

Makine Çevirilerinde Sorun Oluşturan Eş Yazımlı Sözcükleri Bulunduran İfadeler için Kural Matrisli Makine Çevirisi (KMMÇ) Modeli

A Rule Matrix Machine Translation (RMMT) Model for the Phrases Having Homographs That Cause Difficulties in Machine Translation

Nevzat ÇAPOĞLU

nevzatcapoglu@hotmail.com

ORCID: 0000-0002-8922-7951

Öz

Doğal diller tam olarak matematiksel formüllerle ifade edilemediği ve sosyal kabullere dayanan sözcük veya ifadeler barındırdıkları için makine çevirilerinde bazı anlam belirsizliği durumlarıyla karşılaşabilmektedir. Bu belirsizlik durumuna en uygun örnek eş yazımlı sözcük(ler) bulunduran ifadelerdir. Çalışmada bu ifadelerin kural tabanlı bir çeviri modeli ile doğru şekilde İngilizceden Türkçeye çevrilebilmesinin mümkün olup olmadığı sorusu ele alınmıştır. Bu maksatla, Kural Matrisli Makine Çevirisi (KMMÇ) olarak adlandırılan bir model geliştirilmiştir. Önerilen modelin başarı ölçümleri için Google, Bing Yandex ve KMMÇ çıktıları çevirmen değerlendirmesi yoluyla karşılaştırılmış; değerlendirmelerde KMMÇ'nin, eş yazımlı bulunduran ifadelerde dikkate değer başarı elde ettiği görülmüştür.

Anahtar sözcükler: Makine Çevirisi, Doğal Dil İşleme, Kural Tabanlı Makine Çevirisi, Hiyerarşik İfade Tabanlı Makine Çevirisi, Eş yazımlılar

Abstract

For a natural language cannot be expressed in exact mathematical formulas or can harbor words or phrases that are based on social acceptance, it is very likely to face some ambiguities in machine translations. The most appropriate

example would be the phrases with homographs. The purpose of this study is to investigate the question of whether a rule-based translation model can translate the expressions with homonymous ambiguity from English to Turkish language. Therefore, a new model called Rule Matrix Machine Translation (RMMT) has been developed. To determine the translation success of the proposed model; Google, Bing, Yandex and RMMT outputs were compared using human evaluation. In the evaluation process, RMMT has shown a notable success with homonymous ambiguities.

Keywords: Machine Translation, Natural Language Processing, Rule-based Machine Translation, Hierarchical Phrase-based Machine Translation, Homographs

1. Giriş

Her ne kadar bir doğal dil, makine yoluyla bir başka doğal dile eksiksiz ve doğru şekilde çevrilemiyor olsa da bu alandaki çalışmalar halen devam etmekte ve son yıllarda yapay sinir ağlarıyla insan beynini taklit eden Nöral Çeviri (*Neural Translation*) çalışmalarının oldukça başarılı sonuçlar verdiği bilinmektedir. Ancak bu çevirilerde dahi uzun mesafe bağımlılıkları veya uzak yapılarla sahip diller söz konusu olduğunda doğru sözcük karşılıklarının bulunabilmesinde zorluklar yaşanabilmektedir. Özellikle eş yazımlılar gibi birden fazla anlamı olan sözcüklerin görüldüğü cümlelerde bu duruma oldukça sık rastlanmaktadır. Bu hata durumuna örnek olarak "A crane flew above the construction crane." cümlesinin Google, Bing ve Yandex çevirileri Çizelge-1 ile aşağıda sunulmuştur.

Giriş yapılan kaynak dil cümlesinde hem 'turna' hem de 'vinç' manalarına gelen 'crane' sözcüğü bulunmaktadır. Cümlenin hedef dildeki karşılığının "İnşaat vincinin üzerinden bir turna uçtu." olması gerekirken '...vinç uçtu' şeklinde çevrildiği görülmektedir (Çizelge-1).

Çizelge-1: Google, Bing ve Yandex Yanlış Çeviri Örnekleri

Örnek cümle	Google	Bing	Yandex
A crane flew	İnşaat	Bir vinç	İnşaat
above the	vincinin	inşaat vinci	vincinin
construction	üzerinde bir	üzerinde	üzerinde bir
crane.	vinç uçtu.	uçtu.	vinç uçtu.

Çalışmada Chiang'ın [1] hiyerarşik ifade tabanlı istatistiksel makine çevirisine benzer bir model sunulmaktadır. Ancak geliştirilen sistemin eğitiminde Chiang'ın modelinin aksine paralel metinlerden yararlanılmaz. Geliştirilen modelde çekim ekleri, yapım ekleri ve uzun mesafe bağımlılıkları için keskin karar mekanizmalarına sahip kural tabanlı bir yaklaşım tercih edilmiştir. Bu modelde çıktıların tutarlı ve tahmin edilebilir olmasını, çeviri hatalarının kolay tespit edilmesini ve önlem alınması gereken noktaların çabuk belirlenebilmesini sağlayan bir kural matrisi tasarlanmıştır. Dilbilimsel verinin tek bir matriste tutulduğu bu model, Kural Matrisli Makine Çevirisi (KMMÇ) modeli olarak adlandırılmıştır. KMMÇ modeli iki dilli sözlük dışında herhangi bir desteğe gerek duymayıp istatistiksel yöntemlerde olduğu gibi tek veya iki dilli derlemlerden yararlanmaz. Ancak, KMMÇ'de bir aradil görevi yerine getiren kural matrisinin eğitilmesi ve iki dilli sözlük içeriğinin geliştirilebilmesi noktalarında uzman desteğine gerek duyar.

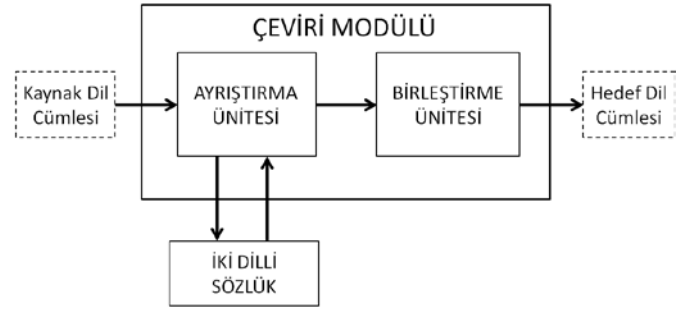
Modelin çeviri başarısı değerlendirmeleri için eş yazımlı sözcükler bulduran İngilizce girdi cümleleri tespit edilmiş; bu girdiler Google, Bing, Yandex ve KMMÇ modellerinde kullanılarak Türkçe çeviri çıktıları elde edilmiştir. Türkçe çıktılar oluşturulan bir değerlendirme ölçeği yardımıyla çevirmen değerlendirmesine tabi tutulmuştur. Elde edilen cümle çeviri dereceleri çeviri hatalarının karşılaştırılabilirliği sağlanmıştır. Ayrıca Google, Bing, Yandex ve KMMÇ çıktılarında göze çarpan çeviri hatalarının üç farklı noktada kümelendiğine dair tespitlerde bulunulmuştur.

1.1 Bilinen Kısıtlar

Önerilen KMMÇ modeli, bütünüyle çeviri sorununa odaklanan ve kendisinden önceki makine çevirisi çalışmalarının kapatamadığı açıkları kapatma düşüncesiyle tasarlanmamıştır. Çalışmada yalnız İngilizceden Türkçeye kural tabanlı bir makine çevirisi modeli geliştirilmiş ve sunulmuştur. Sistemin Türkçeden İngilizceye çeviri yapabilmesi için bu modelin yanında tamamen farklı bir MÇ tasarımına gerek vardır.

2. Kural Matrisli Makine Çevirisi Modeli

Modelin çeviri süreci iki ayrı üniteye gerçekleşir: Ayrıştırma ünitesi ve birleştirme ünitesi. Sistem yalnız İngilizceden Türkçeye çeviri yapabilecek şekilde tasarlanmıştır. KMMÇ'nin ayrıştırma ünitesi, birleştirme ünitesi ve iki dilli sözlüğü içeren genel görünümü Şekil-1 ile sunulmuştur.



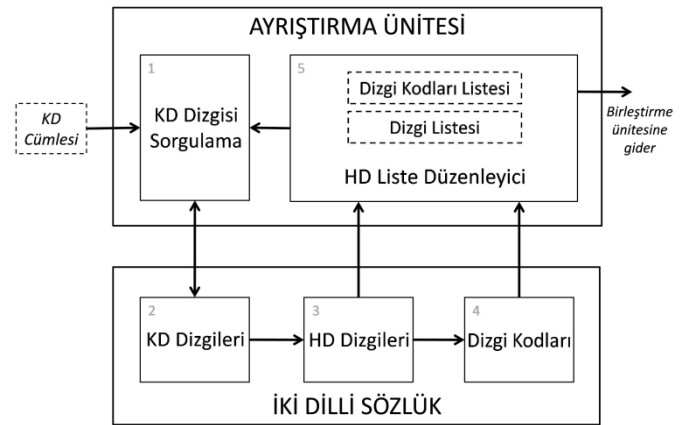
Şekil-1: KMMÇ modelinin genel görünümü

Çeviri modülüne (Şekil-1) girdisi yapılan kaynak dil cümlesi, ayrıştırma ünitesine gelir. Ayrıştırma ünitesi iki dilli sözlük yardımıyla kaynak dil girdisini gerekli ek, sözcük ve ifade dizgilerine (strings) ayırmak ve bu dizgileri devamlılar (nonterminals) veya bu çalışmadaki isimle "kodlar" ile eşleştirmekten sorumludur. Eşleştirme yapılan kodlar listesi ve Türkçe dizgiler listesi bir sonraki aşama olan birleştirme ünitesine gönderilir.

Birleştirme ünitesinde dizgiler kodları üzerinden analiz edilir ve bu dizgilerin hangi sıra ile birleşeceklerine ve hangi yapım veya çekim eki alacaklarına karar verilir. Böylece ayrıştırma ünitesinde kodlar eklenerek parçalara ayrılan İngilizce dizgi, birleştirme ünitesine geldiğinde yine bu kodlar yardımıyla bir araya getirilerek Türkçe dizgi üretimi gerçekleştirilir.

2.1 Ayrıştırma Ünitesi

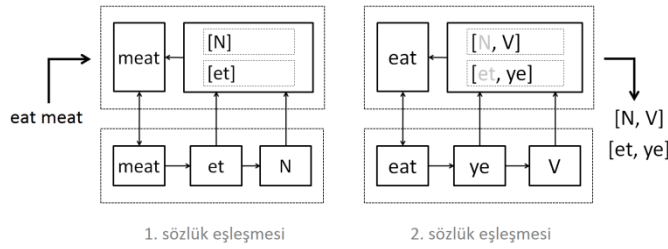
Ayrıştırma ünitesi İngilizce girdiyi gerekli ek, sözcük ve ifade dizgilerine ayırmak ve bu dizgileri kodlar (nonterminals) ile eşleştirmek için üç ayrı birimden yararlanır. Bunlar KD dizgisi sorgulama birimi, iki dilli sözlük ve liste düzenleyici birimidir. Şekil-2 ile ayrıştırma ünitesinin bileşenleri ve aralarındaki ilişkiler görsel hale getirilmiştir.



Şekil-2: Ayrıştırma ünitesi ve bileşenleri

Ayrıştırma ünitesindeki (Şekil-2) KD dizgisi sorgulama birimi İngilizce girdiyi alır ve bu girdideki ifade, sözcük veya ek dizgilerini iki dilli sözlükte arar. KD dizgileri biriminde aranan dizgi bulunduğu İngilizce dizginin iki dilli sözlükteki karşılığı olan Türkçe dizgi HD dizgileri biriminden alınır ve bu Türkçe dizgiye bağlı olan dizgi kodu da dizgi kodları biriminden alınarak HD liste düzenleyiciye gönderilir. Bu birim, hem gelen Türkçe dizgiyi Dizgi listesinde hem de dizgi kodunu Dizgi kodları listesinde aynı indis sırasıyla dizer. Şekil-3 ile "Eat

meat." girdisinin dizgi kodları listesi ve dizgi listesine dönüşümü görsel hale getirilmiştir.



