

Dođal Dil İřleme

Natural Language Processing

Eřref ADALI

İTÜ Bilgisayar ve Biliřim Fakóltesi
adali@itu.edu.tr

Özetçe

İnsanlık tarihinde dil önemli bir yer tutmaktadır. İnsanlar dil aracılığı ile birbirleriyle iletişim kurabilmiş ve anlaşabilmişlerdir. Diller zamanla gelişmiş ve hâlâ gelişmektedirler. Dillere yeni sözcükler eklenmekte ve bazı sözcükler unutulmaktadır. Ancak dillerin yapıları ve kuralları uzun süreler değişmezliğini korurlar.

Birleşmiş Milletler verilerine göre günümüzde 4000'den fazla dilin konuşulduğu anlaşılmaktadır. Ancak bu dillerden bazıları bir avuç insan tarafından konuşulmaktadır. Dillerin gelişmişliği ile o dili konuşan toplumların kültür, sanat ve bilim alanında ürettikleri arasında sıkı bir bağlantı olduğu açıktır.

Biliřim teknolojilerindeki gelişmeler, bilim insanlarını diller üzerinde çalışmaya teşvik etmiştir. İlk dönemlerde bilgisayar ile konuşmayı amaçlayan bilim insanları zamanla, konuşulanları veya yazılanları anlamaya ve değerlendirmeye çalışmıştır. Doğal Dil İşleme adı verilen bu bilim dalı, genel olarak, dillerin bilgisayar yardımıyla işlenmesi üzerinde çalışmaktadır. Bu çalışmaların sonunda, bir konuşma veya yazının anlaşılması, sorulara yanıt verilmesi ve diller arası çeviriler hedeflenmektedir. Bu hedeflere ulaşabilmek için dillerin ses, biçim, dizilim, anlam gibi temel konuları bu bağlamda incelenmektedir.

Bu yazıda DDİ kapsamında yer alan temel konular, dillere ilişkin özellikler ve bazı DDİ uygulamaları tanıtılmıştır.

Anahtar Sözcükler

Dođal Dil İşleme, DDİ, Ses bilimi, biçimbilim, Söz dizimi, Anlambilim, Çeviri, Yanıtlama Dizgeleri,

Abstract

Languages played important role in human history. Human can communicate and understand each other. Languages have been developed in time and still developing. Some new words are imported or created to language and some are faded out. But the rules and structures of a language is more stable.

According to United Nations records, there are more than 4000 languages today. A few people speak some of these. The development level of a language reflects the cultural, scientific and art development level of a community.

Recent developments of the Information Technologies are encouraging researchers to work in the field of natural language processing areas. In early age of this study researcher wanted to talk to computer. By the time, they worked on understanding of a speech and text and process of them. This research field is called as Natural Language Processing NLP. NLP, today deals with understanding of a text or a speech, automatic answering and language translation. In order to reach these goals, NLP researchers do many works on phonology, morphology, syntax and semantics of languages.

In this article, we are going to explain major topics of NLP and present some information about NLP applications

Keywords

Natural Language Processing, NLP, phonology, morphology, syntax, semantics, Machine Translation, Automatic answering systems

1. Giriş

İnsanlık tarihine genel olarak baktığımızda bazı gelişmelerin insanlığın gelişiminde önemli sıçramalara neden olduğunu görürüz. Dil ve yazı bu gelişmelerden biridir. İnsanlar konuşarak birbiriyle iletişim kurabilmiş ve anlaşabilmiştir. Dil sayesinde kültür, sanat ve bilimde ilerlemeler sağlanabilmiştir.

Dillerin gelişimi uzun bir zaman alır ve gelişim devam eder. Zaman içinde, toplumdaki gelişmelere bağlı olarak dile yeni sözcükler eklenir veya bazı sözcükler unutulur. Hatta zaman içinde dilin tümce kurma biçimi de değişebilir. Bu nedenlerle geçmiş yüzyıllarda yazılmış metinler günümüzde anlaşılmaz olabilir. Konuya Türkçe özelinde eğildiğimizde, zaman içinde, Türkçenin söz varlığında değişme olduğunu görürüz. Ancak dilin kurallarında binlerce yıldır bir değişme olmaması çok dikkat çekicidir.

Birleşmiş Milletler verilerine göre günümüzde 4000'den fazla dilin konuşulduğu anlaşılmaktadır. Ancak bu dillerden bazıları bir avuç insan tarafından konuşulmaktadır. Dillerin gelişmişliği ile o dili konuşan toplumların kültür, sanat ve bilim alanında ürettikleri arasında sıkı bir bağlantı olduğu açıktır. Bir başka deyişle dili yetersiz olan toplulukların kültür, sanat ve bilim alanında başarılı olmaları beklenemez.

Dil ve toplum ilişkisini Konfüçyüs şöyle açıklamaktadır:

"... Bir ülkenin yönetimini ele alsaydım, yapacağım ilk iş, hiç kuşkusuz dilini gözden geçirmek olurdu. Çünkü dil kusurlu ise sözcükler düşünceyi iyi ifade edemez. Düşünce iyi ifade edilemezse görevler ve hizmetler gereği gibi yapılamaz. Görev ve hizmetin gerektiği şekilde yapılamadığı yerlerde âdet, kural ve kültür bozulur. Âdet, kural ve kültür bozulursa adalet yanlış yollara sapar. Adalet yoldan çıkarsa şaşkınlık içine düşen halk ne yapacağını, işin nereye varacağını bilemez. İşte bunun içindir ki hiçbir şey dil kadar önemli değildir!.."

Türkçede sözcüklerin zaman içinde değiştiği bilinmektedir. Hatta Osmanlı Türkçesinde, Türkçe sözcüklerin, dilin toplam söz varlığının yarısına kadar düştüğü söylenebilir. Bu dönemde dilbilgisi kurallarında da değişimlerin olduğu bilinmektedir. Ancak 12 Temmuz 1932 yılında Dil Devrimi'yle başlatılan çalışmalar sonunda Türkçeden yabancı sözcükler atılmaya başlanmıştır.

Atatürk'ün dil üzerine söylediklerinden bazıları şunlardır:

"Türk milletinin dili Türkçedir. Türk dili dünyada en güzel, en zengin ve en kolay olabilecek bir dildir. Onun için her Türk dilini çok sever ve onu yükseltmek için çalışır. Bir de Türk dili, Türk milleti için kutsal bir hazinedir. Çünkü Türk milleti, geçirdiği bunca tehlikeli durumlarda, ahlakının, geleneklerinin, anılarının, çıkarlarının, özetle, bugün kendi ulusallığını yapan her şeyin dili aracılığıyla korunduğunu görüyor. Türk dili, Türk milletinin kalbidir, zihnidir."

