hazır->häzir, aht->äht, aşikar->äşgär
3. “G” sessizinin “Ğ” sessiz gibi telaffuz olduğu durumlar
eğilmek->egilmek, ağ->ag, bağ->bag, ağız->agyız, dağ->dag, çağ->çag,
kağıt->kagyz, oğul->ogul, ege->ige, ağa -> aga, sağ->sag
4. “G” sessizinin “K” sessizine dönüştüğü durumlar
kalem->galam, kız->gyz, kale->gala, kavun->gawun, kazık->gazyk,

seksen->segsen, kuzu->guzy, kuşak->guşak, kazmak->gazmak

5. “P” sessizinin “F” sessizine dönüştüğü durumlar

harf->harp, fayton->paytun, vazife->wezipe, Ferhat->Perhat,
vefa->wepa, fırlatmak->pyrlatmak, fakir->pahyr, farz->parz,
felek->pelek

6. “Ñ” sessizinin “NG” sessizine dönüştüğü durumlar

ñ sessizi hiçbir zaman kelimenin başında gelmez.

deniz->deñiz, sonra->soñra, donmak->doñmak, yün->yüñ, bin->müñ

1.7.2. İsim Ekleri

Türk ve Türkmen dillerinde kelimeler, köklere durum ekleri eklenerek oluşturulur. Genel olarak her iki dilde de isim ekleri birbirine benzerdir. Bazı durumlarda eklerin sıralanışı ve bazı eklerin kullanılışı farklı olabiliyor. Aşağıda isim hallerine ilişkin tablolarda iki dilin ek oluşumuna örnekler verilmiştir.

Tablo 5. Türkçe ve Türkmençe isimin yaklaşma hali ekleri

Türkçe		Türkmençe	
- YA	vatan + yA = vatana	-A	watan + A = watana
- y(a)/(e)	aile + yA = aileye	-a/e	maşgala + A=maşgala
- Eğer sonu ünsüz harf ile biten kelimeye bu ek eklendiğinde y harfi düşer.	keçi + yA = keçiye	- Ünlü uyumu vardır. - Sonu a ünlüsüne biten köklere eklendiğinde kelime yazımında değişiklik olmaz söylenirken biraz uzun söylenir	geçi + A = geçä
-Ünlü uyumu vardır.	arı + yA = arıya	- Sonu i ve e ünlülerine biten köklere eklendiğinde, son ünlü ä sesine dönüşür. - Sonu y ünlüsüne biten köklere hal eklendiğinde y ünlüsü a ünlüsüne dönüşür.	ary + A = ara

Tablo 6. Türkçe ve Türkmençe isimin ilgi hali ekleri

Türkçe		Türkmençe	
-(n)Hn	vatan + nHn = vatanın	-(n)Hň	watan + nHň = watanyň
-(n)ın/(n)in/(n)un/(n)ün	deve + nHn = devenin	-(n)yň/(n)iň/(n)uň/(n)üň	düýe + nHň = düýäniň
- Sonu ünsüz harf ile biten kelimelerde n harfi düşer.	gün + nHn = günün	- Sonu ünsüz harf ile biten kelimelerde n harfi düşer.	gün + nHň = günüň
-Ünlü uyumu vardır.		-Ünlü uyumu vardır.	
		- Eğer sonu e seslisine biten kelimeye eklenirse, e sesi ä sesine dönüşür	

1.7.3. Fiil Ekleri

Aşağıdaki Türk ve Türkmen dillerinde birbirine karşılıklı örnekler gösterilmiştir. Tablo Türkçe ve Türkmençe için ikiye bölünmüştür. Her dil için ayrılan bölümdeki ilk sütun zamanın alabileceği eklerini gösterir. Ayrıca, eklerin ünlü uyumuna uyup uymadığını gösterir. İkinci sütun ise o zamana ait örneklerdir. Her iki dilde geçmiş zaman için yapı ve kurallar aynıdır, ama diğer zamanlarda farklılıklar görülmektedir. Her iki dildeki geniş zaman, şimdiki zaman, geçmiş zaman ve gelecek zaman karşılaştırılmaları sırasıyla aşağıdaki tablolarda gösterilmiştir.

1.7.3.1. Basit Kipler

Geniş zaman, Türkmen gramerlerinde “belirsiz gelecek zaman” olarak ele alınsa da Türkçe’deki geniş zaman anlamını vermektedir. Türkçe ve Türkmençe’de geniş zaman eki “ar/er”, ünlü ile biten fiillere geldiğinde ekin ünlüsüyle kelime sonundaki ünlü kaynaşır ve sadece “r” harfi ekler.

Tablo 7. Türkçe ve Türkmençe geniş zaman ekleri

Türkçe		Türkmençe	
- r, Ar, Hr	gel + Ar = gelir	- Ar	gel+Ar = geler
-r/ar/er/ır/ir/ur/ür	kullan + Hr = kullanır	- ar/er	ulan + Ar = ulanar
-Ünlü uyumu vardır.	söyle + r = söyler	- Ünlü uyumu vardır.	sözle +Ar = sözlär
	gizle + r = gizler		gizle + Ar = gizlär

Türkmen dilinde iki çeşit şimdiki zaman vardır. İlk çeşidi, “ýar/ýär” ekleri ile oluşturulur. Bazen şimdiki zaman eki “ýar/ýär”in “r”si düşmekte ve “ýa/ýä” şekline girmektedir. Söz konusu ekin bu türünün ünlüsü de daima uzundur. Her iki şekilde Türkçe’deki şimdiki zaman eki “iyor/ıyor” eklerine karşılık gelmektedirler. Aşağıdaki Tablo 8’de “ýar/ýär” için Türkçe şimdiki zaman karşılaştırılması gösterilmiştir

Tablo 8. Türkçe ve Türkmençe şimdiki zaman ekleri

Türkçe		Türkmençe	
- (H)yor	al + Hyor = alıyor	- ýÄr	al + ýÄr = alýar
- (ı/i/u/ü)yor	iç + Hyor = içiyor	- ýar/ýär	iç + ýÄr = içýär
- Ünlü uyumuna uymaz.	yaz + Hyor = yazıyor	- Ünlü uyumu vardır.	ýaz + ýÄr = ýazýar

Türk ve Türkmen dillerinde bilinen geçmiş zaman ekleri aynıdır. Türkçe’de ünsüz uyumu vardır. Örneğin *git* kelimesi *t* harfsine bittiğinden *ti/tı* eki eklenir. Türkmençe de ise ünsüz uyumu yoktur.

Tablo 9. Türkçe ve Türkmençe bilinen geçmiş zaman ekleri

Türkçe		Türkmençe	
- DH	gel + DH = geldi	- dH	gel + dH = geldi
- di/dı/du/dü/tı/ti/tu/tü	at + DH = attı	- dy/di/du/dü	at + dH = atdy
- Ünlü uyumu vardır.	yürü + DH = yürüdü	- Ünlü uyumu vardır.	ýöre + dH = yöredi
- Ünsüz uyumu vardır.	yaz + DH = yazdı	- Ünsüz uyumu yoktur.	ýaz + dH = ýazdy

Öğrenilen geçmiş zaman Türk dilinde “mış/miş/muş/müş” ekleri ile oluşturulur. Türkmen dilinde ise “ypdyr/ipdir/updyr/üpdır” ekleri ile oluşturulur. Türkmen dilinde sonu ünlü ile biten bir fiile öğrenilen geçmiş zaman eki geldiğinde, fiilin sonundaki ek uzar.

Tablo 10. Türkçe ve Türkmençe öğrenilen geçmiş zaman ekleri

Türkçe		Türkmençe	
- mHş	bil + mHş = bilmiş	- HpdHr	bil + HpdHr = bilipdir
- mış/miş/muş/müş	uç + mHş = uçmuş	- ypdyr/ipdir/updyr/üpdır	uç + HpdHr = uçupdyr
- Ünlü uyumu vardır.	otur + mHş = oturmuş	- Ünlü uyumu vardır.	otur + HpdHr = oturypdyr

Türkmen dilindeki gelecek zaman ile Türk dilindeki gelecek zaman birbirinden farklıdır. Türk dilinde gelecek zamandaki şahıs eki Türkmen dilinde kullanılmaz. Şahıs ekinin yerine, fiilin başına şahıs zamirleri getirilir. Türkmençe’de ünsüz uyumuna girmez.

Tablo 11. Türkçe ve Türkmençe gelecek zaman ekleri

Türkçe		Türkmençe	
- yAcAk	dur + yAcAk = duracak	-jAk	dur + jAk = durjak
- (y)acak/(y)ecek	oku + yAcAk = okuyacak	-jak/jek	oka + jAk = okajak
- Ünlü uyumu vardır.	gel + yAcAk = gelejek eğil + yAcAk = eğilecek	- Ünlü uyumu vardır.	gel + jAk = geljek egil + jAk = egiljek

1.7.3.2 Birleşik Kipler

Türk dillerinde genel olarak birleşik zamanlar bilinen geçmiş zaman veya öğrenilen geçmiş zamandan oluşur. Birleşik zamanlarda Türkçe ile Türkmençe birbirine uymuyor. Türkçe’de kullanılan hikâye ve bilinen geçmiş zaman Türkmençe’de bir birleşik zamana eşittir.

Tablo 12. Türkçe ve Türkmençe birleşik kipler ekleri

Türkçe		Türkmençe	
- mış + di	yazmıştın	- ypdy	ýazypdyň
- Birleşik hikaye zaman		- Hâzirki zaman bilen Öten zaman birleşmesi	

1.7.3.3. Edilgen ve Ettirgen Ekleri

Türkçe ve Türkmençe için edilgenlik ve ettirgen hali, kelimelere ek eklenerek yapılır. Eklerde küçük farklılıklar olabilir. Aşağıdaki tabloda Tablo 13 her iki dilde yapılaş şekline örnekler vardır.

Tablo 13. Türkçe ve Türkmençe edilgen ekleri

Türkçe		Türkmençe	
- Hl, Hn	ver + Hl +di= verildi	- Hl, Hn	ber +Hl +di = berildi
- il/il/ül, in/in/un/ün	yaz + Hl + dı = yazıldı	-yl/il/ul/ül, yn/in/un/ün	ýaz +Hl + dy = ýazyldy
- Ünlü uyumu vardır.	al + Hn + dı = alındı bil + Hn + di = bilindi	- Ünlü uyumu vardır.	al + Hn +dy = alyndy bil + Hn + di = bilindi

1.7.3.4. Soru Cümleler

Türkçe ve Türkmençe cümlelerini “mH” eki kullanılarak soru şekline dönüştürülebilir. Türkmençe’de soru ekleri kelimeye bitişik yazılır, ama Türkçe’de soru eki kelimedenden ayrı yazılır. Türkçe’de bilinen geçmiş zaman hariç bütün zamanların soru şekilleri “fiil + zaman eki + soru eki + sahiplik eki” sırasıyla yapılır. Diğer taraftan geçmiş zaman, “fiil + zaman eki + sahiplik eki + soru eki” şeklindedir. Türkmençe’de gelecek zaman hariç tüm zamanların soru şekilleri “fiil + zaman eki + sahiplik eki + soru eki” şeklindedir. Gelecek zaman sahiplik eki almaz, sahiplik eki yerine zamir kullanılır. Örneğin: *sen gitjekmi?* (sen gidecekmisin?)

