



Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi Sayı: 10/1 2021 s. 192-205, TÜRKİYE

Derleme Makale

MAKİNE ÇEVİRİ YÖNTEMLERİ VE MAKİNE ÇEVİRİSİNİN BUGÜNKÜ DURUMU

Taşpolot SADIKOV*

Kerim SARIGÜL**

Geliş Tarihi: Temmuz, 2020

Kabul Tarihi: Ocak, 2021

Öz

Bu çalışmada makine çevirisinin ortaya çıkışı, gelişimi ve bugün geldiği nokta değerlendirilmekte, popüler makine çeviri sistemleri (Google Çeviri, Microsoft Bing Çeviri, Yandex Çeviri) üzerinden günümüzde yaygın kullanılan makine çevirisi yöntemlerine değinilmektedir.

20 milyar kelime içeren Birleşmiş Milletler belgeleri üzerinde eğitilen Google Çeviri; ilk yıllarda “Kural Tabanlı Makine Çevirisi” yöntemini ve “İstatistiksel Makine Çevirisi” yöntemini kullanmış, son yıllarda ise insan çevirisine daha yakın olan “Sinirsel Makine Çevirisi” yöntemini kullanmaya başlamıştır. Bu yöntemi İngilizce, Fransızca, Almanca, İspanyolca, Portekizce, Çince, Japonca, Korece ve Türkçe olmak üzere 9 dilde kullanmaktadır.

95’ten fazla dilde çeviri hizmeti sunan Yandex Çeviri ise; önceden öğretilmiş dil kurallarına göre değil, üzerinde çalıştığı dilleri istatistiksel yöntemlerle analiz ederek çeviri yapmaktadır. Yandex bu yöntemi son yıllarda geliştirerek hibrit bir model (istatistiksel makine çevirisi ve sinirsel makine çevirisini birlikte) kullanmaya başlamıştır.

Microsoft Bing Çeviri ise 70’tan fazla dilde metin çeviri hizmeti vermekte, bunlardan 60’ında Sinirsel Makine Çevirisi yöntemini, 10 dilde ise istatistiksel makine çevirisi yöntemini kullanmaktadır.

Makalede istatistiksel makine çevirisi Türkçe-Kırgızca, sinirsel makine çevirisi ise İngilizce-Türkçe örnekler üzerinden değerlendirilmektedir.

Anahtar Sözcükler: Makine çevirisi, kural tabanlı makine çevirisi, istatistiksel makine çevirisi, sinirsel makine çevirisi.

MACHINE TRANSLATION METHODS AND CURRENT STATE OF MACHINE TRANSLATION

Abstract

In this study, machine translation’s emergence, development and the point which it stands today are evaluated, the commonly used machine translation methods in our present day are addressed over the popular machine translation systems (Google Translation, Microsoft Bing Translation, Yandex Translation).

* Prof. Dr.; Bişkek Sosyal Bilimler Üniversitesi, Kırgızistan, tash_sadykov@mail.ru

** Dr.; Gazi Üniversitesi, TÖMER, Ankara, kerim@kerimsarigul.com

Google Translation which was instructed over documents of United Nations containing 20 billion of words has used the method of “Rule-Based Machine Translation” and the method of “Statistical Machine Translation” in the early years, it has started to use the method of “Neural Machine Translation” that is closer to the human translation in the recent years. It uses this method in 9 languages as English, French, Deutsch, Spanish, Portuguese, Chinese, Japanese, Korean and Turkish.

Yandex Translation which provides translation service in more than 95 languages makes translation by analyzing the languages it works on by statistical methods, not by the language rules which have been already instructed. Yandex has developed this method in recent years and started to use a hybrid model (in combination of statistical machine translation and neural machine translation).

Microsoft Bing Translation offers text translation service in more than 70 languages, it uses the method of Neural Machine Translation in 60 languages and the method of statistical machine translation in 10 languages.

In this article, the statistical machine translation will be evaluated over Turkish-Kyrgyz examples, the neural machine translation will be evaluated over English-Turkish examples.

Keywords: Machine translation, rule-based machine translation, statistical machine translation, neural machine translation.

Giriş

Bilgisayar teknolojisi hayatımıza girdiğinden beri bir dilden başka dile çeviri yapan makine sistemlerinin var olabileceği mantığı gelişmiş ve bu konularda birçok çalışma yapılmıştır. Günümüzde yaygın olarak kullanılan ve her geçen gün daha da gelişerek yeni özelliklerle kullanıcılara hizmet sunan çeviri sistemlerinin başında Google Çeviri, Yandex Çeviri, Microsoft Bing Çeviri sistemleri gelmektedir.

“Makine Çevirisi” ifadesi temelde iki anlamda kullanılmaktadır: 1) Metin veya konuşmayı bir dilden (kaynak dil) diğerine (hedef dil) çevirmek için yazılım araçlarının kullanımını inceleyen hesaplamalı dil bilimin bir parçasıdır. 2) Makine çevirisi, bir yazılım/uygulama aracılığı ile kaynak dildeki herhangi bir metnin insan müdahalesi olmadan (İnsanın müdahalesi yazılım aracılığıyla makinedir, kaynak dildeki veya hedef dildeki metne değildir.) hedef dile çevrilmesi işlemidir.

Makine çevirisi araştırmaları 1954’te Georgetown Üniversitesi ve IBM ortaklığıyla gerçekleştirilen Georgetown Deneyi ile başlamıştır. Hafızasındaki 6 gramer kuralı ve 250 kelime kullanılarak yapılan Georgetown Deneyinin başarılı sonuçlar vermesi makine çevirisine yönelik çalışmaların artmasına sebep olmuştur. (Hutchins, 2005, s. 1)

Makine çevirisinde kullanılan yöntemlerden bazıları: Sözlük Tabanlı Makine Çevirisi/Doğrudan yöntem (Dictionary Based Machine Translation), Kural Tabanlı Makine Çevirisi (Rule-Based Machine Translation), Örnek Tabanlı Makine Çevirisi (Example Based Machine Translation), İstatistiksel Makine Çevirisi (Statistical Machine Translation), Sinirsel Makine Çevirisi (Neural Machine Translation).