Şekil-3: Ayrıştırma ünitesinde "Eat meat." dizgisinin dizgi kodları listesi ve Türkçe dizgi listesine çevrilmesi

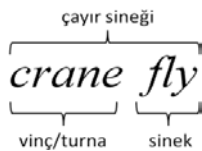
2.1.1 KD Dizgisi Sorgulama Birimi

Ayrıştırma ünitesi (Şekil-2) sınıfı içinde tanımlanmış bir dizgi bulma yöntemi, İngilizce dizgideki ifade, sözcük ve ekleri iki dilli sözlükte aratılmak üzere sorgulara çevirir. Yöntem İngilizce girdiyi sözlükte tararken mümkün olan en az sayıda sorgu döndürecek şekilde çalışır. Bunun için, sözcük arası boşluk karakterler de dâhil olmak üzere aramasını tüm girdi cümlesi karakterlerinden başlayıp her seferinde dizgi başından bir karakter eksilterek yapar. Eğer aradığı dizgi ile sözlükteki dizgi eşleşirse, eşleşen parçayı girdi dizgisinden ayırır ve kalan dizgiyi tekrar sözlükten sorgulamaya devam eder. Her sorgu, girdideki tüm dizgi karakterlerinden başlayarak yapılır ve her seferinde dizgi başından bir karakter eksiltilecek sorguya devam edilir (Şekil-4).

1. Sorgu	2. Sorgu	3. Sorgu	4. Sorgu
Eat meat	at meat	t meat	meat

Şekil-4: "Eat meat." dizgisinin her sorguda baştan bir karakter eksiltilecek sözlükte taratılması ve "meat" sözcüğünün bulunması

KD dizgisi sorgulama birimi bu döngüyü (Şekil-4) İngilizce dizgide, liste düzenleyiciye gönderilecek dizgi kalmayana kadar; yani taratılan dizgide sıfır karakter kalana kadar sürdürür. Bu özellik, analiz edilen girdide birden çok sözcük barındıran ifade dizgilerinin öncelikli olarak tercih edilmesini sağlar. Örneğin 'vinç/turna' ve 'sinek' anlamına gelen 'crane' ve 'fly' sözcüklerinin yan yana bulunması durumunda sistem bu iki farklı sözcüğü 'çayır sineği' anlamına gelen tek bir dizgi 'crane fly' olarak seçecektir (Şekil-5).



Şekil-5: Sözlük taramasında en fazla karaktere sahip dizginin tercih edilmesi yoluyla "crane fly" dizgisinin yakalanması

Bu yöntem birden fazla sözcük içeren ifadelerin gereksiz yere dizgilere parçalanarak yanlış sözcük sorgularına fırsat vermezken tek bir sözcük içerisinde oluşabilecek yanlış çözümlene durumlarına da engel olur. Örneğin 'yaratmak' anlamına gelen 'create' sözcüğünün sonundaki üç harf (ate) sözlükte bulunduğu için ayrı bir sözcük olarak listeye alınır. Ayrıca sözcükten geriye kalan 'cre' dizgisiyle hiçbir sözlük eşleşmesi olmayacağı için doğru çevirinin yapılabilmesi olanaksız hale gelecektir.

2.1.2 Hedef Dil Liste Düzenleyici Birimi

Ayrıştırma ünitesi (Şekil-2) sınıfı içinde tanımlanmış bir başka yöntem olan liste düzenleyici, iki dilli sözlükten elde edilen HD dizgileri ve dizgi kodlarını liste düzenleyicideki listelere (dizgi listesi, dizgi kodları listesi) tüm elemanları sağdan sola olacak şekilde sıralar (Şekil-6). Kod ve Türkçe dizgilerin tersten sıralanması çeviri modülünün ikinci ve son aşaması olan birleştirme ünitesinde gerçekleşen analizlerin sondan başlayarak yapılıyor olmasından kaynaklanır. Aşağıdaki (Şekil-6) "more mischievous cat" örneğinde kod ve dizgiler, köşeli parantezlerle oluşturulan (python) liste veri tipinde görülmektedir.

['cat',	'mischievous',	'more',	' '],
['Na',	'JJ',	'JJ'	' '],
['kedi',	'yaramaz',	'daha',	' ']]

Şekil-6: Ayrıştırma ünitesi çıktısı

Verilen iç içe liste bir ayrıştırma ünitesi çıktısıdır (Şekil-6). Ancak, İngilizce dizgi listesi ünite çıktısında bulunsa da birleştirme ünitesine gönderilmez. HD liste düzenleyici (Şekil-2) HD dizgilerinin ve dizgi kodlarının listelenmesi işlemini yine bir liste (Şekil-6) içinde gerçekleştirir. Bu dış liste sayesinde listeler bir arada tutulur.

Örnekteki (Şekil-6) "more mischievous cat" dizgisi üç parçaya bölündüğü halde iç listeler dört elemanlıdır. Daha sonraki örneklerde de karşılaşılabilecek olan bu kod '|' liste düzenleyici tarafından her listenin sonuna eklenir. Bu yolla ayrıştırma ünitesinden çıkan tek sözcüklü bir listede bile en az iki dizi eleman bulundurulmuş olur. Bunun sebebi sonraki aşama olan birleştirme ünitesinin dil bilimsel analizler için en az iki dizi elemanına ihtiyaç duymasıdır.

Girdiler taranırken sıklıkla eş yazımlı (birden fazla anlamı olan) sözcüklere rastlanır. HD liste düzenleyici, sözcüklerin farklı anlamlarını da kodlarıyla birlikte sözlükten çıkarıp kod listesine ekler. Mesela 'crane' sözcüğü kod listesine 'N', 'Na' ve 'V2'; Türkçe dizgi listesine ise 'vinç', 'turna' ve 'boynunu uzat' olarak işlenir (Şekil-7).

['N',	'Na',	'V2',	' '],
['vinç',	'turna',	'boynunu uzat',	' ']]

Şekil-7: Üç farklı anlamı olan 'crane' eş yazımının ayrıştırma ünitesi çıktısı

Sunulan modelde eş yazımlı bir sözcük en fazla beş farklı çevirisi ile listelere eklenir. Aşağıda "My old car's doors." ifadesinin ayrıştırma ünitesi çıktısı verilmiştir (Şekil-8). Görselde (Şekil-8) dikkat çeken İngilizce dizgiler, yalnız anlatım kolaylığı sağlamak amacıyla gösterilmiştir. Dizgilerin bir araya getirildiği sonraki aşama olan birleştirme ünitesinde İngilizce dizgilere ihtiyaç duyulmaz.

Listede 'old' sözcüğü iki farklı Türkçe çevirisi ile listelenmiştir. Çünkü bu çevirilerden hangisinin seçileceği birleştirme ünitesine bırakılmıştır.

[['s', 's', 'door', 's', 's', 'car', 'old', 'old', 'my', ' '],
 ['PLR', 'TPS', 'N', 'S_', 'BE', 'N', 'J', 'Ja', 'DT', ' '],
 ['_ler', ' ', 'kapı', '_in', '_dir', 'araba', 'eski', 'yaşlı', '-m', ' ']]

Şekil-8: Sağdan sola doğru sıralanmış İngilizce dizgiler, Dizgi kodları ve Türkçe dizgiler

Ayrıştırma ünitesi iki dilli sözlükte sorgulanan sözcük veya ifadelerde bulunan yapım veya çekim eklerini de kodlar. Yani "My old car's doors." (eski arabamın kapıları) ifadesinde geçen "car's" sözcüğündeki sahiplik eki "-s" ile "doors" sözcüğünü çoğul yapan ek "-s", listelerde yerini alır. Ayrıca, "car's" ifadesindeki "-s", hem 'olmak' anlamına gelen "is" sözcüğünün kısa hali "-s" hem de sahiplik bildiren "-s" olabilir. Ekin, bu iki yapıdan hangisi olduğu, komşu sözcüklere göre tespit edilir. Özetle, bu birimde sözlükten çekilen her anlam veya fonksiyon ihtimali, kod ve Türkçe dizgi listelerine eklenir ve analiz edilmek üzere birleştirme ünitesine gönderilir.

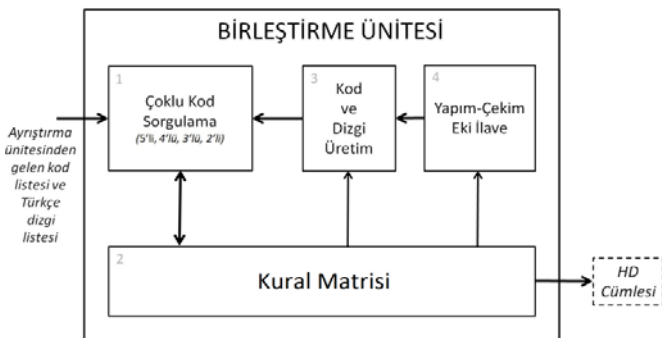
2.1.3 İki Dilli Sözlük

SQLite veri tabanı olan iki dilli sözlük 'kod', 'ingilizce' ve 'turkce' sütunlarını içeren bir tablodur. Sözlükte her İngilizce dizginin bir kodu ve bir çevirisi vardır. İki veya daha fazla sayıda aynı dizginin sözlükte farklı kodlar aldığı durumlar yalnız eş yazımlı ('crane', 'fly' vb.) sözcüklerin bulunduğu ifadelerde veya benzer karakterler içeren dil bilgisi yapılarında görülür. Örneğin "doors" sözcüğündeki "-s" eki, sözlükte karşılık olarak hem geniş zamanda üçüncü şahıslar için fiile eklenen 'TPS' (Third Person Singular) koduyla; hem de çoğul eki olan 'PLR' (Plural) kodu ile eşleştirilir (Şekil-8).

Sözlükteki İngilizce sözcükler için uygun kod ve çeviriler çevirmenlerce seçilirken biçim bilim, sözcük bilim, söz dizim ve sosyal kullanım düzeyinde değerlendirmelerden yararlanılır. Bu nedenle iki dilli sözlük içeriği oluşturma veya geliştirme çalışmaları, alanında uzman ve her iki dilin de kültürüne hâkim kişilerce yapılmalıdır.

2.2 Birleştirme Ünitesi

Çeviri modülünün ikinci ve son aşaması olan birleştirme ünitesi, ayrıştırma ünitesinden gelen kod listesi ve Türkçe dizgi listelerinden HD çıktısını üretebilmek için dört ayrı birimden yararlanır. Bunlar (i) çoklu kod sorgulama, (ii) Kural matrisi, (iii) kod ve dizgi üretim ve (iv) yapım-çekim eki ek birimleridir (Şekil-9).



Şekil-9: Birleştirme ünitesi genel görünümü

Birleştirme ünitesinde görülen (Şekil-9) çoklu kod sorgulama birimi, üniteye gelen kod ve dizgi listelerini kod listesi üzerinden kural matrisinde analiz eder. Çoklu kod sorgulama birimi, dizgilerin hangi yapım veya çekim eki olarak birleşmesi gerektiğini kod ve dizgi üretim ile yapım çekim eki ek birimleri yardımıyla kural matrisinden öğrenir ve bir sonraki analiz için kod ve dizgi listelerine son halini verir.

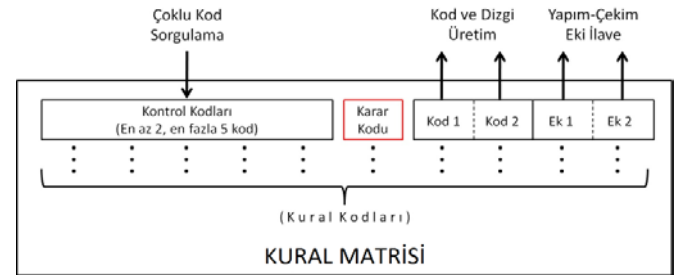
2.2.1 Çoklu Kod Sorgulama Birimi

Ayrıştırma ünitesinden çıkan kod listesi ve Türkçe dizgi listesi ilk olarak çoklu kod sorgulama birimine (Şekil-9) gelir. Birimde tanımlanmış olan çoklu kod sorgulama yöntemi listelerdeki hangi kodların ve dolayısıyla hangi dizgilerin önce birleşeceğini bulmak için kural matrisi içerisinde kod taraması yapar. Kod taraması, kod listesi elemanlarının (2'li, 3'lü, 4'lü veya 5'li gruplar halinde) kural matrisi içindeki kontrol kodlarıyla (Şekil-10) eşleşmelerinin aranması işlemidir. Eşleşen kodlar bulunduktan sonra sıra, kod ve dizgi listeleri üzerinde analiz yapacak olan kural matrisine gelir.