Tablo 14. Türkçe ve Türkmençe zamanların soru şekilleri

Türkçe		Türkmençe	
- mH	al +Hyor mH = alıyor mu	- mH	al + ýAr + mH= alýarmy
- mı/mi/mu/mü	ıç + Ar mH = içerr mu	- my/mi	ıç + Ar + mH = içermi
- Ünlü uyumu vardır.	yaz+AcAk mH = yazacak mı	- Ünlü uyumu vardır.	ýaz + jAk + mH = ýazjakmy

Tablo 15. Türkçe ve Türkmençe gelecek zaman soru şekli

Türkçe		Türkmençe	
-yAcAk mH	dur + yAcAk mH = duracak	-jAk + mH	dur + jAk + mH= durjakmy
-(y)AcAk	oku + yAcAk mH= okuyacak mı	-jAk + my/mi	oka + jAk +mH= okajakmy
mı/mi/mu/mü		- Ünlü uyumu vardır.	
- Ünlü uyumu vardır	gel + yAcAk mH= gelejek mi		gel + jAk +mH= geljekmi
	eğil + yAcAk mH = eğilecek mi		egil + jAk +mH= eğiljekmi

1.7.3.5. Olumsuz Cümleler

Türk ve Türkmen dillerinde cümlelerin olumsuz halleri, olumsuz ek kullanılarak elde edilse bile bazı durumlarda birbirinden farklılık göstermektedirler. Genel olarak her iki dilde de “mA” eki ile cümleler olumsuz hale getirilir. Örneğin: gel + me + di. Her iki dilde de bu kalıptan çıkan durumlar vardır. Türkçe’de geniş zamanın olumsuzluğu için “mH” eki kullanılır. Diğer taraftan, Türkmençe gelecek zamanın bazı durumlarında olumsuzluk eki yerine “däl” kelimesi eklenerek olumsuz cümle yapılır. Aşağıdaki tabloda örnek olarak geçmiş zamanın olumsuzluk hali gösterilmiştir.

Tablo 16. Türkçe ve Türkmençe bilinen geçmiş zaman olumsuzluk eki

Türkçe		Türkmençe	
-mA +dH	gel +mA+di = gelmedi	-mA +dH	gel + mA+di = gelmedi
-ma/me + dH	at +mA+dı = atmadı	-ma/me + dH	at + mA+dı = atmady
- Ünlü uyumu vardır	yürü +me+di = yürümedi	- Ünlü uyumu vardır	ýöre + mA + di = yöremedi
	yaz + mA+dı = yazmadı		ýaz +ma+ dH = ýazmady

1.8. İmla Hatalarının Düzeltilmesi

Günümüzde kelime işleyici programlarının çoğu başarılı bir şekilde imla hatalarını bulabilmekte ve uygun olan çözümleri önerebilmektedir. Microfost'un "*Word*" programı en ünlülerindendir. 2 çeşit imla hatası olabilir; ilki OSR (Optical Character Recognition) diğeri, elle yazılan yazılardan dolayı oluşan hata. Hataların çeşidine göre algoritma üretilmesi önemlidir. OSR'lerde anlamsız karakterler yüze çıkabilirken, elle yazılan metinlerdeki hatalarda iki harfin yeri değişebilmektedir.

Elle yazılan metindeki hataların bulunması da sözlüklü ve sözlüksüz olmak üzere iki çeşide ayrılırlar. Sözlüksüz yöntemlere örnek olarak frekans ve poligram yöntemleri gösterilebilir. Frekans yönteminde hataların algılanması metindeki rastlanan kelimelerin rastlantı sıklığına göre değerlendirilir. Poligram yönteminde ise, kelimedeki harf birleşmelerine göre değerlendirilir.

Sözlük kullananlar yöntemlerde mutlak ve bağıl olmak üzere ikiye ayrılırlar. Mutlak yöntemde, daha önce o kelimenin oluşabilecek hatalı türevleri sözlükte tutular ve ona göre düzeltme yapılır. Bağıl yöntemlerde ise hatalı kelimeye en yakın olan kelime aranır ve hata uzaklığına göre sonuçlar üretilir. Genellikle bu yöntemde birebir eşleme değil, harf düşmesi, harf eklenmesi, harf değişikliği için ön işlemler yapılır. Bu durumda yeni kelimelerin listesinin kısaltılması için frekans ve poligram yöntemleri kullanılmaktadır.

Araştırmalara göre, hataların büyük bir kısmını %86'lık kısmı tekli harf değişmesinden, %8'lik kısmı komşu harflerin yer değişmesinden, %4,5'lik kısmı ise tekli harf düşmesi ve eklenmesinden doğmaktadır. Bu gruplara girmeyen %1,5'lik kısım diğer hata türlerini oluşturmakta ve bu gruplarla kıyaslandığında daha az görülmektedir [17].

Yukarda araştırmalardan da görüleceği üzere, genel hata tekli harf değişmesinden olmaktadır. Eğer sözlüklü bağıl yöntem kullanılsa ve uzaklığı bir vererek sonuçlara ulaşırsa hataların büyük kısmı giderilmiş olacaktır. Bu işlemleri Levenshtein uzaklık algoritması (Levenshtein Distance Algorithm) ile yapabiliriz. Levenshtein algoritması verilen giriş ve hedef kelimelerinin uzaklığını hesaplar. Uzaklık silme, ekleme veya yer değiştirme olabilir. Örneğin:

Eğer kaynak "okul" ve hedef "okul" ise $LU(\text{kaynak}, \text{hedef}) = 0$ olur. Çünkü herhangi bir transfere gerek yoktur. Kaynak ve hedef aynıdır.

Eğer kaynak "okul" ve hedef "okil" ise $LU(\text{kaynak}, \text{hedef}) = 1$ olur. Çünkü i yerine u yer değişmesi olacaktır. Yani tek değişiklik yeterli olmaktadır.

2. YAPILAN ÇALIŞMALAR

2.1. Giriş

Türkçe ve Türkmençe konuşan insanların internet üzerinden birbirleri ile sohbet edebilmeleri için Türkçe ve Türkmençe çeviri yapan sohbet motoru uygulaması geliştirildi. Sohbet uygulaması 2 modülden oluşur. İlk modül Türkçeden Türkmençeye çeviri yapar, ikinci modül ise Türkmençeden Türkçeye. Çeviri yapan sohbet uygulamasının geliştirilmesinde Java programlama dili kullanıldı. Java dilinin internete uyarlamasının kolaylığı, tam platform bağımsızlığı, kolay programlanabilir olması ve güçlü dil olması bu dilin seçilme tercihlerindedir.

Kullanıcı programı kullanmaya başlamadan önce Türkçe veya Türkmençe bildiğini belirtir. Eğer Türkçe seçmişse, sohbet ortamında Türkçe cümle yazar ve karşı taraftaki Türkçe bilmeyen kişi, yazıları Türkmençe okur ve karşılığını Türkmençe yazar. Türkçe seçen kişi, Türkmençe yazılan cümleleri Türkçe okuyabilir.

Girilen Türkçe bir cümle, ilk önce morfolojik analizci ile cümledeki kelimeler teker teker analiz edilir. Cümledeki her kelime kök ve ekler ayrıştırılır. Kök sözlüğünden Türkçe kökün Türkmençe karşılığı alınır, ekler sözlüğünden de eklerin Türkmençe karşılığı alınır. Türkmençe kök ve ekleri sözlüklerden çektikten sonra, Türkmençe kelime üreticiye gönderilir ve Türkmençe kelime üretilir. Bu işlem her kelime için ardışık bir şekilde oluşturulur ve oluşan Türkmençe cümle çıktı olarak gönderilir. Eğer girişteki kelime çözümlenemezse, öneri mekanizması imla hatalarını düzeltilmeye çalışır. Eğer birden fazla çözüm varsa, en uygun olanını alır. Bu iki dilin gramerleri birbirine çok yakın olduklarından, çoğunlukla bir dildeki morfem diğer dildeki morfeme biraz değişikliğe uğrayıp veya hiçbir değişikliğe uğramadan denk düşer. Çoğunlukla bir dildeki belirsizlikler aynen diğer dilde de korunur.

2.2. Sohbet Motorunun Tasarımı

Soketler yardımı ile Java dilinde sohbet sistemi programlandı. Web siteden applet'i yükleyen kullanıcıların birbirleri ile Türkçe ve Türkmençe sohbet edebiliyorlar.

Güvenlik nedeniyle appletler birbirleri ile direkt iletişim kuramazlar. Bu nedenden dolayı sohbet sistemi sunucu/istemci mimarisine göre geliştirildi. Server üzerinde bir uygulama çalışır, devamlı bağlantı bekler ve bağlantı isteklerine cevap verir. Appletin bağlandığı HTML sayfası da bu server üzerindedir. HTML sayfası diğer bilgisayarlar (kullanıcılar) tarafından yüklediğinde, bu sayfanın içinde olan applet çalışacak ve server ile bağlantı kuracaktır. Girilen tüm veriler server'a gönderilecek, server ise gelen bilgileri, kendisine bağlı olan tüm kullanıcılara geri gönderecektir.

Java programlama dilini kullanarak geliştirilen sohbet sistemi toplam 3 sınıftan oluşacak. Birincisi Sohbet Server, server üzerinde çalışacak ve bağlantı taleplerine cevap verecektir. İkincisi Bağlantı Sınıfı, Bağlantı taleplerini karşılayacak ve kullanıcılar arasındaki veri iletişimini düzenleyecek. Üçüncüsü Sohbet Applet, kullanıcının bilgisayarında çalışacak.

Sohbet Server sınıfı kendi başına çalışan bir thread şeklinde olacaktır. Bunun için implements Runnable kullanılır. Soket oluşturabilmek için porta ve ip adrese ihtiyaç var. Aşağıda Server sınıfı için gerekli olan parametreler:

```
public static final int PORT = 8888;
protected ServerSocket dinle;
protected Vector baglantilar;
Thread baglan;
```

PORT sabit bir değişken ve daha önceden karar verdiğimiz port numarası atandı. *dinle* bağlantı taleplerini dinlemek için, *baglantilar* aktif bağlantıların tutulacağı liste için, *baglan* sohbet server sınıfının thread 'ini tutar

```
public void run() {
try { while(true) {
    Socket istemci=dinle.accept();
    baglanti b = new connection(this, istemci);
    baglantilar.addElement(b); }
} catch (IOException e) {
    System.err.println("Bağlantı talebi beklerken hata :"+e);
    System.exit(1); } }
```

run() metodu sonsuz döngü içinde bağlantı taleplerini bekliyor. Bağlantı talebi geldiğinde bu talep için *baglanti* adında yeni bir *connection* nesnesi oluşturuyor. Böylece, sohbet yapmak isteyen istemci serverde bir nesnesi oluşturulmuş ve servere bağlanmış olur.

Baglanti sınıfı sohbet server sınıfı tarafından oluşturulur ve aktif bağlantılar arasındaki iletişimi kontrol eder. Her bağlantı talebi için *baglanti* nesnesi oluşturulur. Bu nesne bağlantı talebini alır ve bağlantı sonuna kadar çalıştırır. Bundan dolayı her bağlantı kendi başına bir thread olur.

```
protected Socket istemci;
protected DataInputStream in;
protected PrintStream out;
protected chatserver server;
```

İstemci Sohbet Server'dan Socket'i alır. *in* ve *out* adlı giriş/çıkış streamleri bu nesne ile bağlantılı olacaklardır. Server Sohbet Server objesini gösterir. *Baglanti* sınıfının başlangıcında (constructor) parametre olarak alınan *server* ve *istemci* değişkene saklanıyor. Sonra bir giriş ve bir çıkış stream i oluşturulmak isteniyor.

```
public void run() {
  String satir;
  try {
    while(true) {
      satir=in.readLine();
      if(satir!=null)
        server.mesajgonder(satir);
    }
  } catch (IOException e) {
    System.out.println("Hata:" + e);
  }
}
```

Baglanti sınıfının asıl çalışan kısmı gelen mesajları aktif bağlantılara gönderen sonsuz bir döngüden oluşuyor. Mesaj göndermek için sohbet server sınıfını mesajgonder metodu kullanılıyor. Mesasgonder metodu gelen cümleyi, Türkçe ise Türkmenceye çevirisi için, Türkmence ise Türkçe çevirisi için gönderir.

java.net içerisinde Socket veya URL gibi internet iletişimde gerekli olan sınıflar bulunur. java.io içerisinde giriş/çıkış sınıfları, java.util içerisinde ise bir sürü gerekli sınıf bulur.

