



Machine translation system from Turkish to Turkish Sign Language

Hüsnü Erten^{1*}, Nursal Arıcı²

¹Department of Computer Engineering, Faculty of Technology, Gazi University, 06570, Ankara, Türkiye

²Department of Management Information Systems, Faculty of Applied Sciences, Gazi University, 06570, Ankara, Türkiye

Highlights:

- Using temporal gaps between translation visuals
- Rule-based and statistics-based hybrid translation method
- Creating a parallel data set of 3000 sentences between Turkish and Turkish Sign Language

Keywords:

- Machine translation
- Statistical translation
- Rule-based translation
- Turkish
- Turkish Sign Language

Article Info:

Research Article
Received: 14.04.2023
Accepted: 18.01.2025

DOI:

10.17341/gazimmfd.1283219

Correspondence:

Author: Hüsnü Erten
e-mail:
husnu.erten@gazi.edu.tr
phone: +90 545 678 8979

Graphical/Tabular Abstract

In this study, an automatic translation system from Turkish texts to Turkish Sign Language (TSL) has been proposed. Within the scope of the study, only the subject of translation from Turkish text to Turkish sign language has been handled. The lack of parallel corpus in languages with few users is also valid for Turkish and TSL. This situation negatively affects the success of statistical-based translations from Turkish to TSL. With this study, a parallel data set of 3000 sentences from Turkish to Turkish Sign Language has been created and contributed to the elimination of the parallel data deficiency in the field. In the study, by using rule-based and statistical-based translation methods together, it has been tried to prevent the low level of translation success due to lack of data.

Table A. Comparison of BLEU scores in all datasets for the sentence 120
(If you want to sit by the window, this table is available)

Sentence 120	BLEU	Length ratio	Text
Source	-	-	Cam kenarında oturmak isterseniz bu masa müsait. siz cam kenar oturmak istemek bu masa müsait olmak
Human	100.00	1.00	cam kenarında oturmak isterseniz bu masa müsait. cam olmak kenar oturmak istemek siz bu masa müsait değil
Machine ST with Raw Data	20.61	0.89	cam kenar oturmak istemek siz bu masa müsait camdaki kız ama kenar oturmak istemek bu masa müsait siz
Machine ST with Morphological Data	29.07	1.11	
Machine Rule Based Translation	45.62	0.89	
Machine ST with Rule Based Data	53.73	1.11	
Machine ST with Rule Based Data (3000 sentences)	88.25	0.89	cam kenar oturmak istemek bu masa müsait olmak

Purpose: In this study, it is aimed to contribute to the solution of the problems by the hearing impaired in communication by developing an application that makes automatic translation from Turkish to TSL.

Theory and Methods: The translation steps start with the identification of sentences from the Turkish text given as input. In the second step, the punctuation marks in the sentences are cleaned. In the third step, each sentence is subjected to morphological analysis one by one with Zemberek NLP library and in the fourth step the most appropriate morphological solutions obtained are selected to apply the translation rules. In the fifth step, some translation rules are applied according to the data obtained from the morphological analysis. In the sixth step, the rule-based semi-translated sentence is given as input to the statistical translation tool (Moses, Giza++) and subjected to statistical translation. In the seventh step, word dependencies are detected using the MaltParser dependency analysis tool. Word groups are formed according to word dependencies. In the eighth step, the signs required for visualization are searched in the dictionary, and any words that cannot be found are marked using the finger alphabet for them. The translation steps end with the visualization of the text obtained from the statistical translation.

Results: The BLEU score obtained from the test of raw data has been 6.58 % and it has been increased to 10.58 % after morphological analysis. With the rule-based translation system made with the information obtained from the morphological analysis, the success of the statistical translation has been increased to the level of 32.92 %. The BLEU score of the sentence "Cam kenarına oturmak isterseniz bu masa müsait (If you want to sit by the window, this table is available.)" across all data sets is given in Table A.

Conclusion: With the parallel data set of 3000 sentences specially produced for this study, it has contributed to the elimination of the parallel data deficiency in the field. As a result of using rule-based and statistics-based translation together, translation success has been increased. In addition, for the first time, a temporal gap has been left between the word groups, and the misunderstandings caused by incorrect positioning have been tried to be reduced during the visualization of the translation in this study.



Mühendislik Mimarlık Fakültesi Dergisi

Journal of The Faculty of Engineering
and Architecture of Gazi University

Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Türkçeden Türk İşaret Diline makine çeviri sistemi

Hüsnü Erten^{1*}, Nursal Arıcı²

¹Gazi Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, 06570, Ankara, Türkiye

²Gazi Üniversitesi, Uygulamalı Bilimler Fakültesi, Yönetim Bilişim Sistemler Bölümü, 06570, Ankara, Türkiye

Ö N E Ç İ K A N L A R

- Çeviri görselleri arasında zamansal boşluk kullanılması
- Kural tabanlı ve istatistik tabanlı hibrit çeviri yöntemi
- Türkçe ve Türk İşaret Dili arasında 3000 cümlelik eşlenik veri seti oluşturulması

Makale Bilgileri

Araştırma Makalesi

Geliş: 14.04.2023

Kabul: 18.01.2025

DOI:

10.17341/gazimmfd.1283219

Anahtar Kelimeler:

Makine çevirisi,
istatistiksel çeviri,
kural tabanlı çeviri,
Türkçe,
Türk İşaret Dili

ÖZ

Bu çalışmada işitme engellilerin iletişimde yaşadığı problemlerin çözümüne katkı sağlamak amacıyla Türkçe metinlerden Türk İşaret Dili (TİD) arasında otomatik çeviri sistemi önerilmektedir. Az kullanıcısı olan dillerin genelinde yaşanan eşlenik veri eksikliği, Türkçe ve TİD için de geçerlidir. Bu durum Türkçeden TİD'e yapılan istatistiksel tabanlı çevirilerin başarısını olumsuz etkilemektedir. Bu çalışma ile Türkçeden Türk İşaret Diline 3000 cümlelik eşlenik veri seti oluşturularak alandaki eşlenik veri eksikliğinin giderilmesine katkı sağlanmıştır. Çalışmada kural tabanlı ve istatistiksel tabanlı çeviri yöntemleri birlikte kullanılarak, veri azlığından kaynaklı çeviri başarısı düşüklüğünün önüne geçilmeye çalışılmıştır. Ham verinin testinden elde edilen BLEU (Bilingual Evaluation Understudy) skoru, % 6,58 olup morfolojik analizden geçirilerek % 10,58 seviyesine çıkarılmıştır. Morfolojik analizden elde edilen bilgilerle yapılan kural tabanlı çeviri sistemi ile de istatistiksel çevirinin başarısı % 32,92 seviyesine çıkartılmıştır. Ayrıca, ilk defa bu çalışmada, çevirinin görselleştirilmesi aşamasında kelime grupları arasında zamansal boşluk bırakılarak yanlış konumlandırmadan kaynaklanan anlam yanlışları azaltılmaya çalışılmıştır.

Machine translation system from Turkish to Turkish Sign Language

H I G H L I G H T S

- Using temporal gaps between translation visuals
- Rule-based and statistics-based hybrid translation method
- Creating a parallel data set of 3000 sentences between Turkish and Turkish Sign Language

Article Info

Research Article

Received: 14.04.2023

Accepted: 18.01.2025

DOI:

10.17341/gazimmfd.1283219

Keywords:

Machine translation,
statistical translation,
rule-based translation,
Turkish,
Turkish Sign Language

ABSTRACT

In this study, an automatic translation system from Turkish texts to Turkish Sign Language (TSL) has been proposed in order to contribute to the solution of the problems experienced by the hearing impaired in communication. The lack of parallel corpus in languages with few users is also valid for Turkish and TSL. This situation negatively affects the success of statistical-based translations from Turkish to TSL. With this study, a parallel data set of 3000 sentences from Turkish to Turkish Sign Language has been created and contributed to the elimination of the parallel data deficiency in the field. In the study, by using rule-based and statistical-based translation methods together, it has been tried to prevent the low level of translation success due to lack of data. The BLEU (Bilingual Evaluation Understudy) score obtained from the test of raw data has been 6.58 % and it has been increased to 10.58 % after morphological analysis. With the rule-based translation system made with the information obtained from the morphological analysis, the success of the statistical translation has been increased to the level of 32.92 %. In addition, for the first time, a temporal gap has been left between the word groups, and the misunderstandings caused by incorrect positioning have been tried to be reduced during the visualization of the translation in this study.

1. Giriş (Introduction)

Çeviriler, farklı diller arasında iletişimin kurulabilmesinin olmazsa olmaz bir şartıdır. Teknolojik gelişmelere bağlı olarak, makine çevirilerinin kalitesinin artmasına rağmen, günümüzde en kaliteli çeviriler, insanlar tarafından yapılan çevirilerdir. Ancak insan çevirilerinin yüksek maliyetli olması ve çeviri işlemlerindeki tekrarlanabilirlik gibi sorunlar nedeniyle, makine çevirileri tercih edilmektedir.

Her dilin kendine özgü yapısı ve kuralları bulunduğu için çevirisi yapılacak diller arasındaki dönüşüm kuralları o iki dil arasındaki çeviriye özel kurallardır. Ancak çeviri sistemlerinde kullanılan yöntemlerde benzerlikler görülebilmektedir.

Çeviriler, konuşulan diller arasında yapıldığı gibi işaret dilleri arasında da yapılabilmektedir. İşitme engelli bireyler iletişimde jest, mimik ve hareketlerden oluşan görsel bir dil olan işaret dili kullanmaktadır. İşaret dillerinin de diğer diller gibi kendine özgü özellikleri vardır. Konuşulan bir dilden işaret diline çeviri yapılabildiği gibi işaret dilinden de konuşulan dile çeviri yapılabilmektedir. Çeviri işlemlerinde her iki dilin de yazı formuna dönüştürülmesi gerekir. Sözlü bir ifade, ses algılama yöntemleri ile yazılı metin haline dönüştürülür daha sonra çevirisi yapılır. İşaret dilinde ise hareketler kameralarla izlenerek algılanıp metne dönüşümü sağlanır. Tersi işlemlerde de yazılı metnin ilgili dildeki forma (ses, görüntü, animasyon) dönüşümü sağlanmalıdır. Çeviri işlemi temelde yazılı metinler üzerinden yapılmaktadır.

Türkiye’deki işitme engelli bireylerin ortak olarak kullandıkları işaret dili, Türk İşaret Dili (TİD)’dir. TİD’i kullanan kişi sayısı hakkında net bilgi bulunmamaktadır. Ancak Aile Çalışma ve Sosyal Politikalar Bakanlığının Ağustos 2021 tarihli “Engelli ve Yaşlı İstatistik Bülteni” [1] raporuna göre ülkemizde 179.867 kişinin işitme engeli bulunmaktadır.

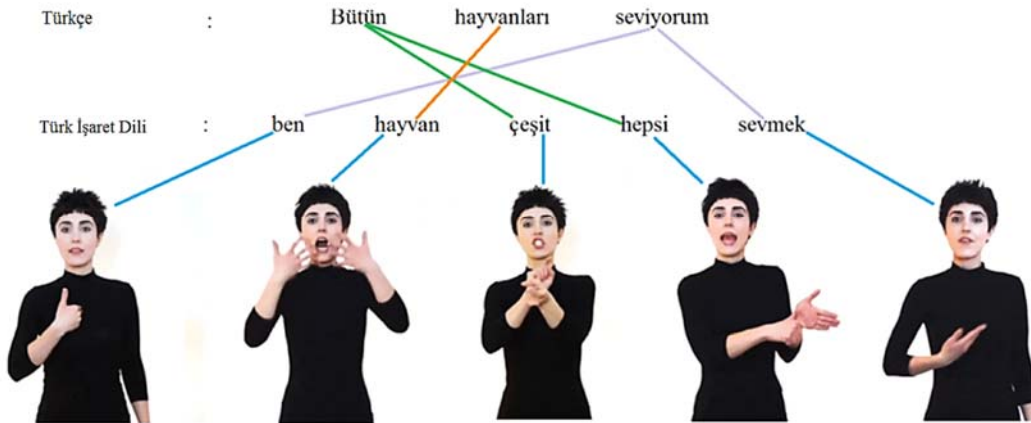
TİD, konuşulan dilin Türkçe olduğu bir coğrafyada ortaya çıktığı için Türkçe ile benzerliklere sahip olsa da dilbilimsel açıdan farklı özelliklere sahip doğal bir dildir [2]. Genel olarak özetlemek gerekirse, TİD’de kelimelere karşılık gelen işaretler vardır. Fakat bu durum, TİD’in, Türkçe cümledeki kelimelere karşılık gelen işaretlerin kullanılmasıyla oluşturulduğu anlamına gelmez. TİD’in kendine has yapısı ve gramer kuralları bulunmaktadır. Öncelikle TİD, Türkçe gibi sondan eklemeli bir dil değildir. Türkçedeki tüm eklerin TİD’de karşılığı yoktur. TİD’de karşılığı olan ekleri alan kelimeler, ayrı

işaretlerle (gitmemek = gitmek+değil) gösterilebildiği gibi temel olarak kelimenin işaretlenmesi sırasında kullanılan mimiklerle de ifade edilmektedir. Diğer bir fark ise Türkçedeki tüm zaman kiplerinin TİD’de olmamasıdır. TİD’de şimdiki zaman, geçmiş zaman ve gelecek zaman kavramları kullanılmaktadır. Vurgu yapılmak istenildiğinde değişiklik gösterse de, Türkçede genellikle cümle yapısı özne+nesne+yüklem şeklindedir. TİD’de ise cümlelerin öğelerinin sıralanışı genellikle özne+yüklem+nesne şeklindedir. Yabancı kelimeler, özel isimler, yeni kullanılmaya başlayan terimler gibi işaret dilinde karşılığı olmayan kelimeler için ise kelimedeki tüm harflerin tek tek parmak alfabesi ile gösterilmesi yoluna gidilir. Kısaca TİD, Türkçeden ayrı doğal bir dildir. Türkçeye benzer yönleri olmakla birlikte kendine has özellikleri ve kuralları vardır. İki dil arasında yapılacak çeviri çalışmalarında hem Türkçenin hem Türk İşaret Dilinin gramer özelliklerinin iyi bilinmesi gerekmektedir.