Makine çevirilerilerinin bugün geldiği noktayı değerlendirmek üzere öncelikle en çok kullanılan yöntemlerden olan Kural Tabanlı Makine Çevirisi, İstatistiksel Makine Çevirisi ve Sinirsel Makine Çevirisi ele alınmıştır.

Kural Tabanlı Makine Çevirisi (Rule-Based Machine Translation)

Kural tabanlı makine çevirisi, otomatik çeviri adına kullanılan ilk yöntem olması açısından önemlidir. Bu yöntemi kullanan makine çevirisindeki temel prensip, kaynak dil verilerinin çeviri motorları vasıtası ile daha önceden programlanmış bilgisayar diline dönüştürülüp sonrasında bilgisayara tanıtılan hedef dil kodlarına aktarılma işlemine dayanır. Kural tabanlı makine çevirisi yaklaşımı ile yapılan çeviriler, kaynak ve hedef dillerin birbirlerine gerek sözcük yapıları gerek dil bilgisi kuralları, gerekse de sözcük dizimi ilkelerinin bulunduğu ortak paydaların çokluğu ile doğru orantılı olarak iyi sonuç vermektedir. Örneğin aynı ya da benzer dil ailesine mensup diller arası yapılan çeviriler, daha olumlu sonuçlar verirken, farklı dil ailelerine bağlı ve yapısal özellikleri ortak olmayan diller arasında yapılan çeviriler olumsuz sonuç vermektedir. “Systran”, “Aptertium” ve “GramTrans” kural-odaklı makine çevirisi programlarına örnek olarak gösterilebilir (Balkul, 2015, s. 23).

Kural tabanlı makine çevirisi; hedef dilde metin oluşturmak için önce kaynak dilde girilen içeriği analiz etmede dil kurallarını kullanır. Analiz sonrası elde edilen kelime ve kurallar hedef dilin kelime ve kuralları ile eşleştirilir, söz dizimi kurallarına göre sonuç elde edilir. Bu işleyiş; yukarıda da belirtildiği gibi aynı veya benzer yapıdaki diller için büyük avantajlar sağlamaktadır. Türkçe ve Kırgızca için basit bir cümle örneği verelim:

Cümle: Misafirler eve geldi.

Aşamalar: 1. Analiz (dilsel analiz) - 2. Aktarma (kelime çevirisi, dil bilgisi çevirisi) - 3. Üretim (Hedef dilde metin oluşturma)

Ek ve köklerinin hedef dille ilişkilendirilmesi için önce ayrıştırılması gerekir.

Misafir-ler ev-e gel-di (kök, gövde, ek, tür, zaman vb.)

misafir: konok (isim)

-ler: +lAr; +dAr, +tAr; +lOr, +dOr, -tOr. Çokluk eki. Hedef dilde işlenirken ilgili kelimeye göre ünlü ünsüz uyumu da dikkate alınacak.

ev: üy (isim)

-e: +gA; +gO; +kA, +kO (yönelme eki)

gel-: kel- (fiil, yüklem)

-di: +dI, +dU; +tI, +tU (belirli geçmiş zaman eki)

Analizi yapılan ve ayrıştırılan, ayrıştırıldıktan sonra hedef dilde karşılıkları oluşturulan, dilin kurallarına göre yeniden düzenlene (ünlü, ünsüz uyumu, istisnalar, söz dizimi vb.) birimler birleştirildiğinde elde edilen sonuç:

Konok-tor üy-gö kel-di.

Bu örnekten de anlaşılacağı gibi kural tabanlı çeviride söz dizimi, kelimeler, kök, gövde ve eklerin yapısal benzerlikleri büyük avantaj sağlamaktadır. Bu yöntemin etkin bir şekilde kullanılabilmesi için iki dilli veya çok dilli zengin bir sözlüğe, dilin tüm kurallarının makineye öğretilmesine (algoritmalar aracılığı ile) ihtiyaç duyulmaktadır.

İstatistiksel Makine Çevirisi (Statistical Machine Translation)

İstatistiksel Makine Çevirisini; kaynak dilde verilen cümlelerin hedef dildeki en olası karşılığının tespit edilmesi olarak ifade edebiliriz. İstatistiksel Makine Çevirisi yönteminde; gerçek kişiler tarafından yapılan çeviriler analiz edilerek kurallar öğrenilir (makine öğrenmesi), hedef dildeki en olası karşılık istatistiklerle tespit edilir ve çeviri işlemi gerçekleştirilir. Örneğin Google Türkçe-Kırgızca çeviride İstatistiksel Makine Çevirisi yöntemini kullanıyor.