2.2.2 Kural Matrisi

Kural matrisi (Şekil-10) listeler üzerinde dört aşamalı bir analiz gerçekleştirerek (i) hangi kodlar ile bu kodlara bağlı dizgiler üzerinde değişiklik (birleştirme, değiştirme, silme vb.) yapılacağına; (ii) birleşecek veya değişecek kodların yerine hangi kodların geleceğine, (iii) dizgilerin hangi yapım-çekim eki alacağına (iv) ve yapım-çekim ekleri almış dizgilerin hangi sıra ile yeni bir dizgi olarak birleşeceğine karar verir.

Kural matrisinde binlercesi bulunan kural kodları, çeviri modülünün HD dizgisi üretiminde gerek duyduğu dil bilimsel verileri tutan listelerdir. Liste veri tipinde tutulan liste elemanları, görevleri itibarıyla dörde ayrılır. Görselin (Şekil-10) solundaki 2-5 adet kod, kontrol kodlarıdır. Bu kodların sağında karar kodu bulunur. Karar kodunun sağında ise iki ikame kodları (Kod 1, Kod 2) ile ek dizgileri (Ek 1, Ek 2) elemanları sıralanır.



Şekil-10: Kural matrisi liste elemanları ve irtibatlı oldukları birimler

Kural matrisinin doğru anlaşılması için kural kodlarının aralarındaki ilişkinin doğru tarif edilmesi gerekir. Bu nedenle matris yapısı 'Kontrol kodları', 'Karar kodu', 'İkâme kodları' ve 'Ek dizgileri' olmak üzere dört alt başlıkta ele alınmıştır.

2.2.2.1 Kontrol Kodları

Ayrıştırma ünitesinde üretilen kod listesi ve Türkçe dizgi listesi ilk olarak çoklu kod sorgulama birimine gelir. Çoklu kod sorgulama birimi, listelerdeki hangi kod ve dizgiler üzerinde değişiklik (birleştirme, değiştirme, silme vb.) yapılması gerektiğini tespitten sorumludur. Birim, bunu yapabilmek için dizgi kodlarından (1) en az iki, en fazla beş adet kodun kural

kodları (2) içindeki kontrol kodlarıyla eşleşip eşleşmediğine bakar.

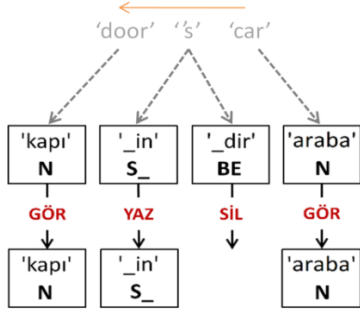
[['_ler', '', 'kapı', 'in', '_dir', 'araba', 'eski', 'yaşlı', '-m', '|'], (1)

['PLR', 'TPS', 'N', 'S_', 'BE', 'N', 'J', 'Ja', 'DT', '|']
 ['N', 'S_', 'BE', 'N', '?X_?', 'S_'] (2)

Örnekte (1) dört kod listesi elemanının, ['N', 'S_', 'BE', 'N'] kural listesindeki (2) dört kontrol koduyla bire bir aynı olduğu görülmektedir. Bu eşleşme, bulunan dört kod ve aynı indise sahip dizgiler ['kapı', 'in', '_dir', 'araba'] üzerinde değişiklik yapılacağı anlamına gelir. Ancak değişikliğin ne olacağı, kontrol kodunun sağındaki karar koduna (?X_?) bağlıdır.

2.2.2.2 Karar Kodu

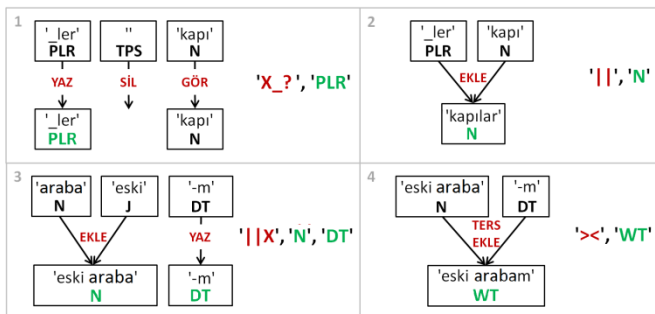
Her karar kodu, solundaki kontrol kodlarının toplam sayısı kadar dizgi karakterine sahiptir. Her karakter, örnekte (2) olduğu gibi ('?'→'N', 'X'→'S_', ' '→'BE', '?'→'N') bir kontrol kodu ile eşleşir. Hangi kod üzerinde ne tür bir ekleme-çıkarma yapılacağı karar kodu karakterlerinin sırasına ve ne olduklarına göre değişir. Örnekteki (2) karar kodu '?X_?', “gör, yaz, sil, gör” sıralamalıdır ve bu sıraya göre kodlar ve dizgiler üzerinde yapılacak değişiklikler Şekil-11 ile görselleştirilmiştir.



Şekil-11: Bir karar kodu ?X_? (gör, yaz, sil, gör) uygulama örneği

Ayrıştırma ünitesi çıktısında (1) “...car’s door...” (...arabanın kapısı...) ifadesinin dizgilere bölünmüş hali sağdan sola 'door', 's', 'car' şeklindedir. Örnekte (Şekil-11) bu dizgi ve kodlar kutularda verilmiştir. Kutulardan iki ve üçüncü sıradakiler 's' dizgisinin karşılığı iki 'S_', 'BE' dizgi kodları ve onlara bağlı '-in', '-dir' Türkçe dizgilerdir. Kural matrisi, dört koda baktıktan sonra 'BE' kodunu irtibatlı olduğu '_dir' dizgisiyle birlikte silmiştir Şekil-11. Çünkü '?X_?' kural koduna göre iki 'N' kodlu isim yapısı arasında, sahiplik ifade eden 'S_' kodlu kutu kalmalı; 'BE' kodlu kutu silinmelidir.

Karar kodlarının farklı uygulama örnekleri için Şekil-12 incelenebilir.



Şekil-12: Karar kodları uygulama örnekleri

Şekil-12 ile gösterilen birinci örnekte 'X_?' (yaz, sil, gör) karar kodu kullanılmış ve bir önceki görselde (Şekil-11) olduğu gibi, silme yapılmıştır. İkame kodu aynı (PLR) kalmıştır. İkinci örneğin karar kodu '||' (ekle) seçildiği için kod ve dizgiler birleşerek 'N' ikame kodlu 'kapılar' dizgisini oluşturmuştur. Üçüncü örnekteki '|X' (ekle, yaz) kodu, 'eski' ile 'araba' dizgilerini 'N' ikame koduyla birbirine eklemiş (eski araba) ve bu kodların önündeki 'DT' kodunu da değiştirmeyip aynı ikame kodunu yazmıştır. Dördüncü örneğin karar kodundaki '>' (ters ekle) karakterleri ise ikame kodu olan 'WT'yi dizgiye ilişkilendirmeden önce, baştaki 'DT' kodlu '-m' ekini 'N' kodlu 'eski araba' dizgisinin sonuna eklemiştir (Şekil-12).

2.2.2.3 İkame Kodları

Karar kodunun 'X' (yaz), '||' (ekle) ve '>' (ters ekle) komut karakterleri, değiştirmek durumunda oldukları kodların yenilerini ikame kodlarından temin eder. Bu iki ikame kodunun kural kodlarındaki yeri, kural listesindeki karar kodu ile ek dizgileri (Ek 1, Ek 2) arasındadır (Şekil-10).

Üçüncü örnekte (Şekil-12) diğer örneklerden farklı olarak iki ikame kodu vardır. Çünkü karar kodu (|X), hem ekleme hem yazma işinde görevlidir ve bu kod birinci değişiklik olan ekleme (|) için birinci 'N' ikame kodunu, ikinci değişiklik olan yazma (X) için de ikinci 'DT' ikame kodunu kullanır.

2.2.2.4 Ek Dizgileri

Ek 1 ve Ek 2 (Şekil-10), Türkçede mevcut olup İngilizcede bulunmayan ekler kullanılır. Örneğin İngilizcede üç dizgiden ('car', 's', 'door') oluşan “car’s door” ifadesi Türkçede dört dizgi ('araba', '-nin', 'kapı', '-sı') ile karşılık bulur. Çünkü Türkçe isim tamlamalarındaki tamlanan eki İngilizce isim tamlamalarında yoktur. İngilizcede iki ismin yapısının yan yana gelmesi bir tamlama oluşturmak için yeterlidir. Bu nedenle “evin kapısı” ifadesindeki ikinci sözcüğe (kapı) eklenecek olan '-sı' tamlanan eki, Ek 2’den alınır. Birinci sözcükler için Ek 1 kullanılır. Örneğin “See İstanbul.” cümlesi, birinci sözcüğe '-u' belirtme eki ilave edilerek “İstanbul+u gör.” şeklinde çevrilir.

2.2.3 Kod ve Dizgi Üretim

Üzerinde değişiklik yapılmasına karar verilen kodların (ve onlara bağlı dizgilerin) birleşme, değişme veya silinme işlemleri kod ve dizgi üretim biriminde (Şekil-9) tanımlanmış birleştirme metoduyla yapılır. Örneğin üç kod içerisinde ilk ikisinin birleşmesini ve sonraki kodun da değişmesini isteyen bir karar kodu (|X) görülmesi durumunda bu metod, birleşecek olan kodların yerine 'Kod 1'i; değişecek olan kodun yerine de 'Kod 2'yi geçirir. Kodların üzerinde yapılan bu değişiklik, kodların indislerle bağlı olduğu dizgilere de yapılır. Ancak değişiklikleri yapılan dizgiler önce yapım-çekim eki ilave biriminde yapım ve/veya çekim eklerini (varsa) almalıdır.

2.2.4 Yapım-Çekim Eki İlave

Türkçede “paket+ten” sözcüğündeki ekin (-ten) ilk sessiz ve seslisi, birleştiği gövdenin (paket) son sessiz ve seslisi ile uyum içinde olmak için 't' ve 'e' olarak seçilmiştir. Bir başka örnek olan “araba+dan” sözcüğü de sesli harf ile bittiği için ekin ilk sessiz harfi 'd' olarak kalmıştır. Ancak ekteki sesli harf uyum için 'a' olmak durumundadır. Sesli (ünlü) uyumu soldan sağa

silsile şeklinde gelişir [2]. Biçimbirimler düzeyinde gelişen bu durum için birleştirme ünitesi sınıfı içerisinde bir ek-ilave yöntemi tanımlanmıştır.

Ek-ilave metodu yalnız ses uyumlarıyla değil; ismin yalın, belirtme, yönelme, bulunma ve ayrılma durumlarına göre değişen hal ekleriyle de ilgilendir. Çizelge-2 ile verilen örneklerde 'home' sözcüğünün farklı fiillerle kullanıldığında Türkçe karşılığının farklı ekler aldığı görülür.

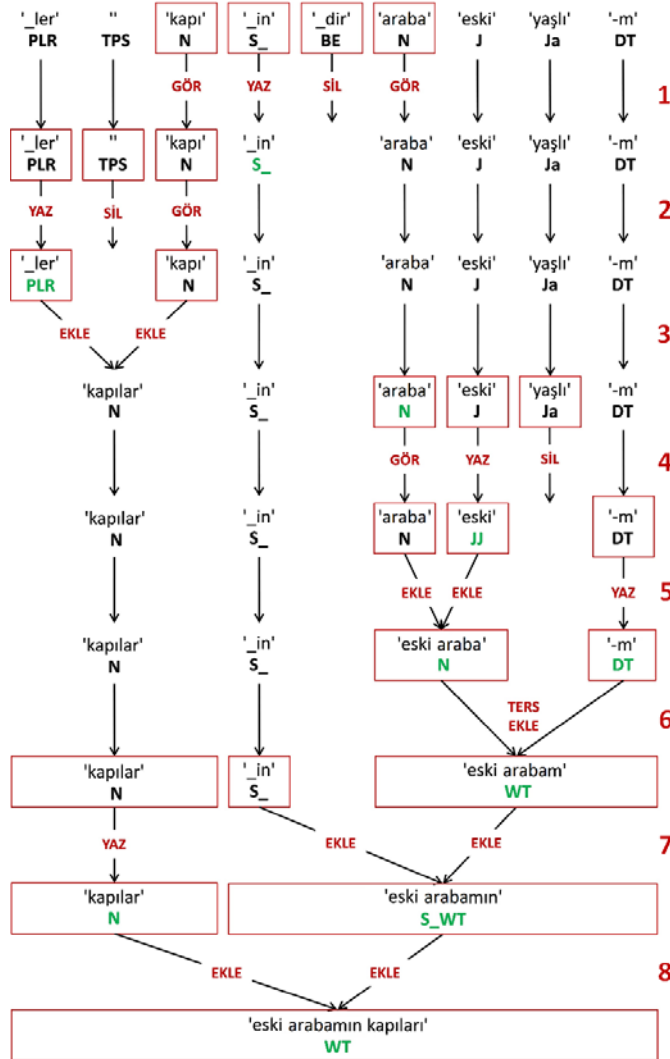
Çizelge-2: Türkçede Fiillere Göre Değişen İsmi Beş Hâli

Örn.	KD Cümlesi	HD Cümlesi	HD İsim Hâlleri
1	Say home.	Ev de.	Yalın hâli
2	Clean home.	Evi temizle.	'-i' hâli (Belirtme)
3	Go home.	Eve git.	'-e' hâli (Yönelme)
4	Stay home.	Evde kal.	'-de' hâli (Bulunma)
5	Leave home.	Evden ayrıl.	'-den' hâli (Ayrılma)

Örneklere (Çizelge-2) 'home' sözcüğünün çevirisi olan 'Ev' yalın durum hariç dört farklı hâl ekiyle (-i, -e, -de, -den) kullanılmıştır.