Şekil 1’de Türkçe bir cümle için TİD’de işaretlenmesi ile ilgili bir örnek, görselleri ile birlikte verilmiştir. Türkçe cümledeki kelimeler, işaret dilinde karşılık gelen kelimeler ile eşleştirilerek gösterilmiştir.

Sağırılar iletişim kurabilmek için işaret dili, dudak okuma ve yazma/çizme gibi yöntemlere başvurmaktadır [3]. Eğitimde yaşadıkları zorluklardan kaynaklı Sağırılar arasında okur-yazarlık oranı oldukça düşüktür. TÜİK’in 2010 yılı Özürlülerin Sorun ve Beklentileri Araştırması sonuçlarına göre Türkiye’de okur-yazar olmayan sağırının oranı %31,6’dır [4]. Yani sağır bireylerin %31,6’sı ne sözlü iletişimi ne de yazılı iletişimi kullanabilmektedir. Bu durumdaki sağır bireyler, iletişim sağlayabilmek için işaret dilini kullanmaya yönelmektedir. Genellikle bu bireylerin yazılı metinleri okuyup anlayabilme kabiliyetleri işitebilen bireylere göre daha düşük olmasından dolayı işitme engellilerin bilgiye erişimine katkı sağlayacak çalışmaların yapılmasına ihtiyaç duyulmaktadır.

İşitme engelliler, ülkemizde topluma uyum konusunda zorluk yaşamamakla birlikte, eğitim, sanat ve sosyal faaliyetler açısından zorluklarla karşılaşmaktadır. Ayrıca sağırının haklarıyla ilgili yasalama sürecinde gecikmeler ve eksiklikler vardır. Son yıllarda, özellikle 2005 engellilik yasasının kabul edilmesinden beri önemli mesafe kat edilmiştir. Son zamanlarda Türkiye’de, işaret dili konusunda toplumda farkındalık artmış, çok sayıda özel ve resmi işaret dili kursu açılmış ve üniversitelere seçmeli işaret dili dersleri konmuştur. Pek çok olumlu gelişme yaşanmasına rağmen alanda yapılacak daha çok çalışmaya ihtiyaç olduğu da bir gerçektir. Sağırının ihtiyaçlarının ve problemlerinin analiz edilmesi ve eğitim, sağlık, sosyal yaşam gibi tüm yaşam alanlarında iyileştirme



Şekil 1. Türkçe-TİD çeviri örneği (Turkish-TSL translation example)

çalışmalarının yapılması gerekmektedir. Özellikle konuşulan dilden işaret diline ve işaret dilinden konuşulan dile otomatik çeviri sistemleri ile ilgili çalışmalar, işitme engelli bireylerin iletişim problemlerinin çözümü için öncelik taşımaktadır.

Çeviri çalışmalarında kullanılan veri setlerinde farklılıklar olabilmektedir. Bu nedenle, başarı karşılaştırması yapabilmek için matematiksel modellere ihtiyaç duyulmuştur. Bu çalışmada, alanda yapılmış çalışmaların başarısını kıyaslayabilmek için tüm çalışmalarda ortak olarak kullanılan BLEU (Bilingual Evaluation Understudy) [5] skoru referans alınmıştır.

BLEU makine çeviri sistemlerinin yapmış olduğu çevirinin, insan çevirisine ne kadar yakın olduğunu otomatik olarak ölçmeye yarayan bir metottur. 2002 yılında yayınlanan BLEU metodu, birçok çeviri çalışmasında başarı ölçümü için kullanılmaktadır. Makine çevirisi ile insan çevirisi metinleri arasında, "n-gram" olarak adlandırılan farklı boyuttaki ardışık sözcük dizileri üzerinden benzerlik ölçümü yapılmaktadır. Genellikle BLEU skoru olarak 4-gram'a kadar olan skorların logaritmik ortalaması ile BP (Brevity Penalty - Kısıklık Cezası) çarpımı kullanılmaktadır. Benzerlik değeri 0 ile 1 arasında bir sayıdır. Kısıklık cezasındaki temel hedef, yapılan çevirideki kelime sayısının, orijinal çeviriden az olduğu durumlarda bir ceza puanı vermektedir. Eş. 1'de BP (kısıklık cezası) formülü verilmiştir. Eş. 1'de c yapılan çevirideki kelime sayılarını, r ise orijinal çevirideki kelime sayısını ifade etmektedir. Eş. 2'de n-gram için kesinlik hesaplama ve Eş. 3'te BLEU hesaplama formülleri verilmiştir. Tüm incelenen çalışmalarda ortak olarak kullanılan BLEU başarı ölçümü, çalışmamız için de başarı ölçütü olarak kullanılmıştır.

$$Brevity Penalty = \begin{cases} 1, & r \leq c \\ e^{\frac{1-r}{c}}, & r > c \end{cases} \quad (1)$$

$$p_n = \frac{\sum_{C \in \{çeviri\ cümleler\}} \sum_{n-gram \in C} Sayı_{kırpılmış}(n-gram)}{\sum_{C' \in \{çeviri\ cümleler\}} \sum_{n-gram' \in C'} Sayı(n-gram')} \quad (2)$$

$$BLEU = BP \cdot \sum_{n=1}^N w_n \log p_n \quad (3)$$

Bu makalede işitme engellilerin iletişimde yaşadığı problemlerin çözümüne katkı sağlamak amacıyla Türkçe metinlerden TİD'e otomatik çeviri yapan bir sistem önerilmektedir. Bu kapsamda makalenin ikinci bölümünde, metinden işaret diline çeviri yöntemleri hakkında literatürden derlenen bilgiler ve örnek çeviri sistemi çalışmaları özetlenerek genel bir değerlendirme yapılmıştır. Üçüncü bölümde, önerilen sistemin işlevleri ve çalışma prensibi ayrıntılı olarak açıklanmış, BLEU skorları hesaplanarak deneysel bulgular verilmiştir. Sonuç bölümünde, bulgular değerlendirilmiş ve önerilerde bulunulmuştur.

2. İlişkili Çalışmalar (Related Works)

Otomatik makine çevirileri için çeşitli yöntemler önerilmiş ve geliştirilmiştir. Önceleri kelime bazlı, sözlük benzeri çeviri çalışmaları yapılırken, daha sonraki çeviriler cümle bazlı hâle gelmiştir. Kural tabanlı çeviri yöntemleri ile başlayan cümle bazlı çeviri çalışmalarına, teknolojik ilerlemelerle birlikte istatistiksel ve yapay sinir ağı yöntemleri kullanılarak geliştirilmiş çeviri sistemi çalışmaları da dâhil olmuştur. Bu yöntemler, metinden işaret diline çeviri kapsamında aşağıda sunulmuştur.

2.1. Kural Tabanlı Çeviri Yöntemi (Rule-Based Translation Method)

Kural tabanlı çevirilerde, cümlelerin belirli kurallara göre çevrilmesi sağlanır. Dil yapısının ensek olduğu dillerde, her duruma kural tanımlamak oldukça güçtür. Bu nedenle bazı kabuller ile işlemler

kısıtlanmaya çalışılır. Kural tabanlı çeviri sistemlerini hazırlamak zor olsa da, önceden çevirimi yapılmış metin verisi gerektirmediği için tercih edilen bir metottur.

İşaret dilleri konuşma dilleri kadar geniş bir kelime çeşitliliğine sahip değildir. Ancak konuşma dilindeki hemen her cümlelerin anlamının işaret diline aktarılabilmesi mümkündür. Çeviri işlemi için tanımlanacak kuralların tamamını burada açıklamak mümkün değildir. Hemen hemen her dilde bulunan gramer kurallarına aşağıda kısaca değinilmiştir.

Zaman: Konuşma dillerinde zamanlar, konuşulan dile göre çeşitlilik gösterebilmektedir. Ancak işaret dillerinde üç temel zaman tanımı bulunmaktadır. Geçmiş, şimdiki ve gelecek zaman için işaretler kullanılabilmektedir. İşaret dilinde, diğer zamanlar bu üç zamana indirgenerek kullanılmaktadır. Her dilde aynı zaman için aynı işaret kullanılsa da bu üç zaman için tanımlama yapabilmek mümkündür. Örneğin Türk İşaret Dilinde geçmiş zaman için önce işareti kullanılırken, Amerikan İşaret Dili ve Seylanca İşaret Dilinde bitti (finished) işareti kullanılarak geçmiş zaman anlamı verilmektedir [5, 6]. İngiliz, Ürdün, Hint ve Arap İşaret dili gibi bazı dillerde arkadan öne doğru olan düzlem, uzak geçmişten uzak gelecek zamana kadar bir zaman dilimi işaret etmek için kullanılmaktadır [7-9].

Eril-dişil, soru: Bazı dillerde kelimelerin cinsiyet kategorileri bulunmaktadır. Arapça, Almanca, Seylanca gibi bazı dillerde kelimelerin eril, dişil ve soru hallerinde değişimler görülmektedir. Çoğu işaret dilinde bu durumların karşılıkları yoktur. Burada kelimenin kök hali tespit edilerek çevrimi yapılır. Ancak bazı çalışmalarda kelimenin önüne ya da arkasına eklenen belirli bir işaret ile bu anlamlar işaret diline de aktarılmaya çalışılmıştır [6].

Tekillik, çoğulluk: Bazı işaret dillerinde çoğulluk anlamı kelimenin birden fazla tekrarlanmasıyla kazandırılırken, bazı kelimelerin çoğullarında farklı işaretler kullanılmaktadır. Türkçe ve Seylanca İşaret Dillerinde çoğulluk için kelimeye çok (a lot) ifadesi eklenmektedir [5, 10]. Her dilde aynı olmamakla birlikte, Arapçada olduğu gibi nesnelerin tek, çift ve çok ile ifade edildikleri durumlar da mevcuttur [8]. Bu gibi durumlarda, dilin yapısına göre kelimelere çoğulluk anlamı vermede belirlenen bir yöntem ile işaret diline çoğulluk anlamı taşınabilmektedir.

Olumsuzluk: Cümlelerde olumsuzluk anlamı taşıyan eklerin işaret diline aktarılabilmesi mümkündür. İşaret dilleri, konuşma dillerinden farklı olarak, bir işaret ile paralel mimik ya da el dışı harekete izin vermektedir. Bir ifadede kasların kaldırılması ve başın geri gitmesi ile ifadenin olumsuz anlamı verilmiş olur [12]. Otomatik makine çeviri sistemlerinde bu durumu animasyon kullanımı dışında sağlamak mümkün olamamaktadır. Videolu çeviri sistemlerinde, bu problemi aşmak için kelimelere olumsuzluk işareti eklenir. Konuşma dillerinde genellikle olumsuzluk fiilden önce gelen bir kelime ile ifade edilirken, işaret dillerinde fiilden sonra gelen ekle ifade edilmektedir [8].

