

Uygur Tümcenin Bilgisayar ile Çözümlemesi

Murat Orhun

Istanbul Bilgi Üniversitesi, Bilgisayar Mühendisliği Bölümü, İstanbul

murat.orhun@bilgi.edu.tr

Özet: Bilgisayarlı çeviride, kaliteli çevirinin yapılabilmesi için kaynak ve hedef diller tümcenin çözümlemesi çok önemlidir. Eğer kaynak ve hedef dillerin tümce yapısı aynı yapıya sahip ise, örneğin Türkü diller ailesinde bulunan Azerice, Türkmençe, Kazakça gibi, tümce üzerinde araştırma yapılmasına gerek kalmıyor. Ancak, kaynak ve hedef dillerin tümce yapısı belirgin bir şekilde farklı ise örneğin, batı Avrupa ve doğu Asya dilleri gibi, tümcelerin ciddi bir şekilde araştırması gerekmektedir. Bu bildiriye, Uygur tümcelerinin bilgisayar ortamında nasıl çözümleneceği ile ilgili öneriler ve bu öneriler kullanılarak elde edilen ilk sonuçlar sunulmuştur. Sonuçlara göre, ileride nasıl bir araştırma yöntemi kullanılması ile ilgili öneriler sunulmuştur.

Anahtar Sözcükler: Bilgisayarlı Çeviri, Uygurca tümce, Uygur gramer, Biçimbilimsel Çözümleyici.

1. Giriş

Uygur dili Türk dillerine ait en eski ve aynı anda en zengin dillerdendir. Teknolojinin, özellik ile bilgisayar bilimlerinin gelişmesiyle, dillerin bilgisayar ortamında daha verimli kullanılabilmesi için bilgisayara dayalı araştırmalar yapılmaktadır. Örneğin otomatik metin düzeltme, otomatik kelime tamamlama, dil bilgisi denetleme, farklı diller arasına otomatik çeviri yapma ve kripto işlemleri gibi. Ancak İngilizce, Japonca, Fince, ve Türkçe [1-4] gibi diller daha yoğun olarak araştırılırken, bazı diller ile ilgili araştırmalar çok kısıtlıdır. Özellikle ile, az gelişen ülkelerde bilgisayara dayalı araştırmalar çok sınırlıdır ya da araştırma aşamasındadır. Örneğin, Kanada'da yasal olarak kullanılan İngilizce ve Fransızca dilleri arasına otomatik çeviri yapan yazılımlar geliştirilirken[5], az gelişen ülkelerde bu çeşit çeviri işlemleri insanlar tarafından yapılmaktadır. Çeviri yapılacak metinlerin boyutu ve güvenilirliği düşünüldüğünde, insanlar tarafından yapılacak işlemlerin maliyeti oldukça yüksektir. Örneğin yetkin çeviri yapabilen bir insan yetiştirme, ister zaman açısından, ister maliyet açısından çok pahalıdır. Bu neden ile, çeviri işlemlerin bilgisayar aracılığı ile yapılması çok

önemlidir. Ancak dillerin doğası gereği, farklı dillerin söz dizim, yani cümle eğelerinin yapısında farklı olmaktadır. Eğer çeviri yapılacak dillerin yapısı bir-birine yakın akraba diller ise [6], genellikle bu dillerin söz dizimi aynıdır ya da çok az farklıdır. Bu neden ile yakın diller arasında çeviri sistemi geliştirmek, başka farklı diller göre oldukça basittir [6]. Dolayısıyla, Türk ailesinde bulunan, Türkçe, Türkmençe, Azerice, Kırım Tatarcısı ve Uygurca gibi diller ilgili çalışmalar yapılmıştır [7-11]. Genel olarak, Türkiye Türkçesi hariç diğer diller ile ilgili tümce yapısı ile ilgili çalışmalar yapılmamıştır [12].

