

Türkçe ve Kırım Tatarcası Arasında Bir Çeviri Sistemi

İlyas Çiçekli

Bilgisayar Mühendisliği Bölümü
Bilkent Üniversitesi
Bilkent 06800, Ankara

ilyas@cs.bilkent.edu.tr

Özet

Yapıları birbirlerine yakın olan diller arasında makina ile bir tercüme sistemi geliştirmek, yapıları farklı olan iki dil arasındakına göre daha kolaydır. Gramerleri ve kelimeleri arasındaki benzerlik, aralarında bir tercüme sistemi geliştirilmesi için harcanması gereken çabayı azaltır. Sadece morfolojik analizcileri, duruma dayalı tercüme sözlükleri ve basit tercüme kural öbekleri kullanan bir tercüme sistemi çoğu zaman yeterli olacaktır. Genelde bir anlamsal analizce gerekmede, sözdizim derleyicilerine ihtiyaç olabilir. Bu makalede Türkçe'den Kırım Tatarca'sına tercüme yapabilen ve sonlu durum yöntemleri kullanan bir tercüme sistemi anlatılmıştır. Türkçe ve Tatarca arasında böyle bir tercüme sistemi geliştirerek, benzer yapılardaki diller içinde prototip bir tercüme sistemi sunmak istedik. Geliştirilen tercüme sistemi bir Türkçe cümleyi girdi olarak alır, o cümledeki bütün kelimeleri bir Türkçe morfolojik analizci kullanarak analiz eder, o cümledeki gramer yapılarını ve duruma bağlı yapıları tercüme kuralları kullanarak Tatarca'ya aktarır, kök kelimeleri Tatarca'ya çevirir, en son olarak ta Tatarca kelimelerin normal yazılış halini bulmak içinde bir Tatarca morfolojik üretici kullanır. Çoğu zaman sistem tarafında üretilen tercüme sonuçlarından biri verilen Türkçe cümlelerin doğru Tatarca tercümesidir.

Anahtar Kelimeler: Doğal Dil İşleme, Makina İle Tercüme, Sonlu Durum Yöntemleri

1. Giriş

Genelde başka bir dildeki sözlü ve yazılı dökümanları anlamada, tercümanlar diğer insanlara yardımcı olmaktadır. Ama her zaman bize yardımcı olacak bir tercüman bulmak mümkün olmayabilir, vede bir tercümanın bize masrafı çok olabilir. Bir dilden diğer dile tercüme edilmesi gereken dökümanlar miktarı çok hızlı olarak artmaktadır, ama yeterli miktarda tercüman bulmakta problem olmaktadır. Bu yüzden şirketler ve insanlar tercüme işlemleri yaptırmanın alternatif yollarını aramaktadırlar. Bu alternatif yollardan biride dökümanların makina ile tercüme sistemleri kullanılarak yapılmasıdır.

Makina ile tercüme alanında çalışmaları genelde İngilizce ve Fransızca gibi batı dilleri üzerine yoğunlaşmıştır. Diğer diller için de çalışmalar yapıldığında, genelde tercüme sistemleri o dille İngilizce arasında geliştirilmeye çalışılmıştır. Yapıları birbirlerine yakın diller arasında tercüme sistemleri üzerinde fazla çalışmamıştır. Türkçe ve Türkük diller ise bu konuda pek fazla ilgi çekmemiştir.

Bu makalede yapıları birbirlerine yakın olan diller arasında sonlu durum yöntemlerine dayalı tercüme sistemlerini tartıştık. Yapıları birbirine yakın diller arasında tercüme sistemleri geliştirmek, yapıları farklı olan

dillere göre biraz daha kolaydır. Tercüme sırasında problemlerin bir kısmı, bu tür tercüme sistemlerinde çalışırken karşımıza çıkmayabilir. Böylece diğer problemlere yoğunlaşmak mümkün olacaktır.