2.2.5 Birleştirme Ünitesinde HD Üretim Aşamaları

Şekil-13 ile "My old car's doors." dizgisi üzerinde, kural kodları (3)-(10) yardımıyla gerçekleştirilen üretim aşamaları görselleştirilmiştir.



Şekil-13: İngilizce dizgi üretimi aşamaları

Kutularda görülen (Şekil-13) 'DT', 'Ja', 'J', 'JJ', 'N', 'BE', 'S_', 'TPS', 'PLR', 'WT', 'S_WT' dizgi kodlarının (tokens) hangi dil bilimsel yapıları tarif ettikleri şu şekilde tarif edilebilir: (i) 'My', 'Your', 'The' gibi niteleyici sözcükler (Determiners) için 'DT'; (ii) hangi anlamının kullanılacağı seçilmemiş sıfatlar (Adjectives) için 'J' veya 'Ja'; (iii) hangi anlamının kullanılacağı belirlenmiş sıfatlar (Adjective) için 'JJ'; (iv) isimler (Nouns) için 'N'; (v) 'olmak' fiili için 'BE'; (vi) sahiplik bildiren 's' eki (Possessive) için 'S_'; (vii) geniş zamanda üçüncü tekil şahıslarda (Third Person Singular) kullanılan 's' takısı için 'TPS'; (viii) isimleri çoğul (Plural) yapan '-s' eki için 'PLR'; (ix) isim ifadeleri (Noun phrases) için WT; (x) sahiplik (Possessive) eki 's' bulunduran isim ifadeleri (Noun phrases) için ise 'S_WT' kullanılmıştır.

Görselde (Şekil-13) sayılarla ifade edilen aşağı yönlü oklar üretim aşamalarını gösterir. Sekiz katmanlı üretim için sekiz adet kural kodlarından (3) - (10) yararlanılmıştır.

- ['N', 'S_', 'BE', 'N', '?X_?', 'S_', "", "",], (3)
- ['PLR', 'TPS', 'N', 'X_?', 'PLR', "", "",], (4)
- ['PLR', 'N', "", '|', 'N', "", "",], (5)
- ['N', 'J', 'Ja', '?X_', 'JJ', "", "",], (6)
- ['N', 'JJ', 'DT', '|X', 'N', 'DT', "", ""], (7)
- ['N', 'DT', '>', 'WT', "", "", ""], (8)
- ['N', 'S_', 'WT', 'X|', 'N', 'S_WT', "", ""], (9)
- ['N', 'S_WT', "", '|', 'WT', "", "", '-si',], (10)

Yukarıdaki listeler (3) - (10) kural matrisinde bulunan binlerce liste içerisinde sadece "My old car's doors." dizgisi üzerinde uygulanmış kural kodlarıdır. Birleştirme ünitesinin Türkçe dizgi üretimi aşamaları bu kural listeleri (3) - (10) üzerinden tarif edilecektir. Aşağıdaki (11) iç içe liste, "My old car's doors." dizgisinin ayrıştırma ünitesinde üretilen ve birleştirme ünitesine gönderilen kod ve dizgilerdir.

Kural kodları (3) - (10) içindeki kodlardan biriyle (3) eşleşmiş olan kontrol kodları bu örnekte (11) kalın harf karakterleriyle gösterilmiştir. Eşleşen kodlar üzerinde yapılacak her değişiklik, aynı indis değerine sahip (yine kalın olarak yazılmış) dizgilere (11) de uygulanır.

- [['PLR', 'TPS', 'N', 'S_', 'BE', 'N', 'J', 'Ja', 'DT', '|'], (11)
- ['_ler', "", 'kapı', 'in', '_dir', 'araba', 'eski', 'yaşlı', '-im', '|']

Dizgi kodları (11) üzerinde birinci kural (3) işletildiğinde aşağıdaki kod ve dizgi listeleri (12) elde edilir. Birinci kuraldaki karar kodu (?X_?) komut karakterlerine (?, X, _ ?) göre kod ve dizgi üzerinde silme işlemi uygulanmış; 'BE' kodu ve ona bağlı '-dir' eki buldukları listelerden kaldırılmıştır.

- ['PLR', 'TPS', 'N', 'S_', 'N', 'J', 'Ja', 'DT', '|'] (12)
- ['_ler', "", 'kapı', 'in', 'araba', 'eski', 'yaşlı', '-im', '|']

Yukarıdaki kodlarda (12) ikinci kural (4) uygulandığında sıradaki iki listenin (13) üretimi gerçekleşmiş olur. Karar kodu (X_?) komut karakterlerine göre kod ve dizgilerde yine silme işlemi uygulanmış; 'PLR', 'TPS', 'N' kodları içinden 'TPS' ve ona bağlı boş liste karakteri (") silinmiştir.

['PLR', 'N', 'S_', 'N', 'J', 'Ja', 'DT', '|'] (13)
['_ler', 'kapı', '_in', 'araba', 'eski', 'yaşlı', '-im', '|']

Kalan kodlarda (13) sıradaki kural (5) uygulanarak aşağıda görülen listelerin üretimi (14) gerçekleşmiş ve karar kodu (||) komut karakterlerine göre kod ve dizgilerde birleştirme yapılmıştır. 'PLR' ve 'N' kodları yerine 'N' ikame kodu yazılmış; irtibatlı oldukları liste karakteri '_ler' ve 'kapı' dizgileri de bir araya getirilerek 'kapılar' dizgisi elde edilmiştir. ('kapı' ve '-ler' dizgileri, birim içinde tanımlanmış yapım-çekim eki ilave metodu yardımıyla 'kapı+lar' şeklinde birleşmiştir.)

['N', 'S_', 'N', 'J', 'Ja', 'DT', '|'] (14)
['kapılar', '_in', 'araba', 'eski', 'yaşlı', '-im', '|']

Üretilen liste kodlarında (14) bir sonraki kural (6) uygulanmış ve aşağıdaki listeler (15) elde edilmiştir. Karar koduna (?X_) göre kod ve dizgilerde silme yapılmış; 'J' ve 'Ja' kodları arasında tercih yapılarak yerlerine 'sıfat tercihi yapıldı' anlamına gelen 'JJ' ikame kodu getirilmiştir. Silinen 'Ja' kodu ile aynı indisteki 'yaşlı' dizgisi de listelerden kaldırılmıştır.

['N', 'S_', 'N', 'JJ', 'DT', '|'] (15)
['kapılar', '_in', 'araba', 'eski', '-im', '|']

Üzerinde silme işlemi gerçekleştirilen kod listesinde (15) bir sonraki kural (7) uygulanmış ve sıradaki listeler (16) üretilmiştir. Komut karakterlerine (||X) göre 'N' ve 'JJ' kodlarının, irtibatlı oldukları dizgilerin tek dizgi haline gelerek 'N' ikame kodunda birleşmesine; 'DT' kodu ve ona bağlı dizginin ise değişmeden kalarak üretilen listeye geçirilmesine karar verilmiştir.

['N', 'S_', 'N', 'DT', '|'] (16)
['kapılar', '_in', 'eski araba', '-im', '|']

Ortaya çıkan yeni liste kodlarına (16) sıradaki kural (8) uygulanmış ve aşağıdaki listeler (17) elde edilmiştir. Karar koduna (><) göre kod ve dizgilerde baştaki öne, öndeki başa gelecek şekilde birleştirme yapılmıştır.

['N', 'S_', 'WT', '|'] (17)
['kapılar', '_in', 'eski arabam', '|']

Birleştirme işlemi gerçekleştirilen kod listesinde (17) müteakip kural (9) uygulanmış ve aşağıdaki listeler (18) üretilmiştir. Karar koduna (X||) göre, 'S_' ve 'WT' kodları ile onlara bağlı dizgilerin 'S_WT' ikame kodunda birleşmesi; 'N' kodu ve ona bağlı dizginin de aynı şekilde yazılması gerekir.

['N', 'S_WT', '|'] (18)
['kapılar', 'eski arabamın', '|']

Birleştirme işlemi yapılmış kod listesinde (18) son kalan kural (10) da işletildikten sonra HD dizgi kodu ve HD Türkçe dizgisini içeren listeler (19) üretilmiş olur. Komut karakterlerine (||) göre 'N' ve 'S_WT' kodları ve bağlı oldukları dizgiler birleşmiş; dizgi kodu olarak 'WT' ikame kodu seçilmiştir. Ancak hemen öncesinde kural kodundaki (10) ek 2'de bulunan '-si' eki, ikinci sıradaki 'kapılar' dizgisine (19) eklenmiş; yapım-çekim eki yöntemi sayesinde bu ek, birleşeceği dizgiye ilave edilmeden önce harf uyumuna göre değişerek 'kapılar+ı' şeklinde eklenmiştir.

['WT', '|'] (19)
['eski arabamın kapıları', '|']

Örnekte (19) elde edilen 'WT' dizgi kodu ile 'eski arabamın kapıları' dizgisinin sağında bulunan '|' karakterinin, en az iki elemana ihtiyaç duyan kural matrisi analizleri için listelerde bulunması gereklidir. Dizgi listesinde cümle başı anlamına gelen bu dizgi karakteri HD dizgi çıktısına (20) dâhil edilmez.

['eski arabamın kapıları'] (20)

Özetle, bu aşamaya kadar KMMÇ modülünün iki ana birimi olan ayrıştırma ve birleştirme üniteleri ele alınmış; (i) ayrıştırma ünitesi başlığı altında KD girdisinin iki dilli sözlük yardımıyla gerekli ek, sözcük veya ifade dizgilerine (strings) ayrıldığı; bu dizgilerin kodlarla (nonterminals [devamlılar]) eşleştirildiği; eşleşen kod ile dizgi listelerinin birleştirme ünitesine gönderildiği ve (ii) birleştirme ünitesi başlığında bu kodların çözümlenmesi yoluyla dizgilerde hangi değişikliklerin yapılarak HD çıktısının üretildiği tarif edilmeye çalışılmıştır.

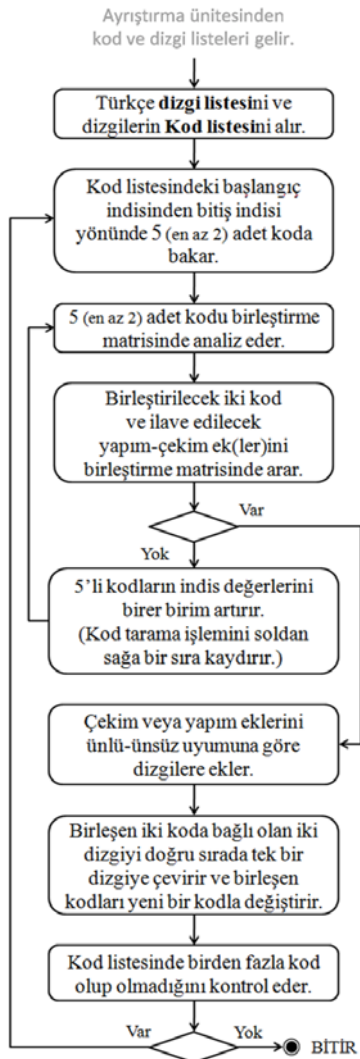
Şekil-14 ile KMMÇ modelinin iki ana biriminden ilki olan ayrıştırma ünitesinde gerçekleşen dizgi ayrıştırma aşamalarını tarif eden akış diyagramı sunulmuştur.