Farklı tümce yapısına sahip diller arasında çeviri sistemi tasarlanırken, tümce yapısının araştırılması çok önemlidir. Örneğin Türkçe ile Çince arasında bir çeviri sistemi tasarlama düşünüldüğünde, ilk önce yapılması gereken çalışma ise, her iki dilini tümce yapısının çözümlenmesidir. Çünkü Türkçenin tümce yapısı ile Çincenin cümle çok farklıdır. Örneğin:

Çince: Ni you mei you chien?
TR (yanlış): Sen var yok var para?
TR (doğru): Sende para varmı ?

Örnekte ise, yanlış çevir, Çince tümce sözcük sırası değiştirilmeden, olduğu gibi Türkçeye

aktarılmıştır. Ancak istenilen çeviri ise bir alta verilmiştir.

Bu bildiriye Uygurcanın tümce özellikleri Türkçe ile karşılaştırılarak araştırılmıştır ve bu özelliklerin bilgisayar ortamında nasıl tespit edileceği ile ilgili öner verilmiştir.

Bu bildirinin 2- bölümünde Uygur tümcelerinin yapısı Türkçe tümceler yapısı ile karşılaştırılarak analiz edilmiştir, 3- bölümünde tümce öğelerinin tespit edilmesi ile ilgili çalışmalar verilmiştir, son bölümde ise, önerilen çalışmalara göre yapılan ilk sonuçlar verilmiştir ve sonuçlar yorumlanmıştır.

2. Uygur Tümcelerinin Yapısı

Sözcükler tek başlarına sadece bir anlam verir. Ancak tam bir düşünce ya da lojik bir anlam verebilmek için, sözcükler belli bir sıraya göre sıralanması gerekiyor. Genelde sözcüklerin sırası, dillerin gramer bilgilerine göre farklı oluyorlar. Bu neden ile tümcelerdeki sözcüklerin sırası çok önemlidir. Çünkü sözcükler tümcelerde değişik biçimlerde gelerek, milyonlarca cümle oluşturur ve anlam üretir [13-16]. Uygur dili bitişken dil olduğundan, üstelik, bir sözcüğün teori olarak, sonsuz ek alabileceğinden, tümce anlamında da sonsuz şekilde değişik anlam oluşturulabilir.

Genel olarak, bir cümle, yüklem, özne, nesne ve tümleç gibi dört öğeden oluşur. Ama bu dört öğenin tümünün olması zorunlu değildir. Özellik ile, nesne ve tümleçler bazen gizli, ya da gelmeyebilir.

Örneğin:

UY: Ramile kiliwatidu.

TR: Ramile geliyor.

Bu cümlede “Ramile” ise bu tümcenin öznesidir ve zamirdir. Yani zamir ise tümcenin öznesidir. İkinci sözcük “kiliwatidu” ise, yüklemidir, yani bir eylem ise yüklem olarak gelmiştir. Bu sözcüğün kökü ise “gel” dir. Alt bölümlerde bu öğeler Türkçe tümceler ile karşılaştırma yapılarak örnekler ile anlatılacaktır.

2.1 Yüklem

Bir cümlede gerçekleşen olayları bildirir. Genel olarak, ad, sıfat, adıl, belirteç ve ilgeç gibi öğeler yüklem olarak geliyorlar.

Örneğin:

UY: Siz bek chiraylıq.

TR: Siz çok güzelsiniz (sıfat).

UY: Dersimiz chüştün buruntı.

TR: Dersimiz öğleden önceydi (belirteç).

Bunlar hariç ad durum ekleri, “ bar (var)”, “yoq (yok)”, “ emes (değil)” sözcüklerinde yüklem olarak gelebilir. Bu durm Türkiye Türkçesine benziyor (Hengirmen- kaynak).

Örneğin:

UY: Bizning öyde elektirk yoq.

TR: Bizim evde elektrik yok.

UY: U bala oqughuchi emes.

TR: O çocuk öğrenci değil

Bazı durumlarda yüklem öznenden önce gelebilir. Bu durum özellik ile şiirlerde gözlenir. Ancak bu bildiri kapsamında, bu durum araştırılmamıştır. Nedeni ise, kurallı düz tümceler temel alınmıştır.