Bu makalede yapıları birbirlerine yakın olan Türkçe ile Kırım Tatarca'sı arasında geliştirilen makina ile bir tercüme sistemini sunmaktayız. Bu iki dildeki kelimelerin büyük bir kısmı aynı, vede gramerleri de birbirlerine çok benzerdir. Gramerlerindeki farklılıkların çoğu bazı gramer yapıları için farklı morfemlerin kullanılmasıdır. Kök kelimelerin çoğu benzerlik gösterir. Geliştirilen tercüme sisteminde sonlu durum teknikleri kullanılmıştır. Bu proje kapsamında geliştirilen yöntemlerin Türkçe ile diğer Türkük diller arasında da uygulanabileceği kanısındayız. Bizim kullandığımız benzer yöntemlerin Çek ve Slovak dilleri arasındaki bir sistemde [5] vede İspanyolca ve Katalan dilleri arasındaki diğer bir sistemde [2] kullanıldığını gözlemlemekteyiz.

Makalenin geri kalan kısmında geliştirdiğimiz tercüme sisteminin detayları bulunmaktadır. 2. bölümde yapıları birbirlerine yakın iki dil arasındaki tercüme işleminin neler gerektirdiği konusu tartışılmıştır. Türkçe'den Tatarca'ya tercüme yapan sistemin detayları 3. bölümde bulunmaktadır. Sistemin nasıl uygulandığı 4. bölümde ve sistemin çalışmasını açıklayan bir örnek 5. bölümde verilmiştir. 6. bölümde sistemimizin sınırlamaları, son bölümde de sonuçlar ve gelecekte yapmayı planladığımız işler sunulmuştur.

2. Yapıları Birbirlerine Yakın Diller Arasında Tercüme

Tercüme birçok değişik sebepten dolayı genel olarak zor iştir. Her topluğun kendine göre kültürleri vardır. Her topluluğun kendisine göre kavramları ve nesnelere verdikleri isimler vardır, vede bunlar çok farklılıklar gösterebilirler. Bazı dillerde olan gramer yapısı diğer dillerde olmayabilir. Örneğin Türkçe'de özel olarak İngilizce'deki "perfect tense" in bir karşılığı yoktur, İngilizce'deki "perfect tense" deki bir cümleyi Türkçe'ye çevirmek problemlere neden olabilir. Tercüme konusundaki en önemli zorluklardan biri, dillerin yapılarındaki belirsizliklerdir. Bir kelimenin birden fazla anlamı olabileceğinden, kelimenin o cümledeki doğru anlamını seçip ve o anlamına denk düşen diğer dildeki karşılığını bulmak zor bir iştir.

İki dilin yapıları ne kadar farklı olursa, tercüme sırasında çıkan problemler o kadar çok olur. Yapıları birbirlerine yakın olan diller arasındaki tercüme sırasında ise farklı diller arasında çıkan problemlerin bir kısmı gözükmez. Genelde dilleri birbirine yakın olan iki topluluğun, kültürleri ve tarihi geçmişleride birbirlerine yakındır. Böyle iki topluluğun kültürleri arasındaki farklılıklarda fazla değildir. Bu iki topluluk arasında genelde yoğun bir iletişim olduğundan, birinin kültüründe olan kavramların diğer kültürde tam karşılığının olma olasılığı çok yüksektir. İki dil ne kadar yakınsa, gramerleri arasındaki farklılıklar o kadar az olacak vede bir dilde kelimenin diğer dilde tam bir karşılığının olmama olasılığı azalacaktır. Bazı belirsizlikler iki dilde mevcut olabileceğinden, bazı belirsizlikler tercüme sırasında aynen korunabilir. Örneğin, "John saw the girl with binoculars" İngilizce cümlesinde dürbünün (binoculars) kızamı (girl) ait olduğu, yoksa John'un kızı bir dürbünlemi gördüğü belirli değildir. Bu cümleyi Türkçe'ye çevirirken bu belirsizlik giderilmelidir, çünkü bu iki değişik anlam Türkçe'de farklı yapılara denk düşecektir. Ama aynı cümle Fransızca'ya bu belirsizlik giderilmeden tercüme edilebilir, çünkü aynı belirsizlik bu cümleye denk düşen Fransızca cümlede de aynen mevcut olacaktır [4]. Sonuç olarak, genelde benzer iki dil arasında tercüme işlemi daha kolay olacaktır.