"Türk demek dil demektir. Ulusallığın çok belirgin özelliklerinden birisi dildir. Türk ulusundanım diyen insanlar, her şeyden önce ve ne olursa olsun Türkçe konuşmalıdır. Türkçe konuşmayan bir insan, Türk ekinine, topluluğuna bağlılığını öne sürerse buna inanmak doğru olmaz."

"Ülkesini, yüksek istiklalini korumasını bilen Türk milleti, dilini de yabancı diller boyunduruğundan kurtarmalıdır."

"Türk dilinin sadeleştirilmesi, zenginleştirilmesi ve kamuoyuna bunların benimsetilmesi için her yayın vasıtasından faydalanmalıyız. Her aydın hangi konuda olursa olsun yazarken buna dikkat edebilmeli, konuşma dilimizi ise ahenkli, güzel bir hale getirmeliyiz."

Osmanlı döneminde, özellikle kamusal alanda Türkçe bozulmuştur. Ancak Türkçenin sağlam olan kuralları bozulmadığından, bugün, Selçuklu ve erken Osmanlı dönemi eserlerini anlayabilmekteyiz. Bu durum Türkçenin kurallarının, yaklaşık bin yıldır bozulmamış olduğunun bir kanıtıdır.

Dilin yapısı, insanların düşünme yeteneklerine de etki etmektedir. Çok basit bir anlatımla, elma sözcüğünü bilmeyen birinin aklına elma yemek gelemez. Özenme sözcüğü olmayan bir dilde, kişi bir şeye özendiğini anlatamaz. Bir dilin sözlük varlığı, o dilin gelişmişliğinin bir ölçüsü olabilir. Ancak bir dildeki sözcüklerin sayısına bakarak bir dilin diğer bir dile göre daha gelişmiş olduğu söylenemez. Dillerin yapıları göz önüne alınarak gelişmişlik derecesine karar verilebilir. Örneğin, İngilizcede *göz*, *gözlük*, *gözlükçü* sözcükleri için birbiri ile ilişkisi olmayan üç sözcük (*eye*, *eyeglasses*, *optician*) gerekirken Türkçede *göz* sözcüğünden türetilmiş üç sözcük (*göz*, *gözlük*, *gözlükçü*) yetmektedir. Türkçe bitişken bir dil olması nedeniyle bir kökten değişik anlamlar ifade eden

sözcükler türetilir. Buna karşın, Hint-Avrupa dillerinin bu yöndeki yetenekleri çok kısıtlıdır. Bu nedenle her şeye yeni bir sözcük bulmak zorundadırlar. Sami dilleri de bir kök sözcükten çok sayıda sözcük türetme özelliğine sahiptir. Bu açıdan değerlendirildiğinde, bitişken ve Sami dillerin yeni sözcük üretme açısından diğer dillere oranla üstün oldukları görülmektedir. Daha önemlisi bu dillerde türetilen sözcüğün anlamı kolayca çıkarılabilmektedir. Örneğin, "kırnak" sözcüğünü ilk duyan kişi bunun kırnak eyleminden türetilmiş olduğunu; kırma işlemi sonunda elde edilen bir nesne olduğunu anlayabilir. Kırnakın İngilizce karşılığı olan "chip" sözcüğü türetilmek istendiğinde ağaçtan kırnak çıkarma veya patatesten parça kesme işlemine benzetilmeye çalışılmıştır. Sözcük türetme yeteneği kısıtlı olan dillerde bazı sözcüklerin ilginç yöntemlerle türetilmesine tanık olmaktadır.

Örneğin "kuşluk vakti" için İngilizcede "late morning" yerine "breakfast" ve "lunch" sözcüklerinin birleşimi olan "brunch" sözcüğü kullanılmaktadır. Nylon sözcüğünün "Now You Lousy Old Nippons" tümcesinde yer alan sözcüklerin ilk harflerinden türetilmiş olduğu söylenmektedir.

Dil bir ulusu oluşturan temel öğelerden biridir. Bunun sonucu olarak her toplum kendi dilini sever ve üstün görür. Belki bu nedenle dil bilimciler dilleri karşılaştırmaya ve derecelendirmeye gerek görmemiştir. Dil bilimciler dilin, dilbilgisi ve zaman içindeki gelişimi üzerinde çalışmaya özen gösterirler.

Bilişim teknolojilerindeki gelişmeler, bilgisayarlı dil bilimi çalışmalarına önemli bir ivme kazandırmıştır. Doğal Dil İşleme (DDİ) adı verilen bu yeni bilim alanı önceleri insan bilgisayar etkileşiminde doğal dillerin kullanılabilmesi amacıyla başlatılmış, zamanla bilgisayarlı dil bilimine dönüşmüştür. Belleğinde bulunan veri ve sonuçları, bilgisayarın doğal dilde konuşarak insanlara aktarılmasına "konuşma", insan tarafından konuşularak bilgisayara yapılan girişlere "konuşmayı anlama" adı verilmiştir. Bir bilgisayarın, insanın konuşmasını anlayabilmesi ve insanın anlayabileceği dilde konuşarak insanlarla etkileşimde bulunabilmesi için bilgisayarın dilin tüm özelliklerini bilmesi gerekir. Bir başka deyişle dilin tüm özelliklerinin bilgisayara öğretilmesi gerekir. Bu nedenle DDİ alanında çalışan bilişimciler dillerin özelliklerini kendi bakış açılarından değerlendirmeye başlamışlardır. Bu bağlamda

özellikle şu konularda çalışmaya ağırlık verilmişlerdir:

- Ses bilimi
- Biçim bilimi
- Söz dizimi
- Anlam bilimi

Dil bilimciler dilin yapısını oluşturan yukarıdaki dört ana konu üzerinde çalışarak dilin özellik ve kurallarını ortaya koymaya çalışırlar. Bu çalışmalara koşturarak dilin zaman içindeki evrimini incelerler. Doğal dil alanında çalışan bilişimciler, dil bilimcilerin ortaya koymuş oldukları sonuçları kullanarak dili bilgisayar yardımıyla işlemeye çalışırlar. Örneğin Türkçenin ses bilimi ve biçim bilimi özelliklerini kullanarak yazılı bir metin içindeki yazım yanlışlarını bulmaya ve düzeltmeye çalışırlar. Dil bilimciler tarafından belirlenen, iki dilin dilbilgisi kurallarını kullanarak diller arası çeviri araçları geliştirmeye çalışırlar. Bir dil bilimciye göre bitişken dillerin ek zenginliği bu diller için olumlu bir özellik iken bilişimciler için çözümü zor ama ilginç bir konudur. Bu örneklerden de anlaşılacağı gibi dil bilimciler, dil veya dillerin özellik ve yapılarını bilimsel olarak ortaya koymaya ve tanımlamaya çalışırken doğal dil işleme ile ilgilenen bilişimciler bu bilgileri kullanarak insanlara bilgisayar aracılığı ile yardımcı olmaya çalışmaktadırlar.