Konuşma dillerinde, bazen yüklemeler düşe bilir. Ancak, bir önceki tümceye bakılarak, yüklemeler tespit edilebilir.

Örneğin:

UY: Nege barisen?

UY: Bazara

TR: Nereye gidiyorsun?

TR: pazara

Burada, “gidiyorum” yüklemi söylenmemiştir.

Bazı cümlelerde birden fazla yüklem bulunur ve yüklemeler arasına, dil kurallarına göre bağlaçlar ya da notlandırma işaretleri konur.

Örneğin:

UY: Bizning wezipimiz, mektepke birish, ders oqush ve kelguside yaramliq insanlar bolush.

TR: Bizim görevimiz, okula varmak, ders okumak ve gelecekte yararlı insanlardan olmak.

2. 2 Özne

Cümlede işi yapan, eylemi oluşturan kişi, hayvan, bitki ya da diğer canlı cansız varlıklara özene denir [13]. Özenler bir Ya da birden fazla sözcüklerden oluşur. Bazen basit bir tümce, özne olarak da gelebilir. Ancak bu bildiri kapsamında tek sözcüklerden oluşan özneler araştırılmıştır. Birden fazla sözcüklerden oluşan özenler ise sadece örnek olarak gösterilmiştir.

Örneğin:

UY: Salih Qeshqerge ketti.

TR: Salih Kaşgara gitti (kişi).

UY: Limon zukamgha payda qilidu.

TR: Limon nezlye iyi gelir (bitki) .

UY: Yamghur yighivatidu.

TR: Yağmur yağıyor (nesne).

Genel olarak, özne olabilecek sözcüklerin kapsamı çok geniştir. Örneğin, kişi, hayvan, bitki, eylemler, basit tümceler gibi. Çoğu zaman özneler, kim ve ne gibi sorulara cevap olarak gelirler. Öznesiz cümle olmaz. Buna rağmen, bazı cümlelerde, özellik ile kısa cümlelerde özne gizli olur. Ancak cümle anlamına bakılarak, özneler buluna bilinir.

Örneğin:

UY: Mekeptin kelduq.

TR: Okuldan geldik.

Bu tümcede özne gizlidir. Yani bu cümlenin öznesi “biz” dır. Burada “kelduq (geldik)” yüklemine bakarak, bu tümcenin öznesinin gizli “biz” olduğu tespit edilebilir. Edilen tümcelerde öznelerin bulunması, eden tümcelere göre daha karmaşıktır. Bu nenden ile bu bildiriye araştırma kapsamına alınmamıştır. Genel olarak, özne ile yüklem teklik, çoğulluk bakımından ve kişi bakımından kesin uyumak zorundadır.

Örneğin:

UY: U ötken yili Qeshqerge biriptu.

TR: O geçen sene Kaşgara gitmiş.

UY: Ular ishlewatidu.

TR: Onlar çalışıyorlar .

Bu örneklerde zamirler haberlere uyum sağlamıştır. Burada Uygurca ile Türkçe arasında belirgin bir fark vardır. Uygurca cümlelerde, yüklemler çoğul eki almazken, Türkçe'de çoğul eki alırlar.

2. 3 Nense

Öznenin yaptığı işten doğrudan etkilenen ve geçişli eylemi tümleyen sözcüklere nesne denir [13]. Nesnenin görevi ise özne ve yüklem yapıları işi belirtmektir.

Örneğin:

UY: Biz kelduq.

TR: Biz geldik.

Bu örneklerde nesne yoktur.

UY: Biz Urumqi'ge kelduq.

TR: Biz Urumqi'ye geldik.

Bu örnekte, nesne sözcüğü “Urumqi”, tümce öznesi “biz” in nereye geldiğimiz ile ilgili bilgi verilmiştir. Yani tümce tam bir anlamı anlatabilecek şekilde tamamlanmıştır. Genelde nesnelere, “neyi”, “kimi” gibi sorular ile bulunur. Bazen, birden fazla sözcüklerden oluşan çift sözler Ya da basit tümcelerde bir tümce nesnesi olarak gelebilir.