Araştırmacılar şimdiye kadar genel olarak birbiri ile fazla ilişkisi olmayan diller arasındaki tercüme sistemlerinde çalıştılar. Ama diğer yönden yapıları birbirlerine yakın diller arasındaki tercüme sistemleride önemlidir. Birbirine yakın diller arasında tercümede bazı problemler olmayacağından, böyle iki dil arasındaki bir tercüme sisteminde çalışırken diğer kalan problemleri çözmeye yoğunlaşılacaktır. Örneğin, Türkçe ve İngilizce arasında bir tercüme sisteminde çalışırken, Türkçe'nin bağımsız kelime sırası kullanan bir dil, ama İngilizce'nin bağımsız kelime sırası kullanan bir dil olmamasından dolayı kelime sırası problemine diğer problemler ile birlikte atak etmeliyiz. Ama Türkçe ve Tatarca ikisi birden bağımsız kelime sırası kullanan diller olduğundan, kelime sırası problemi ile uğraşmadan diğer problemlerin çözümüne odaklanabiliriz.

Yapısal olarak yakın iki dil arasındaki tercüme sistemlerinde, bir anlamsal analiz yapmadan bazı tecüme kuralları ve morfolojik analiz yapan araçlar kullanmak çoğu zaman yeterli olabilir. Böyle bir sistemde kaynak

dil için morfolojik analizci ve morfolojik belirsizlik gidericiye, iki dil arasında bir tercüme sözlüğüne ve tercüme kuralları öbeğine, vede kaynak dil için bir morfolojik üreticiye gerek olacaktır. Genelde çoğu tercüme için kaynak dil için bir sözdizim analizci kullanmadan yapılabilsen de, kapsamlı bir tercüme sistemi için bir sözdizim analizcisine de ihtiyaç olacaktır.

Literatürde yapıları birbirlerine çok yakın diller arasındaki tercüme sistemlerinde çalışılmıştır [2,5]. Çek ve Slovakca arasındaki tercüme sisteminde [5] tercüme hafızası tekniği kullanılmıştır. Tercüme hafızası tekniğine göre önceki tercüme örnekleri tercüme hafızasında saklanır, bir tercüman (bir insan) yeni bir cümleyi tercüme etmeye başlarken, sistem tercüme hafızasında bu cümlenin olup olmadığını kontrol eder, eğer varsa bu tercüme önerir. Tercüman bu öneriyi kabul edip etmeme hakkına sahiptir. İspanyolca ve Katalan dilleri arasındaki tercüme sisteminde [2] kullanılan yöntem bizim kullandığımız yöntemle benzer. Onların sistemi ile bizim sistemimiz arasındaki temel farklılık, onlar kelimelerin tercümelerini yapacak olan kuralları gramer tercüme kurallarından önce uyguluyor olmalarıdır.

3. Tercüme Sistemi

Türkçe'den Kırım Tatarca'sına tercüme yapan sistem genel olarak bir Türkçe kelimeyi karşılığı olan Tatarca kelimeye çevirerek yapan bir tercüme sistemidir. Bu iki dilin grameri birbirlerine çok yakın olduklarından, çoğunlukla bir dildeki morfem diğer dildeki bir morfeme biraz değişikliğe uğrayıp veya hiç bir değişikliğe uğramadan denk düşer. Çoğunlukla da bir dildeki belirsizlikler aynen diğer dilde de korunur. Genel olarak birbirine yakın olan iki dil arasındaki bir tercüme sisteminde sonlu durum dönüştürücülerin kullanılması yeterli olacaktır. Burada sonlu durum dönüştürücülerin görevleri gramer yapılarını, duruma bağlı yapıları ve kelimeleri bir dilden diğerine aktarmaktır.

Tercüme sistemindeki temel adımlar aşağıdaki gibidir:

1. Türkçe metnin morfolojik analizi
2. Türkçe kelimelerdeki morfolojik belirsizliklerin giderilmesi
3. Gramer yapılarını ve duruma bağlı yapıları aktaracak olan tercüme kurallarının uygulanması
4. Kelimelerin bire-bir tercümesini yapan kuralların uygulanması
5. Tatarca kelimelere Tatarca morfolojik üreticinin uygulanması