Günümüzde dil bilimciler ve doğal dil işleme alanında çalışan bilişimciler ayrı ayrı çalıştıkları gibi belli konularda birlikte de çalışmaktadırlar. Birlikte çalışmalarının daha verimli sonuçlar doğuracağı açıktır. Çünkü birinin tanımladığı sonuçları diğeri kullanmaktadır. Beraber çalışmayı sağlamak için dil bilimcilerin bilişim teknolojileri ve bilişimcilerin de dilbilimi konularına eğilmeleri gerekmektedir.

Bu yazıda DDİ alanındaki çalışmalar genel hatları ile tanıtılacak, bu çalışmalar sırasında dilin hangi özelliklerinin bilinmesi gerektiği ve karşılaşılan zorluklar ortaya konacaktır. Yazının sonunda DDİ bakış açısından dillerin değerlendirilmesi yapılacaktır.

2. DDİ'nin İlgili Alanları

DDİ alanında çalışanların yukarıda sözünü ettiğimiz dil biliminin dört temel konusu ile ilgilenmelerinin nedenini anlayabilmek için DDİ konusunda çalışanların ilgi alanlarını ve bu çalışmalar sırasında karşılaştıkları zorlukları bilmemizde yarar vardır.

DDİ çalışmaları kapsamında aşağıda sıralanan konuları görmekteyiz:

- Yazım yardımcı araçlarının geliştirilmesi
- Yazım yanlışlarının düzeltilmesi
- Bul ve değiştir
- Basılı bir metni okuma (optik olarak metin okuma) ve okuma yanlışlarını düzeltme
- Bir metnin özetini çıkarma
- Metnin içerdiği bilgiyi çıkarma
- Bilgiye erişim
- Metni anlama
- Bilgisayarla sesli etkileşim
- Bilgisayarın konuşması (metni seslendirme)
- Konuşmayı anlama (konuşmayı metne dönüştürme)
- Soru yanıt dizgeleri
- Yabancı dil okuma yardımcı araçları
- Yabancı dilde yazma yardımcı araçları
- Doğal diller arası çeviri

2.1 Yazım Yanlışlarının Düzeltmesi

Günümüzde, kitaplar, dergiler ve raporlar bilgisayar ortamında hazırlanmaktadır. XIX. Yüzyılın sonunda önemli bir buluş olarak kabul gören daktilolar artık kullanılmamaktadır. Yazılan her yazıda yazım hatası olabileceği bir gerçektir. İnsanlar yazmış oldukları yazımın yazım kurallarına uygun olmasını isterler. Geçmiş dönemlerde, yanlışları bulan ve düzelter insanların olduğu ve bunların basın kuruluşlarında musahhah sıfatıyla çalıştıkları bilinmektedir. Geçmişte musahhahlar tarafından yapılan işi bugün bilgisayarlarla yaptırmak için çalışmalar oldukça ileri düzeydedir.

Yazma işleminin bilgisayar ortamına geçmesiyle birlikte bilgisayar ortamında bulunan bir metnin yazım hatalarını bulmak ve düzeltmek bilimciler için ilginç bir araştırma alanı olmuştur. Bir metindeki yazım hatalarını bulmak için değişik yöntemler kullanılabilir. Yöntemlerden biri, metnin yazıldığı dilin sözlüğünü bilgisayarda bir veri tabanına yerleştirmek olarak düşünülebilir. Bitişken olmayan diller için olanaklı görülen bu uygulama, Türkçe gibi bitişken diller için ilk seçenek olmamalıdır. Ayrıca Türkçe gibi kurallı bir dilde yazılmış bir metin içindeki yazım hatalarını bulmak için;

- Ses uyum kuralları
- Hece yapısı
- Eklerin uyumu

özellikleri kullanılabilir.

Bir sözcüğü hecelere ayırma ve satır sonunda hece bölmesi işlemi için İngilizcede sözlüğe bakma zorunluluğu vardır. Buna karşın Türkçede bir sözcüğün hecelere ayrılması kuralıdır ve sözlüğe bakılmadan yapılabilir.

Bu kısa bilgilerden de anlaşılacağı gibi yazım yanlışlarını düzeltmek üzere çalışmalarda bulunacak birisi, üzerinde çalışacağı dilin ses bilimi ve biçim bilimi özelliklerini bilmek zorundadır. Hint Avrupa dilleri üzerinde gerçekleştirilecek çalışmalar için dilin söz varlığını veri tabanına yerleştirmek uygun bir yöntem olarak görülmektedir. Türkçe gibi bitişken diller için veri tabanına dayalı çözümlere ek olarak kural tabanlı çözümlerin kullanılması önerilebilir.

Türkçe metinlerde yazım yanlışlarının düzeltilmesi ile ilgili kapsamlı bir çalışma A. Delibaş tarafından yapılmıştır [1]

2.2 Bul ve Değiştir

Bir metin içindeki bir sözcüğün bir başka sözcük ile değiştirilmesi sıkça karşılaştığımız bir uygulamadır. Örneğin, metin içinde geçen tüm "elma" sözcüklerini "erik" sözcüğü ile değiştirmek istediğimizi düşünelim. İngilizce yazılmış bir metinde elma karşılığı olan "apple" sözcüğü ya yalın hâlde "apple" ya da çoğul halde "apples" bulunabilir. Dolayısıyla metnin içindeki tüm "apple"leri "plum" ve "apples"leri "plums" olarak değiştirmek yeterli olacaktır.

Türkçe gibi bitişken dillerde sözcüklerin alabileceği ekler çoktur. Örneğin *elma* ve *erik* sözcükleri aşağıda görülen ekleri alabilirler:

elma	erik
elmalar	erikler
elmacık	erikcik
elmacı	erikçi
elmacılık	erikçilik
elmalık	eriklik
elmam	erikim
elman	erikim
elması	erikçi
elmamız	erikimiz
elmanız	erikiniz
elmaları	erikleri
elmanın	erikimin
elmayı	erikçi
elmaya	erikçi

elmada	erikde
elmadan	erikden
elmanın	eriğin
elmasız	eriksiz
elmacık	erikcik

Dolayısıyla metnin içinde geçen tüm elma sözcüklerini erik ile yer değiştirmek istediğimizde ekleri de göz önüne almamız gerekir. Örneğin “elmacının” yerine “erikçinin” yazılmalıdır. Bir sözcüğün kök ve eklerini bulmak için biçim bilimi çözümlemesinin yapılması gerekmekte ayrıca ses bilimi özelliklerine göre de düzeltmelerin yapılması gerekmektedir.

Türkçe için yetkin bir bul ve değiştir konulu çalışma F. Şentürk tarafından gerçekleştirilmiştir [2].