Toplam üç sınıfımızdan biri olan sohbet appleti, tek grafik ara yüzüne sahip sınıftır. Bu yüzden java.applet paketinin yanında java.awt ve swing sınıf paketlerinin kullanması gerekiyor.

```
public static final int PORT = 8888;
Socket socket;
DataInputStream in;
PrintStream out;
TextField inputfield;
TextArea outputarea;
Thread thread;
```

PORT, *socket*, *in*, *out*, *chatserver* ve *baglanti* da açıklanan ile aynı anlamı taşır. *inputfield* giriş nesnesini *outputarea* ise çıkış nesnesini tutuyor. *thread* kendine ait Thread 'i saklıyor. Chap Applet sınıfımızda esas iş yapacak metotlar şunlardır:

```
public void start()
public void stop()
public void run()
```

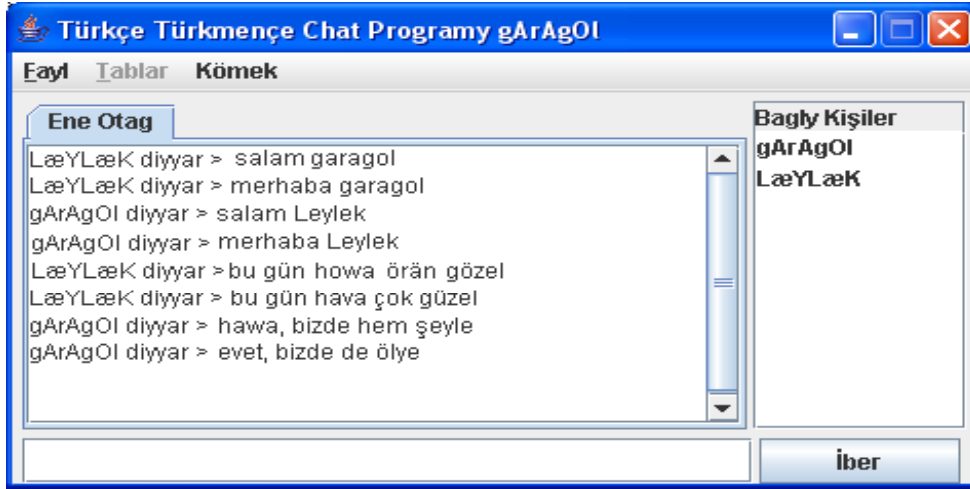
Start() metodunda applet server ile bağlantı kurulmaya çalışılıyor. Gerekli olan server adresi HTML dokümanı tarafından belirleniyor. Bu da daima HTML dokümanın bulunduğu serverın adresidir. Başka serverlar ile bağlantıya izin verilmediği için, başka server olmasında hiç bir anlamı yok. Hata oluşursa veya bağlantı kurulursa, sonuç bir mesaj ile kullanıcıya bildiriliyor.

Stop() Metodu browser tarafından, kullanıcı sayfayı terk ettiği zaman otomatik olarak çağrılır. Socket kapatılıyor ve thread durduruluyor.

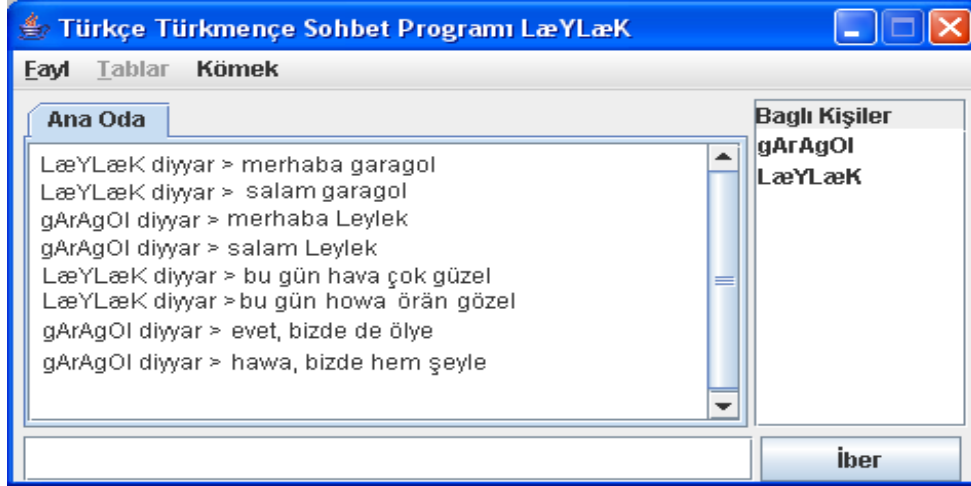
Run() metodunda sonsuz bir döngü içinde giriş sahasından girilen mesajlar, çıkış sahasında gösteriliyor.

Sunucu makinede Server program sonsuz döngü içinde istemcinin kendine bağlamasını bekler. Herhangi bir istemci sohbet ortamına katılmak için daha önce belirlenmiş port numarası, serverin ip adresi ve rumuzunu yazarak, servere bağlanır. İstemci bağlandıktan sonra, bütün mesajlarını servere gönderir. Server de mesajı alır ve ilgili istemci/istemcilere iletir. Bağlı olan istemci çevirim içi olan herkesle özel ve genel olarak mesajlaşabilir.

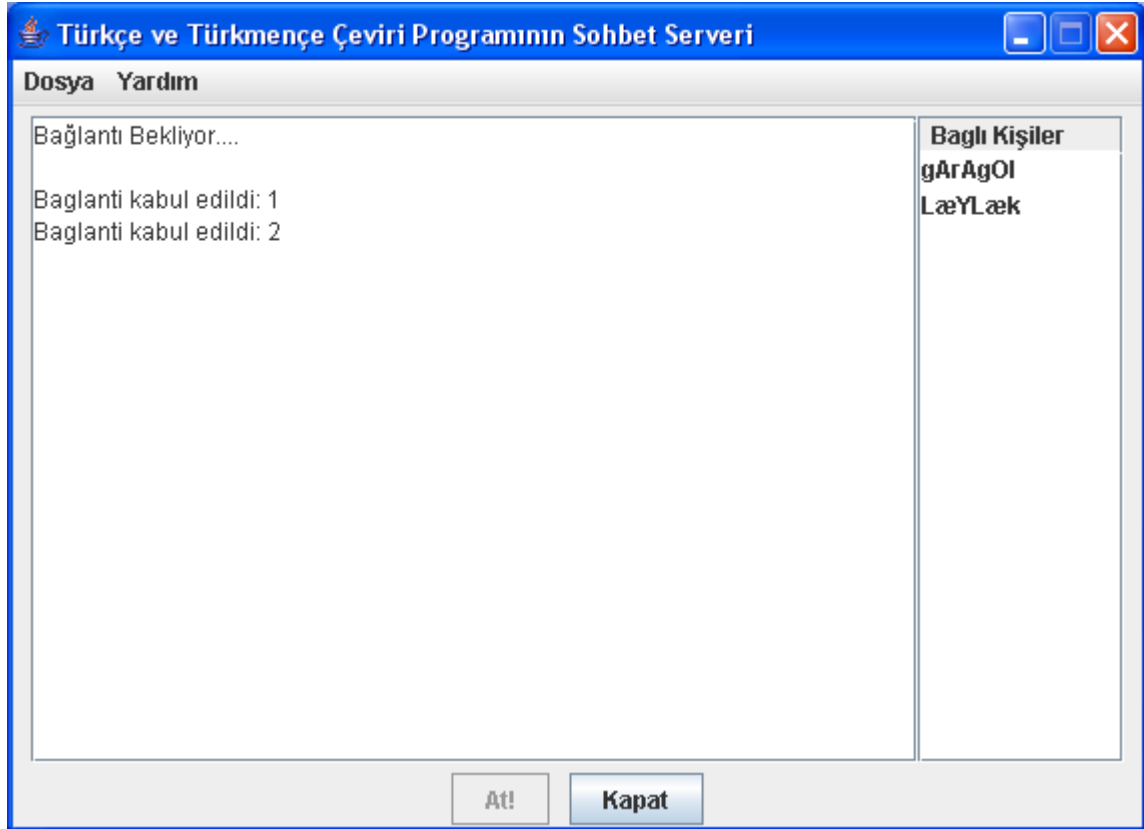
İstemci olarak bağlananlar kendi aralarında özel mesajlaşabilirler. Bunun için, “Bağlı Kişiler” menüsünden istediği rumuzun üstüne çift tıklar ve açılan yeni pencerede o kişi ile yazışabilir. Aşağıda istemci ve server pencereleri gösterilmiştir.



Şekil 5. Garagol rumuzlu Türkmençe bilen istemcinin penceresi



Şekil 6. Leylek rumuzlu Türkçe bilen istemcinin penceresi



Şekil 7. Sonsuz döngüde istemci bekleyen server

2.3. Türkçe ve Türkmençe Morfolojik Analiz Modülü

Türkçe ve Türkmençe özgür kelime sırası kullanan iki doğal dildir. Teorik olarak cümlelerin kelimeleri değişik sırada organize edilebilir. Türkçede bir fiilin nesnesi o fiilin hemen yanında tercih edilse bile bu şart değildir. Türkçe ve Türkmençe birbirlerine çok yakın iki dil olsalar bile, kelime kelime çeviri yapan sistemde bazı sınırlamalar ile karşılaşabilir.

Çeviri için Türkçe ve Türkmençe için ayrı ayrı düzyazı dosyasında sözlük oluşturuldu. Ayrıca, Türkçe – Türkmençe ve Türkmençe – Türkçe kök ve ekler sözlükleri oluşturuldu. Bunlar, morfolojik olarak ayrıştırılmış olan kök ve eklerin karşı (hedef) dildeki çevirisini almak için kullanılır. Türkçe ve Türkmençe sözlüklerde her kelimelerin tipi ve özellikleri sözlükte belirtildi. Aşağıdaki tabloda kelimelerin tiplerini tutmak için kullanılan kısaltmalar gösterilmiştir.

Tablo 17. Sözlükte tutulan kelime tipleri

Türkçe	Türkmençe	Açıklama ve örnek
IS	AT	İsim
FI	ISH	Fiil
SI	SY	Sıfat
SA	SAN	Sayı
OZ	HAS	Özel isim
ZA	CAL	Zamir
YUM	YUM	Yumuşama. Örnek: kitap – kitabı (kitapı değil)
DUS	DUS	Harf Düşmesi nutuk – nutka (nutuka değil)
TERS	TERS	Ters dönüşüm saat – saate (saata değil)
YAL		Kelime sadece yalın olarak kullanılır
GEN		Geniş zaman istisnası

Türkçe sözlükten örnek:

addetmek FI GEN YUM

addolmak FI

adem IS

Adem OZ

adese IS

Türkmençe sözlükten örnek:

äşgär AT

äşgärlik AT YUM

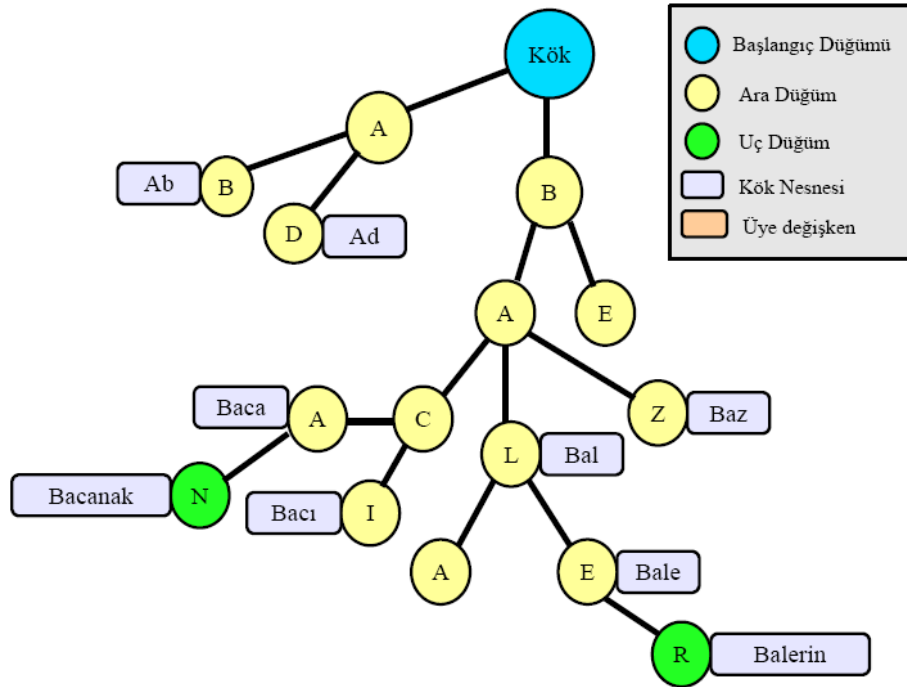
aşgazan AT

aşyr ISH

Uygulamada kökleri bir düzyazı dosyadan okutularak bir Ağaç Sözlük oluşturulur. Ağaç Sözlük kendisine verilen sözlük okuyucu kullanarak tüm kökleri okur ve bir Kök Ağacına yerleştirir. Bili bir kök için aday kökleri bulmak isteyen istemciler Ağaç sözlükten çekerler.