Verilen Türkçe metin morfolojik olarak analiz edildikten sonra, kelimelerin morfolojik yapısındaki belirsizlikler giderilmelidir. Bir sonraki aşamada morfolojik belirsizliklerin giderildiği Türkçe metindeki duruma bağlı yapılar ve deyimler Tatarca'daki yapıya çevrilirler. Türkçe kelimeleri bire-bir Tatarca kelimelere çevirmeden önce, bir kelimedenden fazla oluşan yapıların ve karşılığı önceki veya sonraki kelimelere dayalı olarak kelimelerin çevrilme işlemi yapılır. Böylece durum bilgileri kaybolmadan önce, duruma bağlı çevirme işlemleri yapılmış olacaktır. Daha sonra Türkçe ve Tatarca arasındaki bir sözlük kullanılarak, Türkçe'deki kelimeler bire-bir Tatarca'ya aktarılır. Son aşamada ise, Tatarca'ya çevrilmiş olan morfolojik seviyedeki kelimelerden oluşan metindeki bütün kelimeler bir Tatarca morfolojik üretici tarafından onların normal yazılış şekline çevrilir.

Türkçe morfolojik analizci [7] bir Türkçe kelimeyi alır ve o kelimenin bütün olası morfolojik yapılarını geri döndürür. Bu işlemi yaparken o kelimenin cümle içindeki konum bilgisini kullanmaz. Örneğin Türkçe morfolojik analizci "evlerimizden" kelimesini aldığı anda, o kelimenin morfolojik yapısını ifade eden "ev+Noun+A3PL+P1PL+ABL" sonucu geri döndürür. Bu morfolojik yapı kelimenin kökünün "ev", türünün isim, çoğul halde, sahiplik ekinin birinci çoğul şahıs vede ismin den-halinin kullanıldığını belirtir.

Türkçe morfolojik analizci bir kelime için birden fazla morfolojik yapı geri döndürebilir. Ama o kelimenin cümledeki kullanıma göre bu geri döndürülen morfolojik yapılardan sadece biri o kelimenin verilen cümledeki kullanıma denk düşer. Bir Türkçe morfolojik belirsizlik gidericinin görevi birden fazla morfolojik yapısı olan bir kelimenin, cümledeki kullanım şekline göre o kelime için doğru morfolojik yapıyı seçerek morfolojik belirsizliği gidermektir.

Tercüme sistemin 3. ve 4. adımlarında Türkçe yapıları Tatarca yapılara çeviren tercüme kuralları XEROX Sonlu Sistem Araçları (XEROX Finite State Tools -- XFST) [6] kullanılarak uygulanmıştır. Bu yüzden tercüme kurallarını ifade ederken XFST sözdizimi kurallarını kullanacağız. Tercüme kuralları temel olarak duruma bağlı olarak değişiklik yapan kurallardır. Türkçe'deki bir yapı, o yapının cümledeki konumuna göre bir Tatarca yapı ile değiştirilir. XFST sözdizim kuralları kullanılarak bir tercüme kuralı aşağıdaki gibi ifade edilir:

[kaynak -> hedef || SolDurum _ SağDurum]

Bu kurala göre, eğer "kaynak" verilen durumda gözükyorsa, "hedef" ile değiştirilir. Burada SolDurum kaynaktan hemen önce görülen yapıyı, SağDurum hemen sonra görülen yapıyı ve alt-çizgi karakteri de kaynağın pozisyonu gösterir. Kurallarda durum bilgisi şart değildir. Eğer bir kuralda durum bilgisi verilmemiş ise, kaynağın her zaman hedef olarak değiştirileceğini ifade eder.

Tercüme sisteminde kullanılan kuralları kelimeler üzerindeki yaptıkları değişikliklere göre 8 kategoriye ayırdık. Bu kategoriler aşağıda açıklanmıştır.

1. Hiç Değişiklik Yapılmaması

Türkçe metindeki bazı kelimeler arkalarından gelen morfepler ile birlikte hiç bir değişikliğe uğramadan Tatarca'da aynen kullanılabilirler. Bu durumda hiç bir tercüme kuralı bu kelimelere uygulanmaz, ve kelime morfepleri ile birlikte aynen Tatarca'ya aktarılmış olur.

2. Kök Kelime Değişikliği

Bazı durumlarda Türkçe'deki kök kelime Tatarca'daki bir kök kelime ile değiştirilir, ama o kelimenin kök kelimesi bir değişikliğe uğramaz. Bu kurallar temel olarak Türkçe ve Tatarca arasındaki sözlükten alınan kurallardır. Örneğin

[çocuk %+ Noun -> bala %+ Noun]

kuralı çocuk isminin Tatarca'ya bala olarak çevrileceğini gösterir.