2. 4 Tümleç

Eylemin anlamını tümleyen ad durum ekleri ve ilgeç alan sözcüklere tümleç denir. Tümleçler aldığı eklerle göre kendi içerisinde , dolaylı tümleçler, ilgeç tümleçleri ve belirteç tümleçleri gibi kategorilere ayrılır.

Tümleçler cümlenin en önemli öğelerinden biridir. Dolayısıyla bir tümcede birden fazla tümce buluna bilirlir ve bir tümleç, diğer tümce öğeleri gibi, birden fazla sözcüklerden oluşabilir.

Örneğin:

UY: Biz ete mektepke barimiz.
TR: Biz yarın okula gideceğiz.

Genelde, tümleçler, nereye, kime, neye, kaç, kimde, ne zaman gibi sorulara cevap olarak gelirler.

Bu nenden ile bir önceki örnekte 2 adet tümleç sözcüğü, “yarın, okula” sözcüğü bulunmaktadır.

Bazı özel ilgeçlerde tümcelerde tümleç görevi üstelenir. Bu amaç ile yaygın olarak kullanılan ilgeçler ise, “ bilen(ile), için (için), dek (gibi)” lerden oluşuyor.

Örneğin:

UY: İtaliye'ge ayırılan bilen kettim.
TR: İtalya'ya uçak ile gittim.

Yüklem anlamını zaman, yer, yön, durum, ve nicelik ilgisiyle tamamlayan sözcüklere belirteç tümleçleri denir [13].

Örneğin:

UY: Biz talada pranglishayli.
TR: Biz dışarıda konuşalım.

UY: U öyge astağine kirip keldi.
TR: O eve yavaşça girdi.

Genelde, belirteç tümleçleri, ne zaman, nereye, nasıl, ne kadar gibi sorulara cevap olarak gelir.

Sonuç olarak, Uygur tümcelerinin yapısı Türkçe tümcelerinin yapısı ile genel olarak benziyor. Yani bir tümceyi oluşturan temel öğeler benziyor. Ancak, aldığı ekler yünden azımsamayacak kadar farklılıklarda bulunuyor [12].

3. Uygur Tümce Öğelerinin Belirlenmesi

Tümceler, genelde tümceleri oluşturan öğelerinin sırasına göre, kurallı ve kuralsız cümle olarak iki büyü kategoriye ayrılırlar. Kurallı cümlelerde genellikle özne başta, yüklem sonunda bulunur. Diğer öğeler belirtmek istenen konunun önem durumuna göre cümle içerisinde yer alır. Cümlede yer

alan sözcük sayısının artmasına göre, cümle öğelerinin tespit edilmesi de zorlaşır. Bu durum özellik ile Türkçe gibi bitişken dillerde çok belirgin oluyor. Bunun temel nedeni ise, eklenen bir ekten dolayı, o sözcüklerin yapısı, anlamı, dolayısıyla cümlelerdeki görevinin değişmesidir. Bu bildiriye, öğelerin bulunması için, kısıtlı sayılardan oluşan basit cümleler temel alınmıştır.

Bir önceki bölümde anlatıldığı gibi, cümleleri oluşturan öğeler, bazen aldığı eklerle değişirken, bazen ise, bazı özel sözcükler, örneğin ilgeç ile tespit edilebilir. Ancak Uygurca bitişken dil olduğundan, tüm sözcüklerin listesinin oluşturulması imkansızdır. Bu neden ile sözcüklerin yapısı araştırılması, yani sözcüklerini yapısını araştıran bir yapının geliştirilmesi zorunludur. Bu neden ile sözcüklerin yapısı Uygurca için geliştirilen biçimbilimsel çözümleyici kullanarak çözümlenir [10-11].

Örneğin.

UY: Men mektepke keldim.
TR: Ben okula geldim.