Şekil-14: Ayrıştırma ünitesi akış diyagramı

Diyagramda (Şekil-14) ayrıştırma ünitesine İngilizce (KD) girdi dizgisi alınır. Ayrıştırılacak kod ve dizgiler için boş kod listesi ve boş Türkçe dizgi listesi tutan bir iç içe liste oluşturulur. Üniteye alınan İngilizce dizgi karakterlerinin tümü iki dilli sözlükte aranır. Karakterler sözlükte bulunmadığında dizginin başından bir karakter eksiltilerek yeniden sözlük araması yapılır. Dizgi başından her seferinde bir karakter eksiltilerek yinelenen arama işlemi dizginin sözlükte eşleşeni bulunana kadar devam eder. Bulunan dizginin karşılığı olan Türkçe dizgi ve dizgi kodu sözlükten alınır. Türkçe dizgi, (daha önce oluşturulan) Türkçe dizgi listesine; dizgi kodu da (daha önce oluşturulan) dizgi listesine gönderilir. Gönderilen dizgi, bulunduğu girdi dizgisinden çıkarılır. Kalan KD dizgisi için, başından bir karakter eksilterek yapılan yinelenmeli sözlük taramasına devam edilir. İngilizce dizgide karakter kalmayana kadar sözlük taraması devam eder. Her döngüde sözlükten alınan kod ve Türkçe dizgi elemanları aynı indis değerine sahip şekilde kendi listelerine eklenir. Oluşan bu listeler sonraki birleştirme ünitesine gönderilir.

KMMÇ modelinin iki ana biriminden ikincisi olan birleştirme ünitesinde gerçekleşen üretim aşamalarını tarif eden akış diyagramı (Şekil-15) aşağıda sunulmuştur. Diyagramda dizgi birleştirme işleminin (silme, değiştirme vb. hariç) gerçekleştirildiği aşamalar görsel hale getirilmiştir.



Şekil-15: Birleştirme ünitesi akış diyagramı

Birleştirme ünitesi (Şekil-15) gelen kod ve Türkçe dizgi listelerini (iç içe liste yapısında) aldıktan sonra tüm analizlerini kodlar üzerinden yapar. Kodlarda yapılacak her değişiklik dizgi elemanları için de geçerli olur. Değişiklik için, kod listesindeki bitiş indisinden başlangıç indisine doğru 2 ile 5 arasında değişen sayıda kod seçilerek kural matrisinde analiz edilir. Analiz, seçilen kodlar üzerinden birleşmesi gereken dizgiler ve eklerin aranması işlemidir. Birleşmesi gereken kodlar bulunmadığında seçilen kodlar sondan başa doğru bir sıra kaydırılır (Soldan bir kod çıkarılıp sağdan bir kod eklenir.) ve tekrar kural matrisinde analiz yapılır. Birleşecek kodlar (ve bu kodlara bağlı olan Türkçe dizgiler) bulunduktan sonra eklenmesi gereken ekler (varsa) ilave edilir. Ekleri (varsa) ilave edilmiş iki dizgi doğru sıra ile arada bir boşluk karakteri olacak şekilde birleştirilir. Birleşen dizgilere ait eski iki kodun yerine de yeni bir kod getirilir. Böylece hem kod hem dizgi listelerinde birer eleman eksilmiş olur. Birleştirme işlemi, kod listesinde birleştirilecek kod kalmayana kadar devam eder.

2.2.6 Birleştirme Ünitesinde HD Üretim Aşamaları

Eş adlı, eş sesli ve eş yazımlı olarak bilinen sözcükler bazen birbirleriyle karıştırılabilir [3]. Bu nedenle kısa bir açıklama yapmak yerinde olacaktır.

Eş adlılar (*Homonyms*) farklı anlamları olan ve aynı şekilde telaffuz edilen sözcüklerdir [3]. Örneğin; 'fast' sözcüğü /fæst/ (fest) olarak okunur ve hem "oruç tutmak" anlamında fiil hem "hızlı" anlamında bir sıfattır. Eş sesliler (*Homophones*) farklı anlamlara geldiği gibi farklı yazılıp aynı şekilde telaffuz edilen sözcüklerdir. Bunlara, "deniz" ve "görmek" manalarına gelen ve /si:/ (sii) olarak okunan 'sea' ve 'see' sözcükleri örnek gösterilebilir. Eş yazımlılar (*Homographs*) ise birden fazla anlama sahip olduğu halde yazımlarının aynı olduğu [4] sözcüklerdir ve bu sözcüklerin eş yazımlı kabul edilebilmesi için (*Heteronym* olarak bilinen sözcüklerdeki gibi) farklı telaffuzlarının olması şeklinde bir şart yoktur [5]. Farklı telaffuzu olan eş yazımlılara örnek olarak 'object' sözcüğü gösterilebilir. Bu sözcüğün "nesne" anlamında kullanılması /'a:bdʒekt/ ('abcekt) ve "itiraz etmek" anlamında kullanılması ise /'æb'dʒekt/ (ib'cekt) şeklinde telaffuz edilir.

KMMÇ modelinin başarı değerlendirmelerinde yararlanılan dizgilerde eş sesliler yoktur. Bunun nedeni çeviri değerlendirmelerinin yazılı metinler üzerinden yapıyor olmasıdır. Ayrıca makale içinde; eş adlı ve eş yazımlı kavramları içinden, eş yazımlı tabirinin kullanılması uygun görülmüştür. Çünkü yazımı aynı olup söylenişi farklı (*Heteronym*) sözcükler ile yazımı ve söylenişi aynı (*Homonym*) sözcükler için, "eş yazımlı" (*homograph*) tabiri bir üst kavram niteliğindedir.

2.2.6.1 Eş Yazımlı Bulunduran Dizgi Çevirileri

Önerilen KMMÇ modülünde eş yazımlı sözcük analizleri dil bilimsel verinin tutulduğu kural matrisiyle gerçekleştirilir. Bu analizlerden birine örnek niteliğinde iki eş yazımlı sözcük içeren "Old fly." dizgisinin ayrıştırma ünitesi çıktısı (21) sunulmuştur.

[[['V', 'V1', 'NNa', 'J', 'Ja', '|'],
['uç', 'uçur', 'sinek', 'eski', 'yaşlı', '|']] (21)

Ayrıştırma ünitesine iki sözcüklü bir dizgi girişi yapıldığı halde çıktısında (21), 'old' (eski, yaşlı) sözcüğünün iki; 'fly' (uçmak, uçurmak, sinek) sözcüğünün ise üç farklı anlamı olmak üzere toplam beş farklı sözcük çevirisi (sağdan sola) sıralanmıştır. Hangi iki sözcüğün doğru çeviri olduğu birleştirme ünitesindeki analizler yoluyla belirlenir. Birleştirme ünitesinin, gelen kod listesi (21) analizinde yararlandığı kural listeleri (22), (23), (24), (25) uygulanma sıralarına göre verilmiştir.

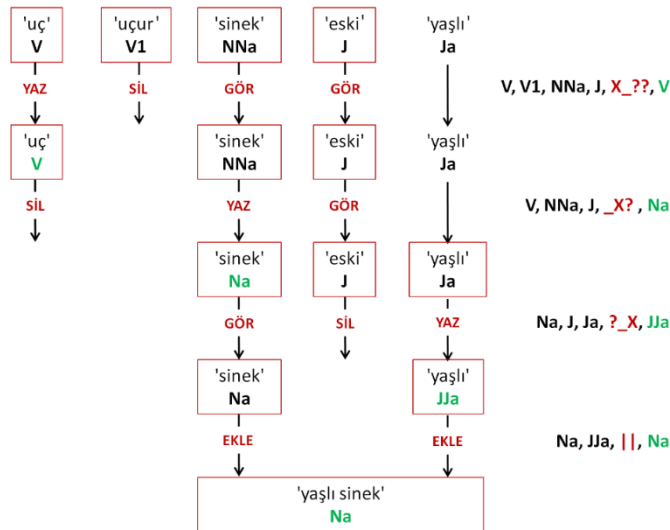
['V', 'V1', 'NNa', 'J', 'X_??', 'V', "", "", "", ""], (22)

['V', 'NNa', 'J', '_X?', 'Na', "", "", "", ""], (23)

['Na', 'J', 'Ja', '?_X', 'JJa', "", "", "", ""], (24)

['Na', 'JJa', "", '|', 'Na', "", "", "", ""], (25)

Bu dört liste (22) - (25) dizgi kodu listesine (21) bakılarak kural matrisinden seçilmiş kural kodlarıdır. Dizgiye ait analizler aşağıda (Şekil-16) görüldüğü gibi şeklin sağındaki kural listelerine göre yapılır. Karar kodlarındaki her komut karakteri (X, _, ?, vb) işlem yaptığı kutu altında yazı ile (yaz, sil, gör, vb) gösterilmiştir.

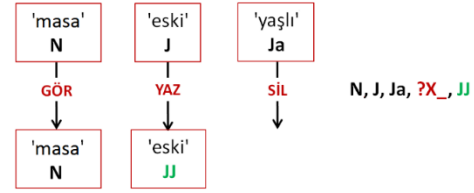


Şekil-16: Eş yazımlı sözcükler için doğru çeviri analizi aşamaları

Görseldeki (Şekil-16) dizgi analizleri kural listeleri (22) - (25) yardımıyla dört aşamada gerçekleşmiştir: (i) Birinci kuralın (22) işletildiği ilk aşamada kırmızı kutular ile gösterilen dört kod analiz edilmiş; 'J' (sifat kodu) ile tanımlanan sözcükten sonra (solunda) doğal dilin kuralı gereği bir isim yapısının bulunması beklendiği için 'V1' (fiil kodu) ve ona bağlı 'uçur' dizgisi silinir. (ii) İkinci kuralın (23) işletildiği ikinci aşamada kırmızı kutular ile gösterilen üç kod analiz edilmiş ve yine 'J' (sifat kodu) ile tanımlanan sözcükten sonra fiil gelemeceği için 'V' (fiil kodu) ve ona bağlı 'uç' dizgisi kaldırılır. Böylece, eş yazımlı 'fly' sözcüğünün üç farklı anlamından geriye sadece 'sinek' dizgisi, ve anlam seçiminin tamamlandığına işaret eden 'Na' kodu kalmıştır. (iii) Analizinin devamı 'old' sözcüğünün hangi anlamda kullanılacağı (yaşlı/eski) ile alakalıdır. Üçüncü sıradaki kural (24) 'Na', 'J', 'Ja' kodlarını yan yana gördüğü yerde "eski" dizgisinin bağlı olduğu 'J' kodunu silip "yaşlı" dizgisinin bağlı olduğu 'Ja' kodunu bırakır. Kalan 'Ja' kodu ise mana tercihinin yapıldığına işaret eden 'JJa' (ikame kodu) ile değiştirilir. (iv) Dördüncü aşamada son kural (25) işletilmiş ve

'yaşlı' ile 'sinek' dizgileri 'Na' kodunda birleşerek HD dizgisinin üretimi gerçekleşmiştir.

Ancak; görselde (Şekil-16) 'Na' kodlu bir 'sinek' dizgisi yerine aşağıdaki (Şekil-17) gibi 'N' kodlu 'masa' (akılsız varlık) dizgisi olmuş olsaydı, seçilecek kural listesi ve ona bağlı yapılan kod-dizgi seçimi de farklı olacaktı (Şekil-17).



Şekil-17: Tamlanana göre tamlayanın seçildiği analiz örneği

Görselin (Şekil-17) sağındaki kural kodlarındaki 'N', 'J', 'Ja' kontrol kodları, kutulardaki dizgi kodlarıyla eşleştiği için karar kudundaki (?X_) "gör, yaz, sil" işlem sırasına göre 'Ja' kodu silinip 'J' kodu bırakılmış ve kalan 'J' kodu ise mana tercihinin yapıldığına işaret eden 'JJ' (ikame kodu) ile değiştirilmiştir.

3. Uygulama

Tasarlanan KMMÇ modeli ile eş yazımlı sözcükler bulunduran cümle çevirileri yapılarak aynı cümlelerin Google, Bing ve Yandex çevirileri ile karşılaştırılması sağlanmıştır. Karşılaştırma için en güvenilir değerlendirme yöntemi olarak kabul edilen çevirmen değerlendirmesinden yararlanılmıştır.