2.3 Basılı Bir Metni Okuma

Bilgisayar ortamında bulunmayan basılı metinlerin bilgisayara aktarılması işlemine metni okuma demektedir. Bu metinler eski dönemlerde basılmış kitap veya belgeler olabileceği gibi kullanıcıların elleriyle veya makineyle doldurdukları bilgi giriş formları olabilir. Bu tür basılı metinleri optik yöntemler ile okuyup bilgisayara aktarmak en temel değerlendirmeye görüntü işleme veya karakter tanıma olarak sayılabilir. Optik olarak yazılı metinleri okuyan ve okuduğu karakterleri bilgisayar karakterlerine çeviren dizgeler oldukça başarılı sonuçlar üretmektedir. Ancak yanlışsız tanıma yaptıkları söylenemez. Kısaca optik karakter okuyucu olarak adlandırdığımız bu dizgelerin okuma yanlışları DDİ teknikleri kullanılarak giderilmeye çalışılmaktadır. Konuya açıklık getirmek amacıyla Şekil-1'deki örnek verilmiştir. Bu örnek bir bankanın müşterilerine doldurulmak üzere verdiği bir form olabilir. Bu formlar elle doldurulacağı gibi daktilo veya bilgisayar ile doldurulabilir. Optik karakter okuyucunun bu formu okurken iki yanlış yaptığını varsayabiliriz. Bu yanlışların kaynağı formu

Adı	E R E N
Soyadı	M U T L U
Mesleği	M E M U R
Hesap numarası	1 2 3 4 - 9 8 5 7 4
Adresi	
İlçe	Ü S K Ü D A R
İl	İ S T A N B Ü L

Şekil-1: Optik karakter okuyucuda yanlış okuma

dolduran olabileceği gibi form üzerine yapışmış olan tozdan da kaynaklanabilir. Forma dikkatli baktığımızda kişi adındaki üçüncü ve ilçe alanındaki birinci harfler yanlış okumaya açıktır. Kişi adındaki üçüncü harf "C" ve ilçenin ilk harfi "I ve J" olarak okunabilir. Ancak bu yazıyı bir insan okuduğunda, kişi adını "Eren" ve ilçe adını Üsküdar olarak okuyacaktır. Çünkü metin Türkçedir. Türkçe isimler içinde "Ercn" diye bir isim yoktur. En yakın isim "Eren"dir. İstanbul ili içinde "Ijsküdar" adında bir ilçe yoktur; Üsküdar vardır. Bu örnekten de anlaşılacağı gibi metnin dili bilinir, ayrıca verilerin nitelikleri bilinirse optik karakter okuyucusunun yanlış okumaları düzeltilebilir.

2.4 Bir Metnin Özetini Çıkarma

Yayın organlarının gelişmesi, özellikle İnternet üzerindeki bilgi kaynaklarının artması, insanlara çok büyük olanaklar sunmaktadır. Bilgi kaynaklarının çokluğu ve bilginin genişliği, bir başka sorunu beraberinde getirmektedir: Erişilen bir kaynağın, gerçekten yararlanılmak istenilen kaynak olup olmadığına karar verebilmek için tüm kaynağı okumak gerekmektedir. Bu ise bazen zaman kaybına neden olmaktadır. Bu sorunu gidermek üzere ve insanlara araştırdıkları bilgiye en hızlı erişebilmelerini sağlamak amacıyla metinlerin özetini çıkarmak üzerine çalışmalar yapılmaktadır.

Özetleme üzerinde yapılan çalışmalarda öncelikle yazının varsa içindekiler kısmına yoksa başlıklarına bakılmaktadır. Bundan sonra metin içinde sıkça geçen sözcükler taranmaktadır. Bu araştırmaların ardından özet yazılmaya çalışılmaktadır. Örneğin şu anda okumakta olduğunuz yazı İnternet üzerinde bulunsaydı ve bir özetleme aracı tarafından özetlenmiş olsaydı, özeti şöyle olabilirdi:

Doğal dil işleme ve dilbilimi hakkında temel bilgiler içeren, DDİ'nin ilgi alanlarını tanıtan, DDİ'nin karşılaştığı sorunlara yer veren ve dilleri değerlendiren bir yazı.

Özete yayın yeri, yazar adı ve yayın tarihi de eklenebilir.

Özet çıkarma konusunda yapılan çalışmalar, bir metnin konusunu ortaya çıkarmayı bu sonucu düzgün tümcelerle ifade etmeyi hedeflemektedirler.

2.5 Metnin İçerdiği Bilgiyi Çıkarma

Doğal dilde yazılmış metinlerin belli kurallara uygun olarak yazıldığı bir gerçektir. Kurallar dile

ilişkin dilbilgisinden kaynaklanmaktadır. Her dilde tümce içindeki sözcüklerin sıralanışı belli kurallara uyar. Bu kurallar dilden dile değişmektedir. Bazı dillerde sözcüklerin dizilişi katı kurallara bağlı olmasına karşın bazı dillerde diziliş çok esnektir. Bir metnin okuyucuya vermek istediği bilgi metin içindeki bir tümce ile verilebileceği gibi metnin tümü ile de verilebilir. Konuya açıklık getirmek üzere aşağıdaki metni okumamızda yarar vardır.

"Hava çok bulutlu ve dışarıda sağanak hâlinde yağmur yağıyor. Bu zor hava koşulunda, sabah on sularında Ahmet odama geldi ve Ayşe ile evleneceklerini söyledi. Düşün günü olarak 22 Kasım seçmişler. Düşün Boğaziçi'nde yapılacakmış. Bu mutlu haberin üzerine oturup biraz sohbet ettik ve kahve içtik."

Bu metin daha uzatılabilir ancak metin incelendiğinde çıkarılacak bilgi şöyledir:

- Taraflar : Ahmet ve Ayşe
- Eylem : Evlenme
- Zaman : 22 Kasım
- Yer : Boğaziçi

Bilgi çıkarma işlemi sadece bir tümceye bakılarak çıkarılamaz. Örnek metinden de görüldüğü gibi komşu tümcelere veya tüm metne de bakmak gerekebilir. Bu sırada metnin asıl konusu dışında kalan bilgiler ayıklanmalıdır.

2.6 Bilgiye Erişim

Bilgiye erişim konusu, İnternet'in yaygınlaşması sonunda ortaya çıkmış bir konudur. İnternet'te bir konuda yapılan araştırmalar için geliştirilmiştir. Günümüzde İnternet'te bir konu hakkında araştırma yapmak istediğimizde anahtar sözcüğü yazarız. Bunun üzerine yüzlerce hatta binlerce bağlantı karşımıza gelmektedir. Araştırmacı olarak bunlardan hangisinin gerçekten işimize yarayacak yazı olduğunu bulmak bizlere düşmektedir.