Çeviri Yöntemleri ve Örnekleri

Örnek cümleler (Türkçe)	Türkçe-Kırgızca İstatistiksel Makine Çevirisi (Google)	Türkçe-Kırgızca İnsan Çevirisi	Türkçe-İngilizce Sinirsel Makine Çevirisi (Google)
Arabada bir adam var.	машине менен бир жигит бар. (<i>Araba bir delikanlı var</i>)	Машинедө бир киши бар.	There's a man in the car.
Arabada uzun boylu bir adam var.	машине бир узун бойлуу жигит бар. (<i>Araba bir uzun boylu delikanlı var</i>)	Машинедө узун бойлуу бир киши бар.	There's a tall man in the car.
Bizim arabamızda uzun boylu bir adam var.	Биздин унаа бийик адам турду. (<i>Bizim araç büyük adam kalktı</i>)	Биздин машинебизде узун бойлуу бир киши бар.	We have a tall man in our car.
Bizim arabamızdaki uzun boylu adamın 2 kızı var.	Биздин машинабыз бизди бийик жигит эки кыздуу. (<i>Bizim arabamız büyük delikanlı iki kızı var</i>)	Биздин машинебиздеги узун бойлуу кишинин 2 кызы бар.	The tall guy in our car has two daughters.
Bizim kırmızı renkli arabamızı kullanan uzun boylu adamın 2 büyük kızı var.	Биздин Машинени кызыл болгон узун бойлуу киши 2 чоң кыз бар. (<i>Bizim arabayı kırmızı olan uzun boylu kişi iki büyük kız var</i>)	Биздин кызыл түстөгү машинебизди айдаган узун бойлуу кишинин 2 чоң кызы бар.	The tall man who uses our red car has 2 older daughters.
Örnek cümleler (Türkçe)	Türkçe-Kırgızca İstatistiksel Çeviri (Google)	Türkçe-Kırgızca İnsan Çevirisi	Türkçe-İngilizce Sinirsel Makine Çevirisi (Google)
Ali İstanbul'a gidiyor.	Элис Стамбулга барат. (<i>Alis İstanbula gider</i>)	Али Стамбулга бара жатат.	Ali is going to Istanbul.
Ali, İstanbul'a otobüsle gidiyor.	Али Стамбул мерт кетет. (<i>Ali İstanbul mert gider.</i>)	Али Стамбулга автобус менен бара жатат.	Ali goes to Istanbul by bus.
Ali, yarın İstanbul'a otobüsle gidiyor.	Али, эртең Стамбул мерт менен бара жатат. (<i>Ali, yarın İstanbul mert ile gidiyor</i>)	Али Стамбулга эртең автобус менен кетет.	Ali travels to Istanbul tomorrow by bus.
Ali, yarın İstanbul'a büyük otobüsle gidiyor.	Али, эртең Стамбулга бараткан ири автобус. (<i>Ali, yarın İstanbul'a giden büyük otobüs</i>)	Али Стамбулга эртең чоң автобус менен кетет.	Ali, goes to Istanbul by big bus tomorrow.
Ali'nin arkadaşı, yarın İstanbul'a otobüsle gidiyor.	Али досум, эртең Стамбул мерт менен бара жатат. (<i>Ali arkadaşım, yarın İstanbul mert ile gidiyor</i>)	Алинин досу Стамбулга эртең автобус менен кетет.	Ali's friend is taking a bus to Istanbul tomorrow.

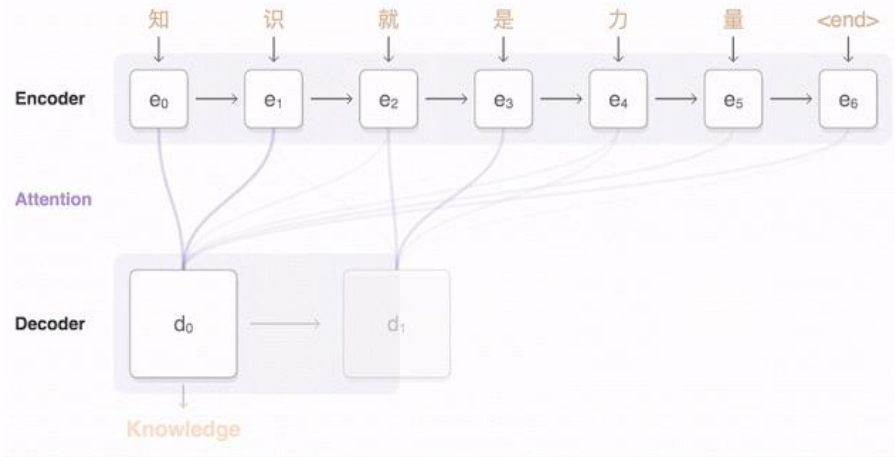
İstatistiksel Makine Çevirisi; tek kelimeli çeviriler yapmak yerine belirli bir kelimenin etrafındaki bağlamı kullanarak daha önce yapılan tercümelemlerle eşleştirir ve en olası sonucu elde etmeye odaklanır. İstatistiksel makine çevirisinin olumsuz yanlarından biri, kelimeleri tercüme ederken kendinden önceki ve sonraki birkaç kelime bağlamında çeviriyor olmasıdır. Yukarıda verilen örnek cümlelerden de anlaşılacağı gibi istatistiksel makine çevirisi kısa cümlelerde kısmen başarılı iken uzun cümlelerin çevirilerinde bu başarının düştüğü görülüyor. Bununla birlikte aşağıda değineceğimiz Sinirsel Makine Çevirisi örneğindeki (Türkçe-İngilizce) başarıyı da görmek mümkün.

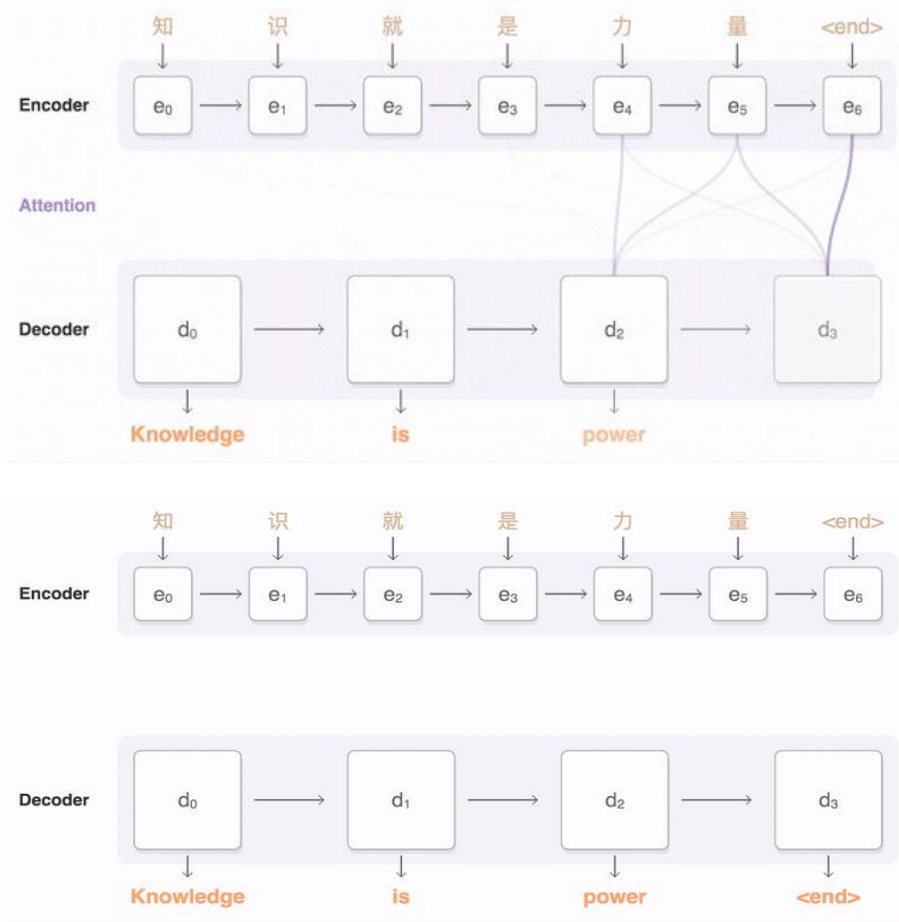
Sinirsel/Yapay Makine Çevirisi (Neural Machine Translation)