2. Morfem Değişikliği

Bazı durumlarda Türkçe'deki morfem Tatarca'daki bir morfem ile değiştirilir, ama o kelimenin kök kelimesi bir değişikliğe uğramaz. Örneğin

[FeelLike -> FutPart]

kuralı Türkçe'deki FeelLike morfeminin Tatarca'ya FutPart morfemi olarak çevrileceğini gösterir.

4. Kök Kelime ve Morfem Değişikliği

Bazı durumlarda Türkçe'deki kök kelime Tatarca'daki bir kök kelime ile değiştirilirken, o kelimeyi takip eden bazı morfeplerde bir değişikliğe uğrarlar. Örneğin

[sakın %+ Adj %^ DB %+ Verb %+ Become -> tın %+ Verb]

kuralı "sakinleş" kelimesinin Tatarca'ya "tın" fiili olarak çevrileceğini gösterir.

5. Nesnelere Etkileyen Fiiller

Bazı fiillerin nesnelere Türkçe ve Tatarca'da ismin farklı halleri ile kullanılır. Örneğin "sormak" fiilinin nesnesi Türkçe'de ismin e-hali (dative) ile kullanılırken, Tatarca'da ismin den-hali (ablative) ile kullanılır.

6. Önceki ve Sonraki Kelimeleri Etkileyen Gramer Yapıları

Bu kategorideki kurallar önceki ve sonraki kelimelerin yapılarını etkiler. Örneğin Türkçe'deki fiilere eklenen –dik morfemi Tatarca'da –gen morfemine denk düşer, vede Türkçe'de –dik morfemi almış bir fiile bir sahiplik morfemi eklenirken, Tatarca'da sahiplik morfemi –gen morfemi almış bir fiili izleyen isim eklenir. Böylece Türkçe'deki "oturduğum yer" Tatarca'ya "oturğan yerim" olarak tercüme edilir.

7. Birden Fazla Kelimenin Bir Kelime İle Değiştirilmesi

Bazen Türkçe'deki birden fazla kelime Tatarca'da bir kelimeye karşılık gelir. Örneğin Türkçe'deki "şarkı söylemek" Tatarca'da "yırlamak" ile ifade edilir.

8. Bir Kelimenin Birden Fazla Kelime İle Değiştirilmesi

Bazen Türkçe'deki bir kelime Tatarca'da birden fazla kelimeye karşılık gelir. Örneğin Türkçe'deki "sunmak" Tatarca'da "taqdim etmek" ile ifade edilir.

Tercüme sisteminde yukarıdaki kategorilerde olan bir çok tercüme kuralı bulunur. Normalde bu kuralların hangi sırada uygulanacağı fazla önemli olmasada, kök kelime değişikliği yapan kurallar en son uygulanır. Sıranın önemli olduğu durumlarda, tercüme kurallarının sırası o duruma göre ayarlanmıştır. Bazen birden fazla kural paralel olarak uygulanır.

Bütün tercüme kuralları uygulandıktan sonra, elimizde verilen Türkçe metine karşılık düşen ve Tatarca kelimelerin morfolojik yapıda olduğu bir Tatarca metin olacaktır. Son adımda morfolojik yapıda olan Tatarca kelimeler bir Tatarca Morfolojik İşlemci [1] kullanılarak normal yazım hallerine getirilir. Bu tercüme sisteminde kullanılan Tatarca Morfolojik İşlemci hem Tatarca kelimelerin analizinde, hem de morfolojik yapıları verilen kelimelerin üretilmesinde kullanılabilir.