Bir yazının ana konusunu saptamak için kullanılacak yöntemlerden ilk akla geleni, yazı içinde en çok kullanılan sözcükleri bulmaktır. Bu varsayımına göre bir yazı içinde en çok geçen sözcük o yazının ana konusu ile ilişkilidir. Örneğin bir yazı içinde en çok geçen sözcük "İstanbul" ise bu yazının İstanbul'la ilişkili olduğu varsayılmaktadır. Bu varsayımın doğru olma olasılığı vardır ancak kesin değildir. Yazı İstanbul hakkında olmayabilir, İstanbul'da geçen bir olayı anlatıyor olabilir. Her iki

durumda da İstanbul ile ilişkili olduğu sonucuna varılabilir.

Bir yazı içinde sık kullanılan sözcükleri arama ve bunun sonucu olarak metnin içerdiği bilgiye erişme yöntemi bükümlü diller için uygulanabilir bir yöntemdir. Türkçe gibi bitişken dillerde yazının konusunu belirlemek için aynı yöntemi kullanmak oldukça zordur. Örneğin "göz" ile ilgili bir araştırma yaptığımızda "göz, gözlük, gözlükçü, gözlükçülük" gibi konular da karşımıza gelebilir. Benzer bir araştırmayı İngilizce için yaptığımızda "eye" yazdığımızda sadece "göz" ile ilgili yazılar gelecektir. Çünkü "gözlük" ve "gözlükçü" sözcüklerinin İngilizce karşılıkları çok farklı sözcüklerdir: (*eye glass, optometrist*)

Özellikle bitişken dillerde yazılmış metinlerde bilgiye erişim amaçlı geliştirilecek çalışmalara katılacak olanların dilin biçim bilimsel özelliklerini bilmeleri gerekir. İnternet'teki arama çalışmalarını kolaylaştırmak ve daha anlamlı kılmak üzere "*anlamsal web*" adıyla yeni çalışmalar başlatılmıştır. Anlamsal web de metinlerin ilgili oldukları konuyu belirten etiketlerin bulunması hedeflenmektedir.

Şu anda okuduğunuz yazı anlamsal web'de yer alsaydı şöyle etiketlenebilirdi:

- **Konu** : Doğal dil işleme
- **Alt konu** : DDİ ilgi alanları, DDİ'nin karşılaştığı sorunlar, Dillerin karşılaştırılması

Bilgiye erişim konusunda çalışmak isteyenler anlam bilimi dâhil, dile ilişkin tüm özellikleri bilmek durumundadır.

2.7 Metni Anlama

Yazılı metinlerin anlamını çıkarmak ve bu anlamı eyleme dönüştürmek, doğal dil işleme alanındaki bir diğer araştırma konusudur. Bu çalışmaların amacı kısa metinlerin veya tümcelerin tam ve kesin olarak anlaşılmasını hedeflenmektedir. Bir tümcenin anlamını çıkarmak, değişik alanlarda uygulanabilecek bir konudur. Örneğin, bankalara gelen yazılı istekler, günümüzde insanlar tarafından okunmakta ve istek insan tarafından bilgisayara aktarılmaktadır. Genellikle bir tümce halinde olan isteğin anlamını çıkararak ve uygulamaya geçen çözümler yakın zamanda görülmektedir. Bankalara gönderilen isteklere bir örnek aşağıda verilmiştir:

"Bankanızın Merkez şubesindeki 1234-56789 numaralı hesabımdan 1500 TL nin, bankanızın

Ankara şubesindeki 1255-54321 numaralı Bartu Yiğit hesabına havale edilmesini rica ederim."

Bu tür istek yazıları örnekte olduğu gibi, genellikle belli bir kalıba uygun olarak yazılır. Metnin anlamını çıkarırken sırasıyla;

- Kaynak ve hedef hesapların ad ve numaraları belirlenir.
- Aktarılabacak para miktarı belirlenir.
- İşlemin türü öğrenilir.

Bu temel bileşenlerin belirlenmesinde çok dikkatli olunması gerekir. Çünkü metinden çıkarılacak anlam daha sonra eyleme dönüştürülecektir. Küçük bir yanlış anlama önemli maddi kayıplara neden olabilir. Bu nedenle, bilgisayarın anladığı yönergenin bir insan tarafından onanması uygun olabilir. Bu alanda bir çalışma için Ş. Adalı'nın çalışması örnek olarak verilebilir [3].

İnsanların isteklerini yazılı olarak anlatabilmeleri, DDİ'nin ilk zamanlarından beri amaçlanan bir konudur. Günümüzde bu yöndeki uygulamaların İnternet üzerinden yapılan işlemlerde kullanılması hedeflenmektedir. İnsanların isteklerini düz bir tümce ile belirtmeleri beklenmektedir. Örneğin bir e-ticaret sayfasında,

"Bu akşam katılacağım toplantıda giymek üzere bir takım elbise almak istiyorum"

tümcesini yazarak, kendisine en uygun elbiseyi satın alabilmek, yeni kuşak insanların beklentisidir. Bir metni tam olarak anlamak için, sesbilim, biçimbilim, sözdizimi ve anlambilimini bilme gerekir.

2.8 Metni Seslendirme

Yazılı bir metni seslendirme bir başka deyişle okuma bilişim dünyasının üzerinde uzun süredir çalıştığı bir konudur. Çok basit bir gösterimle bilgisayarın bir metni sesli olarak okuması diye özetleyebiliriz.

İnsanlar konuşurken sürekli ses çıkardıkları düşünülebilir. Ancak konuşmanın sözcükler ve sözcüklerin de harflerden oluştuğu bir gerçektir. Bu açıdan değerlendirildiğinde bir metni seslendirmek için ses bilimini bilmek gerekir. Ses bilimi, bir dildeki sesleri inceler ve dildeki sesleri ortaya koyar.

Türkçe, yazıldığı gibi okunan bir dil olarak tanımlanmasına karşın yabancı dillerden alınan sözcükler ile eski dilden kalan sözcüklerde bazı

harflerin birden fazla söyleniş biçimleri bulunmaktadır. Örneğin kalın " a " ve ince " a " gibi.

Yazılı bir metni seslendirmede ilk adım, metni sesçil abeceye dönüştürmektir. Bunun ardından sesçil abece ile yazılmış olan metni dilin vurgu özelliklerine uygun olarak seslendirmektir.

Bir metni seslendirmek üzere iki temel yöntemden biri kullanılmaktadır:

- Doğal sesleri ekleme yöntemi
- Yapay ses üretme yöntemi

Doğal sesleri ekleme yöntemi, insan tarafından üretilmiş ses kayıtlarının kullanılması ilkesine dayanır. Bu yöntemde en büyük parçadan en küçük parçaya kadar ekleme yöntemi kullanılmaktadır.