Her dilin yapısına göre tanımlanması gereken kurallar olacaktır. Kurallar tanımlanırken dikkat edilmesi gereken temel nokta, oluşturulacak kuralın her iki dilin tüm durumlarına uygun olması gerektiğidir. İyi bir çeviri sistemi için dile özgü her kuralın tespit edilmesi, modellemedeki her kuralın iyi şekilde tanımlanması ve son olarak istisnaların ele alınması gerekir [13]. Kural tabanlı çeviri yönteminin uygulandığı bazı çalışmalar aşağıda özetlenmiştir.

Dangsaart vd. Tayca metinleri Tay İşaret Diline çeviren kural tabanlı bir sistem geliştirmiştir. Çalışmada, Tayca'da cümle sonu işareti bulunmadığı için cümle bölümlenici ile metin parçalara ayrılmaktadır. Cümleler kelimelere ayrılarak morfolojik ve söz dizimsel analizler yapılmakta ve işaret dilinde karşılığı bulunmayan kelimeler metinden

atılmaktadır. Analizden elde edilen kelime bilgileri işaret dilinin kurallarına göre sıralanıp kelimelerin işaretlerinin sözlükten bulunması ile çeviri tamamlanmaktadır. Geliştirilen çeviri sisteminin F-skor başarıları %96,9 olarak ölçülmüştür [14].

Punchimudiyanse ve Meegama çalışmalarında Seylanca (Sinhala) metinleri Seylanca İşaret Diline çeviren bir sistem geliştirmişlerdir. Kelime kelime çeviri yerine kelimelerin üç zamanda eril, dişil ve soru hallerine göre işaret edilebileceği bir öneri sunmuşlardır. Kelimenin önüne ve arkasına getirilen parmak alfabeti ekleriyle kelimelerin farklı zaman ve kişi çekimlerini tanımlamışlardır. Çevirisi yapılacak cümleler 3-gram'dan 1-gram'a parçalar halinde sözlükte aranmaktadır. Böylece en uzun kelime öbeğinin öncelikli çevrilmesi sağlanmaktadır. Eğer kelime öbeği sözlükte yoksa kelimenin fonetik karşılığının parmak alfabetinde gösterimi ile çeviri yapılmaktadır. Seylanca işaret dilinin yapısına uygun sıralanan kelimelerin 3D avatar ile görselleştirilmesini sağlamışlardır. 30 isim ve 20 fiil içeren 100 cümlelik test metni ile yapılan testte %70,83 kelime eşleşmesi yapılmış ve çevirinin doğruluk oranının %100 olduğu ifade edilmiştir [6].

2.2. İstatistik Tabanlı Çeviri Yöntemi (Statistics-Based Translation Method)

İstatistik tabanlı çeviri sistemlerinde, kural tabanlı yöntemlerden farklı olarak, öncelikle çevirisi yapılmış veri havuzuna (eşlenik çeviri metinlerine) ihtiyaç duyulmaktadır. Sistemin kaliteli çeviri yapabilmesi veri havuzunun genişliği ile yakından ilgilidir [15].

Bu yaklaşımda, geniş bir çeviri örnekleri koleksiyonu (corpus) analiz edilerek ve öğrenilerek otomatik olarak çeviri kuralları oluşturulmaya çalışılır. Veri miktarı ne kadar yüksek olursa, sistemden öğrenilen kurallar da o kadar iyi olur. Bu yöntemin en önemli avantajı, sistemin çeviri kurallarını öğrenirken, insan etkileşimi gerektirmemesidir. Bu nedenle çeviri örnek havuzunun mevcut olduğu durumlarda, kural üretim maliyetleri daha ucuz olur [13].

İstatistiksel yaklaşımda 3 temel analiz aşaması bulunmaktadır. Bunlar kelime karşılıklarının tespit edilmesi, kelime sıralarının tespit edilmesi ve karşılığı bulunmayan kelimelerin tespit edilmesi aşamalarıdır [16]. Bu analizlerde çeşitli istatistiksel yöntemler kullanılmaktadır. Tüm yöntemlerde en yüksek olasılık değerini alan ifade çeviri sonucu olarak sunulmaktadır [17].

İstatistiksel yöntemlerle çeviri sistemi geliştirme, kural tabanlı sistemlere göre daha az emek gerektirmektedir. Ancak sistemin kaliteli çeviri yapabilmesi, öğrenme aşamasında kullanılacak veri havuzunun yeterli büyüklükte olmasına bağlıdır. Yeterli büyüklükte verinin mevcut olduğu durumlarda tercih edilmelidir.

IBM'in çalışmalarıyla hız kazanan istatistiksel makine çeviri çalışmaları, dil özelliklerinin bilinmediği durumlar için özellikle tercih edilmektedir. İstatistiksel yöntemler, paralel derlem (corpus) içinden dil modellerinin tespit edilebilmesini mümkün kılmaktadır. Paralel veri setlerinin artması ile istatistiksel çeviriler halen popüler araştırma alanları arasında yer almaktadır [18].

İstatistiksel çevirilerde, önceleri kelime tabanlı çeviri birimleri kullanılırken, zamanla çeviri birimi olarak kelime grupları kullanılmaya başlamıştır [19]. Son zamanlarda ise kelimenin kökü, türü ve morfolojik özellikleri gibi bilgilerin de sisteme dâhil edilebildiği vektörlü modeller kullanılmaktadır [17, 19, 20].

İstatistiksel makine çevirilerinde eğitim veri setinin boyutu yapılacak çevirinin kalitesini etkilemektedir. Veri setinin çeviride kullanılacak tüm kelime çeşitliliğini kapsaması beklenir. Morfolojik olarak zengin

dillerde, morfolojik olarak daha basit olan dillere göre, kelimelerin birden çok formunun olmasından dolayı, istatistiksel çevirinin kalitesi daha düşük olmaktadır [22]. Burada diğer önemli bir nokta ise morfolojik çözümlemenin doğru yapılmasıdır. Hatalı ayrıştırma sonuçları, oluşturulacak kuralların doğruluğunu düşürmektedir [23]. İstatistiksel çeviri modelinin kullanıldığı bazı çalışmaların özetleri aşağıda verilmiştir.

Kouremenos vd., Yunanca metinleri Yunan İşaret Diline çeviren istatistiksel bir çeviri sistemi geliştirmişlerdir. Ellerinde hazır bir çeviri havuzu bulunmadığından, öncelikle veri havuzunu oluşturabilmek için kural tabanlı çeviri tasarlamışlardır. Kural tabanlı sistemden elde ettikleri verileri, insan çevirmenle kontrol ettikten sonra da istatistiksel modelin eğitim ve testinde kullanmışlardır. Kural tabanlı sistem ile çevirisi yapılan 1015 cümlelerin 100'ü test için ayrılmış ve geri kalanı ise sistemin eğitimi aşamasında kullanılmıştır. 100 cümleyle yapılan testin ortalama başarıları 0,61 olarak ölçülmüştür [15]. Yunan işaret dilinin gramer yapısı tam olarak ortaya konulmadığı için kural tabanlı çeviri sisteminin çevirileri yetersiz kalmıştır. Ancak kural tabanlı çeviri sisteminin çıktılarının istatistiksel metotlarla analiz edilmesiyle daha başarılı çeviriler elde edilmiştir.

Stein, Almanca metinleri istatistiksel yöntemlerle Alman İşaret Diline çeviren bir sistem tasarlamıştır. İstatistiksel analizin daha doğru karar verebilmesi için, metin üzerinde biçimsel ve sözdizimsel işlemler yapmıştır. Kelimeleri çoğulluk, cinsiyet gibi özelliklerle analiz edip bunlardan işaret dilinde karşılığı olanlar için yeni bir ifade ekleme, kullanılmayanlar için kullanılan halindeki ifadeyle değiştirme yoluna gidilmiştir. Yani işaret dilindeki kelimelerde cinsiyet bulunmadığından, Almanca kelimeler cinsiyetten arındırılarak işleme alınmıştır. Cümlelerin özne, nesne ve dolaylı nesne gibi özelliklerinin tespit edilmesinin yararlı olacağını ancak veri havuzu olarak kullandığı hava durumu verilerinde bu öğeler bulunmadığından kendisinin kullanmadığını belirtmiştir. Veri havuzunu önışlemeden geçirmeden elde ettiği hata oranı % 45,2 olurken verileri önışlemeden geçirmenin hata oranını % 6,3 azalttığını ifade etmiştir. İstatistiksel çevirinin başarılı bir yöntem olduğunu belirtmiş, ancak çevrimi yapılacak diller arasında gramer dönüşümlerinin önceden yapılmasının çeviri kalitesine katkı sağlayacağını vurgulamıştır [13].

2.3. Yapay Sinir Ağı Tabanlı Çeviri Yöntemi (Artificial Neural Network Based Translation Method)

Yapay sinir ağı, insan beyninin çalışma prensiplerini taklit ederek tasarlanmış bir modeldir. İnsan beyninde sinir hücreleri gelen sinyali elektriksel ve kimyasal işlemlerden geçirerek diğer hücrelere iletir ve beynin ilgili biriminde bu sinyalin yorumlanması ile bilgi oluşur [24]. Yapay sinir ağları da giriş ve çıkış katmanları olarak adlandırılan belirli sayıda hücreler ve bu iki katmanı birbirine bağlayan ara katman hücrelerinden oluşur. Giriş katmanından girilen parametrelerin hücreden hücreye aktararak çıkış katmanında oluşturduğu değerlerin yorumlanması ile bilgi elde edilmiş olur.

Yapay sinir ağı tabanlı çeviri sistemlerinde de istatistik tabanlı sistemler gibi öğrenme aşaması için hazır veriye ihtiyaç vardır. Yeterli eşlenik veri olmaksızın hazırlanan sistemlerin düzgün çeviri yapabilmesi çok zordur. Sistemin hazır verileri defalarca kullanarak en uygun sonuca yakınsaması ile eğitimi yapılır [25].

Yapay sinir ağı tabanlı sistemlerin, istatistik tabanlı sistemlerden ayrıldığı nokta, girdiye göre bir sonuç vermesidir. Cümleden cümleye bir çeviri sağlamaktadır. Bu yöntem, istatistiksel yöntemlere göre daha iyi sonuç vermekle birlikte, yapay sinir ağlarında sistemin eğitilmesi uzun, karmaşık ve maliyetli bir süreçtir. Yapay sinir ağları çeviri modeli kullanılarak yapılan bazı çalışmaların özetleri aşağıda verilmiştir.

Brour ve Benabbou, Arapça metinlerden Arap İşaret Diline çeviri yapan, sinir ağı tabanlı makine çeviri sistemi geliştirmişlerdir. Yapay sinir ağı girişlerinde, her kelimeden maksimum bilgiyi elde etmek için kelimenin morfolojik özellikleri kullanılmıştır. Kelimelerin tür, zamir, ön ek, son ek, kalıp, sayı, cinsiyet, çekim ve zaman özellikleri bir vektör olarak kodlanarak sisteme verilmiştir. Geliştirilen sistemin yüksek doğrulukta çeviri yapabildiği ifade edilmiştir. Farklı zaman türlerinde, ünlem, soru ve normal cümlelerden oluşan 73 cümlelik veri seti ile sistem test edilmiştir. Daha önce kendilerinin geliştirdikleri kural tabanlı sistemin BLEU skoru 0,37 de kalırken, yeni geliştirdikleri sistemin BLEU skorunu 0,79 olarak ölçmüşlerdir [25].

Stoll vd. Almanca metinleri Alman İşaret Diline çeviren yapay sinir ağı tabanlı bir sistem geliştirmişlerdir. Bu çeviri sisteminde diğer çalışmalardan farklı olarak işaretin gösteriminde doğrudan video ya da görsel kullanılmamıştır. Sistem eğitilirken işaret dili video görüntülerinden iskelet hareketleri çıkarılmıştır. Her kelimeyle birlikte iskelet hareket bilgileri de kullanılmıştır. Sistemin eğitiminden sonra, girilen cümlelerin işaret dili karşılığı metin oluşturulmuş, daha sonra bu metinden iskelet hareketleri bilgileriyle hareketli görüntü elde edilmiştir. 8257 cümlelik veri seti ile sistem eğitilmiştir. BLEU skoru 0,5 olurken kelime hata oranı 0,048 olarak ölçülmüştür [26]. Sistemin hazır video ya da animasyon kullanmayıp, eğitim videolarındaki görüntüleri kullanarak hareketli videolar hazırlaması alana farklı bir bakış açısı kazandırmıştır.