Bu tümcede yer alan sözcükler çözümlendiğinde altta olduğu gibi çözümlenmeler oluşur.

ben+Noun+A3sg+Pnon+Nom
ben+Pron+Pers+A1sg+Pnon+Nom

okul+Noun+A3sg+Pnon+Dat

gel+Verb+Pos+Past+A1sg

Bu tümce analiz edilirken, öncelik ile bu tümcenin kurallı olduğunu kabul ediyoruz. Kurallı cümlelerin temel prensibi ise, Özne her zaman cümle başında gelir. Özne kendi açısından bazen gelmeye bilir, yani gizli olarak gelebilir. Bu ön koşulları var sayarak, yukarıda çözümlenen sözcüklere bakıldığında, “ben” sözcüğü “isim (Noun)” ve “zamir (Pron)” olarak çözümlenmiştir, üstelik ek almamıştır. Bu neden ile, bu sözcük bu tümcenin “öznesi” olarak etiketlenir. Bir sonraki sözcük “okul” ise, “isim (Noun)”

olarak çözümlenmiştir, ayrıca yönelme eki almıştır ve “Dat” almıştır. Bu neden ile bu sözcük, tümleş olarak çözümlenir. Bu sözcük burada, “nesne” olarak çözümlenemez. Çünkü kurallı cümlelerde, nesne olana sözcük, “yönelme” eki almaz.

En son sözcük, “gel” ise, eylem olarak çözümlenmiştir ve tümcenin sonunda gelmiştir. Bu nedenle bu sözcük yüklem olarak kabul edilir ve etiketlenir.

UY: Men mektepke poyiz bilen keldim.

TR: Ben okula tren ile geldim.

Bu tümce biçim bilimsel çözümleyici ile çözümlendiğinde, bir önceki örnekte çözümlenen sözcüklere ek olarak, “ poyiz (tren) ve “ bilen (ile)” sözcükleri de çözümlenir ve alttaki çözümlenmeleri üretir.

poyiz+Noun+A3sg+Pnon+Nom
bilen+Conj

“Poyiz” sözcüğü “isim (Noun)” olarak çözümlenmiştir. Bu sözcüğün bu tümcede, özne, nesne ve tümleş gibi üç farklı öge olarak çözümlenebilir. Özne olamaz çünkü tümce ortasında gelmiştir. Nesne olamaz neden ise bu sözcükten sonra ilge sözcüğü “bilen” gelmiştir. Bu neden ile son öge çeşiti “tümleş” olacaktır.

UY: Men ukamni mektepke poyiz bilen ekeldim.

TR: Ben kardeşimi okula tren ile getirdim.

Bu tümce çözümlenirken, bir önceki örnekte bulunan sözcüklere ek olarak, “ukamni (kardeşimi)” sözcüğü çözümlenecektir ve alttaki çözümlenmeler üretilecektir.

ukam+Noun+A3sg+P1sg+Acc

“Ukam (kardeş)” sözcüğü isim olarak çözümlenmiştir. Bu sözcük özne, nesne ve tümleş gibi 3 farklı öge olarak çözümlenebilir. Ancak, bu sözcük cümle başında gelmediğinden, özne olamaz, tümleş de olamaz, nedeni ise, tek sözcüktür ve aynı

anda yükleme hali ile gelemez. Bu nedenle bu sözcük, nesne olarak çözümlenir. Aynı yöntem ile, bir den fazla özne içeren ve birden fazla nesne için tümceler bile çözümlenebilmektedir. Ancak, tümleşlerin gerekirse yapısı, gerekirse aldığı eklerden dolayı çözüm başarımları çok düşük olmaktadır.