Şekil-2: Bir metnin seslendirilmesi

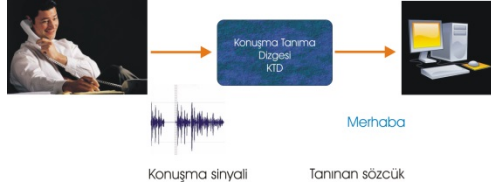
Örneğin bilgisayarlı yanıtlama dizgelerinde belli sorulara karşı hazırlanmış yanıt tümcenin tamamı insan tarafından seslendirilebilir. Örneğin; "Kuruluşumuzu aramış olmanızdan dolayı teşekkür ederiz."

Bazı uygulamalarda, tümcenin tamamı yerine tümce içinde yer alması olası sözcük kalıpları seslendirilerek veri tabanında saklanmaktadır. Böylece veri tabanındaki seslendirilmiş kayıt sayısı azaltılmaya çalışılır. Örneğin; "Ankara'ya gidecek yolcuların 212 numaralı bekleme salonuna gelmeleri beklenmektedir." tümcesinde koyu olarak görülen kent adı ve salon numaraları bir yerde diğer sözcükler bir başka yerde saklanır. Koyu yazılı sözcükler değişkendir; diğerleri sabit kabul edilebilir. Değişken sözcükler değiştirilerek aynı bilgi şöyle de okunabilir: Konya'ya gidecek yolcuların 344 numaralı bekleme salonuna gelmeleri beklenmektedir.

Yazılı metni seslendirme, sesli kitap, sesli gazete veya görme özürlüler için yararlı bir uygulama alanıdır. Bu konuda T. Şentürk'ün çalışmasından yararlanılabilir [4]

2.9 Konuşmayı Yazıya Dökme

İnsan konuşmasını yazıya dökme doğal dil işleme üzerinde çalışan araştırmacılar için ilgi çekici bir araştırma konusudur. İnsan ağzından çıkan seslerin mikrofona ile elektriksel sinyale dönüştürüldüğü bilinmektedir. Bu elektriksel sinyallerin işlenmesi ile metne dönüştürmek araştırmanın temel bileşenidir. Ancak bu dönüştürme sürecinde dilin özelliklerinin bilinmesi gerekir. Sesçil bir dil olan Türkçe için konuşmayı yazıya dönüştürme işleminin diğer dillere oranla daha kolay olacağı söylenebilir. Şekil-3'te konu ana hatları ile gösterilmiştir.



Şekil-3: Konuşmayı yazıya dökme

Konuşmayı yazıya dökme işlemi sadece bir konuşmayı yazıya dökme için düşünülmemekte; yazıya dönüşen metnin anlaşılabilir eyleme dönüştürülmesi de amaçlanmaktadır. Sadece bir konuşmayı metne dönüştürmek,

- Konuşmacının, konuşmasını yazıya dökme,
- Toplantı tutanaklarını hazırlamak,

gibi örnek uygulamalarda yararlı katkılar sağlayabilir. Konuşmayı yazıya çevirmenin basit bir uygulaması kısa emirlerin anlaşılmasıdır.

Örneğin;

- Terliğimi getir
- Kapıyı kapat
- Sol damar tıkalı
- İki yüz on iki, iki yüz seksen beş, otuzaltı, seksen altı

İlk iki örnek bir robota emir vermek için kullanılabilir. Üçüncü örnek bir anjiyo işlemi doktorun tanısını ifade eder. Son tümce ise bir telefonu sesli olarak aramaya yarar. Bu örneklerde konuşmayı yazıya dönüştürmek amaçlanmamaktadır. Ancak konuşma söz kalıpları ile karşılaştırılmakta ve uyduğu kalıba göre eylem yerine getirilmektedir.

2.10 Soru Yanıtlama

Bankalar, gezi düzenleyicileri ve oteller müşterilerine daha iyi hizmet vermek üzere çağrı merkezleri kurdukları bilinmektedir. Bu merkezlerde çalışanlar, müşterilerin isteklerini dinler ve onlara en uygun hizmeti vermeye çalışırlar. Çağrı merkezlerinde çalışanların yüksek maliyeti, çağrı merkezlerinin işlevini bilgisayar ile çözmek üzere çalışmalar yapılmasını gerekli kılmıştır. Kısaca soru yanıtlama dizgesi olarak adlandırabileceğimiz bu dizgede, çağrı merkezini telefonla arayan kişinin isteği bilgisayar tarafından anlaşılacak ve kendisine en uygun yanıt verilecektir.

Günümüzde gerçekleştirilen uygulamalarda müşterinin konuşmasından sözcükler yakalanmaya çalışılmakta ve yakalanan sözcüğe bağlı olarak soru sorulmaktadır. Soru yanıtlama dizgesi dolayısıyla konuşmayı metne çevirme ve metni seslendirme yöntemlerini içerir.

Günümüzde kullanılan bir başka sesli yanıt verme dizgesi örneği bankacılık, çağrı merkezi gibi değişik alanlarda kullanılmaktadır. Bu alanlarda kullanılan sesli yanıt dizgeleri akıllı dizgeler değildir. Genellikle önceden yapılmış olan ses kayıtlarını kullanarak arayanlara yardımcı olmaya çalışmaktadır. Günümüzde aşağıda örneğini verdiğimiz düzeyde yanıtlama dizgeleri kullanılmaktadır.

Yanıtlayan: İyi günler size nasıl yardımcı olabilirim?

Müşteri: Bir kişilik oda bakıyorum

Yanıtlayan: Bir kişilik oda mı istiyorsunuz?

Müşteri: Evet

Yanıtlayan: Hangi gün geleceksiniz?

Müşteri: 10 Ekim Pazartesi ve sonrasındaki üç gün için

Yanıtlayan: 10 Ekimde giriş mi yapacaksınız?

Müşteri: Evet, 10 Ekim öğlen giriş yapacağım. Söyledim ya!

Yanıtlayan: Kaç gün kalmayı planlıyorsunuz?

Müşteri: 13 Ekimde ayrılacağım

Yanıtlayan: Sizin için odamız var. Göl manzaralı oda isterseniz 10 TL fazla ödeyeceksiniz ama manzaralı. İster misiniz?

Müşteri: Olabilir

Yanıtlayan: Anlayamadım. Evet mi dediniz

Müşteri: Evet, olabilir dedim

Yanıtlayan: Sizin odanızı ayırdım. Şimdi kimlik bilgilerinizi öğrenebilir miyim?

...

Bu karşılıklı konuşmadan anlaşılacağı gibi bilgisayar daha önceden kaydedilmiş soruları sormakta ve belli kalıp yanıtlar beklemektedir. Bu kalıplar içinde belli sözcükleri yakalamaya çalışmaktadır.