4. Uygulama

Tercüme sistemi XEROX Sonlu Durum Araçları (XEROX Finite State Tools - XFST) kullanılarak uygulanmıştır. XFST Xerox tarafından sonlu durum ağları ile hesaplama yapmak için geliştirilmiş genel amaçlı bir sistemdir. XFST kullanılarak sonlu durum makinaları ve dönüştürücüleri verilen düzenli ifadelerden (regular expressions) ve iki-seviyeli kurallardan değişik işlemler kullanılarak yaratılır. Yaratılan sonlu durum makinaları ve dönüştürücüleri kütüklerde saklanabilir. XFST içindeki TWOLC sistemi verilen iki-seviyeli kurallardan deterministik ve en küçük boyuttaki sonlu durum dönüştürücülerini yaratan bir derleyicidir. XFST içindeki LEXC sistemi ise sözlük ve sözlük dönüştürücüleri yaratılmasında kullanılan bir derleyicidir ve TWOLC tarafından üretilen dönüştürücülerle birlikte kullanılır. Biz bu XFST derleyicilerini iki-seviyeli kurallar olarak düzenlenmiş olan tercüme kurallarının derlenmesinde vede Tatarca morfolojik işlemcinin geliştirilmesinde kullandık.

Tercüme sisteminin arayüzü Java programlama dili kullanılarak gerçekleştirilmiştir. Bu arayüz bir Türkçe metni bir kütükten okur ve kelimelerine ayrıştırır. Kelimeler organize edilerek bir dış uygulama olarak çalışan Xerox derleyicileri kullanılarak yaratılan dönüştürücülere girdi olarak verilir. Her bir dönüştürücünün çıktısı düzenlenerek diğer bir dönüştürücünün girdisi olarak kullanılır. Son dönüştürücünün çıktısı kullanıcıya tercümenin sonucu olarak sunulur.

Tercüme sistemi tarafından girdi olarak alınan her Türkçe cümle, ilk önce kelimelerine ayrılır. Her Türkçe kelime Türkçe morfolojik analizciye verilerek, morfolojik yapısı bulunur. Her kelimenin birden fazla morfolojik analizi olabilir. Cümledeki bütün kelimelerin olası morfolojik yapıları kullanılarak kelimeleri morfolojik seviyede olan değişik Türkçe cümleler elde edilir. Kelimeleri morfolojik seviyede olan her cümle tercüme kurallarını uygulayacak olan sonlu durum dönüştürücüsüne girdi olarak verilir. Böylece kelimeleri morfolojik seviyede olan Tatarca cümleler elde ederiz. Bu cümledeki her bir kelimenin normal şeklini bulmak için kelimenin morfolojik yapısı Tatarca Morfolojik Üreticisine verilir. Tatarca morfolojik üreticiden gelen

kelime sonuçları kullanılarak Tatarca cümleler yaratılır ve sonuç olarak bütün olası tercüme kullanıcıya sunulur. Tercüme sistemin son halinde morfolojik belirsiz giderici kullanılmadığı için, birden fazla Tatarca cümle tercüme sonucu olarak verilebilir. Bu yanıtlardan bir kısmı hatalı olabilecektir.

5. Örnek

Burada bir Türkçe cümle için Tercüme sistemi tarafından nasıl Tatarca'ya çevrildiğini göstereceğiz. Örnek olarak "akşam eve geleceğiz" Türkçe cümlesini kullanacağız. Bu cümle ilk önce "akşam", "eve", "geleceğiz" kelimelerine ayrıştırılır, ve bu bütün kelimeler Türkçe Morfolojik Analizci tarafından analiz edilerek aşağıdaki sonuçlar bulunur.

- akşam: 1. akşam+Noun+A3sg+Pnon+Nom
eve: 1. ev+Noun+A3sg+Pnon+Dat
geleceğiz: 1. gel+Verb+Pos+Fut+A1pl
2. gelecek+Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Pres+A1pl
3. gelecek+Adj^DB+Verb+Zero+Pres+A1pl

Burada son kelimenin üç olası analizi olduğundan, verilen Türkçe cümle morfolojik seviyede aşağıdaki üç değişik cümle ile ifade edilir.

1. akşam+Noun+A3sg+Pnon+Nom ev+Noun+A3sg+Pnon+Dat gel+Verb+Pos+Fut+A1pl
2. akşam+Noun+A3sg+Pnon+Nom ev+Noun+A3sg+Pnon+Dat gelecek+Noun+A3sg+Pnon +Nom^DB+Verb+Zero+Pres+A1pl
3. akşam+Noun+A3sg+Pnon+Nom ev+Noun+A3sg+Pnon+Dat gelecek+Adj^DB+Verb +Zero+Pres+A1pl