Günümüzün yeni ve en önemli makine çeviri yöntemlerinden olan Sinirsel Makine Çevirisi veya Yapay Makine Çevirisi diye adlandırabileceğimiz yöntem otomatik çeviri için uçtan uca bir öğrenme yaklaşımıdır. Sinirsel makine çevirisi yönteminde cümleler parça parça değil, cümledeki tüm parçalar eş zamanlı olarak çevrilir. Cümlelerin bir bütün olarak ele alındığı yöntemde çeviriler geniş bir bağlamda değerlendirilmekte, yaygın kullanımlar göz önüne alınarak uygun dil bilgisi ile yeniden düzenlenerek konuşma diline yakın hâle getirilmektedir.

Yukarıda, istatistiksel makine çevirisinin kelimeleri tercüme ederken cümleyi kendinden önceki ve sonraki birkaç kelime ile değerlendirdiğini ve bunun hatalı sonuçlar doğurduğunu, kısa cümlelerdeki başarısının uzun cümlelerde elde edilemediğini ifade etmiştik. Sinirsel Makine Çeviri yönteminde ise cümle içindeki tüm kelimeler daha geniş bağlamda değerlendirilir ve ilgili kelimenin cümle içindeki diğer tüm kelimelerle olan kombinasyonları yani modelleri oluşturulur. Yapay sinir ağı tarafından oluşturulan bu modeller cümle bağlamında yeniden değerlendirilerek hedef dile çevrilir. Bu özellik sayesinde İstatistiksel makine çevirisinde yaşanan sorunlar en aza indirilir ve daha doğru sonuçlara ulaşılır.

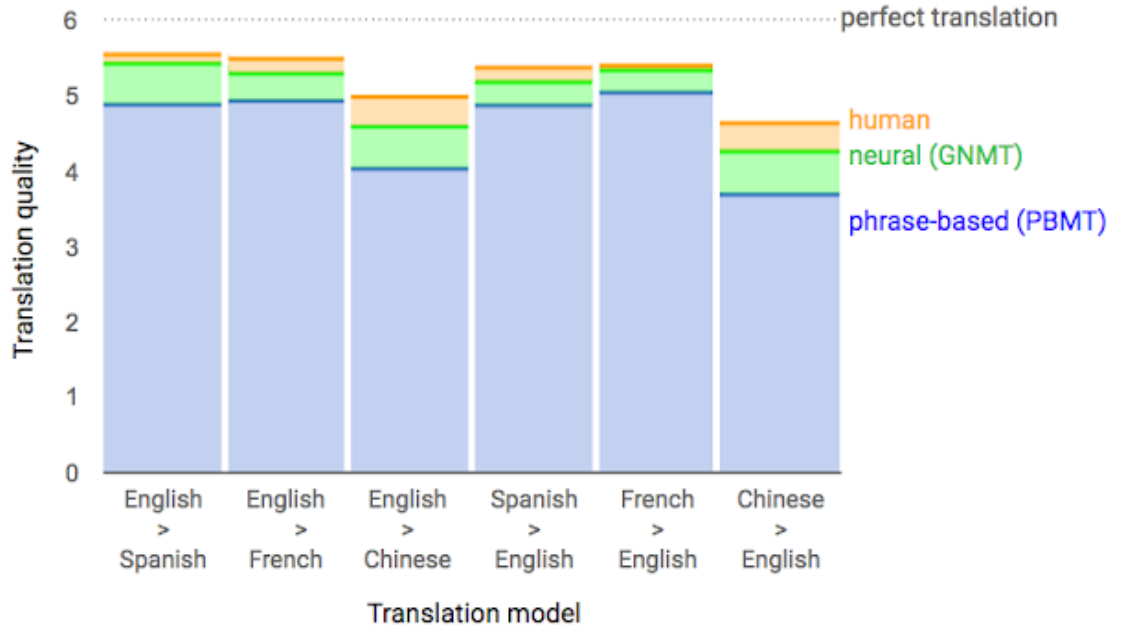
Aşağıda Google Sinirsel Makine Çeviri Sisteminin çalışma aşamaları görselleştirilmiştir.





Hareketli Resim Açıklaması: Çince bir cümleyi İngilizceye çevirirken Google Sinirsel Makine Çeviri Sisteminin ilerlemesini göstermektedir. İlk olarak, ağ Çince sözcükleri vektörler listesi olarak kodlar, burada her vektör şimdiye kadar okunan tüm kelimelerin anlamını temsil eder ("Encoder"). Tüm cümle okunduktan sonra, kod çözücü, İngilizce cümlenin her bir kelimesini tek seferde üretmeye başlar ("Decoder"). Kod çözücü İngilizce en uygun sözcüğü üretmek için her adımda kodlanmış Çince vektörler üzerindeki ağırlığa dikkat eder. (Not; mavi bağlantı şeffaflığı, kod çözücünün kodlanmış bir kelimeye ne kadar önem verdiğini gösterir. (Quoc ve Schuster, 2016)

Google, bu yöntemle bugün gelinen noktada makine çevirisindeki hataların %55-85 arasında azaldığını ve insan çevirisi kalitesine yaklaşıldığını ifade ediyor.