4. Sonuç ve Öneriler

Bu bildiriye, Uygurcanın biçimbilimsel çözümleyicisi kullanarak, kurallı cümlelerde öğelerin bulunması ile ilgili kural tabanlı bir çözüm önerisi sunulmuştur. Önerilen yöntem, tek sözcüklü öğeler üzerinde denenmiştir. Yani tek sözcüklerden oluşan, özne, nesne, yüklem ve tümleş gibi öğelerin tespit edilmesi için denemeler yapılmıştır. Deneme sonucu, farklı sözcüklerden oluşan 100 adet tümce üzerinde yüzde 60 üzerinde doğru sonuç vermiştir. Deney başarımının düşük olmasının temel nedenleri ise, biçimbilimsel çözümleyicinin oluşturduğu belirsizlikler, biçimbilimsel çözümleyicinin çözümlene başarımları, tümce öznelerin bazen gizli olması gibi durumlar çok etkili olmuştur. Birden fazla sözcüklerden oluşan öğeler için düşünüldüğünde, daha kapsamlı kuralların tanımlanması gerekmektedir. Bu Çalışma Uygur dili tümce öğeleri tespiti için yapılan ilk çalışmadır. Bu çalışmanın ileride yapılacak kapsamlı ve gelişmiş çalışmalara temel oluşturulması hedeflenmektedir.

5. Kaynaklar

[1] Uchida, H. ve Sugiyama, K. “A Machine Translation System From Japanese Into English Based On Conceptual Structure, Proceedings of the 8th conference” on Computational linguistics, Tokyo, Japan. (1980).

[2] Sagay, Z., A computer translation from English to Turkish. Yüksek Lisans Tezi, Bilgisayar Mühendisliği Bölümü, ODTÜ. Ankara (1981).

- [3] Lun, S., A Two-Level Morphological Analysis of French, in Texas Linguistic Forum, Texas, USA. (1983)
- [4] Koskenniemi, K., 1985. An Application of the Two-Level Model to Finnish, University of Helsinki Department of General Linguistics (1985) .
- [5] Chandiox, J., Météo: Un système opérationnel pour la traduction automatique des bulletins météorologiques destinés au grand public (1976).
- [6] Hajič, J., Hric, J., ve Kuboň, V., Machine translation of very close languages, in Proceedings of the sixth conference on Applied natural language processing, Morgan Kaufmann Publishers Inc, (2000).
- [7] AltıntaĖ, K.,. Turkish to Crimean Tatar Machine Translation System, in Bilgisayar Mühendisliđi Bölümü. MSc Ankara: Bilkent Üniversitesi (2000).
- [8] Hamzaođlu, Ė., Machine translation from Turkish to other Turkic languages and an implementation for the Azeri languages, in Institute for Graduate Studies in Science and Engineering. MSc Thesis İstanbul: Bogazici University, (1993).
- [9] Tantuđ, A. C., Adalı E., ve Oflazer K., Computer Analysis of The Turkmen Language Morphology, FinTAL Lecture Notes in Computer Science, Vol.4139, Springer, (2006).
- [10] Orhun, M., Tantuđ, A. C. ve Adalı, E. Rule Based Analysis of the Uyghur Nouns, Proceedings of the International Conference on Asian Language Processing (IALP). Chiang Mai, Thailand, (2008).
- [11] Orhun, M., Tantuđ, A. C. ve Adalı, E. Rule Based Tagging of the Uyghur Verbs, Proceedings of the Fourth International Conference on Intelligent Computing and Information Systems, Cairo, Egypt, (2009).
- [12] Tantuđ, A. C. Akrafa ve Bitiřken Diller Arasında Bilgisayarlı Çeviri İçin Karma Bir Model. . Bilgisayar Mühendisliđi Bölümü. Doktora Tezi. İstanbul: İstanbul Teknik Üniversitesi (2007).
- [13] Hengirmen, M., Türkçe Dilbilgisi, Engin Yayınevi, Ankara, (2005).
- [14] Tömür, H.,. Modern Uygur Grammar (Morphology). Yıldız Teknik Üniversitesi, Fen-Ed Fak. T.D.E Bölümü. İstanbul. (İngilizce baskı, 1997).
- [15] Osmanof, M., Hazırqı Zaman Uyghur Edebiy Tilining İmla ve Teleppuz Lughiti. Shin Jiang Xeliq Neshiryatı. (Uygurca baskı, 1997).
- [16] Öztürk, R., “Yeni Uygur Türkçesi Grameri”. Türk dil kurumu yayınları: 593, (1993).