3.1 KMMÇ Modelinin Test Aşaması ve Kullanılan Yöntem

Makine çevirilerinde girdi-çıkı arasındaki ilişkinin değerlendirilme şekli, ölçümün kara kutu (black box) testi ya da cam kutu (glass box) testi olup olmadığını belirleyen faktördür. Kara kutu incelemeleri, çeviri modelinin mekanizmasını hesaba katmadan yalnız girdi ve çıktıları değerlendirirken; cam kutu incelemeleri, MÇ modelinin içine bakarak her bir bileşenin çeviri sürecinde kendisinden bekleneni yapıp yapmadığına bakar. İki yöntemin de farklı avantajları vardır. Cam kutu araştırmaları daha çok sistemlerin kapsamının genişletilebilirliği gibi durumların tespitinde kullanılır. Kara kutu testleri, araştırılan sistemin yapısını veya tasarım felsefesini dışarıda tuttuğu için farklı sistemlerin karşılaştırılması veya belirli bir modelin çeviri başarısının tespiti için daha uygundur [6]. Dolayısıyla önerilen KMMÇ modelinin başarı ölçümleri için, karşılaştırıldığı diğer modellerden farklı olarak kural tabanlı yapı arz etmesi ve birbirinden uzak dil yapılarına sahip dillerin çevirisi için tasarlanmış olması nedeniyle, bir makro değerlendirme yaklaşımı olan kara kutu testi uygun görülmüştür. Testin uygulanabilmesi için öncelikle Google, Bing Yandex ve KMMÇ modellerinin çeviri başarılarının karşılaştırılmasında kullanılmak üzere eş yazımlı bulunduran cümleler tespit edilmiştir.

3.2 KMMÇ Modelinin Değerlendirilmesinde Kullanılan Çeviri Girdileri İçin Eş Yazımlı Sözcük Bulunduran Örnek Cümleler

Eş yazımlılar ve bu sözcüklerin kullanıldığı cümlelerin tespitinde Oxford, Cambiridge, Macmillan, Merriam-Webster, Collins, Yourdictionary gibi sözlüklerin çevrimiçi sürümlerinden yararlanılmış; kısa ve karmaşık olmayan sözlük örnek cümleleri seçilmiştir. Örneklerin özne veya nesne gruplarında eklemeler yapılmıştır. Ancak eklemelerde cümle veya ifadenin yapısal özelliğine sadık kalınmıştır. Bu şekilde 80 eş yazımlı bulunduran cümle bir araya getirilmiştir. Eş yazımlı sözcüklerin isim ifadelerinde (*noun phrase*) kullanılmış olanları Çizelge-3; fiil ifadelerinde (*verb phrase*) kullanılmış olanları ise Çizelge-4 ile sunulmuştur.

Çizelge-3: Eş Yazımlı Sözcüklerin İsim İfadeleri İçinde Kullanıldığı ÇM Girdileri

İsim İfadeleri İçinde Kullanılan Eş Yazımlılar	
1	She was holding the baby in her arms .
2	We saw the little bear cub.
3	My friend is a beef producer.
4	The boil on his arm was painful.
5	He is reading a book about American history.
6	The mailman kept the letter in a box in his room.
7	Carp is a fish.
8	A clip from that video.
9	That bird is a crane .
10	An express shipment.
11	The last drop of my coffee.
12	My friends at work love fast cars.
13	A green flag indicates the start of the race.
14	A kid in high school.
15	The children love to play in the leaves .
16	She was very mean to her sister.
17	Every move was painful for the old man.
18	A collection of precious objects .
19	Fill in the form on the page .
20	A pine forest in the city.
21	At the press of a button.
22	The burning desert sun.
23	Build a ship for the navy.
24	I hate wearing a pink tie .
25	They don't know the tie between my grandmother and the town.
26	Dinner will be a cold buffet .
27	My uncle and my father are very close .
28	The antique console will be in the living room.
29	This is the new word record .
30	I swatted the fly with a newspaper.
31	The leaves rustle in the wind .
32	The does are walking in the forest.
33	They sat on the grass in the park .
34	The gas stove was in the old house in the village.
35	They will be live on the city radio.
36	That person is really not your type .
37	My roommates ate two jars of jam in two days.
38	Endurance was the most important attribute in a teacher.
39	The old man's old watch is fast.
40	There is a huge wave of divorce in the developed countries.

Çizelge-4: Eş Yazımlı Sözcüklerin Fiil İfadeleri İçinde Kullanıldığı ÇM Girdileri

Fiil İfadeleri İçinde Kullanılan Eş Yazımlılar	
41	She arms herself with knowledge.
42	They bear the pain.
43	They are going to beef .
44	They boil the goat milk in a pan.
45	We book flights to big cities.
46	He is going to box in the Olympics in summer.
47	The people carp about junk mail.
48	Her mom is going to clip her hair.
49	They crane forward.
50	I'd like to express my gratitude to everyone.
51	She is going to drop the book.
52	Muslims fast during Ramadan.
53	Flag the potential problems in our proposal.
54	Don't kid yourself.
55	Their father leaves home for work.
56	I didn't mean any harm.
57	They move the flag up and down.
58	People at my work object to the book.
59	Page the man in the hospital.
60	They still pine for their lost wealth.
61	Press the button for the third floor.
62	He deserted his wife and children for another woman.
63	The company ships its goods to twenty countries.
64	The robbers were going to tie the men to a pipe with cable.
65	The buildings house the young students.
66	The waves are going to buffet the shore.
67	We close the office at noon on Fridays.
68	She is going to console herself with the cat.
69	These cameras record 25 frames per second.
70	He was the first person to fly .
71	The little girls were going to fly the kite.
72	My daughter does her homework every night.
73	They park their cars in the building.
74	We are going to park in front of the TV.
75	He wants to live the last years of his life with his dogs.
76	I type over 100 words per minute.
77	Their man's gun will jam .
78	She will attribute her success to her late father's hard work.
79	Our kids only watch YouTube videos at weekends.
80	The people on the bus wave to the villagers.

Oluşturulan bu girdiler (Çizelge-3, Çizelge-4) Google, Bing, Yandex ve KMMÇ modellerinde KD girdisi olarak kullanılmış; elde edilen HD çeviri çıktılarıyla, çevirmen derecelendirme ölçütüne (Çizelge-5) dayanan bir çeviri değerlendirme ölçeği hazırlanmıştır.

3.3 Çeviri Çıktıları ve Çevirmen Değerlendirme Ölçeği

Tüm makine çevirisi çıktıları A ve B çevirmenleri tarafından değerlendirilmek üzere bir çeviri ölçeğinde bir araya getirilmiştir. Dört farklı çeviri çıktısının hangi makine çevirisine ait olduğu, değerlendirici olan A ve B çevirmenlerine bildirilmemiştir. Çevirmenler tarafından Çizelge-5 ile gösterilen derecelere göre değerlendirilen çıktıya karşılık gelen ağırlık puanı, cümlenin değerlendirme puanı olarak atanmış ve bu ağırlık puanlarının aritmetik ortalaması alınarak

her çeviri modeli için 0 ile 100 aralığında belirlenmiş bir çeviri skoru elde edilmiştir.

Öznel değerlendirme olarak da bilinen insan değerlendirmesi yönteminde özellikle şu üç noktaya dikkat edilmelidir: (i) HD çevirisinin KD cümlesine göre anlamsal ve sözdizimsel doğruluğu; (ii) cümlenin anlamsal olarak doğruyken, sözdizimsel olarak yanlış olması; (iii) ve cümlenin hem anlamsal hem de sözdizimsel olarak yanlış olması [7].

Değerlendirme ölçeklerindeki çeviri değerlendirme sıralaması yaklaşımında, akıcılık (*fluency*) ve yeterlik (*adequacy*) ölçümleri gerekli açıklamalarla birlikte isteğe göre yedi katmana kadar artırılabilir [8]. Karşılaştırılan çevirilerin akıcılık ve yeterlik değerleri, söz dizimsel ve anlamsal doğruluğa bağlı olarak değişmektedir. Bahsedilen kriterlere göre bu çalışma için tasarlanmış olan dört katmanlı çevirmen derecelendirme ölçütü Çizelge-5 ile aşağıda sunulmuştur.

Çizelge-5: Çevirmen Derecelendirme Ölçütü

Derece	Açıklama	Ağırlık
Tam	Anlamsal ve sözdizimsel olarak doğru.	100
Makul	Anlamsal olarak doğru; sözdizimsel olarak yanlış.	70
Eksik	Anlamsal olarak yanlış; sözdizimsel olarak doğru.	35
Yanlış	Anlamsal ve sözdizimsel olarak yanlış.	0

Yukarıdaki derecelendirme ölçütüne (Çizelge-5) göre yapılan çevirmen değerlendirmelerinde eş yazımlı bulunduran isim veya fiil ifadelerinin (i) okuyucu tutumunu belirleyen “kabul edilebilirlik”, (ii) kendisinden önce kullanılanagelen cümle veya ifadelerin bilgisine bağlı kılan “metinler arasılık” ve (iii) makul düzeyde “bilgilendiricilik” özelliklerine sahip olması durumunda *anlamsal* olduğu kabul edilmiştir. Ayrıca; bu cümle veya ifadeler, sahip oldukları sözcüklerin doğru sıra ile dizilmesi ve yapım-çekim eklerinin gerekli yerde kullanılmış olması gibi dilbilgisine uyumluluk durumlarında ise *sözdizimsel* olarak doğru sayılmıştır.

Buna göre, bir cümle veya ifadenin anlamsal ve söz dizimsel olarak doğru çevrilmiş olması durumunda çeviri skoru 100 (yüz); anlamsal olarak kabul edilebilir olduğu halde sözdizimsel olarak yanlış veya genel kabul gören sözcük(ler) kullanılmamış ise çeviri skoru 75 (yetmiş beş); anlamsal olarak yanlış ancak sözdizimsel hatası olmayan sözcük(ler) kullanılmış ise çeviri skoru 35 (otuz beş); hem anlamsal hem de sözdizimsel olarak yanlış çevrilmiş ise çeviri skoru 0 (Sıfır) olarak belirlenmiştir.

3.4 Çevirmen Değerlendirme Puanları

Google, Bing, Yandex ve KMMÇ modeli çıktıları için çevirmen değerlendirme puanlaması yoluyla elde edilen çeviri dereceleri ve her bir modelin çeviri başarı değerlendirmeleri iki ayrı çizelge halinde Çizelge-6 ve Çizelge-7 ile sunulmuştur.

Çizelge-6 çeviri skorlarına bakıldığında birinci sırada 96 ile Bing’in, ikinci sırada 93,5 ile Google’ın, üçüncü sırada 92,63 ile KMMÇ’nin ve dördüncü sırada 90 ile Yandex’in olduğu; Çizelge-7 çeviri derecelerine bakıldığında ise birinci sırada 87,88 ile KMMÇ’nin; ikinci sırada 87,63 ile Google’ın; üçüncü

sırada 76,25 ile Yandex’in ve dördüncü sırada 74,38 puan ile Bing’in olduğu görülmektedir.

Çizelge-6: İsim İfadelerinde Google, Bing, Yandex Ve KMMÇ Modeli Çevirmen Değerlendirme Puanları Ve Her Bir Modelin Çeviri Skorunu Gösterir Çizelge

Madde	Eş Yazımlı Sözcüklerin İsim İfadelerinde Kullanıldığı Cümle Çevirisi Puanlamaları			
	Google Derecesi	Bing Derecesi	Yandex Derecesi	KMMÇ Derecesi
1.	100	100	100	100
2.	100	100	100	100
3.	100	100	100	100
4.	100	100	100	100
5.	100	100	100	100
6.	100	100	100	70
7.	100	100	100	100
8.	100	100	0	100
9.	100	100	0	100
10.	100	100	100	100
11.	100	100	100	100
12.	100	100	100	100
13.	100	100	100	100
14.	100	100	100	100
15.	100	100	100	100
16.	70	100	100	100
17.	100	100	100	100
18.	100	100	100	100
19.	100	100	100	100
20.	100	100	100	100
21.	100	100	100	0
22.	100	100	100	100
23.	100	100	100	100
24.	100	70	100	0
25.	100	100	100	100
26.	100	100	100	100
27.	100	100	100	100
28.	100	100	100	100
29.	100	100	100	35
30.	0	100	0	100
31.	100	100	100	100
32.	0	0	0	100
33.	100	100	100	100
34.	100	100	100	100
35.	70	70	100	100
36.	100	100	100	100
37.	100	100	100	100
38.	100	100	100	100
39.	100	100	100	100
40.	100	100	100	100
AO:	93.5	96	90	92,63

Her iki çizelge (Çizelge-6, Çizelge-7) incelendiğinde 80 maddeden 49’unda Google, Bing, Yandex ve KMMÇ çevirilerinin tam başarı gösterdiği, 29 maddede çeşitli hataların bulunduğu ve 2 maddede ise dört makine çevirisinin de başarısız olduğu gözlenmiştir.