Bu üç cümle ayrı ayrı tercüme kurallarını uygulayacak olan sonlu durum dönüştürücüsüne verilir. Burada ilk önce gramer kurallarını ve de duruma bağlı kuralların olduğu dönüştürücü, sonrada kök kelimelerin nasıl tercüme edileceğini gösteren kuralların olduğu dönüştürücü çalıştırılır. En son olarak ta bazı kelimelerdeki k harfini q harfine çevirecek olan dönüştürücü çalıştırılır. Türkçe'deki bazı kelimeler içindeki bazı k harfleri q harfleri değiştirilerek Tatarca'da aynen kullanılır. Bu kural sadece "a, ı, o, u" harflerini izleyen bütün k harflerini q harflerine çevirir. Böylece kelimeleri morfolojik seviyede olan üç Tatarca cümle elde ederiz.

1. aqSam+Noun+A3Sg+Pnon+Nom ev+Noun+A3Sg+Pnon+Dat kel+Verb+Pos+Fut+A1PI
2. aqSam+Noun+A3Sg+Pnon+Nom ev+Noun+A3Sg+Pnon+Dat istiqbal+Noun+A3Sg+Pnon+Nom^DB+Verb+Zero+Pres+A1PI
3. aqSam+Noun+A3Sg+Pnon+Nom ev+Noun+A3Sg+Pnon+Dat kelecek+Adj^DB+Verb+Zero+Pres+A1PI

Bu cümledeki her bir kelime Tatarca morfolojik işlemciye verilerek kelimenin normal hali bulunur. Burada birinci ve üçüncü cümleden aşağıdaki ilk Tatarca cümle ve de ikinci cümleden ise aşağıdaki ikinci Tatarca cümle üretilir.

1. aqSam evge kelecekmiz *
2. aqSam evge istiqbalmız

Bu cümlelerden birincisi (yanında * olan) doğru tercümedir.

6. Sistemin Sınırlamaları

Türkçe’de birçok kelimenin birden fazla morfolojik yapısı vardır. Ama genelde verilen bir cümlede bir kelimenin morfolojik yapılarından sadece biri doğrudur. Bir kelimenin hangi morfolojik yapısının bir cümle içinde doğru olduğu o cümledeki durum bilgisine dayanır. Morfolojik belirsizlik gidericiler cümledeki durum bilgisini vede istatistiksel yöntemler kullanarak bir kelimenin o cümledeki doğru morfolojik yapısını tahmin etmeye çalışırlar. Şu anda Türkçe-Tatarca tercüme sisteminde bir Türkçe morfolojik belirsizlik giderici kullanmadığımızdan, tercüme sistemi doğru çeviriye ek olarak yanlış çevirileri de sonuç olarak vermektedir. İyi bir Türkçe morfolojik belirsizlik giderici sonuçlar arasından birçok yanlış çeviriyi yok etmeye yardımcı olacaktır.

Türkçe ve Tatarca birbirlerine çok yakın iki dil olsalar bile, sadece sonlu durum yöntemleri kullanan tercüme sistemleri bazı sınırlamalarla karşılaşabilir. Türkçe ve Tatarca özgür kelime sırası kullanan iki doğal dil olmaları, teorik olarak cümlelerin kelimeleri değişik sırada organize edilebilir. Türkçe’de bir fiilin nesnesi o fiilin hemen yanında tercih edilse bile bu şart değildir. Önceki bölümlerde belirttiğimiz gibi Türkçe’den Tatarca’ya tercüme yapılırken bazı fiillerin nesnelere isim halleri değişikliğe uğrayabilir. Sistemizde yazdığımız bazı tercüme kuralları fiilin nesnesinin fiilden hemen önce geleceğini ve fiil ile nesnesi arasında başka bir isim olmayacağı varsayar. Ama fiilin nesnesi fiilden hemen önce gelmezse bu tercüme kuralı doğru sonuç üretmeyecektir. Örneğin tercüme sistemimiz “Rus kızıyla evlendi” cümlesi Tatarca’ya fiilin nesnesi olan “kız” kelimesinin ile-ekini Tatarca’da e-eki olarak çevirerek doğru tercüme olan “Rus kızına evlendi” elde edebilir. Bu örnekte fiilin nesnesi fiilden hemen önce gelmiştir. Ama “Rus kızıyla Moskava’da evlendi” cümlesi nesne fiilden hemen önce gelmediği için doğru olarak tercüme edilmeyecektir. Yine benzer olarak “Rus kızıyla memnuniyetle evlendi” cümlesi tercüme kuralları tarafından “memnuniyet” kelimesi fiilin nesnesi olarak görüleceğinden doğru olarak tercüme edilmeyecektir. Böyle durumlar ile baş edebilmek için bir tercüme sistemi sonlu durum yöntemlerine ek olarak bir Türkçe sözdizim analizcisi kullanması gerekebilir.