2.4. Çeviri Yöntemlerinin Değerlendirilmesi ve TİD (Evaluation of Translation Methods and TSL)

Literatür genel olarak ele alındığında, metinden işaret diline çeviri çalışmalarının sayısının, işaret dilinin kullanıcı sayısı ile yakından ilişkili olduğu görülmüştür. Kullanıcı sayısının az olduğu işaret dillerinde, yeterli çalışma yapılmadığı ve kaynak eksikliği olduğu vurgulanmaktadır. Metinden işaret diline çeviri çalışmalarında karşılaşılan zorlukların en başında yeterli eşlenik metinlerin bulunmadığı ifade edilmiştir [27]. Var olan eşlenik metin derlemlerinin büyük çoğunluğu akademik araştırmalar için oluşturulmuştur. Erişilebilir daha büyük eşleştirilmiş derlemlere ihtiyaç duyulmaktadır [28]. Sözlü dilin esnek bir yapısının bulunması, işaret diline çeviri çalışmalarında karşılaşılan diğer bir zorluktur. Dilin esnek olması oluşturulacak çeviri kurallarının artmasına neden olmaktadır. Dilin kurallarının iyi şekilde tanımlanamaması yapılacak çevirinin kalitesini düşürmektedir. İşaret dillerinde eş zamanlı hareketlerin yapılabilir olması ancak sözlü dilin ardışık zamanlarda ifade ediliyor olması, aşılması gereken başka bir problemdir. İşaret dili görüntüsünün doğrudan işaret dili metnine çevrilmesi sırasında eş zamanlılığın kaybolması söz konusu olmaktadır. Bunun için farklı yöntemlerin kullanılması gerekmektedir [12]. Çevirisi yapılacak diller arasında tüm kelimelerin tam karşılıklarının bulunmaması, problemlere neden olmaktadır. Çeviri işleminden önce, karşılığı olmayan kelimelerin, karşılığı olan yakın anlamlarına çevrilmesiyle, bu problemin önüne geçilmeye çalışılmaktadır [16].

Açık kaynak erişimi olan dillerde yapılan çalışma sayıları ve çeviri başarı oranları her geçen gün artmaktadır. Elinde kaynağı olmayan araştırmacılar, zamanının büyük bir kısmını veri hazırlamaya ayırmak zorunda kalmaktadır. Bu zorluktan dolayı alana ilgi duyan araştırmacıların da alandan uzaklaşma ihtimali bulunmaktadır.

Özetleyecek olursak çeviri yöntemi olarak; kural tabanlı, istatistiksel tabanlı ve yapay sinir ağı tabanlı yöntem olmak üzere üç metot kullanılmaktadır. Kahlon ve Singh'in 2021 yılında yayımlanan çalışmasında, metinden işaret diline yapılan makine çeviri çalışmaları incelenmiş ve çalışmaların yaklaşık %55'nin kural tabanlı olduğunu ortaya konmuştur [29]. İstatistiksel tabanlı ve yapay sinir ağı tabanlı yöntemlerin, kural tabanlı yöntemlere göre tasarlanmasında kolaylık

bulunmaktadır. Ancak bu modellerde sistemin kural çıkarımı yapabilmesi için hazır veri setine ihtiyaç duyulmaktadır. Kural tabanlı yöntemde ise veri setine ihtiyaç duyulmaz. Ancak çevirimi yapılacak diller arasındaki tüm kuralların tespit edilerek tanımlanması gerekmektedir. Elinde çeviri metni olmayan araştırmacıların; istatistiksel tabanlı ve yapay sinir ağı tabanlı modelleri kullanabilmek için kural tabanlı yöntemlerle veri üretme yoluna gittikleri de görülmüştür [29].

Türkçeden Türk İşaret Diline yapılan çeviri çalışmaları verilen Tablo 1 incelendiğinde, genellikle kural tabanlı çeviri yönteminin kullanıldığı görülmektedir [30-32, 35]. Araştırmacılar bunun nedeninin, Türkçeden TİD'e yapılmış çeviri verisine ulaşamamaları olduğunu ifade etmişlerdir. Kullanıcı yoğunluğu az olan diller üzerine yapılan çalışmaların da az olduğu tartışılmaz bir gerçektir. TİD gibi kullanıcı sayısı sınırlı olan bir dil üzerine yapılacak çalışmaların önünü açabilmek için öncelikle açık erişim olan kaynaklara ihtiyaç duyulmaktadır. Çeviri sistemlerinin, daha hızlı ilerleyebilmesi ve artması kaynak paylaşımıyla doğrudan ilgilidir.

İncelenen çalışmalarda, Türkçe metinden Türk İşaret Diline yapılan yedi çeviri çalışmasından üç tanesinde istatistiksel tabanlı çeviri yöntemi kullanılmıştır [33-35]. Ancak eşlenik veri yetersizliğinden dolayı kural tabanlı çeviri yönteminin de birlikte kullanıldığı görülmüştür. Yapay sinir ağı tabanlı model istatistiksel tabanlı modele göre daha başarılı sonuç vermesine rağmen TİD'de yapay sinir ağı tabanlı modelde yapılmış bir çalışmaya rastlanılmamıştır. Alanda yapılan çalışmaların ve erişilebilir kaynakların artmasıyla yapay sinir ağı tabanlı modelin de çeviri çalışmalarında kullanılabileceği düşünülmektedir.

3. Materyal ve Yöntem (Material and Method)

Bu çalışmada, Türkçe metinlerin Türk İşaret Dili görüntü videolarına dönüşümünü sağlamak amacıyla bir sistem geliştirilmiştir. Çeviri yöntemi olarak kural tabanlı ve istatistiksel tabanlı çeviri yöntemleri birlikte kullanılmıştır. Az kullanıcısı bulunan dillerin genelinde karşılaşılan eşlenik veri eksikliği problemi Türk İşaret Dili için de geçerlidir. Çalışmada, eşlenik veri sayısı istatistiksel tabanlı çeviri başarısının düşük olmasına neden olduğu için kural tabanlı çeviri sistemi ile desteklenmiştir.

3.1. Veri seti (Data set)

Paralel çeviri için ihtiyaç duyulan eşlenik veri setinin Aile ve Sosyal Politikalar Bakanlığı'nın kamuya açık olarak paylaşılan TİD Sözlük sayfasından [36] alınması planlanmıştır. Fakat bu veriler paralel bir derlemden ziyade örnek kelimenin kullanım alanlarını göstermek amaçlı hazırlanmıştır. Genellikle birden fazla Türkçe cümle içeren metinlerden oluşmaktadır. Cümleler grubu TİD'e çevrilirken ayrı ayrı cümleler olarak değil, tüm cümlelerin anlamlarını kapsayacak tek bir işaretleme dizisi olarak çevrilmiştir. Bu durum, istatistiksel çeviride paralel eşleştirmenin başarısını azaltacağı için Aile ve Sosyal Politikalar Bakanlığı'nın veri seti kullanılamamıştır.

Türkçe-TİD için hazırlanmış erişime açık eşlenik veri seti olmadığından, özgün bir veri setinin oluşturulması ihtiyacı oluşmuştur. Bu çalışma kapsamında, iki dil eğitimi uzmanı tarafından, kavramlara ve diyaloglara dayalı, 50'den fazla kategoride 1084 cümle örneği hazırlanmıştır. Renkler, şekiller, mevsimler, hayvanlar, meyve-sebzeler, eşyalar gibi kavramlar ile hastanede, okulda, pazarda, mağazada, uçakta, otobüste, telefonda, sosyal medyada kısaca günlük yaşamımızda ihtiyaç duyabileceğimiz farklı alanlarda diyaloglar içeren cümleler oluşturulmuştur. Farklı kategorilerdeki 1084 cümlelerin işaret dili karşılıkları için iki işaret dili uzmanından yardım alınmıştır. 1084 cümlelik veri setinin içerdiği kelime kalıpları

Tablo 1. Türkçeden TİD'e çeviri çalışmaları (Translation studies from Turkish to TİD)

Yazar	Yıl	Yöntem	Katkı
Yasan M. C.	2014	Kural tabanlı çeviri	Türkçe cümleleri, 83 kelime kökü ve 41 son eki kural tabanlı olarak işaret dili karşılıklarına çevirmek için bir sistem geliştirilmiştir.
Şimşek M. S.	2016	İstatistiksel çeviri	2068 cümleden oluşan veri seti kullanılarak, TR=>TİD ve TİD=>TR şeklinde çift yönlü çeviri yapabilen örneğe dayalı çeviri sistemi geliştirilmiştir.
Eryiğit C.	2017	Kural tabanlı çeviri	Türkçe metinlerden makine okunur TİD karşılıklarını oluşturan kural tabanlı bir sistem geliştirilmiştir. Çeviri sistemi için 990 işaret içeren bir sözlük oluşturulmuştur.
Karaca M. F.	2018	Kural tabanlı çeviri	İşitme engelliler okullarında okutulan 1., 2., 3. Sınıf Türkçe ders kitaplarındaki cümlelerden oluşturulan veri seti kullanılarak Türkçe metinleri kural tabanlı yöntemlerle 3 boyutlu animasyon hareketlerine çeviren uygulama geliştirilmiştir.
Buz B., Güngör T.	2019	İstatistiksel çeviri	1.Sınıf hayat bilgisi kitabındaki cümlelerden oluşturulan veri seti ile istatistiksel çeviri modeli geliştirilmiştir.
Kayahan D., Güngör T.	2019	Kural tabanlı + İstatistiksel çeviri	Aile bakanlığının verileri kullanılarak, kural tabanlı ve istatistiksel çeviri yönteminin birlikte kullanıldığı çeviri sistemi geliştirilmiştir.

Tablo 2. Veri setinin genişletilmesi örneği (Example of expanding the dataset)

No	TİD	Türkçe
1	ben bu akşam uçmak	Ben bu akşam uçuyorum.
2	ben yarın uçmak	Ben yarın uçuyorum.
3	ayşe bu akşam uçmak	Ayşe bu akşam uçuyor.
4	sayın yolcu yol devam bir saat onbeş dakika son	Sayın yolcularımız, yolculuğumuz 1 saat 15 dakika sürecektir.
5	sayın yolcu yol devam iki saat otuz dakika son	Sayın yolcularımız, yolculuğumuz 2 saat 30 dakika sürecektir.
6	sayın yolcu seyahat devam iki saat otuz dakika son	Sayın yolcularımız, seyahatimiz 2 saat 30 dakika sürecektir.
7	lütfen sen bavul koymak o tartı üst	Lütfen, bavulunuzu tartının üzerine yerleştirin.
8	lütfen sen çanta koymak o tartı üst	Lütfen, çantanızı tartının üzerine yerleştirin.
9	lütfen sen bavul koymak o masa üst	Lütfen, bavulunuzu masanın üzerine yerleştirin.

korunarak çoklanmış ve 3126 cümlelik Türkçe cümlelerin metin olarak işaret dili karşılığı oluşturularak eşlenik veri seti elde edilmiştir. Böylece, alanda yaşanan Türkçe-TİD eşlenik veri seti eksikliğinin giderilmesine katkı sağlanmıştır.

1084 cümlelik veri seti farklı kategorilerde ve benzerleri bulunmayan cümlelerden oluşturulmuştur. Bu durum istatistiksel çevirinin başarımının düşük çıkmasına neden olmaktadır. Cümlelerin kelime kalıpları korunarak yine uzmanlar tarafından yeni cümleler oluşturulmuştur. Tablo 2'de verilen cümlelerden No alanı mod 3'e göre 1 olan cümleler ilk (1084 cümlelik) veri setindeki cümlelerdir. Bu cümlelerin ardından gelen diğer cümleler ise yeni oluşturulan cümlelerdir. 1084 cümlelerin her biri için 1 veya 2 yeni cümle eklenerek veri seti 3126 cümleyle çıkarılmıştır.

Bu çalışmada 1084 cümleden oluşan ilk veri seti 1000 cümlelik, 3126 cümleden oluşan ikinci veri seti 3000 cümlelik veri seti olarak anılacaktır. 1000 ve 3000 cümlelik her iki veri setinin %75'i istatistiksel modelin eğitimi ve %25'i oluşturulan modelin testi için ayrılarak, test işleminde kullanılan cümlelerin, modelin eğitiminde bulunmayan cümlelerle yapılması sağlanmıştır.

Cümle içindeki kelimelerin bağımlılıklarını tespit edebilmek için, Sabancı Üniversitesi ve Orta Doğu Teknik Üniversitesi ortaklığında corpus araştırma grubu tarafından hazırlanan 5635 cümleden oluşan Türkçe ağaç yapılı derlem verisi kullanılmıştır.