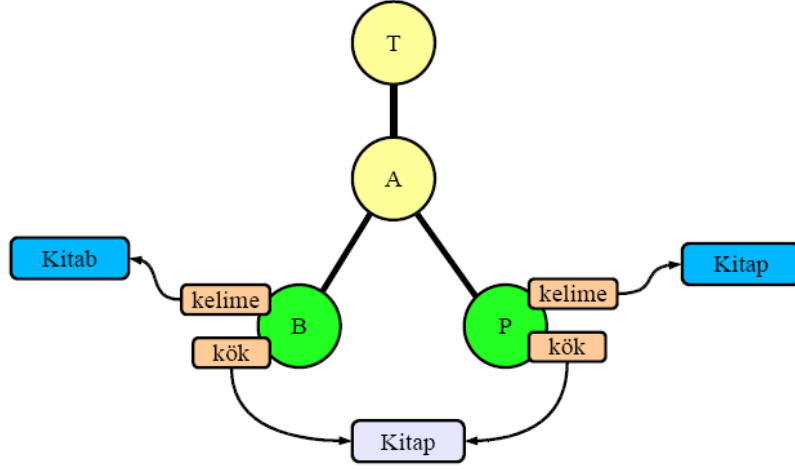
Kök Ağacı çeviri sisteminin temel veri taşıyıcılarından biridir. Kök sözlüğünden okunan tüm kökler bu ağaca yerleştirilirler Ağaca eklenen her kök harflerine göre bir ağaç oluşturacak şekilde yerleştirilir. Bir kök'ü bulmak için ağacın başından itibaren kök'ü oluşturan harfleri temsil eden düğümleri izlemek yeterlidir. Örneğin aşağıdaki şekilde “Bale” kökünü bulmak için önce ağacın en başındaki “b” harfine, sonra “a” , sonra “l” ve son olarak ta “e” harfini temsil eden alt düğüme doğru ilerlemek yeterlidir. Eğer bir kök'ü ararken erişmek istediğimiz harfe ait bir alt düğüme gidemiyorsak kök ağaçta yok demektir.

Ağacın bir özelliği de boşuna düğüm oluşturmamasıdır. Eğer bir kök'ün altında başka bir kök olmayacaksa tüm harfleri için ayrı ayrı değil, sadece gerektiği kadar düğüm oluşturulur. Aşağıdaki şekilde “balerin” kökü için sadece B-A-L-E-R düğümleri oluşmuş, daha aşağıda kök olmadığı için Balerin kökü doğrudan “R” düğümüne eklenmiştir.



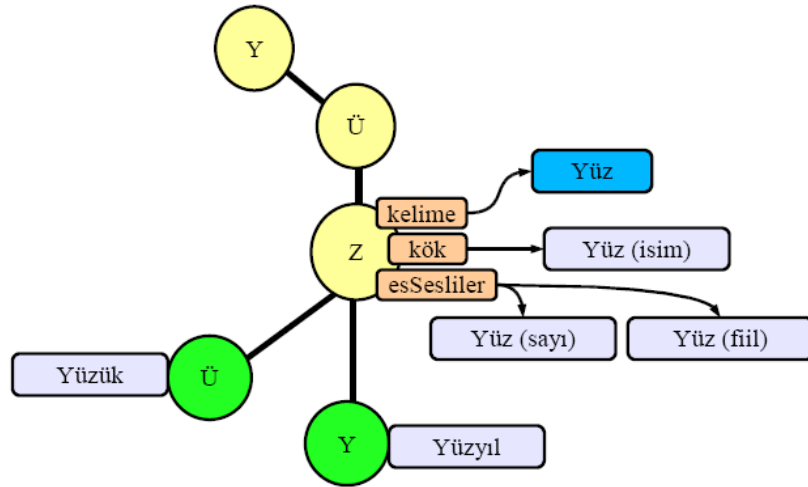
Şekil 8. Ağaçtan köke ulaşma

Çeşitli nedenlerle değişikliğe uğrayabilecek olan kökler ağaca eklenirken değişmiş halleri ile beraber eklenirler. Örneğin kitap kökü hem “kitab” hem de “kitap” hali ile sözlüğe eklenir, ancak bu iki kelime için oluşan düğüm de aynı kök'ü gösterirler. Böylece “Kitabına” gibi kelimeler için kök adayları aranırken “kitap” köküne erişilmiş olur.



Şekil 9. Köklerin değişmiş halleri ile Ağaç sözlüğüne eklenmesi

Eş Sesli olan kökler aynı düğüme bağlanırlar. Aşağıdaki örnekte “Yüz” kökünün üç hali de aynı köke bağlanmış şekilde duruyorlar. Ağacın oluşumu sırasında ilk gelen kök düğümdeki “kök” değerine, sonradan gelenler de “esSesliler” listesine eklenirler. Arama sırasında bu üç kökte aday olarak döndürülür.



Şekil 10. Eşsesli köklerin Ağaç sözlüğüne eklenmesi

Türkçe ve Türkmençe ekler belirli kuralla göre xml dosyasında tutuluyor. Xml dosyasında Türkçe için isim ve fiil ekleri belirlenerek, üretim kuralları isim veya fiil ekine hangi kurallara göre nasıl oluşacağını belirler.

Kurallar:

(A):son seslinin ince, kalın olmasına göre a ya da e harfi üretilir.

(I):son seslinin durumuna göre i,ı,u ya da ü üretilir.

?x: son harf sesli ise x harfi (ya da kural ise ilgili harfler) eklenir.

^x: son harf sert ise göre x harfinin sert hali kullanılır.

Küçük harfler: doğrudan üretime eklenir.

Büyük harfler: Türkçeye özel karakterlerin benzerleri. G yumuşak g gibi doğrudan eklenir. Aşağıda bir ekin nasıl üretildiğine bir örnek verilmiştir.

```
<ek ad="ISIM_KUCULTME_CIK" uretim="'^c+(I)+k">
  <ozel-durum ad="SON_HARF_YUMUSAMA" uretim="'^c+(I)+G"/>
  <ardisil-ekler>
    <kume>ISIM_HAL</kume>
    <kume>ISIM_SAHİPLİK</kume>
    <kume>ISIM_KISI</kume>
    <kume>İMEK_ZAMAN</kume>
    <aek>ISIM_COĞUL_LER</aek>
    <aek>ISIM_YOKLUK_SİZ</aek>
    <aek>ISIM_BULUNMA_LİK</aek>
    <aek>ISIM_DURUM_LİK</aek>
    <aek>ISIM_BULUNMA_Lİ</aek>
    <aek>ISIM_BİRLİKTELİK_LE</aek>
  </ardisil-ekler>
</ek>
```

Sıfat eki “*cık*” ekinin adı ISIM_KUCULTME_CIK olarak tanımlanmış. ^c bu işaretin anlamı, eğer “*cık*” eki sonu sert harf ile biten kelimeye eklendiğinde “*çık*”a dönüşür demektir. “*cık*” ekinin ardından ünlü harf ile başlayan ek geldiğinde sondaki *k* harfi *g*

harfine dönüşüyor. Bunu da özel durumla belirtildi. Ardışıl ekler bloğunda da “*cık*” ekinden sonra gelebilecek ekler belirtilmiştir. Birden fazla özel durum olabilir, bunları ardışık olarak eklenebilir.

Bu xml dosyasını programa okutuluyor ve yukarda belirtilen kurallara göre belleğe ekleniyor. Üretim kuralları bir ekin hangi kurallara göre nasıl oluşacağını belirler. Ek yazılım tarafından okunduğunda bu kurallara göre olası tüm üretimler tespit edilir.

Çeviri işlemi için her dilin kendisine has çözümleyiciye ihtiyaç vardır. Çözümleme işlemi için sistem başlamadan tüm kökler ve ekler belleğe yüklenir. Kökler Ağaç yapısında belleğe yüklenir. Eklerde kendine has kuralları ile, yani hangi durumda ne ekleri üretileceği belirtilerek, o ekten sonra hangi eklerin gelebileceği belirtilerek ve varsa özel durum özellikleri ile belleğe yüklenir.

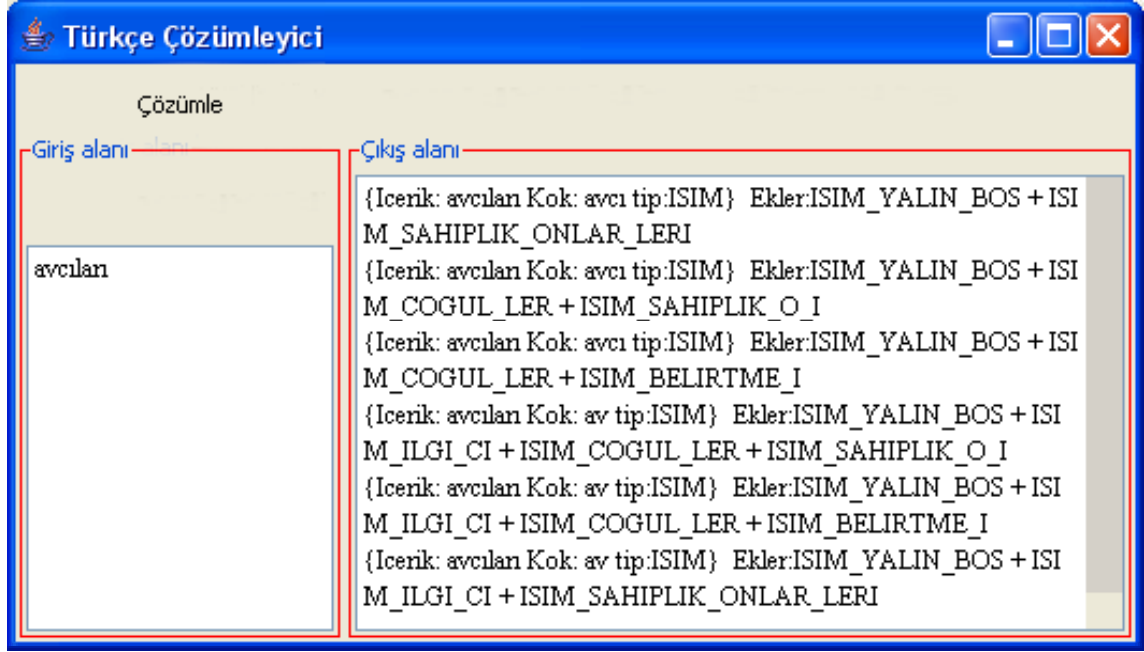
Çözümleyiciye bir cümle veya kelime girildiğinde, çözümleyici kelimeleri sırasıyla alır ve her kelimenin içinde Türkçe veya Türkmençe’ye uymayan karakter olup olmadığını kontrol eder. Eğer başarılı bir şekilde geçerse, ağaç sözlükten o kelimeye en uygun kökleri çekip alır. Örneğin: *balerinlerin* kelimesini çözümlenmesi için giriş olarak verdiğimiz, ağaç sözlükte *balerinlerin* kelimesine uyuşan tüm kökleri getirir. *Balerinlerin* kelimesi için ağaç sözlükten *bal*, *bale*, *balerin* kökleri döner. Bu gelen köklere, sistemdeki ekleri ekleyerek, giriş kelimesi olan *balerinlerin* kelimesi ile uyup uymadığını kontrol ederiz. Bu kelimenin çözümü:

{Icerik: *balerinlerin* Kok: *balerin* tip:ISIM} Ekler:ISIM_YALIN_BOS + ISIM_COGUL_LER + ISIM_SAHİPLİK_SEN_IN

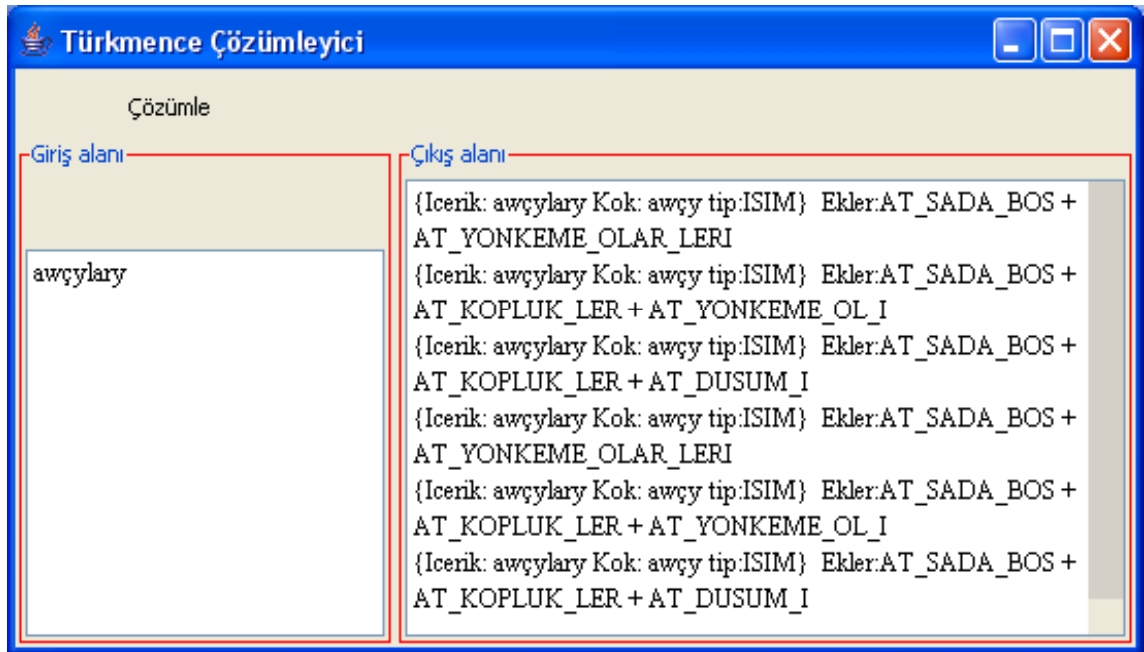
{Icerik: *balerinlerin* Kok: *balerin* tip:ISIM}Ekler:ISIM_YALIN_BOS + ISIM_COGUL_LER + ISIM_TAMLAMA_IN

olarak döner.

Aşağıda *avcıları* (awçylary) kelimesinin Türkçe ve Türkmençe çözümleyicilerinde nasıl çözümlendikleri gösterilmiştir.



Şekil 11. Türkçe Çözümleyici



Şekil 12. Türkmençe Çözümleyici

Bir kelimenin birden fazla çözümü olabilir. Yukarıdaki çözüm gibi bazı çözümlerden herhangi birini Türkmençe kelime üreticiye gönderirsek doğru çeviriyi elde edilebilir. Ama maalesef birden fazla çözümün hangisinin doğru olduğunu belirleyen bir

mekanizma yoktur. Ağaç sözlükten gelen köklerden en uzun olanını alınırsa, doğruluk ihtimalini yükselttiği görülmüştür.

2.4. Türkçe ve Türkmençe ve Üretim Çeviri Modülü

Türkçe ve Türkmençe için ekler ve kökler sözlüğü tutulmuştur. Türkçe bir kelimeye karşı Türkmençe bir veya birden fazla kelime karşı gelebiliyor. Türkmençe içinde birden fazla Türkçe kelime karşılık gelebiliyor.

Aşağıda Türkçe Türkmençe sözlükten örnek verilmiştir:

abone abonent

acele gyssag

acele tiz

aceleci gyssanmaç

acıma rehim

acımasız wagşyça

acımasız zabun

acil gyssagly

aç aç

açgöz açgöz

açı burç

Türkmençe Türkçe ekler bir Map'te tutulur; tutuluşuna örnek:

ekTnTr.put(TurkmençeEkAdlari.AT_DUSUM_DE, TurkceEkAdlari.ISIM_KALMA_DE);

ekTnTr.put(TurkmençeEkAdlari.AT_DUSUM_I, TurkceEkAdlari.ISIM_BELIRTME_I);

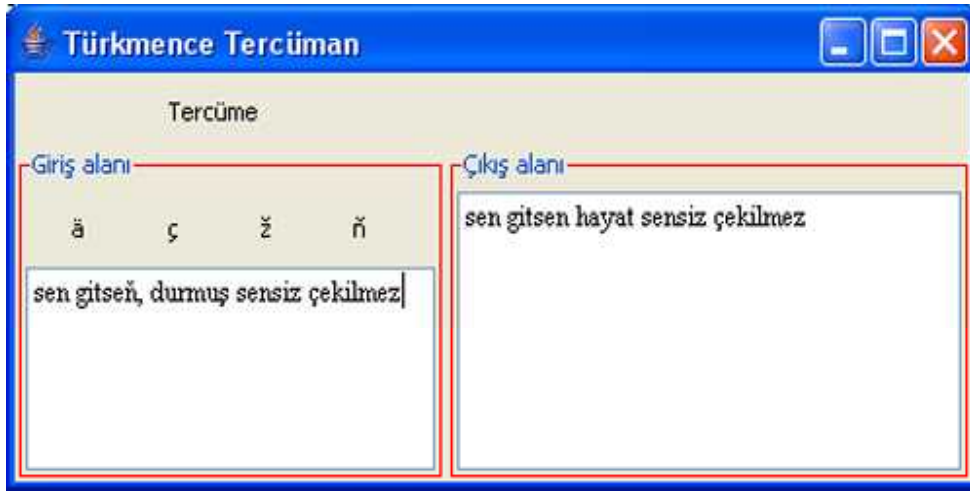
ekTnTr.put(TurkmençeEkAdlari.AT_DUSUM_E, TurkceEkAdlari.ISIM_YONELME_E);

ekTnTr.put(TurkmençeEkAdlari.AT_SADA_BOS, TurkceEkAdlari.ISIM_YALIN_BOS);

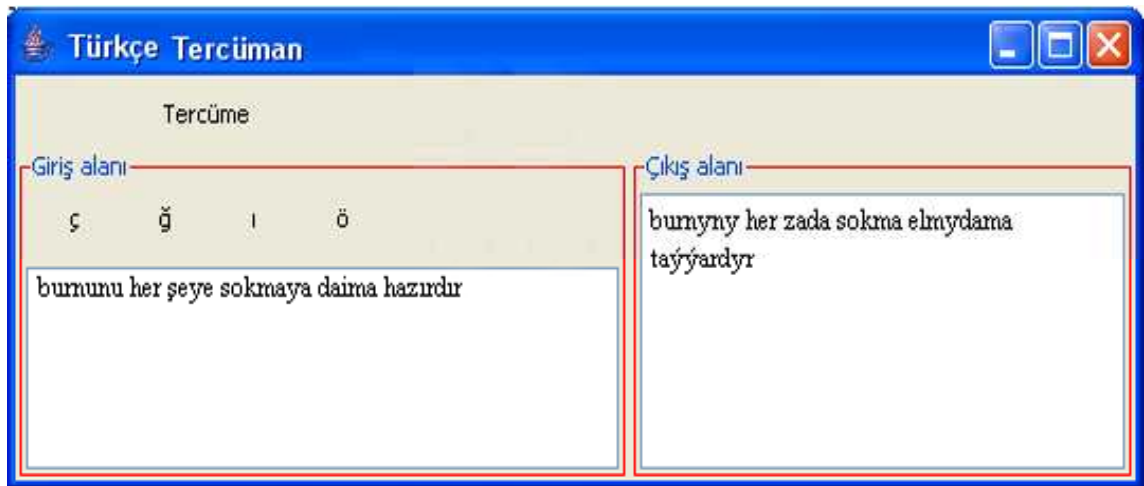
Türkçe çözümlenmiş olan kelimeler, çeviri için Türkmençe çeviri sınıfına kelimeyi gönderir. Türkmençe çeviri sınıfında kelimenin kökü ve ekleri ayrıştırılır. Köke göre sözlükten Türkmençe kök alınır. Eklere göre eklerin çevirileri ekler sözlüğünden alınır.

Türkçe ile Türkmençe'nin eklerinin sıralanışı her zaman aynı değildir. Örneğin Türkçe al+mış+sın+dır kelimesinin Türkmençe karşılığı al+an+dyr+syň olur. Türkçe'de "sın" eki "dır" ekinde önce gelirken, Türkmençe'de tam tersidir. Bu yüzden, çözümlenmiş olan kök ve ekleri, o dilin kendisine has bir sıralama yapılmalıdır. Türkmençe kelime üreticiye gönderilmeden önce, ekler Türkmençe'ye göre sıralanır. Kök ile sıralanmış ekler Türkmençe kelime üreticiye gönderilir. Türkmençe üreticiden gelen sonuçlar çıktı olarak gönderilir.

Aşağıda, basit bir cümlenin çevirisi gösterilmiştir.



Şekil 13. Türkmençe girilmiş bir cümlenin Türkçe çevirisi



Şekil 14. Türkmençe girilmiş bir cümlenin Türkçe çevirisi

2.5. İmla Hatası Düzeltme Modülü

Sohbet ortamında genellikle kısaltılarak ve hızlı yazıldığından, imla hataları yüze çıkar. Buda doğru çeviri olasılığını azaltmaktadır. Herhangi bir şekilde hatalı yazılmış olan bir kelime çözümlenmeden geri döner. Bu durumda, hatalı yazılmış kelimeyi imla denetleyicisine göndeririz.

İmla hatalarını düzeltmek için Levenshtein algoritması kullanılır. Levenshtein algoritmasında s (kaynak) kelime ve t (hedef) kelime olarak tanımlanır. Levenshtein algoritmasını aşağıdaki adımlarla tanımlamam mümkün.