Tatarca’da geniş zaman yerine şimdiki zamanın kullanılması yaygındır. “Siz giderseniz ben de gelirim” Türkçe cümlesindeki geniş zaman eki Tatarca’daki karşılığı olan “Siz ketseñiz men de kelirim” cümlesinde aynen korunur. Ama “Ben de bazen gelirim” cümlesindeki geniş zaman eki karşılığı olan “Men de kimerde kelem” Tatarca cümledeki gibi şimdiki zaman eki olarak tercüme edilir. Hangi cümlelerde geniş zaman yerine şimdiki zamanın kullanılacağını belirten bir kural Tatarca’da tam olarak tanımlanmamıştır. Buda cümleyi bir Türkçe sözdizim derleyicisiyle derlese bile bu problemi nasıl çözeceğimiz tam belli değildir.

7. Sonuçlar

Türkçe ile Kırım Tatarca’sı arasında geliştirdiğimiz prototip tercüme sisteminin, Türkçe ile diğer Türkük diller arasında da uygulanabileceğini düşünmekteyiz. Eğer ilk önce o diller için bir morfolojik analizci geliştirilse, Türkçe ile bu diller arasında tercüme sistemlerin benzer şekilde kolayca geliştirilebileceğini düşünmekteyiz. Burada temel olarak yapılması gerekenler, bu diller için morfolojik analizcilerin geliştirilmesi, Türkçe ile bu diller arasında tercüme sözlüklerin hazırlanması vede tercüme kurallarının geliştirilmesidir.

Türkçe ve Tatarca gibi yapıları birbirine yakın olan iki agulanatif dil arasındaki temel farklılıklar genelde morfem seviyesinde oluşur. Kaynak dildeki hangi kelimenin veya hangi morfemin sonuç dildeki hangi yapılar karşılık geleceği genelde sonlu durum yöntemleri ile bulunur. Kaynak dildeki morfolojik ve anlamsal seviyedeki belirsizliklerin bir çoğunda sonuç dilde de korunduğundan, tercüme sırasındaki harcanan çabayı azaltır. Üretilen sonuçların doğruluğunu artırmak için kaynak dil için bir morfolojik belirsizlik giderici kullanmakta fayda vardır. Çoğu zaman sonlu durum yöntemleri yeterli olsada, bazı durumlarda kaynak dil için bir sözdizim derleyicisinin kullanılması iyi olacaktır. Ama bir anlamsal analizci kullanılmadan başarılı bir tercüme sistemi geliştirebileceği kanısındayız.

Kaynakça

- [1] Altintas, Kemal, Cicekli, Ilyas: 2001, A Morphological Analyser for Crimean Tatar. In *Proceedings of Turkish Artificial Intelligence and Neural Network Conference (TAINN2001)*, North Cyprus
- [2] Canals, Raül, Esteve, Anna, Garrido, Alicia et.al.: 2000, interNOSTRUM: A Spanish-Catalan Machine Translation System, *Machine Translation Review*, Issue No.11, December 2000 - pp 21-25.
- [3] Cicekli, Ilyas, Guvenir, H. Altay: 2001, Learning Translation Templates from Bilingual Translation Examples, *Applied Intelligence*, Vol. 15, No. 1, pp: 57-76.
- [4] Jurafsky, Daniel, Martin, James H.: 2000, *Speech and Language Processing*, Prentice Hall.
- [5] Kubon, Vladislav, Hajic, Jan, Hric, Jan: 2000, Machine Translation of Very Close Languages, in *ANLP-NAACL2000*, Washington.
- [6] MLTT Finite State Homepage, (<http://www.xrce.xerox.com/research/mltt/fst/home.en.html>)
- [7] Oflazer, Kemal: 1994, Two-level Description of Turkish Morphology, *Literary and Linguistic Computing*, Vol. 9, No:2.