3.2. Kullanılan Araçlar (Tools Used)

Türkçe sondan eklemeli dillerden olup, morfolojik olarak zengin bir dildir. Bir kelime, yeni ekler alarak çok sayıda formda bulunabilmektedir. Kelime çeşitliliği, istatistiksel çevirinin başarısını düşüren bir faktördür. İstatistiksel çevirinin başarımının artması için Türkçe metinlerdeki kelimelerin morfolojik çözümlemelerinden elde edilen, kelime köklerinin ve eklenen eklerin isimlerinin kullanılmasının yararlı olduğu birçok çalışma ile ortaya konulmuştur [17, 34, 37]. Türkçe cümlelerin morfolojik analizi için Zemberek-NLP [38] kütüphanesi kullanılmıştır. Zemberek Türk dilleri için oluşturulmuş doğal dil işleme kütüphanesidir. Türkçe "Annemle eve gidiyorum." cümlesinin Zemberek DDİ aracı ile morfolojik analizi aşağıda verilmiştir:

Annemle => [anne:Noun]: anne(Noun)+A3sg+m(P1sg)+le(Ins)
eve => [ev:Noun]: ev(Noun)+A3sg+e(Dat)
gidiyorum => [gitmek:Verb]: gid(Verb)+iyor(Prog1)+um(A1sg)
. => [.:Punc] :.Punc

Paralel hizalama için ise Moses [39] istatistiksel çeviri aracı kullanılmıştır. Moses, istatistiksel makine çeviri araçlarını içeren ücretsiz bir yazılımdır. Bu çalışmanın konusu olan uygulamada, paralel hizalama ve istatistiksel çeviri için Giza++ [40] ve Moses istatistiksel çeviri aracı kullanılmıştır. İstatistiksel dil modelinin

oluşturulması için KenLM [41] algoritması 2 ngram parametresi ile kullanılmıştır. Dil modelini oluşturmada kullanılan KenLM aracı için ngram parametresinin 3-gram kullanılması önerilmiş olmasına rağmen yaptığımız deneylerde 2-gram kullanıldığında daha yüksek başarı elde edildiği görülmüştür.

İstatistiksel çeviri ile Türkçe metinlerin işaret dili metin karşılıkları oluşturulmaktadır. Çeviri metinlerinde anlamsal karmaşa yaşanabilen durumların giderilmesi için kelimeler arasındaki bağımlılıklardan yararlanılmıştır. TİD’de sıfatlar bazen tanımladığı kelimeden önce bazen de sonra gelebilmektedir. “Kız küçük defter almak” gibi çeviri yapıldığında cümlede küçük kızın defter aldığı mı kızın küçük defter aldığı mı karmaşası olabilmektedir. Bağımlılık analizi için MaltParser uygulaması kullanılmıştır. ODTÜ’den alınan veri seti ile bağımlılık modeli oluşturulmuştur. Böylece girdi cümlesinin kelimeleri arasındaki bağımlılıklar tahminlenerek kelime grupları arasında zamansal boşluklar bırakılmıştır.

3.3. Uygulama ve Deneysel Sonuçlar (Application and Experimental Results)

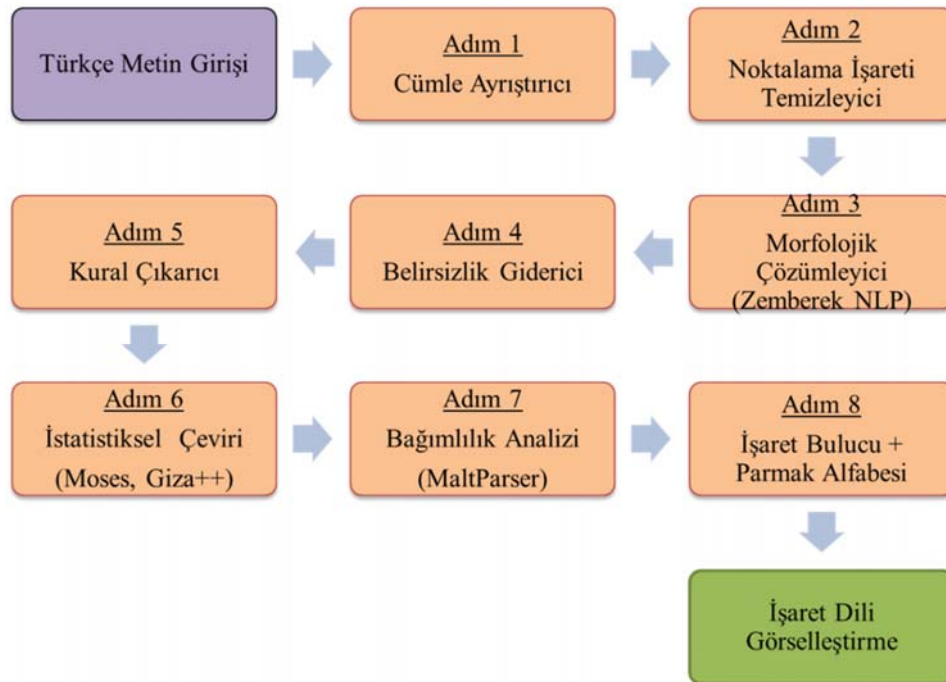
Türkçe metnin Türk İşaret Diline çevrilmesi için uygulanan işlem adımlarını gösteren akış şeması Şekil 2’de verilmiştir. Çeviri adımlarına, girdi olarak verilen Türkçe metinden cümlelerin tespit edilmesi ile başlanmaktadır. İkinci adımda cümlelerdeki noktalama işaretleri temizlenir. Üçüncü adımda, her cümle tek tek Zemberek DDİ kütüphanesi ile morfolojik analize tabi tutulur ve dördüncü adımda elde edilen morfolojik çözümlerden en uygun olanı çeviri kuralları uygulanmak üzere seçilir. Beşinci adımda, morfolojik çözümlemelerden elde edilen verilere, ayrıntıları çalışmanın ilerleyen bölümlerinde verilen çeviri kuralları uygulanır. Altıncı adımda, kural tabanlı yarı çevrilmiş cümle, istatistiksel çeviri aracına (Moses, Giza++) girdi olarak verilerek istatistiksel çeviri işlemine tabi tutulur. Yedinci adımda, MaltParser bağımlılık analizi aracı kullanılarak kelime bağımlılıkları tespit edilmektedir. Kelime bağımlılıklarına göre kelime grupları oluşturulmaktadır. Sekizinci adımda, görselleştirme için gerekli olan işaretler sözlükte aranır, bulunamayan

kelime varsa bunlar için parmak alfabesi kullanılarak işaretlenir. Çeviri adımları, istatistiksel çeviriden elde edilen metnin görselleştirilmesiyle sona erer.

Ekleme dillerde morfolojik çözümlemenin kullanılması ile istatistiksel çevirinin başarısının arttığı bilinmektedir. Ancak Türkçedeki birçok ekin işaret dilinde karşılığı yoktur. Çeviri karşılığı olmayan ekler istatistiksel çevirinin başarısını olumsuz yönde etkilemektedir. İşaret diline doğrudan çevrilebilecek kurallarla istatistiksel çevirinin başarısına katkı sağlanabilmektedir. İşaret dilinde genellikle kelimeler ek almadan yalın olarak kullanıldığından dolayı kural tabanlı çeviri, kelimenin kök hâlini vermektedir. Bu amaçla Türkçe metinlerin morfolojik çözümlemelerinden yararlanılarak kelime kökünün önüne ya da sonrasına, eklerin anlamını karşılamaya yarayan, yeni kelimeler getirilerek bazı çeviri kuralları tanımlanabilmiştir. TİD’de karşılığı olmayan ekleri barındıran kelimeler ise kök haliyle kalmaktadır.

TİD, Türkçe ile aynı coğrafyada kullanılmasına rağmen gramer olarak farklılıklara sahiptir [42]. TİD’in gramer yapısı henüz tam olarak ortaya konulmasa da bazı özellikleri bilinmektedir. Alanda yapılmış çalışmalar incelendiğinde, Türkçeden işaret diline çeviride kullanılabilecek bazı özelliklerin tanımlandığı görülmüştür. Ancak bu tanımlamaların tümü ortak kabul görmüş değildir. Örneğin, yönelme hâl eki (-e), bulunma hâl eki (-de), ayrılma hâl eki (-den) gibi ekler işaret diline aktarılırken kullanılabilecek tanımlamalar yapılmıştır [35, 44]. Zaman için Türkçedeki zaman eklerinin TİD’e aktarılabilceğini ifade eden çalışmalar bulunmakla beraber genel görüş Türkçedeki zaman eklerinin TİD’de karşılığı bulunmadığı yönündedir [43]. Bu nedenle çalışmada ortak kabul görmüş tanımlamalardan yola çıkarak yaptığımız kural tanımları aşağıda örneklerle açıklanmıştır.

Olumsuzlama: Türkçede bir ekle verilen anlam, TİD’de farklı bir kelimenin kullanılması ile sağlanabilmektedir. Örneğin Türkçede yap+mak kelimesi olumsuzlanmak istendiğinde yap+ma+mak şekline oluşurken, İşaret dilinde “yapmamak” kelimesinin anlamı, “yapmak”



Şekil 2. Çeviri sürecinin adımları (The steps of the translation process)

kelimesinin işaretiyle eş zamanlı olarak kafanın geriye doğru atılması ve kaşların yukarı kaldırılması gibi el dışı hareketlerle verilmektedir. Ancak elimizde sabit görüntü verileri olduğu için, işaret dilinde “yapmamak” kelimesinin anlamı “yapmak” ve “değil” işaretlerinin ardışık şekilde gösterilmesi ile ifade edilebilmektedir.

İyelik: Türkçede sahiplik anlamı kelimelere eklenen iyelik ekleri ile verilirken, TİD’de “benim, senin, onun” gibi kelimelerin kullanılması ile sağlanabilmektedir. Örneğin Türkçede “kitabım” kelimesi işaret dilinde “benim” ve “kitap” işaretlerinin ardışık şeklinde gösterilmesi ile ifade edilmektedir.

Tamlama: Türkçe için hazırlanmış morfolojik çözümleme araçları tamlanan ekini, 3. tekil şahıs iyelik eki olarak çözümlenmektedir. Dolayısıyla bu morfolojik çözümlemeyi kullanarak çeviri kuralı uygulandığında “kuşun tüyü” ifadesi “kuş onun tüyü” olarak çevrilmektedir. Bu da yanlış bir çeviriye neden olmaktadır. Bu problemi çözmek için bağımlılık analizi ile elde ettiğimiz kelime birliktelikleri incelenmekte, bu kelimenin bağımlı olduğu kelimeler içerisinde tamlayan kelimenin varlığına bakılarak 3. tekil şahıs ekinin kullanılıp kullanılmayacağına karar verilmektedir.

Çoğulluk: Türkçede –lar, –ler ekleri ile verilen çokluk anlamı, TİD’de bazen kelimenin farklı konumlarda tekrarlanması ile bazen de

kelimedden sonra çok işaretinin yapılması ile sağlanmaktadır. Bu çalışmada, önceden çekilmiş görüntüler üzerinden işaretleme yapıldığı için, görüntüdeki hareketin konumunu değiştirme imkânı bulunmamaktadır. Bu nedenle, tüm çoğul kelimeler için, Türkçedeki çoğulluk ekinin bulunduğu kelimenin kökünden sonra çok işaretinin kullanılması yoluna gidilmiştir. elmalar => elma çok, evler => ev çok

Şahıs Ekleri: Türkçede yükleme eklenen şahıs ekleri ile eylemi yapan kişi ifade edilmektedir. İşaret dilinde ise eğer özne olarak kişi kullanılmamışsa yüklemden sonra şahıs zamiri işareti kullanılmaktadır, eğer cümle başında şahıs zamiri geçmişse yüklemden sonra tekrar kullanılmamaktadır.

Aşağıda öznenin açık olarak kullanıldığı bir cümle için yapılan çeviriden ekran görüntüsü Şekil 3’te (Ben yemek yaptım. => ben yemek yapmak), gizli öznenin kullanıldığı bir örnek cümle ve çeviri videosundan bir karenin yakalandığı ekran görüntüsü ise Şekil 4’te (Yemek yaptım. => yemek yapmak ben) verilmiştir.