Tablo 18. Levenshtein Algoritması

Adım	Tanımlama
1	n'yi s'nin uzunluğu olarak tanımla m'yi t'nin uzunluğu olarak tanımla Eğer n=0 ise, m'yi döndür ve bitir Eğer m=0 ise, n'yi döndür ve bitir. Başlangıç matrisimiz 0..m satır ve 0..n sütunludur
2	İlk satırı 0..n'ye kadar doldur. İlk sütunu 0..m'ye kadar boldur.
3	s'nin her bir harfini teker teker al. (i=1'den n'e kadar)
4	t'nin her bir harfini teker teker al. (j=1'den m'e kadar)
5	Eğer s[i] eşit t[j], ise maliyeti 0 olur Eğer s[i] t[j]'ye eşit değilse, maliyeti 1 olur
6	Matristeki d[i,j] hücrelerini aşağıdaki durumlardan en küçük olanına eşitle: a.: d[i-1,j] + 1. b. d[i,j-1] + 1. c. d[i-1,j-1] + maliyet.
7	(3, 4, 5, 6) adımları tamamlandıktan sonra d[n,m] hücrelerinde uzaklık belli olur.

Öneri mekanizmasında, öneriler 3 şekilde yapılır.

1. Kökte 1, ekte 1 mesafeye kadar olmak üzere Levenshtein düzeltme mesafesine uyan tüm öneriler
2. Deasciifier'den dönüş değeri olarak gelen öneriler
3. Kelimenin ayrık iki kelimedenden oluşması durumu için öneriler

3. BULGULAR

3.1. Giriş

Türkçe ve Türkmençe'nin alfabelerinde bazı harflerin yazılışları birbirinden farklıdır. Türkçe “ağız” kelimesi Türkmençe “agyž” yazılmaktadır. Türkçe harflerini tanımayan biri Türkçe “ağız” kelimesinin Türkmençe aynı anlamda olan “agyž” kelimesi ile aynı anlamında olduğunu bilemeyebilir. Türkçe ve Türkmençe'nin kökleri aynı olduğundan kelimelerin belli bir oranı birbirine benzerlik göstermektedir. Her iki dilde söylenişi aynı veya çok yakın olup yazılışları farklı olan çok kelime vardır. Bu sadece köklerde değil, aynı zamanda eklerde de vardır. Örneğin: Türkçe çoğul eki olan “lar/ler” eki Türkmençe’de de aynı özellikte kullanılır ve yazılışları aynıdır. Türkçe’deki bilinen geçmiş zaman eki olan “dı/di” eki Türkmençe de kelimeye aynı anlam katmasına rağmen, “dy/di” yazılır.

Türkçe’de ve Türkmençe’de bazı kelimeler söylenişleri aynı olmalarına rağmen zıt anlamda kullanılır. Böyle kelimeler çok az olmalarına rağmen yanlış anlamaya sebep olabilmektedir. Tercüman programı böyle yanlışlıkların düzeltilmesinde büyük rolü vardır. Örneğin: Türkçe “ham” kelimesinin Türkmençe’de insan derisi anlamını vermektedir. Türkmençe “burç” kelimesinin, Türkçe köşe ve biber anlamları vardır.

Türkçe’de kullanılan bazı ekler Türkmençe’de biraz farklı kullanılan veya, hiç kullanılmaya durumları vardır. Genel olarak eklerin sıralanışları aynı olmalarına rağmen bazı yerlerde sıralarında da farklılıklar görülmektedir. Örneğin: Türkçe “görünüyorlardı” kelimesinde bilinen geçmiş zaman eki olan “dı” eki çoğul ekinden sonra gelmektedir. Ama Türkmençe’de bilinen geçmiş zaman “di” eki, çoğul eki olan “ler” ekinden önce gelir: “görünüyördiler”.

3.2. Ekler

Türkçe ve Türkmençe’de, isimlere ve fiillere eklenen ekler farklıdır. Bu çalışma için 29 isim eki ve 38 fiil eki kullanılmıştır. Bu eklerin bir kısmı her iki dilde çok benzerdirler. Bir kısmı da biraz değişikliklere uğramışlardır. Diğer bir kısmı ise, birbirinden tamamen farklıdır. Örneğin : Türkçe’deki git+erek kelimesindeki erek

eki Türkmençe’de tam karşılığı yoktur. Bu yüzden, Türkçe “erek” ekine, Türkmençe en yakın anlam veren “yp” eke karşılık getirilmiştir. Aşağıdaki tabloda Türkçe ve Türkmençe’deki eklerin benzerlik istatistiği verilmiştir:

Tablo 19. Türkçe ve Türkmençe eklerin benzerlik istatistikleri

Aynı Ekler (%)	Benzer Ekler (%)	Benzemeyen Ekler (%)
53	23	24

Aşağıdaki tablolarda isimler ve fiiller için kullanılan eklerin Türkçe ve Türkmençe şekilleri gösterilmiştir. İlk sütun programdaki morfem temsili, ikinci sütun morfem temsile karşı üretilen Türkçe ekler. 3. sütun ise Türkmençe ekleri göstermektedir. Tablolardaki morfem temsillerinin açıklamaları şunlardır :

A: Türkçe için a,e. Türkmençe için a,e, ä harflerinden birini üretebilir. Örneğin: (y)A ile temsil edilmiş ek, *su* kelimesine eklendiğinde *su +yu -> suyu* olur.

H : Türkçe için ı,i,u,ü. Türkmençe için y,i,u,ü harflerinden birini üretebilir. Örneğin: (H)m ile temsil edilmiş ek, *bardak* kelimesine eklendiğinde *bardak + ım -> bardağım* olur.

d: Türkçe için d,t. Türkmençe için d harflerinden birini üretebilir. Örneğin: dA ile temsil edilmiş ek, *ok* kelimesine eklendiğinde *ok+ta->okta* olur

() : son harfin duruma göre gelebilirde gelmeye de bilir. Örneğin : (H)yor ile temsil edilmiş ek, *ağla* kelimesine eklendiğinde, *ağla + yor -> ağlıyor*

Tablo 20. Türkçe ve Türkmençe isim çekim ekleri

Morfemik temsil	Türkçe Ekler	Türkmençe Ekler
-lAr	lar,ler	lar, ler
-(H)m	ım, im, um, üm	ym, im, um, üm
-HmHz	ımız, imiz, umiz, ümiz, mız, miz	ymyz, imiz, umiz, ümiz, myz, miz
-(H)n	ın,in,un,ün,n	yñ, iñ, uñ, üñ, ñ
-(H)nHz	ınız,iniz,unuz,ünüz,nız,niz	yñyz, iñiz, uñyz, üñiz, ñyz, ñiz
-(s)H	ı,i,sı,si	y, i, sy, si
-lArH	ları,leri	lary,leri
-(y)H	yı,yi,yu,yü,i,i,ü,u	y, i, ny, ni

Tablo 20'nin devamı

-(y)A	ya,ye,a,e	a, e, na, ne
-DA	da,de,nda,nde	da, de, nda, nde
-Dan	dan,den,ndan,nden	dan, den, ndan, nden
-lAs	laş,leş	laş,leş
-ki	kı,ki	ki, ky
-dir	dır,dir	dyr, dir
-cHk	cık,cik,cıg,cig	jyk,jik
-cAgHz	çağız	jagaz
-lHk	lık,lik,luk,ük	lyk,lik,luk,lük
-(H)msH	ımsı,umsu,ümsü	ymsy,imsi,umsy,ümsi
-cA	ca,ce	ja,je
-(s)H	sı,si,su,sü,i,i,u,ü	-
-lH	lı,li,lu,lü	ly,li,lu,lü
-lA	la,le	la,le
-sAl	sal,sel	-
-sHz	sız, siz, suz, süz	syz,siz,suz,süz
-(y)Hm	ım,im,um,üm,yım,yim	-
-(y)Hz	ız, iz, uz, üz, yız, yiz	-
-(s)Hn	ın,in,un,ün,sın,sin	-
-(s)HnHz	sınız,siniz,sunuz,sünüz	-
-lAr		-

Tablo 21. Türkçe ve Türkmençe fiil Çekim Ekleri

Morfemik temsil	Türkçe Ekler	Türkmençe Ekler
-Hk	ık,ik,uk,ük	yk, ik, uk, ök
-Hş	ış,iş,uş,üş	yş, iş, uş, üş, ş
-DHr	dır,dir,dur,dür,tır,tir,tur,tür	dyr, dir, dur, dür
-t	t	t
-(H)r	ır,ir,ur,ür,r	yr, ir, ur, ür
-(H)l	ıl,il,ül,ül,l	yl, il, ül, ul, l
-(H)n	ın,in,un,ün,n	yn, in, un, ün, n
-mA	ma,me	ma, me

Tablo 21'in devamı

-(A)r	ar,er,r	ar, er
-(H)yor	ıyor,iyor,yor	ýar, ýär
-DH	dı,di,du,dü	dy, di, du, dü
-mHş	miş,miş,muş,müş	ypdyr, ipdir, ypdy, ipdi
-(y)AcAk	acak,ecek,cak,cek	jak, jek
-(y)A	ya,ye,a,e	a,e,ä,na,ne,nä
-mAlH	malı,meli	maly, meli
-(y)Abil	abil,ebil,bil	
-mAz	maz,mez	maz,mez
-(y)An	an,en,yan,yen	an,en,ýan,ýän
-(H)m	ım,im,um,üm,m	ym,im,um,üm
-HntH	ıntı,inti,untu,üntü	yntı,inti,unti,ünti
-(A)sH	ası,esi	asy,esi
-HcH	ıcı,ici	yjy,iji
-(A)lH	alı,eli	aly,eli
-sHn	sın,sin,sun,sün	syn,sin,sun,sün
-sHnlAr	sınlar,sinler,sunlar,sünler	synlar,sinler,sunlar,sünler
-(H)nHz	ımız,iniz,unuz,ünüz,nız,niz	ňyz,ňiz,ňuz,ňüz
-mAk	mak,mek	mak,mek
-Hp	ıp,ip,up,üp	yp,ip,üp,ip
-dHk	dık,dik,duk,dük,tık,tik	dyk,dik,duk,dük
-sAnA	sana,sene	sana,sene
-sAnHzA	sanıza,senize	saňyzla, seňizle
-(y)sA	sa,se,ysa,yse	sa,se,ýsa,ýse
-(y)ken	ken,yken	kan,kän
-ArAk	arak,erek	yp,ip
-cAs(H)nA	casına,cesine	-
-(H)z	ız,iz,uz,üz,z	ys,is,us,üs,s
-k	k	
-(sH)nHz	sınız,siniz,sunuz,sünüz	syňyz,siňiz,suňyz,siňiz

3.3. Kökler

Eskiden bir Türkçe ve Türkmençe bir dildi. Asırlarca Türkiye Türkleri ile Türkmenistan Türkmenlerinin ayrı coğrafyada ve siyasi sistemde yaşamaları, dillerinin birbirinden uzaklaşmalarına sebep olmuştur. Asırlar geçmesine rağmen hala morfolojik yapı benzerdir ve kelimelerinin %25'i aynı kalmıştır.

Türkçe ile Türkmençe arasındaki kökleri karşılaştırdığımızda 4 sınıfa bölebiliriz. Birincisi her iki dilde de aynı yazılan ve aynı anlamda kullanılan kelimeler. İkincisi, kelimeler çok yakın olmalarına rağmen bazı değişikliklere uğramış kelimeler. Üçüncü sınıfa yazılışları aynı ama, farklı anlamlarda kullanılan kelimeler. Dördüncü sınıfta ise, her iki dilde birbirine hiçbir benzerlikleri olmayan kelimeler. Aşağıdaki tablolarda dört farklı sınıfa örnekler verilmiştir.