Resim: İnsan değerlendiricilerinin belirli bir kaynak cümleden yapılan çevirileri değerlendirmesi sonucu elde edilen veriler. Değerler 0 ile 6 arasında değişir, 0 “tamamen anlamsız çeviri”, 6 ise “mükemmel çeviri” anlamına gelir. (Quoc ve Schuster, 2016)

Google (translate.google.com) bugün 100’den fazla dilde 4000’den fazla kombinasyonda çeviri hizmeti sunmaktadır ve bunlardan İngilizce ile Fransızca, Almanca, İspanyolca, Portekizce, Çince, Japonca, Korece ve Türkçe olmak üzere 8 dilde, Sinirsel Makine Çevirisi yöntemini kullanmaktadır.

Microsoft Bing (bing.com/translator) ise 70’tan fazla dilde metin çeviri hizmeti vermekte, bunlardan 60’ında Sinirsel Makine Çevirisi yöntemini kullanmaktadır. Bununla birlikte Microsoft 11 dilde de Microsoft Speech teknolojisi aracılığıyla konuşma çevirisi hizmeti de sunmaktadır. (Microsoft, 2020).

İstatistiksel Makine Çeviri yönteminde ara dil kullanılırken (L1-EN-L2 / Kaynak dil - İngilizce - Hedef Dil) Sinirsel Makine Çeviri yönteminde ara dil kullanmadan doğrudan hedef dile (L1-L2) çeviri yapılabilmektedir (Wikipedia, 2020).

Günümüzde Google Çeviri ve Microsoft Bing Çeviri ile birlikte en çok kullanılan çeviri sistemlerinden biri de Yandex Çeviri’dir (ceviri.yandex.com.tr). Bugün, 95’ten fazla dilde çeviri hizmeti sunan Yandex Çeviri, Yandex’in geliştirdiği özel bilgisayarlı çeviri teknolojisine dayalı ve sözcüklerin, metinlerin ve ayrıca web sayfalarının metin içeriklerinin çevrilmesi için kullanılabilen otomatik çeviri servisidir. Yabancı dillerin karmaşık kurallarını bilmeyen bilgisayar destekli Yandex Çeviri, önceden öğretilmiş dil kurallarına göre değil, üzerinde çalıştığı dilleri istatistiki yöntemlerle analiz ederek çeviri yapar. Yandex, İstatistiksel Makine Çeviri yönteminin gelişim potansiyelini “İstatistiğe dayalı bilgisayar destekli çeviri sistemlerinin tartışılmaz bir avantajı, bu sistemlerin zaten sürekli gelişen bir olgu olan dil ile birlikte gelişebilmeleridir. Dilde yeni bir şey ortaya çıktığı zaman sistem bu değişiklikleri üzerinde çalıştığı yeni belgeler sayesinde görür. Yenilikler dilde ne kadar çabuk yayılırsa

internette bulunan yeni yazılı metinlerde de o kadar çabuk ortaya çıkacaktır.” ifadeleriyle açıklar (Yandex, 2020).

Bilgisayar destekli Yandex Çeviri: (1) çeviri modeli, (2) dil modeli ve (3) dil şifre çözücüsü (decoder) olmak üzere üç bölümden oluşur.

Çeviri modeli, sistemin “bildiği” bir dilde bir sürü kelime, ifade ve cümleyle bunların diğer bir dilde her türlü karşılığının ve bu karşılıkların her birine dilde rastlanma sıklığı indeksinin belirlendiği bir tablodur. Böyle bir tablo her iki dil için farklı olur. İki dile özgü bir çeviri modeli önce birbirinin karşılıkları olan metinlerin bulunması, ardından cümlelerin ve son olarak da ifade ve kelimelerden oluşan özel anlam cetvellerinin hazırlanması olmak üzere üç aşamalı bir süreç sonucunda ortaya çıkar.

Dil modeli, bilgisayarlı çeviri sisteminin diğer bir parçasıdır. Sistem bir dil modeli oluşturmak için ilgili dildeki yüzbinlerce metni inceleyerek kullanılan tüm kelime ve kombinasyonları içeren ve bunların kullanım sıklığının belirtildiği bir liste hazırlar. Bilgisayarlı sistemin, çevirinin yapılacağı dil hakkında bilgisi bu kadardır.

Sistemdeki çevirilerin hazırlanmasından sorumlu olan asıl kısım dil **şifre çözücüsüdür**. Çeviri modelinin sunduğu veri tabanını inceleyen şifre çözücü, çevrilen dildeki her bir cümle için tüm mevcut çeviri seçeneklerini belirleyince bunları olasılık değerine göre sıralar. Örneğin çevrilmesi gereken İngilizce cümle Shakespeare’in “Hamlet” trajedisindeki ünlü monologdan “To be or not to be” satırı olsun. Olasılık sıralaması sonuçlarına göre “Olup olmaması” kelime kombinasyonu en olası seçenek çıkarırken doğru çeviri olan “Olmak ya da olmamak” seçeneği ikinci veya üçüncü sırada çıkar.

Var olan tüm kombinasyon seçenekleri dil şifre çözücüsü tarafından dil modeline uygun olarak incelenir. Dil modeli ise şifre çözücüye, “Olmak ya da olmamak” ifadesinin mevcut bağlamda var olan kaynaklarda daha sık kullanılan bir seçenek olduğunu gösterir. Sürecin sonunda dil şifre çözücüsü hem çeviri modelinin hazırladığı olasılık önceliği listesini hem de dil modelinin sunmuş olduğu kullanım sıklığı bilgilerini göz önünde bulundurarak bu iki unsurun en uygun kombinasyonunu tespit eder ve böylece en iyi çeviri sonucuna ulaşmaya çalışır (Yandex, 2020).

Burada örnekleriyle açık bir şekilde ifade edilen yöntem istatistiksel makine çevirisinin başarısını ortaya koymaktadır. Ancak örnekte ortaya konan başarılı sonuca istatistiksel yöntemde özellikle uzun ve karmaşık cümlelerde ulaşmak her zaman mümkün olamamaktadır. Bu sebeple Yandex de Google ve Microsoft gibi çevirilerine Sinirsel makine çevirisini dâhil ettiğini “Bir model ikiden iyidir. YandexTranslate Bir Hibrit Makine Çeviri Sistemini Başlattı” başlığıyla duyurmuştur. Yandex, Hibrit Makine Çevirisi Sisteminde sinirsel ve istatistiksel yöntemlerin tamamlayıcı gücünü kullandığını bu şekilde daha kaliteli ve insanlar gibi çeviriler yapabilen bir sistem oluşturduğunu ifade etmektedir (Yandex, 2017).