Zorunluluk: Türkçe kelimelerdeki zorunluluk eki –meli, –malı için işaret dilinde “lâzım” ya da “mecbur” işareti kullanılmaktadır. Bu çalışmada “lâzım” işareti kullanılması tercih edilmiştir. Yapmalısın => yapmak lâzım sen



Şekil 3. “Ben yemek yaptım.” cümlesinin uygulama görüntüsü (“Ben yemek yaptım.”- I cooked.- application image of the sentence)



Şekil 4. “Yemek yaptım.” cümlesinin uygulama görüntüsü (“Yemek yaptım.”- I cooked.- application image of the sentence)

İmkânsızlık: Türkçede imkânsızlık anlamını vermek için –e(memek), –a(mamak) ekleri kullanılırken işaret dilinde fiilden sonra imkânsız işareti kullanılmaktadır. Yapamayız => yapmak imkânsız biz

Birliktelik: Türkçede birliktelik anlamı için kendinden önceki kelimeye birleşik ya da ayrı yazılabilen “ile” sözcüğü kullanılmakta iken işaret dilinde “birlikte” işareti kullanılmaktadır. elmayla armut => elma birlikte armut

Benzetme: Türkçede –(i)msi, –(i)msi ekleri kullanılarak yapılan benzetimler için işaret dilinde “gibi” işareti kullanılmaktadır. kırmızımsı => kırmızı gibi

Yokluk: Türkçede –sız, –siz ekleri kullanılarak ifade edilen bir şeyin yokluğu anlamını vermek için işaret dilinde “hiç” işareti kullanılmaktadır. tuzsuz => tuz hiç

Zaman: Zaman eklerinin TİD’e aktarılabilceğini ifade eden çalışmalar olsa da, bu çalışmada, zaman eklerinin TİD’de karşılığının olmadığını savunan görüş dikkate alınarak, Türkçe tarafındaki zaman ekleri çıkarılmıştır. Zaman eklerinin çıkarılması, çeviri modelinin oluşmasını hızlandırarak ve başarısını artırmıştır.

İş çıkışında eve gideceğim. => ben iş paydos ev gitmek

İş çıkışında eve gittim. => ben iş paydos ev gitmek

Örneklerde de görüldüğü üzere geçmiş ve gelecek zamanda verilen iki cümlelerin de işaret dili çıktıları aynı olmaktadır.

Sayılar: Türkçede rakamlarla verilen bir sayının TİD’e çevrilebilmesi için veri setinde tüm sayıların örneklerinin bulunması gerekmektedir. Veri setinin, tüm sayıların görüntülerini içermesi ise mümkün değildir. Tüm sayıların işaret diline çevrilebilmesi için sayıların okunuşlarına göre çözümlenmesi gerekmektedir. Bu amaçla rakamlarla yazılmış sayıların okunuşları için, sayı çözümleme fonksiyonu tanımlanmıştır. Sayı çözümleme fonksiyonu milyarlar seviyesine kadar olan sayıları çözümleyebilmektedir. 10’dan 20’ye kadar olan sayılar ve 200, 300 gibi yüzlü sayıların Türkçedeki yazılışları iki kelimeden oluşmaktadır. Ancak bu sayılar işaret dilinde tek işaretlerle gösterildiğinden, işaret dili karşılıkları bitişik şekilde (onbeş, ikiyüz gibi) yazılarak verilmiştir. Böylece örneğin 15 sayısı işaretlenirken on ve beş işaretlerinin ayrı ayrı gösterilmesi yerine ‘on

beş’ işaretinin gösterilmesi sağlanmıştır. 1234567 => bir milyon ikiyüz otuz dört bin beşyüz altmış yedi

Alanda yapılmış çeviri çalışmalarında sayıların işaret diline çevrilmesi ile ilgili bilgi bulunmamaktadır. Bu çalışma ile ilk defa Türkçeden TİD’e çeviride sayıların dönüştürülmesi konusu ele alınmış ve milyarlar seviyesine kadar olan sayıların dönüşümü sağlanmıştır.

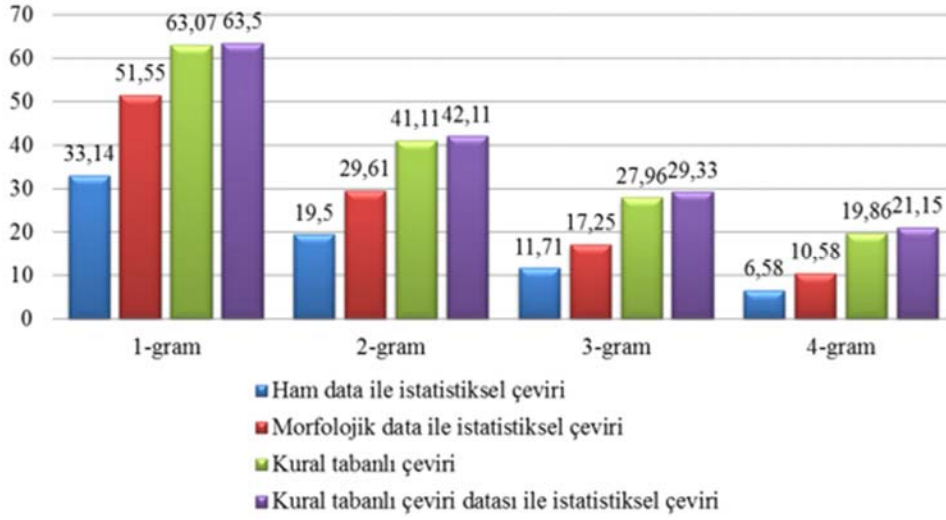
Yukarıda açıklanan çeviri kurallardaki ekler ve TİD’de bu eklerle karşılık kullanılan kelimelerin listesi Tablo 3’te verilerek, kullanılan kurallar özetlenmiştir. Yukarıda verilen kurallara uyan kelimeler ve ekleri için işaret dili karşılıklarını kullanan kural tabanlı çeviri sistemi oluşturulmuştur. Veri setindeki Türkçe cümleler, kural tabanlı çeviri sisteminin geçirilerek, yarı çevrilmiş cümleler elde edilmiştir. Bu yarı çevrilmiş cümlelerle işaret dili metinleri istatistiksel çeviri için Mosses’a girdi olarak verilerek çeviri modeli oluşturulmuştur.

İstatistiksel çeviriden önce kural tabanlı çeviri yapılmasının nedenleri; işaret dilinde karşılığı olmayan ekleri çıkartmak, karşılığı olan ekler için ise o işaretlerin kullanılmasını sağlamak ve kelimeleri kök hâllerine yalınlaştırılabilir. İstatistiksel çevirilerde, veri setinde örneği bulunmayan kelimeler çevrilmeden aynen bırakılmaktadır. Bu da ek almış kelimelerin hatalı çevrilmesine neden olmaktadır. Türk İşaret Dilinde, kural tabanlı çeviriyle kelime köklerinin kullanması ile büyük oranda anlam korunmuş olmaktadır.

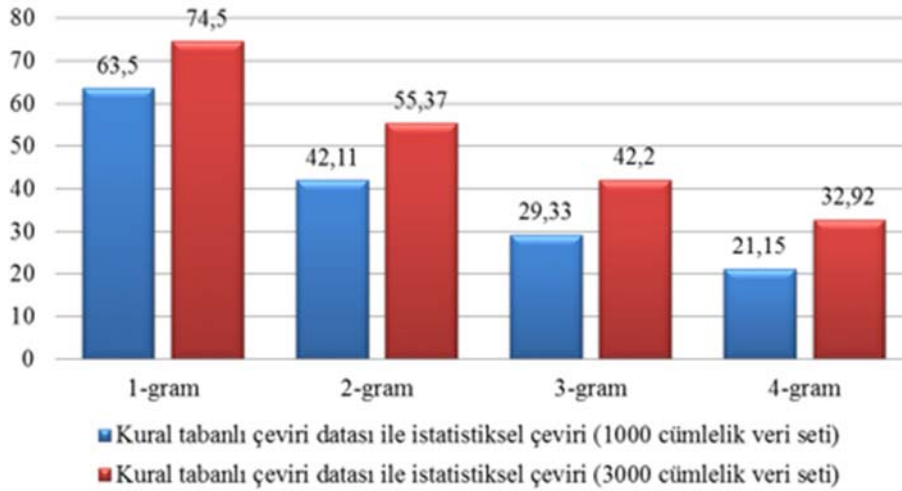
Kural tabanlı çevirinin istatistiksel tabanlı çeviriye katkısını belirleyebilmek için, 1084 cümlelik veri setinin yaklaşık % 25’i (270 cümle) test verisi, kalanlar (814 cümle) da eğitim verisi olarak kullanılmıştır. Eğitim verileri; öncelikle ham halde, sonra morfolojik olarak çözümlenmiş haliyle, son olarak da kural tabanlı çeviriden geçirilmiş haliyle istatistiksel model oluşturulmuştur. Bu verilerle oluşturulan üç çeviri modeli için, modelin eğitime dâhil edilmemiş olan test verileriyle test edilerek BLEU skorları hesaplanmıştır. Ayrıca kural tabanlı çevirinin başarısını ölçümleyebilmek adına test verileri kural tabanlı çeviriden geçirildikten sonra orijinal çeviri ile karşılaştırılarak BLEU skoru hesaplanmıştır. Aşağıda Şekil 5’te dört duruma ait 1-4 gram BLEU skorları verilmiştir.

Tablo 3. Kurallar Listesi (Rules List)

Morfolojik Analiz	Ek	TİD
Çoğulluk	-ler -lar	... çok
Olumsuzluk	-me, -ma	... değil
1. Tekil Şahıs	-(i)m	... ben
2. Tekil Şahıs	-(i)n, -sin	... sen
3. Tekil Şahıs	-	... o
1. Çoğul Şahıs	-iz , -(y)iz	... biz
2. Çoğul Şahıs	-sınız	... siz
3. Çoğul Şahıs	-ler	... onlar
1. Tekil Şahıs İyelik Eki	-(i)m	benim ...
2. Tekil Şahıs İyelik Eki	-(i)n	senin ...
3. Tekil Şahıs İyelik Eki	-i	onun ...
1. Çoğul Şahıs İyelik Eki	-(i)miz	bizim ...
2. Çoğul Şahıs İyelik Eki	-(i)niz	sizin ...
3. Çoğul Şahıs İyelik Eki	-leri	onların ...
Zorunluluk	-meli	... lazım
İmkânsızlık	-(e)memek	... imkansız
Birliktelik	-le	... birlikte
Benzerlik	-(i)msi	... gibi
Yokluk	-siz	... hiç
Rakamsal sayılar	-	sayının okunuşu



Şekil 5. BLEU skor grafiği (BLEU score chart)



Şekil 6. 1000 ve 3000 cümlelik veri setlerinin BLEU skorları (BLEU scores of 1000 and 3000 sentence datasets)

Ham haldeki verilerle Moses istatistiksel çeviri aracı kullanılarak yapılan çevirinin BLEU (4-gram) skoru % 6,58 olarak hesaplanmıştır. Sonraki modelde, kelimeler morfolojik çözümlemeleriyle istatistiksel çeviri için kullanılmıştır. Morfolojik çözümlemeleriyle yapılan istatistiksel çevirinin BLEU skoru % 10,58 olmuştur. Kural tabanlı çeviriden elde edilen yarı çevrilmiş cümlelerin çevirisinin BLEU skoru % 19,86 olmuştur. Kural tabanlı çeviriden elde edilen yarı çevrilmiş cümlelerin, istatistiksel çeviri için girdi olarak kullanılmasıyla bir çeviri modeli oluşturulmuştur. Bu istatistiksel çeviri modelinin BLEU skoru ise % 21,15 olarak hesaplanmıştır. Morfolojik çözümlemenin yerine, kural tabanlı çeviriden elde edilen verilerin kullanılması ile istatistiksel çeviri başarısı % 10,58'den % 21,15'e yükselmiştir.