Yapay us konusundaki gelişmeler, bilgisayarların anlamlı tümceler kurarak insanlar ile konuşabileceğini müjdelemektedir. Bu tür dizgeler sesli olabileceği gibi yazılı da olabilir. Örneğin yazılı olan dizge kullanılarak bir hasta tanılama dizgesi geliştirilebilir. Böyle bir dizgede, bilgisayar mantıklı ve amaca uygun sorular sorar. Hastanın vereceği yanıtlara bağlı olarak yeni sorulacak soruyu oluşturur. Böylece hastalığa tanı koymaya çalışılır.

Soru yanıtlama çalışmaları DDİ'nin ilk dönemlerinin düşüdüğü ve günümüzde uygulanabilir duruma gelmiştir.

2.11 Çeviri

Diller arası çeviri insanlığın önemli düşlerindenidir. Düşlenen şey; bir kişinin ana dilinde konuşması, karşısındaki kişinin bunu kendi dilinde dinlemesidir. Diller arası çeviri konusu iki dilin ses bilimi, biçim bilimi, söz dizimi ve anlam bilimi özelliklerini bilmeyi gerektirir. Günümüzde bilgisayar desteğiyle yapılan çeviriler en basit olandan en karmaşık olana doğru sıralanabilir: Sözcük çevirisi, tümcecik çevirisi, tümce çevirisi. Gerçekleştirilen metinden metne çeviri dizgelerinin bazıları insan gözetimi gerektiren bazıları insan yardımı olmaksızın çeviri yapabilmektedir.

Bir dilden diğer bir dile çeviri yapan kişinin iki dili iyi şekilde bilmesinin yetmeyeceği çeviri yaptığı metnin konusu hakkında da bilgi sahibi olması gerekir. Söz gelimi tıp konusundaki bir kitabı çevirecek kişinin tıp konusunda uzman olmasının gerekeceği açıktır. Bilgisayar yardımıyla yapılan çevirilerde de benzer bir güçlüğün olacağı kolayca söylenebilir.

Günümüzde dünyada konuşulan diller belli dil ailelerinin üyeleridir. Aynı dil ailesinin üyesi olan iki dil arasında çeviri yapmak, doğal olarak farklı iki dil ailesinin üyesi olan iki dil

arasında çeviri yapmaya göre daha kolaydır. Bilgisayarlı çeviri dizgeleri gerçekleştirilirken kullanılacak yöntemler dillerin aynı dil ailesinden olup olmadığına göre değişmektedir.

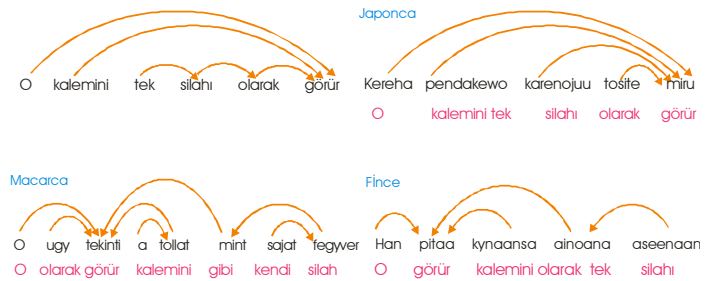
Bilgisayarlı çeviri amacıyla kullanılan yöntemler kural temelli ve istatistiksel temelli olmak üzere genel iki kümeye ayrılabilir. Kural temelli çeviri yöntemleri, yabancı dil eğitiminde izlenen yöntemlere benzer. Öncelik her iki dilin dilbilgisi kurallarını bilgisayara öğretmektir. Buna ek olarak çeviri sözlüğü bilgisayara yüklenir. İstatistiksel temelli çeviride, çevirmenler tarafından hazırlanmış metinler üzerinden yararlanır. Birbirinin karşılığı olarak iki dilde hazırlanmış, koşut derlemeler içinde bir tümce veya tümcecik karşılığı bulunmaya çalışılır.

Bilgisayar kullanarak yapılan çevirilerin başarımlarını karşılaştırabilmek için çeviri dizgesinden beklentilerin tanımlanmış olması gerekir. Bir bilgisayarlı çeviri dizgesi aşağıdaki özellikleri sağlamalıdır:

- **İnsan desteksiz** : Çeviri dizgesi insan katkı ve desteği olmadan çalışabilmelidir.
- **Kaliteli** : Çeviri dizgesinin ürettiği sonuçlar aslına uygun ve anlaşılır olmalıdır.
- **Konu bağımsız** : Çeviri dizgesi her türlü konuyu içeren metinleri çevirebilmelidir.

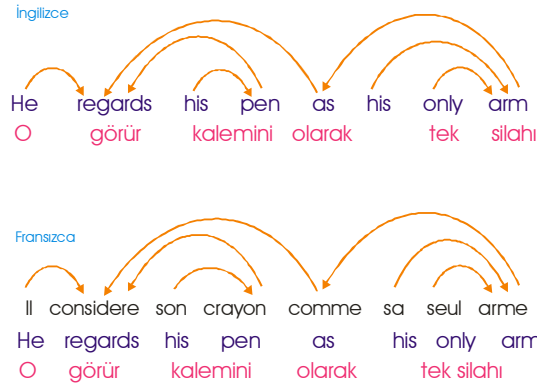
Yeni bir dili öğrenmeye çalışanların ilk yaptıkları sözcüklerin karşılığını ezberlemektir. Ardından öğrenmeye çalıştıkları dilde tümce kurabilmektir. Bazı dillerde sözcükler yalın hâlde bulunmasına karşın bazılarında sözcükler ekler olarak anlamlarını değiştirmektedir. Tümce içinde sözcüklerin niteliklerine göre dizilişi de dilden dile benzerlik ve farklılık göstermektedir.

Şekil-4'te Altay dil ailesine üye dillerinin ve Şekil-



Şekil-4: Türkçe, Japonca, Macarca ve Fince tümce yapıları

5'te Hint-Avrupa dil ailesine üye dillerin söz dizimi biçimleri gösterilmiştir.



Şekil-5: İngilizce ve Fransızca tümce yapısı

Bu şekillerden görüldüğü gibi akraba dillerde tümce içindeki sözcüklerin sıralanışı benzerlik göstermektedir.

Bitişken dillerde sözcükler çok sayıda ek alabilmektedir. Özellikle yapım ekleri sözcüğün anlamını değiştirmektedir. Hint-Avrupa dillerinde ek sayısı bir veya ikiyi geçmemektedir.

Türkçe ve akraba dillerde ses uyumu vardır. Başka dillerde ses uyumu diye bir kural bulunmamaktadır.