Bugün makine çevirilerinde kullanılan en son ve en gelişmiş yöntem olan Sinirsel Makine Çevirisinin gelmiş olduğu noktayı görmek açısından aşağıda İngilizce uzun bir cümlenin Bing, Google, Yandex makine çevirisi ve insan çevirisi örnekleri birlikte sunulmuştur.

İngilizce - Türkçe Çeviri Örnekleri (Sinirsel Makine Çevirisi, Hibrit Makine Çevirisi ve İnsan Çevirisi)	
İngilizce Metin	Yunus Emre Institute, since its establishment, has been carrying out activities to promote Turkey, Turkish language and its history and art in line with its law of establishment, and developing materials to make them available to who are in need especially those who teach and learn Turkish as a foreign language.
Bing Çeviri	Yunus emre Enstitüsü, kuruluşundan bu yana, Türkiye, Türk Dili ve tarihi ve sanatını kurulması Yasası doğrultusunda tanıtmak ve onları özellikle ihtiyaç duydukları kişiler için kullanılabilir kılmak için malzeme geliştirme faaliyetleri gerçekleştirmektedir İngilizce öğretmek ve yabancı dil olarak öğrenmek.
Google Çeviri	Yunus Emre Enstitüsü kuruluşundan bu yana Türkiye Cumhuriyeti'ni, Türk dilini ve tarihini ve sanatını kuruluş kanunu ile uyumlu hâle getirmek için faaliyetler yürütmekte ve özellikle öğrenen ve öğrenenlere ihtiyaç duyanlara sunmak için materyaller geliştirmektedir. Yabancı dil olarak Türkçe.
Yandex Çeviri	Yunus Emre Enstitüsü, kuruluşundan bu yana, Türkiye, Türk Dili ve tarihi ve sanatını, kuruluş Kanunu doğrultusunda tanıtmak ve özellikle de yabancı dil olarak Türkçe öğrenen ve öğrenenler için ihtiyaç duyan kişilere ulaştırmak için malzemeler geliştirmek için faaliyetler yürütmektedir.
İnsan Çevirisi	Yunus Emre Enstitüsü, kurulduğu günden itibaren kuruluş kanununda belirtildiği üzere Türkiye'yi, Türk dilini, tarihini, kültürünü ve sanatını tanıtmak üzere faaliyetler yürütmekte, özellikle yabancı dil olarak Türkçe öğrenenlerin ve öğrenenlerin ihtiyaç duyduğu materyalleri geliştirerek alanın istifadesine sunmaktadır.

Yukarıdaki çeviriler incelendiğinde her 3 çevirinin de (Google, Bing, Yandex) insan çevirisine çok yakın olduğu ancak bazı küçük hataların bulunduğu görülmektedir. Bu başarının temelinde her üç şirketin de çevirilerde Sinirsel Makine Çevirisi yöntemini doğrudan veya dolaylı olarak kullanması yatmaktadır.

Bugün gelinen noktada hem Google, hem Microsoft hem de Yandex çevirilerinde sinirsel makine çevirisi yöntemini henüz bazı dillerde kullanmaya başladıklarını, zamanla bu yöntemi tüm dillere uygulayacaklarını açıklamıştır. Günümüzün en önemli gelişmelerinden biri olan makine öğrenmesi ile sinirsel makine çevirisi yönteminin veya geliştirilecek olan yeni yöntemlerin makine çevirilerini daha kaliteli duruma getireceğini, çevirilerin insan çevirisine eş değer hâle geleceğini tahmin etmek zor değildir. Bugün makine çevirilerinin kalitesini insan çevirileri ile karşılaştırırken, yakın gelecekte insan çevirilerinin kalitesini makine çevirileriyle karşılaştırır duruma gelme ihtimalimiz de çok büyüktür.

Makine çevirilerinin ve sinirsel makine çeviri yöntemlerinin kullanımının artması ile birlikte bu sistemlerin nasıl çalıştığına yönelik bilgiler ve özellikle de bu yöntemlerin algoritmaları açık kaynak kodlu olarak yayımlanmaktadır. Onlardan biri olan OpenNMT, sinirsel makine çeviri modellerinin eğitimi ve dağıtımı için oluşturulmuş açık kaynak bir kütüphanedir. Sistem, Harvard'da geliştirilen seq2seq-att'nin devamı niteliindedir. Verimliliğinin, okunabilirliğinin ve genellenebilirliğinin artırılması ve kolay kullanılabilmesi amacıyla tamamen yeniden yazılmıştır (Klein, G., Kim, Y., Deng, Y., Senellart, J., and Rush, A. M. 2017). Dileyen araştırmacı veya uygulayıcılar bu algoritmaları kullanarak çalışmalar yapabilmekte, kendi projelerini geliştirebilmekte, ilgili algoritmaların geliştirilmesine katkı da sunabilmektedir.

Sonuç ve Değerlendirme

Bugün gelinen noktada popüler çeviri şirketlerinden Google'ın 8 dil için Sinirsel Makine Çeviri yöntemini, diğer diller için istatistiksel makine çeviri yöntemini kullandığını; Microsoft'un hizmet sunduğu 70 dilden 60'ında Sinirsel Makine Çevirisi yöntemini kullandığını; Yandex'in ise Hibrit bir yöntem (İstatistiksel ve Sinirsel Makine Çevirisi yöntemini birlikte) kullandığı görülmektedir.

Ele alınan örneklerden elde edilen sonuçlar incelendiğinde, sinirsel makine çevirisi yöntemi sayesinde kural tabanlı çeviri ve istatistiksel çeviri yönteminde yaşanan sorunların tamamının olmasa da büyük bir kısmının çözüldüğü görülmektedir.