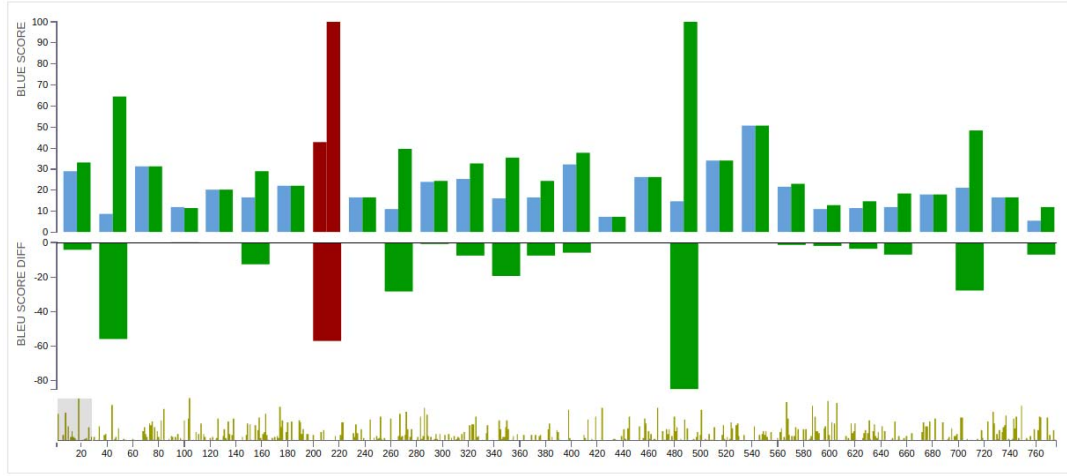
Veri seti 1084 cümleden 3126 cümleye genişletilmiş ve verilerin yaklaşık % 25'i (776 cümle) test, kalanları (2350 cümle) eğitim verisi olarak belirlenmiştir. 2350 cümle ile yeniden istatistiksel model oluşturulmuş ve 776 cümlelik test verisiyle test edilmiştir. Verilerin çoklanmasından sonra hesaplanan BLEU skoru % 32,92'ye çıkmıştır. Aşağıda Şekil 6'da 1000 cümlelik ve 3000 cümlelik veri setleri ile kural tabanlı çeviri sonrası istatistiksel çevirinin BLEU skorları verilmiştir.

Her veri setinde bulunan "Cam kenarında oturmak isterseniz, bu masa müsait." cümlesinin bahsi geçen tüm çeviri modellerindeki çeviri metinleri ve BLEU skorları aşağıda Tablo 4'te karşılaştırmalı olarak verilmiştir. Tabloda da görüldüğü üzere istatistiksel çevirinin başarısı, ham datanın kullanıldığı durumda 20,61 olurken, kelimelerin morfolojik olarak çözümlemiş halinin kullanıldığı durumda 29,07, kural tabanlı çeviriden alınan datanın kullanıldığı durumda 53,73 olmuştur. Bu örnek cümlelerin, kural tabanlı çevirisinin başarısı ise 45,62 iken istatistiksel çeviri modeli ile 53,73'e çıkmıştır. Bu da kural tabanlı çevirinin, istatistiksel çeviriye katkısını göstermesi açısından önemlidir. Ayrıca veri setinin 1084 cümleden 3126 cümleye çıkarılması ile istatistiksel modelin aynı girdilerle yaptığı çevirinin başarısı 53,73'ten 88,25'e çıkmıştır. Bu durum, veri seti boyutunun istatistiksel çevirinin başarısını etkilediğinin göstergesi olmuştur.

İstatistiksel çeviri modeli her cümle için, yukarıda verilen örnekte olduğu gibi olumlu katkı sağlamamaktadır. Aşağıda Şekil 7'de kural tabanlı çeviri ile kural tabanlı çeviri verisinin kullanıldığı istatistiksel çeviri örneği verilmiştir. İstatistiksel çeviri ile kural tabanlı çeviride bulunmayan, ancak çeviri modelinden eşleyerek eklenen "sen" kelimesi ile % 75,98 olan başarının % 30,74'e düşmesine neden olmuştur.

Tablo 4. Tüm veri setlerindeki bulunan örnek bir cümlelerin BLEU skoru karşılaştırması
(Comparison of the BLEU score of a sample sentence in all datasets)

Veri Seti	BLEU	Uzunluk Oranı	Metin
Kaynak Türkçe Cümle	-	-	Cam kenarında oturmak isterseniz bu masa müsait.
İnsan Çevirisi	100,00	1,00	siz cam kenar oturmak istemek bu masa müsait olmak
Ham Verilerle İstatistiksel Çeviri	20,61	0,89	cam <i>kenarında</i> oturmak <i>isterseniz</i> bu masa müsait .
Morfolojik Verilerle İstatistiksel Çeviri	29,07	1,11	cam olmak kenar oturmak istemek siz bu masa müsait <i>değil</i>
Kural Tabanlı Çeviri	45,62	0,89	cam kenar oturmak istemek siz bu masa müsait
Kural Tabanlı Verilerle İstatistiksel Çeviri	53,73	1,11	<i>camdaki kız ama</i> kenar oturmak istemek bu masa müsait
Kural Tabanlı Verilerle İstatistiksel Çeviri (3000 Cümlelik Veri Seti ile)	88,25	0,89	siz cam kenar oturmak istemek bu masa müsait olmak



<<

<

8

>

>>

From 1 to 28 of all 776 sentences.

Cümle 138	BLEU	Uzunluk Oranı	Metin
Kaynak Cümle	-	-	Kuşum için salıncak istiyorum.
İnsan Çevirisi	100,00	1,00	benim kuş için salıncak istemek
Kural Tabanlı Çeviri	75,98	1,20	benim kuş için salıncak istemek ben
Kural Tabanlı Çeviri			
Verileriyle İstatistiksel Çeviri	30,74	1,40	benim kuş için sen salıncak istemek ben

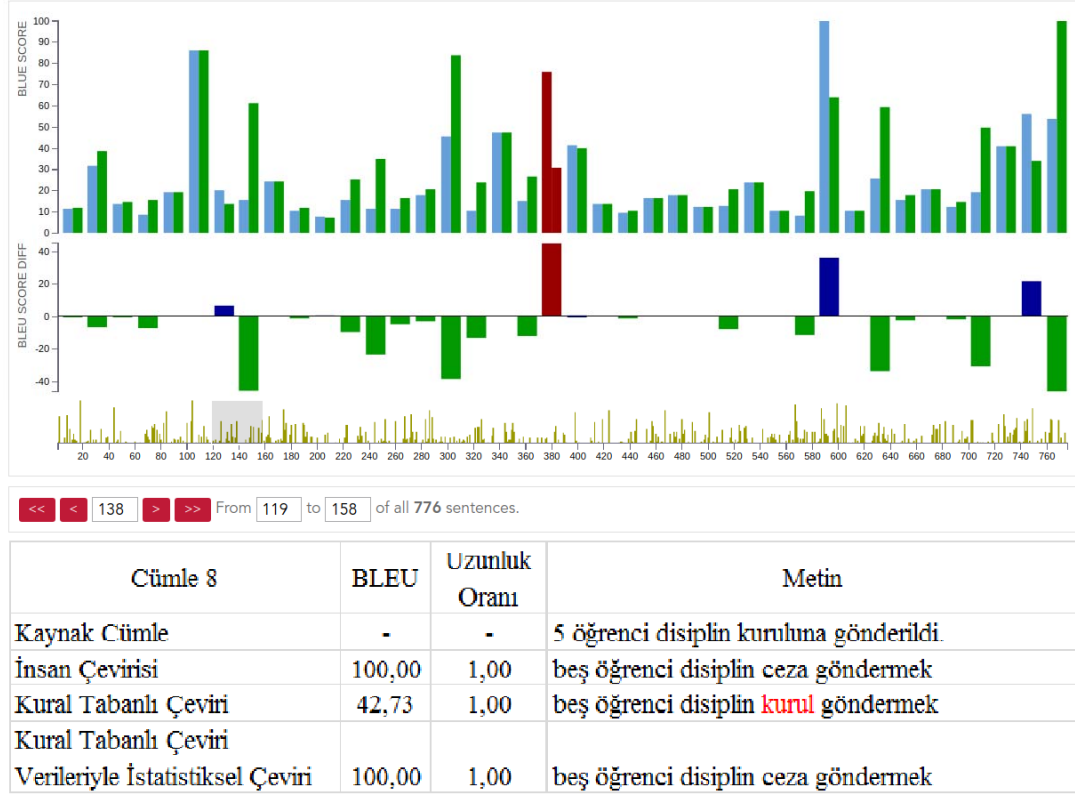
Şekil 7. Çeviri başarısının düşürülmesi örneği (Example of reduced translation success)

İstatistiksel çevirinin bazı cümlelerde çeviri başarısını düşürmesine rağmen tüm veri setinde pozitif yönde etkisi olduğu görülmektedir. BLEU skoru yaklaşık % 19 civarında olan kural tabanlı çeviri verilerinin kullanılması ile istatistiksel çeviri başarısı % 32,92'ye çıkmıştır. Şekil 8'de bu pozitif katkının gösterildiği cümle örneği verilmiştir.

Ayrıca uygulamanın metinden metine çeviri işleminden sonra görselleştirilmesi kısmında; kelimelerin yanlış hizalanmasından kaynaklanan anlam bozukluğunu engelleyebilmek adına kelime birlikteliğinden yararlanılarak kelime grupları arasında zamansal boşluklar bırakılmıştır.

“İki adam ekmek aldı.” cümlesi işaret diline “adam iki ekmek almak” olarak çevrilmiştir. Ancak geliştirilen kural tabanlı çeviride, Türkçede olduğu gibi sıfatlar kelimelerden önce kullanılmıştır. Çünkü hangi durumlarda sıfatların kelimedden önce hangi durumda sonra

kullanılması gerektiği konusunda net bilgi bulunmamaktadır. Uygulamada sıfatların kelime öncesinde kullanılması tercih edilmiştir. Yukarıda verilen cümle işaret diline çevrildiğinde “iki adam ekmek almak” şeklinde çevrilecektir. Bu çeviride sıfatın yeri yanlış olsa da anlam düzgün olarak aktarılabilir. Ancak “Kız küçük defter aldı” gibi bir cümle işaret diline çevrildiğinde “kız küçük defter almak” olarak çevrilmiştir. Bu durumda işaret dilindeki küçük kelimesinin kızı mı yoksa defteri mi nitelediği anlaşılamamaktadır. Verilen örnek cümlelerdeki gibi durumlarda anlamı düzgün aktarabilmek için kelime gruplarının tespiti ile kelime grupları arasında zamansal boşluklar bırakılmıştır. Türkçe cümle, ODTÜ'den alınan Türkçe ağaç yapılı derlem veri seti ile eğitilen MaltParser aracı ile bağımlılık analizinden geçirilmektedir. Bağımlılık analizden cümle içindeki kelime grupları tespit edilmektedir. Çeviri metninin görselleştirilmesi sırasında bu kelime grupları arasında geniş zaman boşlukları kullanılarak anlamın düzgün



Şekil 8. Çeviri başarısının artırılması örneği (Example of increasing translation success)

aktarılmasına yardımcı olunmuştur. “kız küçük defter almak” çeviri metni görselleştirilirken, “kız / küçük+defter / almak” şeklinde gösterilmektedir. “/” kelime grupları arasındaki geçiş zamanını, “+” ise kelime grubu içindeki kelimeler arası geçiş zamanını temsil etmektedir. Kelime grupları arasında “/” için 0,5 sn, kelime grubu içinde geçişlerde “+” 0,1 sn boşluklar bırakılmıştır. Türkçe cümledeki bağımlılıktan yararlanarak küçük ve defter kelimeleri aynı grupta gruplandırılmıştır. Kız kelimesinden sonra daha uzun bir bekleme süresi kullanılmış daha sonra küçük+defter kelime grubunun işaretlenmesine geçilmiştir. Küçük ve defter kelimeleri arasında çok daha az bir süre kullanılarak, kız kelimesinin başka, küçük+defter kelimelerinin başka kelime olduğu anlamı verilmeye çalışılmıştır. Böylece ‘küçük’ kelimesinin defter kelimesini nitelendirdiği anlamı güçlendirilmiştir.

4. Sonuçlar ve Tartışmalar (Results and Discussions)

Türkçe ve TİD arasında yeterli eşlenik verinin olmamasından dolayı Türkçeden işaret diline yapılan çeviri çalışmaları genellikle kural tabanlı çeviri yöntemiyle yapılmıştır. Son zamanlarda oluşturulan veri setleriyle az da olsa istatistiksel yöntemle çeviri çalışmaları yapılmaya başlanmıştır.

Bu çalışma, Türkçeden TİD’e çeviri çalışmalarına kural tabanlı ve istatistiksel tabanlı yöntemleri birlikte kullanarak katkı sağlamıştır. Alanda yapılan diğer çalışmalarda farklı veri setleri kullanıldığı için önerdiğimiz sistemle tam bir başarı karşılaştırması yapılamamaktadır. Ancak BLEU başarı skoru hesaplanarak bir değerlendirme yapılması mümkündür. Türkçeden TİD’e yapılan istatistiksel çeviri çalışmalarında BLEU puanları; Şimşek % 17 [33], Kayahan % 12,64 [45], Buz % 46,46 [46] şeklinde elde edilmiştir. Bu çalışmada % 32,92 BLEU skoru ile yapılan çalışmalardaki en iyi başarıya yakın bir başarı

değeri elde edilmiştir. Kullanılan veri setlerinin farklı olması tam bir karşılaştırma yapılabilmesini engellemiştir.