Çeviri başarılarının veya başarısızlıklarının benzerlikler gösterdiği çizelge verilerine göre Çizelge-6’daki 6, 8, 9, 16, 21, 24, 29, 30, 32, 35’inci ve Çizelge-7’deki 41, 42, 43, 44, 47, 49, 52, 55, 56, 57, 59, 60, 62, 63, 65, 66, 67, 73, 74, 77, 78’inci maddelerin çevirilerinde farklılıklar olduğu ve bu farklılıkların

ekseri küçük çeviri hataları olmakla birlikte maddelerde dikkate değer yanlış çevirilerin de olduğu gözlemlenmiştir. Bazı yanlış çeviri çıktıların çizelgeler eşliğinde ele alınarak çeviri hatalarının incelenmesi çizelgelerin (Çizelge-6, Çizelge-7) daha anlamlı hale gelmesinde yararlı olacaktır.

Çizelge-7: Fiil İfadelerinde Google, Bing, Yandex Ve KMMÇ Modeli Çevirmen Değerlendirme Puanları Ve Her Bir Modelin Çeviri Skorunu Gösterir Çizelge

Eş Yazımlı Kelimelerin Fiil İfadelerinde Kullanıldığı Cümle Çevirisi Puanlamaları				
Madde	Google Derecesi	Bing Derecesi	Yandex Derecesi	KMMÇ Derecesi
41.	100	0	70	70
42.	35	35	35	100
43.	0	0	0	100
44.	100	100	100	70
45.	100	100	100	100
46.	100	100	100	100
47.	70	0	0	100
48.	100	100	100	100
49.	0	0	0	100
50.	100	100	100	100
51.	100	100	100	100
52.	100	100	70	100
53.	100	100	100	100
54.	100	100	100	100
55.	100	35	100	100
56.	100	35	100	0
57.	100	100	100	35
58.	100	100	100	100
59.	100	100	0	100
60.	100	0	100	100
61.	100	100	100	100
62.	100	100	100	70
63.	100	70	100	100
64.	100	100	100	100
65.	100	100	70	100
66.	100	0	0	100
67.	100	100	70	100
68.	100	100	100	100
69.	100	100	100	100
70.	100	100	100	100
71.	100	100	100	100
72.	100	100	100	100
73.	100	100	100	70
74.	0	0	0	0
75.	100	100	100	100
76.	100	100	100	100
77.	0	0	0	0
78.	100	100	35	100
79.	100	100	100	100
80.	100	100	100	100
AO:	87,63	74,38	76,25	87,88

3.5 Çeviri Hataları Üzerinde İnceleme

Değerlendirme puanları (Çizelge-6, Çizelge-7) düşük veya sıfır olan maddeler incelendiğinde çeviri hatalarının üç farklı noktada kümelendiği görülür. Bu hatalar, sırasıyla (i) eş yazımlı sözcüğün yaygın kullanılan yanlış çevirisinin az kullanılan doğru çevirisine tercih edildiği; (ii) eş yazımlı sözcük belirsizliğinin (*Homonymous Ambiguity*) olduğu ve (iii) ek belirsizliğinin (*Attachment Ambiguity*) bulunduğu durumlarda

ortaya çıkmaktadır. Bu üç farklı hata durumu alt başlıklarda tablolar eşliğinde incelenmiştir.

3.5.1. Yaygın İfadenin Az Kullanılan İfadeye Tercih Edilme Hatası

Alsohybe'ye [9] göre sözcük grubu veya ifadeler eğitim verisinde daha çok kullanılmışsa çeviri motoru yanlış çeviri de olsa çok karşılaşılan sözcükleri doğru çeviriye tercih edecektir. Google, Bing, Yandex ve KMMÇ modeli çıktılarında bu duruma örnek gösterilebilecek cümleler için Çizelge-8 incelenebilir.

Çizelge-8: Eş Yazımlı Sözcüğün Yaygın Kullanılan Anlamının Seyrek Kullanılan Anlamına Tercih Edildiği Çeviri Örnekleri

Madde 32	İNGİLİZCE	The <i>does</i> are walking in the forest.
	GOOGLE	Ormanda yürüyorlar.
	BING	Ormandakiler yürüyor.
	YANDEX	Does ormanda yürüyor.
	KMMÇ	Ormanda geyikler yürümekte.
Madde 43	İNGİLİZCE	They are going to <i>beef</i> .
	GOOGLE	Sığır edecekler.
	BING	Sığır etiyle kavga ediyorlar.
	YANDEX	Sığır eti yiyecekler.
	KMMÇ	Sızlanacaklar.
Madde 47	İNGİLİZCE	The people <i>carp</i> about junk mail.
	GOOGLE	İnsanlar önemsiz posta hakkında konuşuyorlar.
	BING	İnsanlar önemsiz postalar hakkında sazani.
	YANDEX	İnsanlar önemsiz posta hakkında sazani.
	KMMÇ	İnsanlar istenmeyen posta hakkında mızızlanır.
Madde 66	İNGİLİZCE	The waves are going to <i>buffet</i> the shore.
	GOOGLE	Dalgalar kıyıya vuracak.
	BING	Dalgalar kıyıya açık büfe olacak.
	YANDEX	Dalgalar kıyıya açık büfe olacak.
	KMMÇ	Dalgalar sahili dövecek.
Madde 74	İNGİLİZCE	We are going to <i>park</i> in front of the TV.
	GOOGLE	Televizyonun önüne park edeceğiz.
	BING	Televizyonun önüne park edeceğiz.
	YANDEX	Televizyonun önüne park edeceğiz.
	KMMÇ	Televizyon önünde park edeceğiz.

Çizelge-8 için Google, Bing, Yandex ve KMMÇ çıktıları arasından söz konusu çeviri hatasına örnek gösterilebilecek maddelerden bazıları (32, 43, 47, 66, 74) seçilmiştir. Verilen İngilizce girdilerde eş yazımlı sözcükler yatık; çevirisi doğru olan çıktılar ise kalın yazılmıştır.

Çizelge-8 incelendiğinde madde 32'de "geyikler" anlamına gelen "does" sözcüğünün "does" şeklinde; madde 43'te "sızlanmak" olarak çevrilmesi gereken "beef" sözcüğünün "sığır", "sığır eti" ve "sığır etiyle kavga" şeklinde; madde 47'de "mızızlanmak" veya "söylenmek" anlamına gelen "carp" sözcüğünün "sazani" ve "sazani" şeklinde; madde 66'da "vuracak" veya "dövecek" anlamında kullanılan "buffet" sözcüğünün "büfe olacak" şeklinde ve son olarak madde 74'te "yerleşmek" veya "kurulmak" şeklinde çevrilmesi gereken "park" sözcüğünün "park etmek" şeklinde çevrildiği görülmüştür. Benzer çeviri hataları Google, Bing, Yandex ve

KMMÇ çıktılarındaki 9, 24, 41, 49, 51, 65, 77 ve 78'inci maddelerinde de tespit edilmiştir.

Bu hata durumunun önüne geçmek için eğitim verilerinin artırılması yoluna gidilse de sözcük anlamlarının yaygın veya seyrek kullanım oranlarında çeviri çıktılarına etki edecek düzeyde bir değişiklik oluşmaz. Yani "does" eş yazımlısının metin içinde "geyikler" anlamında görülme sıklığı ile "yapar" anlamında görülme sıklığı, sistemin eğitilmesi için gerekli olan çeviri çiftlerinin artırılması durumunda da benzer oranda artış gösterecektir.

3.5.2. Eş Yazımlı Belirsizliği (*Homonymous Ambiguity*)

Aslında bu belirsizlik durumu (*Homonymous Ambiguity*) eş yazımlı sözcüklerin geneli için geçerli olmakla birlikte çeviri motorlarının eğitim verilerinin sınırlı olmasından kaynaklanır. Okpor'a [10] göre bu tür çeviri hataları, doğru çevrilmesi beklenen sözcük gruplarının eğitim verisinde büyük oranla yan yana görülmemiş olmasından kaynaklanır. Google, Bing, Yandex ve KMMÇ modeli çeviri çıktılarında bu duruma örnek gösterilebilecek cümleler Çizelge-9 ile aşağıda sunulmuştur.

Çizelge-9: Eş Yazımlı Sözcüğün İlişkilendirilecek Anlamlarının Fazla Olmasından Kaynaklanan Yanlış Çeviri Örnekleri

Madde 42	İNGİLİZCE	They bear the pain.
	GOOGLE	Acı çekiyorlar.
	BING	Acı çekiyorlar.
	YANDEX	Acıyı taşıyorlar.
	KMMÇ	Acıya katlanırlar.
Madde 59	İNGİLİZCE	Page the man in the hospital.
	GOOGLE	Hastanedeki doktora çağrı yapın.
	BING	Hastanedeki doktoru çağrıla.
	YANDEX	Hastanede doktor Page.
	KMMÇ	Hastanedeki doktoru çağır.
Madde 60	İNGİLİZCE	They still pine for their lost wealth.
	GOOGLE	Hala kayıp servetleri için özlemlerini sürdürüyorlar.
	BING	Hala kayıp servetleri için tırnıyorlar.
	YANDEX	Hala kayıp servetleri için can atıyorlar.
	KMMÇ	Kayıp servetleri için hala özlem çekerler.

Çizelge-9 için dört makine çevirisi çıktıları arasından 42, 59 ve 60'uncu maddeler seçilmiştir. Verilen İngilizce girdilerde eş yazımlı sözcükler yatık; çevirisi doğru olan çıktılar ise kalın yazılmıştır.

Yukarıdaki Çizelge-9 incelendiğinde madde 42'de "katlanmak" olarak çevrilmesi gereken "bear" sözcüğünün "taşımak" veya "çekmek" şeklinde; madde 59'da "çağrı yapmak", "çağrılmak" veya "çağırılmak" anlamında kullanılan "page" sözcüğünün "page" şeklinde; madde 60'ta "özlem çekmek" manasında kullanılan "pine" sözcüğünün "tırnıyor" şeklinde çevrildiği görülmektedir. Ayrıca; Çizelge-9'da verilmemiş olsa da benzer çeviri hatası Google, Bing, Yandex ve KMMÇ modeli çıktılarının 25, 29, 52 ve 63'üncü maddelerinde de tespit edilmiştir.

3.5.3. Ek Belirsizliği

Çeviri çıktılarında belirsizlik, sözcük veya cümlede birden fazla farklı anlamların olması durumudur. İki temel belirsizlik türü vardır. Bunlar sözcüksel (tek bir sözcükte iki veya daha fazla

anlam) ve söz dizimsel (tek bir cümle veya ifadede iki veya daha fazla anlam) belirsizliklerdir [11]. Sözcüksel belirsizlik için Nordquist'in [12] makalesinde şu örneğe yer verilmiştir: "We saw her duck." cümlesi hem "Onun eğildiğini gördük." hem de "Onun ördeğini gördük." şeklinde çevrilebilir [13]. Burada eklemek gerekir ki eş sesliler veya eş yazımlılar belirsizliği (*Homonymous Ambiguity*) olarak bilinen bu durum makale çalışmasının genel problem sorusunu ilgilendirmektedir.

Ancak, ek belirsizliği başlığıyla alakalı hata yukarıda bahsedilen sözcüksel belirsizliğe değil; sözdizimsel belirsizlik türüne işaret etmektedir. Ek belirsizliği (*Attachment ambiguity*) olarak da bilinen bu tür anlam karmaşıklıkları, farklı görevlerde kullanılabilen 'in', 'on', 'at' gibi edatlardan (*Prepositions*) kaynaklanmaktadır. Örneğin; "The girl found a book on Main Street." cümlesindeki 'on' edatı belirsizlik içerir: Kitabın Main Street'te bulunduğu veya kitabın konusunun Main Street üzerine yazılmış olup olmadığı konusunda kesin bir hükme varılamaz [11]. Google, Bing, Yandex ve KMMÇ çevirilerinde bu hata durumuna örnek gösterilebilecek anlam değişiklikleri için Çizelge-10 incelenebilir.