Sinirsel makine çevirisinin temelinde makine öğrenmesi mantığı bulunduğu için makinenin analiz edeceği ve öğrenme elde edeceği metin sayısının artması ile birlikte sinirsel makine çevirisinin daha da geliştiği ve bundan sonra daha da gelişeceği; kural tabanlı makine çevirisi ve istatistiksel makine çevirilerinde ara dil kullanımı (L1-EN-L2) sonucu oluşan sorunların ara dil kullanmadan doğrudan iki dil arası çeviri yapabilen (L1-L2) sinirsel makine çevirisi yöntemi ile kısmen çözüldüğü; özellikle benzer yapıdaki veya aynı dil ailesine sahip diller arası çeviri işlemlerinde sinirsel makine çevirisinin çok daha başarılı olduğu görülmektedir.

Kural tabanlı ve istatistiksel makine çevirileri yöntemlerinde dil kurallarının bilgisayara öğretilmesine, iyi bir sözlük ve derleme ihtiyaç duyulmaktadır. İki dilli zengin sözlüklerin, gelişmiş yazılı ve sözlü derlemelerin eksikliği bazı dillerin makine çevirilerinde dezavantajlı duruma gelmelerine sebep olmaktadır. Sinirsel makine çevirisinin bu sorunlara kısmen çözüm olacağı, ilgili dillerdeki metinlerin oranının artması ile makine öğrenmesinin daha da gelişeceği ve çok daha doğru, insan çevirisine yakın çeviriler yapılabileceği düşünülmektedir.

Günümüzde; özellikle yazılım teknolojisinin hızlı bir şekilde gelişmesine ve yaygınlaşmasına sebep olan açık kaynak kod mantığının makine çeviri yöntemleri için de kullanılıyor olması, sinirsel makine çeviri algoritmalarının tüm kullanıcılara açık kaynak olarak sunuluyor olması bu alandaki çalışmaların daha da artacağını göstermektedir.

Hutchins ve Somers'in 1992 yılında makine çevirisi üzerine yaptıkları çalışmada ifade ettikleri "*insan müdahalesi olmaksızın herhangi bir dildeki cümleyi başka bir dile mükemmel derecede çevirisini yapabilecek bir çeviri makinesi yoktur*" (Hutchins ve Somers, 1992, s. 1) çıkarımın, sinirsel makine çevirisi yönteminin geldiği nokta açısından bakıldığında değiştiğini; yapay zekâ alanındaki çalışmaların artması ve makine öğrenmesi ile birlikte yakın zamanda "*insan müdahalesi olmaksızın herhangi bir dildeki cümleyi başka bir dile mükemmel derecede çevirisini yapabilecek bir çeviri makinesi vardır.*" cümlesini rahatlıkla kullanabileceğimizi ifade edebiliriz.

Çeviri sistemleri, doğrudan dil öğretimi veya öğrenimini desteklemek amacıyla geliştirilmemiş olsa da; ses tanıma, metin seslendirme araçlarının gelişmesi ve makine çevirilerinin iyi düzeye ulaşması ile birlikte dil öğretiminde ve öğreniminde ders materyalleri hazırlama, dersi daha eğlenceli hale getirme, öğrencilerin merak ve ilgilerini canlı tutma (Aslan, 2018, s. 102) vb. amaçlarla yardımcı kaynak veya yardımcı bir araç olarak zamanla daha fazla daha etkin bir biçimde kullanılabilir.

Kaynaklar

- Aslan, E. (2018). Otomatik çeviri araçlarının yabancı dil öğretiminde kullanımı: Google çeviri örneği. *Selçuk Üniversitesi Edebiyat Fakültesi Dergisi*, 87-104.
- Balkul, H. İ. (2015). *Türkiye’de akademik çeviri eğitiminde çeviri teknolojilerinin yerinin sorgulanması: müfredat analizi ve öğretim elemanlarının konuya ilişkin görüşleri üzerinden bir inceleme*. Yayımlanmamış Doktora Tezi, Sakarya: Sakarya Üniversitesi.
- Bing Çeviri: <https://www.bing.com/translator/>
- Google Çeviri: <https://translate.google.com/>
- Hutchins J. (2005). *The first public demonstration of machine translation: The Georgetown-IBM system, 7th January 1954*. <http://www.hutchinsweb.me.uk/GU-IBM-2005.pdf> (Erişim: 10.09.2018).
- Hutchins, WJ., & Somers, HL. (1992). *An introduction to machine translation*. Londra: Academic Press.
- Klein, G., Kim, Y., Deng, Y., Senellart, J., and Rush, A. M. (2017). *OpenNMT: Open-source toolkit for neural machine translation*. 2017, ArXiv <https://arxiv.org/abs/1701.02810> (Erişim: 25.09.2019).
- Microsoft (2020). *What languages does Microsoft translator support?*. Web: <https://goo.gl/EbYg9g> (Erişim 15.10.2018)
- Quoc V. Le ve Mike Schuster (2016), *A neural network for machine translation, at production Scale*. Web. 27 Eylül 2016 Google, <https://goo.gl/3RRj1p> (Erişim. 20.10.2019)
- Wikipedia, (2020). *Google çeviri*. Web: <https://goo.gl/xqcBzx> (Erişim: 27.02.2020)
- Yandex (2017). *One model is better than two. Yandex.Translate launches a hybrid machine translation system*. Web: <https://goo.gl/U9zoqR> (Erişim. 20.10.2019).
- Yandex Çeviri: <https://ceviri.yandex.com.tr/>
- Yandex, (2020). *Bilgisayar destekli Yandex çeviri*. Web: <https://goo.gl/j9AYVs> (Erişim: 26.02.2020)

Extended Abstract

A logic stating that machine systems which make translation from one language to another since computer technology has come into our life may exist has developed and many studies have been conducted in this respect. Google Translation, Yandex Translation, Microsoft Bing Translation systems are the leading ones among the translation systems that are being commonly used in our present day and offering service to the users with new features by developing further every day.

The expression of “Machine Translation” is being basically used in two meanings: 1) It is a part of a computational linguistics examining the use of software tools in order to translate the text or speech from a language (source language) into another (target language). 2) Machine translation is a process of translating any text in source language to target language by means of a software/application without any human interference.