Tablo 22. Türkçe - Türkmençe aynı anlamda yazılan ve kullanılan kelimeler

Türkçe	Türkmençe
açgöz	açgöz
açık	açyk
ada	ada
ağır	agyr
ağız	agyz
ak	ak
altı	alty
biz	biz
boş	boş
boz	boz
can	jan

Tablo 23. Türkçe - Türkmençe yakın anlamda yazılan ve kullanılan kelimeler

Türkçe	Türkmençe
çekirge	çekirtge
tercüme	terjime
deli	däli
beş	bäş
diñç	dynç
domuz	doňuz
elma	alma
göl	köl
güç	güýç

Tablo 24. Türkçe - Türkmençe birbirinden farklı yazılan kelimeler

Türkçe	Türkmençe
kavanoz	banka
kavram	düşünje
kedi	pişik
kayısı	erik
kaybet	ýtir
kaybol	azaş
kaygan	taýgançak
kavga	uruş

Tablo 25. Türkçe - Türkmençe zıt anlamda yazılan ve kullanılan kelimeler

Türkçe	Türkmençe	Türkmençe Anlamı
tapmak	tapmak	bulmak
burç	burç	biber, aç
ham	ham	insan derisi
düşmek	düşmek	inmek
sıçan	syçan	fare

Bu çeviri çalışması için Türkçe ve Türkmençe köklerin tutulduğu iki farklı sözlük tutulmuştur. Türkçe için yaklaşık 24,000 kök sözlük, Türkmençe için yaklaşık 3,000 kök sözlük kullanılmıştır. Ayrıca, Türkçe Türkmençe ve Türkmençe Türkçe yaklaşık 3000 kelimedenden oluşan sözlük oluşturulmuştur. Yukarıdaki tablolarda gösterilen örnek kelimelerde oluşturulmuş sözlüklerden alınmıştır. 3000 kökten oluşan sözlükten yararlanılarak bu köklerin istatistikleri çıkartılmıştır. Bu istatistiklerde aşağıdaki tabloda verilmiştir.

Tablo 26. Türkçe ve Türkmençe köklerinin benzerlik istatistiği

Aynı yazılan kökler %	Yakın yazılan kökler %	Yazılışı aynı ama zıt anlamlı kökler %	Diğer kökler %
25	20	1	54

3.4. Çeviri İstatistikleri

Oluşturulmuş çeviri sisteminde bir köke karşılık, birden fazla anlamı olsa bile, tek kelime yazılmıştır. Örneğin Türkçe “*diz*” köküne Türkmençe “*düz*” ve “*diz*” kökleri karşılık gelmektedir. Bu yüzden, yanlış olsa da herhangi birine göre tercüme olur. Türkçe “*diz*” köküne Türkmençe “*düz*” kökü sözlüğe yazılmış olsun. Bu durumda bazı cümleler hatalı bazı cümleler doğru olacaktır. Aşağıda ilki doğru ve ikinci yanlış cümledir.

- Kutuları dizmek sıkıcı (Kitaplary düzmek içgysgynç)
- Futbol oynarken dizimi incittim. (Futbol oynaýarka düzümi aýyrdym)

Türkçe ve Türkmençe çeviri sisteminin testi için cümleleri “doğru”, “anlaşılabilir”, “az anlaşılabilir” ve “yanlış” olmak üzere dörde sınıflandırıldı.

Doğru : Girilen cümle karşı tarafa eksiksiz çeviri yapılmışsa doğru olarak kabul edilir. Bazı cümlelerde, Türkmençe’de veya Türkçe’de az kullanılan kelimeler bulunabilmektedir. Çeviri doğru yapıldığından cümlede biraz düşüklük olsa bile anlamı değiştirmedığından doğru olarak kabul edilmiştir. Aşağıda Türkçe ve Türkmençe doğru kabul edilmiş cümlelere örnekler verilmiştir.

Türkçe : Dokuz yaşında iri saçları olan küçük bir kızdır o

Türkmençe :Dokuz ýaşynda iri saçlary bolan kiçi bir gyzdyr ol

Türkmençe : Kiçiliginden bári surata düşmäni örän söýer

Türkçe : Küçüklüğünden beri resime düşmeyi çok sever.

Anlaşılabilir : Türkçe’de veya Türkmençe’de olan bazı eklerin karşı dilde olmaması veya bazı köklerin karşı dilde tam karşılığı olmaması cümlenin doğruluğunu düşürmektedir. Özellikle birden fazla anlamı olan kelimelerde, hepsine karşılık tek kelime kullanılması cümlenin doğruluğunu etkilemektedir. Aşağıda Türkçe ve Türkmençe anlaşılabilir kabul edilmiş cümlelere örnekler verilmiştir.

Türkçe : Ama yine de ne bakışı ne gülümsemesi Louvre’de gördüğü La Jocondeninkinin aynısı değildi

Türkmençe : Emma yine diýnäme bakysy näme ýylgyrmasy Louvrede gören La Joconde deňi dälidi

Türkmençe : Serdar ýerinden turup gitmekçi boldy

Türkçe :Serdar yerinden kalkıp gitmek oldu.

İlk cümlede Türkçe ne ve demek kökleri Türkmençe çevirisinde cümleye farklı anlam katmıştır. İkinci cümlede ise Türkmençe’de kullanılan “mek+çi” eki Türkçe’de farklı kullanılmaktadır.

Zor Anlaşılır : Bazı durumlarda, bir veya birkaç kelime cümleyi anlaşılmasız hale getirebilmektedir. Bu durumda, cümle zor anlaşılabilir olarak kabul edilir.

Türkçe : Bu daha güzel değil mi sence?

Türkmençe : Bu ýene owadan däl mi sençe?”

Türkmençe : Aýal maşgala näzik bolmaly

Türkçe : Kadın aile narin olmalı

İlk örnekte “*daha*” kelimesinin karşılığı olarak “*has*” yerine “*ýene*” kullanılmıştır. Bu cümlenin anlaşılmasını zorlaştırmaktadır. İkinci cümlede, Türkmençe’de “*Aýal maşgala*” kelimesi bayan anlamındadır. Ama kadın aile olarak tercüme edilmiştir.

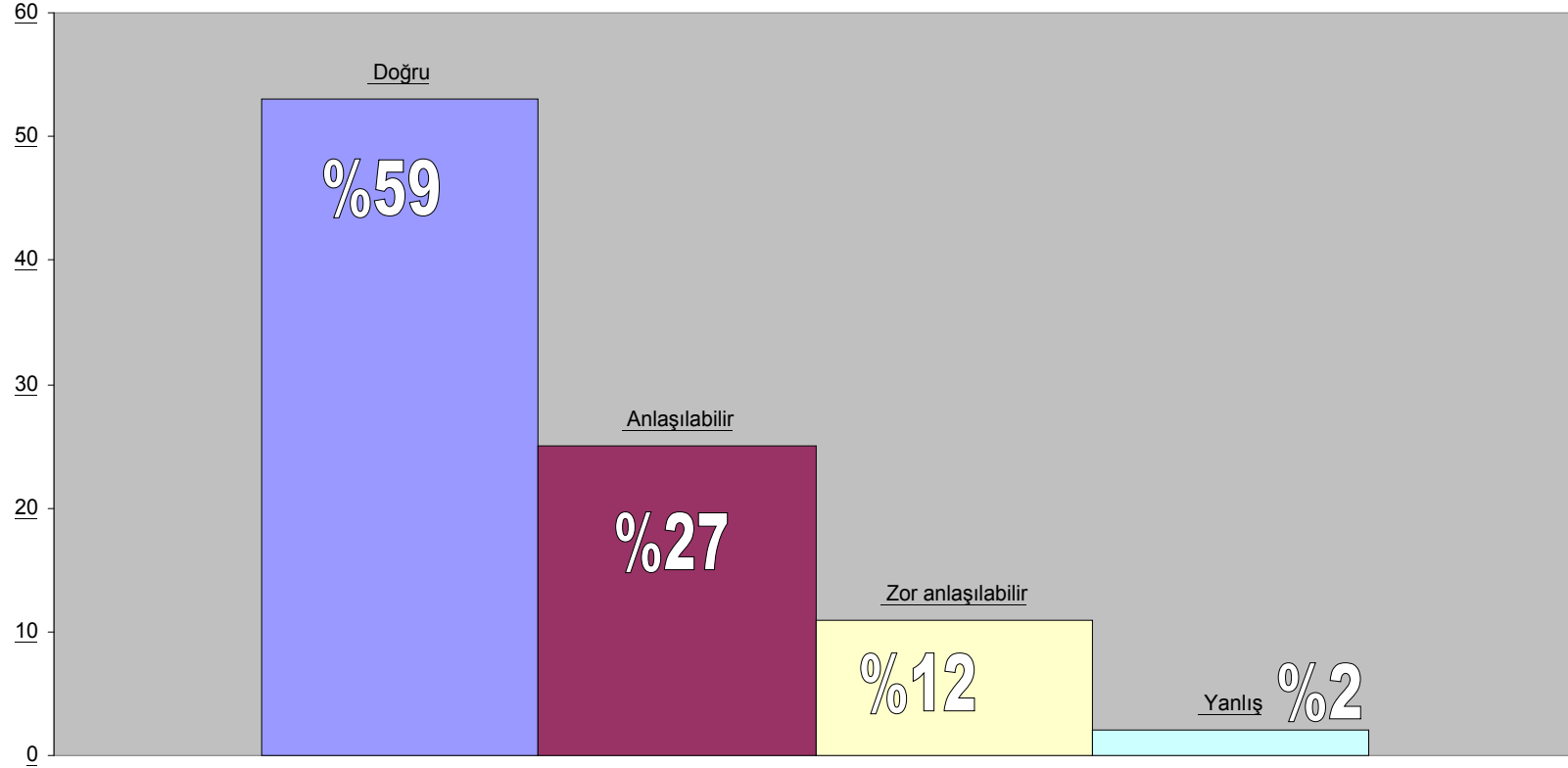
Yanlış : Cümle gerçek anlamından çok farklı bir çeviri yapılmışsa yanlış cümle olarak kabul edilir. Aşağıdaki örnekte Türkçe artık kelimesinin “içildikten, yenildikten veya kullanıldıktan sonra geriye kalan” Türkmençe anlamı kullanılmıştır. Bu da yanlış olmaktadır.