Çizelge-10: Ek Belirsizliklerine (*Attachment Ambiguity*) Dayanan Yanlış Çeviri Örnekleri

Madde 6	İNGİLİZCE	The mailman kept the letter in a box in his room.
	GOOGLE	Postacı mektubu odasında bir kutuda sakladı.
	BING	Postacı mektubu odasındaki bir kutuda saklıyordu.
	YANDEX	Postacı mektubu odasındaki bir kutuda sakladı.
	KMMÇ	Postacı odasındaki bir kutudaki mektubu tuttu.
Madde 44	İNGİLİZCE	They boil the goat milk in a pan.
	GOOGLE	Keçi sütünü tavada kaynatırlar.
	BING	Keçi sütünü tavada kaynatıyorlar.
	YANDEX	Keçi sütünü bir tavada kaynatırlar.
	KMMÇ	Bir tavadaki keçi sütünü kaynatırlar.
Madde 73	İNGİLİZCE	They park their cars in the building.
	GOOGLE	Arabaları binaya park ediyorlar.
	BING	Arabaları binaya park ediyorlar.
	YANDEX	Arabaları binaya park ediyorlar.
	KMMÇ	Binadaki arabaları park ederler.

Çizelge-10 için dört makine çevirisi çıktıları arasından 6, 44 ve 73'üncü maddeler seçilmiştir. Verilen İngilizce girdilerde eş yazımlılar yatık; çevirisi doğru olan çıktılar ise kalın yazılmıştır.

'in' edatı yukarıdaki (Çizelge-10) İngilizce girdilerde ek belirsizliği durumuna neden olmuş ve bu yüzden KMMÇ modelinin madde 6 çevirisinde "...kutuda mektubu..." şeklinde olması gereken ifade "...kutudaki mektubu..."; madde 44 çevirisinde "...bir tavadaki..." şeklinde olması gereken ifade "...bir tavadaki..." ve madde 73 çevirisinde "Arabaları binaya..." şeklinde olması gereken ifade "Binadaki arabaları..."; olarak çevrilmiştir. Bu hata türü KMMÇ modeli çeviri hatalarının 1/3'üne (Madde 6, 21, 44 ve 73) tekabül ettiği; Google, Bing ve Yandex modellerinde ise bu hata türüne rastlanmadığı tespit edilmiştir.

4. Sonuç

Sunulan KMMÇ modelinde, başarı ölçümleri için, eş yazımlı sözcükler (*homonymous ambiguity*) bulunduran cümle çevirileri yapılmış; ardından ise elde edilen çıktıların Google, Bing ve Yandex çevirileri ile karşılaştırılması işlemi gerçekleştirilmiştir. KMMÇ sisteminin karşılaştırıldığı diğer modellerden farklı olarak kural tabanlı olması ve İngilizce Türkçe gibi birbirinden uzak dil yapılarına sahip dillerin çevirisi için tasarlanmış olması nedeniyle kara kutu (*black box*) sınaması uygulanmıştır. White'a [6] göre, araştırılan sistemin yapısını veya tasarım felsefesini dışarıda tuttuğu için farklı çeviri sistemlerinin karşılaştırılmasında kara kutu sınamaları tercih edilmelidir. Karşılaştırılacak çeviri modelleri için en iyi değerlendirme yöntemi olarak kabul edilen çevirmen değerlendirmesi kullanılmış; A ve B çevirmenleri yardımıyla her ayrı model için çeviri derecesi elde edilmiştir (Çizelge-11).

Çizelge-11: Google, Bing, Yandex, KMMÇ Modeli Çeviri Skorları Ve KMMÇ İle Diğer Modellerin Çeviri Skoru Karşılaştırması

Madde 1 – 80	KMMÇ'nin Diğer Modellere Göre Başarı	
	Çeviri Skorları	KMMÇ Farkı Yüzdeleri
Google	90,57	Google ile %0,34
Bing	85,19	Bing ile %5,95
Yandex	83,13	90,26 Yandex ile %8,58

Eş yazımlı sözcüklerin hem isim hem fiil ifadelerinde kullanıldığı cümle çevirilerinin modellere göre çeviri dereceleri (Çizelge-11) incelendiğinde birinci sırada 90,57 puan ile Google'ın; ikinci sırada 90,26 puan ile KMMÇ'nin; üçüncü sırada 85,19 puan ile Bing'in ve dördüncü sırada 83,13 puan ile Yandex'in olduğu görülür. Bu veriler ışığında, çalışmada önerilen KMMÇ modeli başarı farkı yüzdelerinin Google ile mukayese edildiğinde %0,34; Bing ile mukayese edildiğinde %5,95 ve Yandex ile mukayese edildiğinde ise %8,58 olduğu tespit edilmiştir.

KMMÇ modelinin çeviri derecesinin diğer makine çevirileri arasında 90,26 ile ikinci sırada olması dikkate değerdir. Ancak modelin eş yazımlı hatalarında minimum hata ile en iyi performansı gösterdiğini ve sıralamadaki ikinciliğinin, eş yazımlı hata türünden farklı bir hata türü olan ek belirsizliği durumundan kaynaklandığını belirtmek gerekir.

Modellerin hedef dil çıktılarında kontrol edilmesi gereken "yeterlilik" ve "akıcılık" noktalarının ölçülmesinde [14][15] çevirmen değerlendirmesinden yararlanılmıştır. Çok daha az güvenilir olmasına rağmen makine çevirilerinde otomatik değerlendirme yöntemlerinin yaygın şekilde kullanılıyor olması hızlı ve düşük maliyet tercihlerinden kaynaklanmaktadır [16]. Bu çalışma gibi hız veya maliyet kaygısı olmayan ölçümlere sahip araştırmalarda gerçekçi sonuçlar için insan değerlendirmesi yönteminin kullanılması yerinde olacaktır. Çünkü otomatik çeviri değerlendirmelerinde kalite puanlandırmaları, makine çıktıları ile çevirmenler tarafından çevrilen bir veya birden fazla referans metinlerin karşılaştırılması yoluyla yapılır. Ancak, Türkçe gibi morfolojik zenginliğe sahip eklemeli (*Agglutinative*) bir dilde, tek bir çıktı için farklı sözcük sıralamalı referans metinler olabilir. Örneğin; "Ben makaleyi

okudum.", "Okudum ben makaleyi." ve "Makaleyi okudum ben." cümlelerinden hiçbirinde yanlış sözcük sıralaması yoktur. Bu nedenle Türkçe, Macarca ve Japonca gibi dillerin bulunduğu MÇ'lerde çevirmen değerlendirmesinden yararlanmak yerinde olacaktır.

Günümüzde makine çevirisine en fazla yatırım yapan şirketlerden biri Google'dır. 2016 yılında hizmete sunduğu nöral çeviri ile Türkçeyi de sistemine dâhil eden şirket, yüzden farklı dilde makine çevirisi hizmeti sağlamakta ve geleceğe dair bu yöndeki çalışmaları ile büyük umutlar vaat etmektedir [17].

Nöral makine çevirilerinde sürece dışarıdan müdahale etmek oldukça zordur. Bu nedenle KMMÇ modeli, yeni geliştirilebilecek bir nöral çeviri modelinde girdi öncesi ön denetleme veya çıktı sonrası son düzenleme fonksiyonu icra edebilir. Bu tür bir ekleme işleminin insan müdahalesi olmaksızın uygulanabilmesi için eş yazımlı sözcükler bulunan dizgilerin girdi öncesi ayıklanması ve KMMÇ modeli yardımıyla çevrildikten sonra sistem çıktısına eklenmesi yolu tercih edilebilir.

Bugüne kadar teknik imkânların elverdiği ölçüde yapılan tüm MÇ çalışmalarında, ancak insan çevirilerine yakın çıktılar elde edilebilmiştir. Özellikle eş yazımlı bulunduran cümlelerin çevirisi yalnız her iki dilin kültürel derinliğine hâkim kişilerce doğru şekilde yapılabilir. Bu nedenle makine çevirilerinde hızlı ve kaliteli sonuçlar için ön düzenleme (*pre-editing*) ve son düzenleme (*post editing*) yardımıyla eksikliklerin kapatılması, halihazırda en doğru çözüm yolu olarak görünmektedir.

Kaynakça

- [1] Chiang, D. Hierarchical Phrase-Based Translation, *Computational Linguistics*, 2007, 33(2), s. 201-228.
- [2] Oflazer, K. *Türkçe ve Doğal Dil İşleme*, Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 2016, Cilt 5, Sayı 2, 24.06.2016 <https://dergipark.org.tr/tr/pub/tbbmd/issue/22245/238795>
- [3] Samedova, E. Eş Adlılık, Eş Adlılığa Benzer Dil Olayları ve Türkiye Türkçesindeki Basit Fiillerin Eş Adlılığı, Hazar Üniversitesi, 2008, <http://dSPACE.khazar.org/handle/20.500.12323/798> [Erişim: 20.05.2021]
- [4] Vardar, B. *Açıklamalı Dilbilim Terimleri Sözlüğü*. İstanbul: Multilingual, 2002.
- [5] Sakar, E. *Türkçede Eş Biçimlilik Meselesi, Dil ve Edebiyat Araştırmaları (DEA)*, 2021, Bahar, 2021; (23) 405-426 ISSN: 1308-5069 - E-ISSN: 2149-0651
- [6] White, J. *How to Evaluate Machine Translation*. in H. Somers (Ed.), *Computers and Translation*, 2003, s. 211 - 242. Amsterdam: John Benjamins B. V.
- [7] Bangalore B., Bordel G. ve Riccardi G. *Computing consensus translation from multiple machine translation systems*, IEEE Workshop on Automatic Speech Recognition and Understanding, 2001. ASRU '01, Madonna di Campiglio, Italy https://scholar.google.com.tr/citations?view_op=view_citation&hl=th&user=k9pXsKoAAAAJ&citation_for_view=k9pXsKoAAAAJ:9yKSN-GCBOIC [Erişim: 20.03.2022]
- [8] Snover, M. Madnani, N. Dorr, B. ve Schwartz, R. *TER-Plus: paraphrase semantic and alignment enhancements to Translation Edit Rate*, Proceedings of the Fourth Workshop on Statistical Machine Translation, 2009, s. 259-268, Athens,

Greece, 30 March – 31 March 2009. c 2009 Association for Computational Linguistics.

- [9] Alsohybe N. T., Dahan N. A. ve Ba-Alwi F. M. *Machine-Translation History and Evolution: Survey for Arabic-English Translations*, Current Journal of Applied Science and Technology, 2017, 23(4) s. 1-19; Article no.CJAST.36124.
- [10] Okpor, M. D. *Machine Translation Approaches: Issues and Challenges*, IJCSI International Journal of Computer Science Issues, 2014, Vol. 11, Issue 5, No 2, September 2014 / ISSN (Print): 1694-0814
- [11] Jha S. K. Exploring Ambiguous Structures In English, Global Journal of Interdisciplinary Social Sciences, 2017, G.J.I.S.S., 6(3):1-4, ISSN: 2319-8834
- [12] Nordquist, R. *Ambiguity (language)*, 2016 <https://www.thoughtco.com/ambiguity-language-1692388> [Eriřim: 21.03.2022]
- [13] James, R. H., Brendan H., and Michael B. S. *Semantics: A Coursebook*, 2007, 2nd ed. Cambridge University Press
- [14] E. H. Hovy. *Toward finely differentiated evaluation metrics for machine translation*. In Proceedings of the Eagles Workshop on Standards and Evaluation, 1999, Pisa, Italy.
- [15] J.S. White and T. O'Connell, Francis E. O'Mara. *The ARPA MT Evaluation Methodologies: Evolution, Lessons, and Future Approaches*. In Proceedings of the First Conference of the Association for Machine Translation in the Americas, 1994, Columbia, Maryland. <https://aclanthology.org/1994.amta-1.25/> [Eriřim: 25.06.2022]
- [16] Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). *BLEU: A Method for Automatic Evaluation of Machine Translation*. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (Sayfa: 311-318). Philadelphia: ACL. https://www.researchgate.net/publication/2588204_BLEU_a_Method_for_Automatic_Evaluation_of_Machine_Translation [Eriřim: 25.06.2022]
- [17] Aslan, E. Otomatik Çeviri Araçlarının Yabancı Dil Öğretiminde Kullanımı: Google Çeviri Örneđi, SEFAD, 2018 (39): 87-104 - e-ISSN: 2458-908X