Some of the methods used in the machine translation are: Dictionary Based Machine Translation/Direct method, Rule-Based Machine Translation, Example Based Machine Translation, Statistical Machine Translation, Neural Machine Translation.

We will primarily discuss the Rule-Based Machine Translation, Statistical Machine Translation and Neural Machine Translation of the most commonly used methods in order to evaluate the point where machine translations have come today.

Rule-Based Machine Translation

Rule-based machine translation is important in terms of being the first method used on behalf of the automatic translation. The fundamental principle in the machine translation using this method is based on the process of converting the source language data into pre-programmed computer language by means of translation engines and transmitting to the target language codes which are already introduced to the computer. Translations made through the rule-based machine translation approach yield good results in direct proportion to either vocabulary structure, grammar rules of the source and target languages or the redundancy of common grounds where word syntax principles gather.

The rule-based machine translation primarily uses the language rules in analyzing the content entered in source language in order to create a text in target language. Words and rules obtained after the analysis are matched with the words and rules of the target language, a result is attained by the syntax rules. This process provides great advantages for the languages having the same or similar structure, as stated above.

Statistical Machine Translation

We can express the Statistical Machine Translation as identification of the most probable equivalent of the sentence given in source language, in the target language. In the method of the Statistical Machine Translation, translations made by natural persons are analyzed and rules are learned (machine learning), the most probable equivalent in the target language is identified by statistics and translation process is performed.

Statistical Machine Translation matches with the translations that are previously made by using the context around a particular word instead of making single-word translations and focuses on attaining the most probable result. One of the negative aspects of the statistical machine translation is that it translates the words in the context of the several preceding and subsequent words while translating. For this reason, while the statistical machine translation is partially successful in short sentences, it cannot show this success in translation of long sentences. Nevertheless, it has caused the emergence of the method of Neural Machine Translation which we will discuss below.

Neural/Artificial Machine Translation

The method which we can call as Neural Machine Translation or Artificial Machine Translation that is one of the new and the most important machine translation methods of our present day is an end-to-end learning approach for the automatic translation. In the neural machine translation method, sentences are translated not by part by part, all parts in the sentence are simultaneously translated. In the method where sentences are dealt with entirely, the translations are evaluated in a broad context, texts are rendered close to the colloquial language by taking the common usages into account and reorganizing with the correct grammar.

We have already said in the foregoing that the statistical machine translation evaluates the sentence with the several preceding and subsequent words while translating the words and this is resulted in faulty translations, its success in short sentences could not be attained in the long sentences. In the method of the Neural Machine Translation, all words in the sentence are evaluated in broader context and combinations of the related word with all other words in the sentence, in other words, models are created. These models created by the artificial neural network are translated into target language by being re-evaluated in the context of the sentence. Thanks to this feature, problems encountered in the statistical machine translation are minimized and more proper results are attained.

Google (translate.google.com) provides translation service in more than 4000 combinations in more than 100 languages today and it uses the Neural Machine Translation in 8 languages, English, French, Deutsch, Spanish, Portuguese, Chinese, Japanese, Korean and Turkish.

Microsoft Bing (bing.com/translator) offers text translation service in more than 70 languages, it uses the method of Neural Machine Translation in 60 of these languages. Besides, Microsoft also provides the speech translation service by means of Microsoft Speech technology in 11 languages (Microsoft, 2020).

While an interlanguage is used in the method of Statistical Machine Translation (L1-EN-L2 / Source language – English – Target Language), it is directly translated into the target language in the method of Neural Machine Translation without using an interlanguage (L1-L2).

One of the most commonly used translation systems along with Google Translation and Microsoft Bing Translation is Yandex Translation in our present day (ceviri.yandex.com.tr). Today, Yandex Translation providing translation service in more than 95 languages is an automatic translation service that is based on a special computer translation technology developed by Yandex and can be used for translation of words, texts and text contents of the web pages. The computer-aided Yandex Translation that does not know the complicated rules of foreign languages makes translation by analyzing the languages it works on with statistical methods, not by the pre-taught language rules. However, it is not always possible to take successful result particularly in the long and complicated sentences in the statistical method. For this reason, Yandex has included the Neural machine translation into its translations like Google and Microsoft and called this method as Hybrid Machine Translation. Yandex notes that it uses the power of neural and statistical methods together in the Hybrid Machine Translation System, it has created a system which can make more qualified translations like natural persons in this way.

Conclusion and Evaluation

It is seen at the point where we stand now that Google, one of the popular translation companies, uses Neural Machine Translation method for 8 languages, but statistical machine translation method for other languages; Microsoft uses Neural Machine Translation method in 60 of 70 languages for which it provides service; Yandex uses a Hybrid method.

It is seen that since the logic of machine learning underlies the neural machine translation, the neural machine translation has further developed with the increase of text number which the machine will analyze and attain learning, and will further develop thereafter; problems encountered as a result of the use of interlanguage in the rule-based machine translation and statistical machine translations (L1-EN-L2) are partially solved with the method of neural machine translation which can directly make translation between two languages (L1-L2) without using an interlanguage; the neural machine translation is more successful in translation processes of languages particularly having similar structure or originating from the same language family.

There is a need in the rule-based and statistical machine translation methods for teaching the computer the language rules, a good dictionary and collection. Lack of bilingual rich dictionaries, improved written and oral collections leads some languages to be put at disadvantage in the machine translations. It is contemplated that the neural machine translation will be partial solution for these problems, the machine learning will further develop with the increase of the rate of texts in related languages and translations that are more correct and closer to natural person's translation can be attained.

In our present day, the use of open-source code logic leading to the fast development and proliferation of particularly software technology also for the machine translation methods and the presentation of the neural machine translation algorithms to all users as open-source show that studies conducted in this field will further increase.

We can say that the inference of Hutchins and Somers expressed in the study they conducted over the machine translation in 1992 “*there is no translation machine which can translate sentence in any language to another language at an excellent degree without any human interference*” (Hutchins and Somers:1992:1) is not correct when it is considered from the point when the neural machine translation method stands; we will readily use the sentence saying “there is a translation machine which can translate sentence in any language to another language at an excellent degree without any human interference” in the near future with the increase of studies in artificial intelligence field and machine learning.