Kural tabanlı çeviriden elde edilen verilerin kullanıldığı istatistiksel çevirinin başarısı % 21,15 seviyesindeyken, veri setindeki cümle sayısı artırıldıktan sonra, kural tabanlı çevirinin başarısının BLEU puanı % 19 civarında sabit kalırken, yarı çevrilmiş cümlelerin kullanıldığı istatistiksel çevirinin başarısı % 32,92’ye yükselmiştir.

TİD’de, sıfatlar bazı durumlarda kelimeden önce, bazen de kelimeden sonra gelebilmektedir. Sıfatların hangi durumlarda önce, hangi durumlarda sonra geldiği ortaya konulmamış olduğundan, bu çalışmada sıfatların kelimelerden önce kullanılması yoluna gidilmiştir. Çevrilmiş metinde, bir sıfat iki kelime arasındaysa, bu sıfatın hangi kelimeyi tanımladığı hakkında anlamsal bir karışıklık olabilmektedir. Özellikle sıfatların kelimeden önce veya sonra gelmesinden kaynaklanan hataların giderilmesi ve çevirinin insanlar tarafından doğru bir şekilde anlaşılması için zamansal boşluklardan yararlanılmıştır. Çevirinin görselleştirme aşamasında, bağımlılık analizi kullanarak kelime grupları arasında zamansal boşluklar bırakılmıştır. Bu nedenle, her ne kadar otomatik değerlendirme sistemleri kelimeleri ve konumlarını birlikte değerlendirerek bunu bir hata olarak hesaplarsa da, geliştirilen uygulama ile çevirinin anlam açısından insanlara daha doğru bir şekilde aktarılmasına katkıda bulunulmuştur. İlk defa bu çalışmada, Türk İşaret Dili çevirilerinde zamansal boşluk kullanılması gerçekleştirilerek alana katkı sağlanmıştır.

İleriki çalışmalarda sistemin geliştirilmesi için bazı ek çalışmalar yapılabilir. Uygulamada kullanılan sözlüğün genişletilmesi, çevirinin görselleştirilmesi sırasında parmak alfabesinin kullanımının en aza indirilmesine katkıda bulunacaktır. Kural listesine yeni kurallar

ekleyerek iyileştirmeler yapılabilir. Karmaşık cümlelerin çeviri kalitesini artırmak için iyileştirmeler yapılabilir. Örneğin, işaret dilindeki sıralı cümlelerde, özne tekrar tekrar işaretlenmez. Ancak, geliştirdiğimiz uygulamada, her cümle bağımsız olarak ele alınmakta ve her cümlelerin öznesi işaretlenmektedir. Çevirinin performansını artırmak için, çevrilmiş cümleler depolanabilir ve sonraki çeviriler için örnek veri seti olarak kullanılabilir. Böylece, cümleyi yeniden analiz etme ve çevirme adımlarına ihtiyaç duyulmadan çeviri yapılabilir. Uygulamanın ana hedef kitlesi olan işitme engellilerin değerlendirmeleri doğrultusunda iyileştirmeler yapılabilir.

Teşekkür (Acknowledgement)

Bu makale, “İşitme Engelli Bireylere Yönelik Türkçe Metinlerin Türk İşaret Diline Çevrilmesi” adlı tez çalışmasından faydalanılarak hazırlanmıştır. Çalışmada kullandığımız verileri bizimle paylaşan Aile Çalışma ve Sosyal Politikalar Bakanlığı’na, işaret dili videolarını kullanmamıza izin veren isaretce.com sitesi sahiplerine, bağımlılık analizi için Türkçe ağaç yapılı derlem datasını bizimle paylaşan ODTÜ corpus araştırma grubuna, Türkçe metinlerin işaret dili karşılıklarının oluşturulmasındaki destekleri için Ferat KAYA hocamıza teşekkür ederiz.

Kaynaklar (References)

1. Aile ve Sosyal Hizmetler Bakanlığı. Engelli ve Yaşlı İstatistik Bülteni Ağustos-2021. https://www.aile.gov.tr/media/89297/eyhgm_istatistik_bulteni_agustos2021.pdf. Erişim tarihi Aralık 12, 2021.
2. Kın Z. B., Türk İşaret Dili Alfabesinin Derin Öğrenme Yöntemi ile Sınıflandırılması, Master Tezi, Başkent Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2019.
3. Karaca M. F., Bayır Ş., Türk İşaret Dili İncelemesi: İletişim ve Dil Bilgisi, Ulusal Eğitim Akademisi Dergisi, 2 (2), 35–58, 2018
4. TÜİK. Engellilerin Sorun ve Beklentileri Araştırması. <https://data.tuik.gov.tr/Bulten/Index?p=Ozurulerin-Sorun-ve-Beklentileri-Arastirmasi-2010-6370>. Erişim tarihi Aralık 12, 2020.
5. K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, BLEU: A Method for Automatic Evaluation of Machine Translation, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 371 (23), 311–318, 2002.
6. Punchimudiyanse M., Meegama R. G. N., Computer Interpreter for Translating Written Sinhala to Sinhala Sign Language, OUSL Journal, 12 (1), 70–90, 2017.
7. Alkoby K., A Survey of ASL Tenses, Proceedings of the 2nd Annual CTI Research Symposium, 1–11, 1999.
8. El Alfi A. E. E., El Basuony M. M. R., El Atawy S. M., Intelligent Arabic Sign Language to Arabic text Translation for Easy Deaf Communication, International Journal of Computer Applications, 92 (8), 22–29, 2014.
9. Hendriks H. B., Jordanian Sign Language: Aspects of grammar from a cross-linguistic perspective, LOT, Utrecht, 2008.
10. Sugandhi, P., Kumar, S. Kaur, Sign Language Generation System Based on Indian Sign Language Grammar, ACM Transactions on Asian and Low-Resource Language Information Processing, 19 (4), 2020.
11. Eryiğit C., Köse H., Keleşir M., Eryiğit G., Building machine-readable knowledge representations for Turkish sign language generation, Knowledge-Based Systems, 108, 179–194, 2016.
12. Eryiğit G., Eryiğit C., Karabüklü S., Keleşir M., Özkul A., Pamay T., Torunoğlu-Selamet D., Köse H., Building the first comprehensive machine-readable Turkish sign language resource: methods, challenges and solutions, Lang. Resour. Eval., 54 (1), 97–121, 2020.
13. Stein D., Jan B., Ney H., Morpho-syntax based statistical methods for automatic sign language translation, EAMT-2006 - 11th Annual Conference of the European Association for Machine Translation, 169–177, 2006.
14. Dangsaart S., Naruedomkul K., Cercione N., Sirinaovakul B., Intelligent Thai text - Thai sign translation for language learning, Computers and Education, 51 (3), 1125–1141, 2008.
15. Kouremenos D., Ntalianis K., Siolas G., Stafylopatis A., “Statistical machine translation for Greek to Greek sign language using parallel corpora produced via rule-based machine translation”, CEUR Workshop Proc., 2252, 28–42, 2018.
16. Dreuw P., Stein D., Deselaers T., Rybach D., Zahedi M., Bungeroth J., Ney H., Spoken language processing techniques for sign language recognition and translation, Technology and Disability, 20 (2), 121–133, 2008.
17. Bungeroth J., Ney H., Statistical Sign Language Translation, Workshop on representation and processing of sign languages, LREC, 4 (2003), 105–108, 2004.
18. El-Kahlout İ. D., Oflazer K., Exploiting Morphology and Local Word Reordering in English-to-Turkish Phrase-Based Statistical Machine Translation, IEEE Transactions On Audio, Speech, And Language Processing, 18 (6), 1313–1322, 2010.
19. Och F. J., Ney H., The alignment template approach to statistical machine translation, Computational Linguistics, 30 (4), 2004.
20. Koehn P., Hoang H., Factored translation models, EMNLP-CoNLL 2007 - Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 868–876, 2007.
21. Yeniterzi R., Syntax-to-morphology alignment and constituent reordering in factored phrase-based statistical machine translation from English to Turkish, Master Tezi, Sabancı Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2009.
22. Yang M., Kirchhoff K., Phrase-based backoff models for machine translation of highly inflected languages, EACL 2006 - 11th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference, 41–48, 2006.
23. Wu C. H., Su H. Y., Chiu Y. H., Lin C. H., Transfer-based statistical translation of Taiwanese sign language using PCFG, ACM Transactions on Asian Language Information Processing, 6 (1), 1–18, 2007.
24. Ersoy E., Karal Ö., Yapay Sinir Ağları ve İnsan Beyni, İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 1 (2), 188–205, 2012.
25. Brour M., Benabbou A., ATLASLang NMT: Arabic text language into Arabic sign language neural machine translation, Journal of King Saud University-Computer and Information Sciences, 33 (9), 1121–1131, 2021.
26. Stoll S., Camgoz N. C., Hadfield S., Bowden R., Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks, International Journal of Computer Vision, 1–12, 2020.
27. Kaczmarek M., Filhol M., Computer-assisted sign language translation: a study of translators’ practice to specify CAT software, Machine Translation, 35 (3), 305–322, 2021.
28. Sidig A. A. I., Luqman H., Mahmoud S., Mohandes M., KArSL: Arabic Sign Language Database, ACM Transactions on Asian and Low-Resource Language Information Processing, 20 (1), 1–19, 2021.
29. Kahlon N. K., Singh W., Machine translation from text to sign language: a systematic review, Universal Access in the Information Society, 22 (1), 1–35, 2023.
30. Eryiğit C., Yazılı Türkçe Dilinden Türk İşaret Diline (TİD) Makine Çeviri Sistemi, Doktora Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2017.
31. Yasan M. C., Türkçe Metni Türk İşaret Diline Dönüştürme, Master Tezi, Başkent Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2014.
32. Karaca M. F., Üç Boyutlu Sanal Model ile Türk İşaret Dili Simülasyonu, Doktora Tezi, Karabük Üniversitesi, Fen Bilimleri Enstitüsü, Karabük, 2018.
33. Buz B., Güngör T., Developing a Statistical Turkish Sign Language Translation System for Primary School Students, IEEE International Symposium on Innovations in Intelligent Systems and Applications, 2019.
34. Kayahan D., Güngör T., A Hybrid Translation System from Turkish Spoken Language to Turkish Sign Language, IEEE International Symposium on Innovations in Intelligent Systems and Applications, 2019.
35. Şimşek M. S., Türkçe ile Türk İşaret Dili Arasında Örneğe Dayalı Makina Çeviri Sistemi, Master Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2016.
36. Makaroğlu B., Dikyuva H., Güncel Türk İşaret Dili Sözlüğü. <https://tidsozluk.ailevecalisma.gov.tr/>. Erişim tarihi Mayıs 01, 2021.
37. Dastgheib M. B., Fakhrahmad S. M., Jahromi M. Z., A hybrid accurate alignment method for large persian-english corpus construction based on statistical analysis and lexicon/persian word net, International Journal of Information Science and Management, 14 (2), 97–106, 2016.

38. Akın A. A., Zemberek, <https://github.com/ahmetaa/zemberek-nlp>. Erişim tarihi Eylül. 10, 2021.
39. Moses. <http://www.statmt.org/moses/index.php?n=Main.HomePage>. Erişim tarihi Eylül. 10, 2021.
40. Och F. J., Ney H., A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics*, 23 (1), 19–51, 2003.
41. Heafield K., Pouzyrevsky I., Clark J. H., Koehn P., Scalable Modified Kneser-Ney Language Model Estimation, 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 690–696, 2013.
42. Karabüklü S., Strategies to Express Time in a Tenseless Language: Turkish Sign Language (TİD), *Dilbilim Araştırmaları Dergisi*, 111, 87–118, 2018.
43. Karal H., Şılbır L., Bahçekapılı E., Atasoy M., İşitme Engelli Bireyler için Grafik Sembollerle Oluşturulmuş Öğrenme Materyalleri, *Journal of Instructional Technologies & Teacher Education*.3 (3), 9-19, 2014.
44. Dikyüva H., Makaroğlu B., Arık E., *Türk İşaret Dili Dilbilgisi Kitabı*. Aile ve Sosyal Politikalar Bakanlığı Yayınları, Ankara, 2015.
45. Kayahan D., Hybrid Translation System From Turkish Spoken Language To Turkish Sign Language, Master Tezi, Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2019.
46. Buz B., Developing A Statistical Turkish Sign Language Translation System For Primary School Students, Master Tezi, Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2019.