Türkçe : *Artık alıştım*

Türkmençe : *uwundy ögrenişdim*

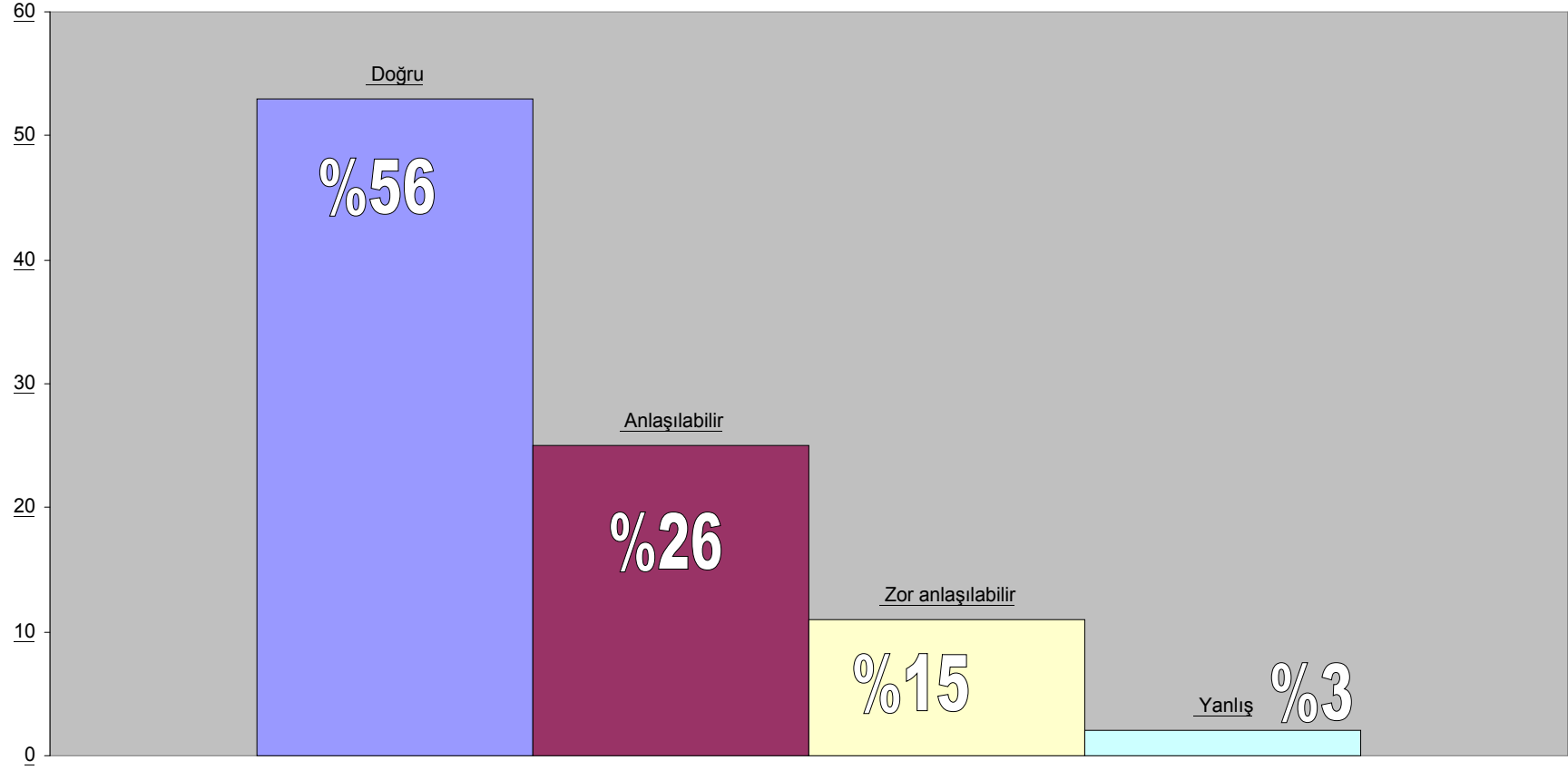
Aşağıdaki tabloda Türkçe ve Türkmençe için çevrisi istatistiği verilmiştir:

Türkçe ve Türkmençe Çeviri İstatistiği Grafiği



Şekil 15. Türkçe'den ve Türkmençe'ye Çeviri İstatistiği

Türkmençe ve Türkçe Çeviri İstatistiği Grafiği



Şekil 16: Türkmençe'den ve Türkçe'ye Çeviri İstatistiği

4. SONUÇLAR VE ÖNERİLER

Bu yüksek lisans tez çalışmasında, Türkçe ve Türkmençe için bilgisayarlı çeviri sistemi tanıtıldı. Türkiye Türkleri ile Türkmenistan Türkmenleri herhangi bir eğitimden geçmeden birbirlerini zor anlayabilmektedirler. Bu çalışma, Türkçe veya Türkmençe bilen kişilerin birbirleri ile Internet tabanlı sohbet etmelerini amaçlamış olsa bile, o dili öğrenen kişi için de yardımcı olacağına inanılmaktadır. Bu çalışmadaki metotları diğer Türk dilleri içinde uygulanabilir.

Geliştirilmiş çevirisi sistemi, Türkçe veya Türkmençe cümleyi alır ve morfolojik analizci ile kelimeleri teker teker köklere ve eklere ayrıştırır. Morfolojik ayrıştırılmış kelimenin köklerin ve eklerin karşı anlamını sözlükten çeker. Sonra, kökleri ve ekleri kelime üreticiye gönderilir. Oluşan kelimeyi sonuca gönderilir.

Türkçe bir cümlenin Türkmençe'ye çevirisinde veya Türkmençe bir cümlenin Türkçe'ye çevirisinde anlaşılabilirliği iki dil arasındaki farklılıklar ile ters orantılıdır. Eğer cümlede morfolojik olarak ne kadar az fark varsa anlaşılabilirlikte o kadar artar. Sözlükte bir köke karşılık tek bir kök tutulmuştur. Ama gerçek hayatta, bir köke, birden fazla kök karşılık gelir. Böyle durumlarda cümlede belirsizlik veya anlam eksikliği yüze çıkmaktadır. Genel olarak her iki dildeki belirsizlikler benzerlik gösterse bile, köklerdeki ve eklerdeki belirsizlikleri giderici bir sistem ile çok daha başarılı bir çeviri sistemi geliştirilebilir.

Türk ve Türkmen dillerinin gramer yapıları benzerlik gösterse bile, çeviri sistemi geliştirmek kolay bir iş değil. Diğer bir deyişle, aradan geçen asırlarca zaman her iki dili birbirinden uzaklaştırmıştır. Bu uzaklaşma daha çok köklerde olmuştur.

Çalışma, ilkokul kitaplarından test edilmiş ve çok iyi sonuçlar elde edilmiştir. İlkokul kitaplarında birleşik kelimeler ve birleşik kipler az kullanıldığından test sonuçlarının başarısını artırmıştır. Her ne kadar bu çeviri sistemi kullanışlı olsa bile daha yapılması gereken çok iş vardır. Çeviri kuralları kök ve eklere başarılı uygulansa bile 3000 kök sözlüğü çok yetersiz kalmaktadır. Birleşik kelimeleri de bulduran, birden fazla çözümün en uygununu seçebilen bir sistem geliştirilmesi gerektir. Bu çalışmadaki morfolojik analizden yararlanılarak, sözdizimi ve anlamsal analiz geliştirilebilir. Geleceki çalışma planı, bu çeviri sisteminin sözdizimi ve anlamsal analizi yapan sisteme geliştirilmesidir!

5. KAYNAKLAR

1. Kubon, V. ve Hajic, J., Machine Translation of Very Closed Languages, ANLP NAACL 2000, (2000) 12-15.
2. Oflazer, K., Say, B., Hakkani, T., Dilek, Z., ve Tür G., Building a Turkish Treebank, chapter in Building and Using Parsed Corpora, Kluwer Academic Publishers, (2003).
3. Tomas, J., Lloret, J., Casacuberta, F., Phrase-based alignment models for statistical machine translation, Lecture Notes In Computer Science, (2005) 605-613.
4. Canals, R., Esteve, A. ve Garrido, A., A Spanish-Catalan Machine Translation System, Machine Translation Review, inter NOSTRUM, 11 (2000) 21-25.
5. Canals, R., Esteve, A. ve Garrido, A., A Spanish-Catalan Machine Translation System, Machine Translation Review, inter NOSTRUM, 11 (2000) 21-25
6. Kara, M., Türkmençe, Kültür Bakanlığı Yayınları, Ankara, 2000.
7. Li, H., Japkowicz, N. ve Barriere, C., English to Chinese translation of prepositions, Lecture Notes In Computer Science, 3501 (2005) 412-416.
8. Nabiyev V, Vasif, Yapay Zeka: Problemler-Yöntemler-Algoritmalar, 2. Baskı, Seçkin Yayınevi, Ankara, 2005.
9. Wang, X. ve Ren, F., Chinese-Japanese clause alignment, Lecture Notes In Computer Science, 3406 (2005) 400-412.
10. Altıntaş, K. ve Çiçekli, İ., A machine Translation System Between a Pair of Closely Related Languages, Proceedings of the 17th International Symposium on Computer and Information Sciences--ISCIS 2002, (2002) 192-196.
11. Allen, J., Natural Language Understanding, The Benjamin/Cummings Publishing Inc., Redwood City, California, 1995.
12. Söyegow, M., Saryhanow, M., Türkmen Diliniň Grammatikasy Morfologiýa, Ruh, Aşgabat, 1999.
13. Hengirmen, M., Türkçe Dilbilgisi, Engin Yayınevi, Ankara, 2002.
14. Hakkani, D. Z., Oflazer K. ve Tür, G., Statistical morphological disambiguation for agglutinative languages. In Proceedings of co LING 2000 - ICCL, (2000).
15. Oh, J. ve Choi, K., Machine learning based English-to-Korean transliteration using grapheme and phoneme information, IEICE Transaction on Information Systems E88D, (2005) 1737-1748.

16. Hakkani, D., Oflazer, K. ve Türk, G. Statistical morphological disambiguation for agglutinative languages. Computers and the Humanities, 4 (2002) 36-42.
17. Altıntaş, K., ve Cicekli, İ., A Morphological Analyzer for Crimean Tatar, Proceedings of Turkish Artificial Intelligence and Neural Network Conference, 7 (2001) 156-166.
18. Atalay, B., Oflazer, K. ve Say, B., The Annotation in The Turkish Treebank, Proceedings of the EACL Workshop on Linguistically Interpreted Corpora - LINC, (2003) 13-14.
19. Boguslavsky, I., Iomdin, L., ve Lazursky, A., Interactive resolution of intrinsic and translational ambiguity in a machine translation system, Lecture Notes In Computer Science, 3406 (2005) 388-399.
20. Çiçekli, I. ve Güvenir, A., Learning Translation Templates from Bilingual Translation Examples, Applied Intelligence, 15 (2001) 57-76.
21. URL-1, <http://zemberek.dev.java.net>, Zemberek projesi, 2006.
22. Turban, E., Expert Systems and Applied Artificial Intelligence, Prentice Hall Inc., New Jersey, 1992.
23. Gibbon, D., Paula, A., Simoes, Q. ve Matthiesen, M., An optimized FS pronunciation resource generator for highly inflecting languages, Proceedings of LREC, (2000).
24. Russel, S. ve Norvig, P., Artificial Intelligence, A Modern Approach, Prentice Hall, 2nd Edition, 2003.
25. Hanser, O., Türkmençe Elkitabı, Yıldız Teknik Üniversitesi Basımevi, İstanbul, 2003.
26. Hinrichs, E. ve Simov, K., Treebanks and Linguistic Theories, Proceedings of the First Workshop, (2002) 24-32.
27. Kara, M., Türkmen Türkçesi Grameri, Gazi Kitapevi, Ankara, 2005.
28. Nabiyev, V., Vasif., Yazıcı, R. ve Ulutaş, M., Akriba diller için bilgisayar destekli çeviri sistemi, 9. Türk zeka ve sinir ağları sempozyumu, (2000) 393-397.
29. Nabiyev, V., Vasif., ve Sayan, N., Doğal dillerin analizinde hipotez mantığına dayalı karar verme, V. Bilgisayar-haberleşme sempozyumu, (1998).
30. Oflazer, K., Two-level Description of Turkish Morphology, Literary and Linguistic Computing, 9,2 (1994) 137-148.
31. Sary, B. ve Güder, N., Türkmençe'nin Grammeri (II Morfolojiya: Şekil Bilgisi) MYB, Ankara, 1998.

32. Say, B., Say, Z., D., Oflazer, K., ve Özge, U., Development of a corpus and a treebank for present-day written Turkish, 11th International Conference on Turkish Linguistics, 4,3 (2002) 201-213.
33. URL-2, <http://www.xrce.xerox.com/research/mltt/tSt/home.-en.html>, 2006.
34. Solak, A. ve Oflazer, K., Parsing agglutinative word structures and its application to spelling checking for Turkish, In Proceeding of the 16th International conference on Computational Linguistic, 1 (1992) 39 – 46.

ÖZGEÇMİŞ

Guychmyrat AMANMYRADOV, 15.08.1980 tarihinde Türkmenistan'ın Mary ilinde doğdu. İlköğretim öğrenimini Vekilbazar ilçesinin 34. Orta Mektebinde, lise öğrenimini Bayramali Türkmen Türk Kolejinde tamamladı. 1997 yılında, Karadeniz Teknik Üniversitesi Fen Bilimleri Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü'nü kazandı ve 2001 yılında mezun oldu. 2002 yılının Şubat ayında Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda yüksek lisans öğrenimine başladı. Ana dili Türkmençe ve Türkçe, İngilizce ve Rusça dillerini bilmektedir.