Diller arası çeviri yapılacağında aşağıda sıralanan tüm bilgilerin her iki dil için de bilinmesi gerekmektedir:

- Ses bilimsel
- Biçim bilimsel
- Söz dizimsel
- Dilimleme
- Anlatım biçimi

Günümüzde diller arası çeviri metinden metne yapılmaya çalışılmaktadır. Sesten sese çeviri yöntemleri üzerinde de çalışmaların olduğu bilinmektedir.

Bilgisayarlı çeviri konusunda A.C. Tantuğ [5], M. Orhun [6] ve Ö. Çetinoğlu'nun [7] çalışmaları örnek olarak verilebilir.

3. DDİ Çalışmalarında Karşılaşılan Zorluklar

Önceki bölümde DDİ alanında çalışanların ilgi alanları kısa açıklamalarla sunulmuştur. Aslında, sunulan her konu bir kitap olabilecek kadar geniştir. Bu bölümde DDİ konuları işlenirken karşılaşılan güçlükler ve bu güçlükleri aşmak için sürdürülen çabalardan özet biçiminde söz edilecektir. Karşılaşılan sorunların başlıkları aşağıda sıralanmıştır:

- Kurationsız ve anlaşılmasız konuşmalar
- Kurationsız ve bozuk yazılar
- Konuşmayı dilimleme
- Metni dilimleme
- Sözcük niteliklerini belirleme
- Anlam belirsizliklerini giderme
- Söz dizimsel belirsizliklerin giderilmesi
- Konuşma planı

3.1 Kurationsız ve Anlaşılmasız Konuşmalar

Her dilin en güzel ve kuralları çerçevesinde konuşulduğu bir yer vardır. Bu yer o dil için ölçünlü dil olarak kabul edilir. Konuşmayı anlama ve yazıya dökme üzerinde çalışanlar bu çalışmaları ölçünlü dil üzerinde sürdürürler. Her dil yöresel olarak farklı biçimde konuşulabilir. Aynı dili konuşup birbirini anlamayan yöre insanlarına da rastlanmaktadır. Aşağıda, konuya açıklık getirmek için Türkçe bazı örnek tümceler verilmiştir.

- *Abim bugün okula gelmicek*
- *Napıyon len?*
- *Gelcen mi?*

Türkçenin seçil bir dil olması yukarıdaki örnekleri yazıya dökme sürecinde önemli bir soruna neden olmaz. Sözcükleri ağızdan çıktığı gibi yazarsak Türkçe konuşanlar bunu anlayacaklardır. Seçil olmayan dillerde kural dışı veya aksanlı söyleyiş biçimleri için ayrı bir yazım kuralı geliştirmek gerekmektedir.

Yukarıda verilen örnekler yöresel ağızlardan alınmış örneklerdir. Bu tür örneklerin dışında kalan çoğunlukla gelip geçici konuşma biçimleri için çözüm bulmanın çok gerekli olmadığı düşünülebilir. Çünkü konuşmayı metne çevirme konusundaki çalışmaların temel hedefi;

- Konuşma tutanaklarını oluşturmak,
- İnsanın isteklerini anlayabilmek,

- İnsanın söylediğinden bilgisayar için veri üretmek
- İnsanın verdiği sözlü emirler ile bilgisayarlı dizgeleri yönetmek

biçiminde özetlenebilir.

Toplantı, bilimsel, ticari veya siyasi olabilir. Böyle bir toplantıda konuşulanların genellikle düzenli ve kurallı olduğu varsayılabilir.

Özellikle bilgisayarlı yanıtlama dizgelerinde müşterinin söylediklerini anlamak önemli olmaktadır. Her müşterinin ölçünlü dili kullandığı söylenemez. Bu durumda müşteriye kesin yanıt vereceği sorular sorularak konuşması anlaşılır hâle getirilir.

Bazı konuşma tanıma dizgeleri sadece veri girişi amaçlıdır. Örneğin sesle telefon numarasının çevrilmesi, hekimin hastaya koyduğu tanıyı söylemesi gibi. Bu örneklerde bilgisayarın beklediği sözcük veya sözcük öbeği bellidir. Bu sözcüklerin ve sözcük öbeklerinin bilgisayara öğretilmesi ile sorunun üstesinden gelinir.

Robot gibi akıllı dizgeleri sözlü emirlerle yönlendirmek istediğimizde, belli sözcükleri kullanırız. Bu sözcükleri anlaşılır biçimde söylediğimizde robot tarafından anlaşılacaktır.

Konuşmayı anlama çalışmalarında, anlamayı zorlaştıran başka etkenler de bulunmaktadır.

- Gürültü sorunu
- Kişiyeye bağlılık
- Konuya bağlılık

Metne çevrilecek olan tüm konuşmaların gürültüden arınmış ortamlarda dinlenilmiş veya kaydedilmiş olması her zaman olanaklı değildir. Konuşmaların gerisinde çeşitli gürültüler olabilir. Bu durumda konuşmayı gürültüden ayırt etmek başlı başına bir sorundur. Konuşma ile gürültüyü ayırt etmenin zorluk derecesini konuşmanın genliği ile gürültünün genliği arasındaki oran belirler. Bir başka deyişle gürültünün genliği büyüdükçe ayırt etme işlemi zorlaşır hatta olanaksız hâle gelebilir.

Bir dilde özellikle ünlülerin temel frekansları bellidir. Bu temel frekans insandan insana değişmez. Ancak harmoniklerinin genliği insandan insana değişir. Bu özellik konuşan kişinin kimliğini anlamamızı sağlar. Konuşmayı anlayan dizgeyi belli bir insan sesi için ayarlamak daha kolaydır. Aynı

dizgeyi konuşmacıdan bağımsız hâle getirmek daha gelişmiş bir çözüm gerektirir.

Konuşma metni içinde geçen sözcüklerin çokluğu anlamayı zorlaştırır. Bir başka deyişle sözcük dağarcığı sınırlı olan bir konuşma tanıma dizgesi geliştirmek daha kolaydır.

3.2 Kuralsız ve Bozuk Yazılar

Günlük yaşamımızda karşılaştığımız yazıların tamamında dilbilgisi kurallarına uygun yazıldıklarını ve noktalama işaretlerinin doğru kullanıldığını söylemek zordur. Basılı metinlerin bilgisayara aktarımı optik karakter okuyucu ile gerçekleştiriliyor ise bazı karakterlerin yanlış okunduğu da bir gerçektir. Yazılı metinler içinde kısaltmalar da bulunabilir. Hatta bazı sözcük kümeleri hem açık hem de kısaltma biçiminde yer almış olabilir. Metin günlük veya yöresel konuşmaların yazılmış hâli de olabilir.

Böylesine sorunlu metinler üzerinde önceki bölümde anlatılan DDİ çalışmalarını gerçekleştirebilmek için metin hatalarından arındırılması gerekir. En sık karşılaşılan hatalar yazım hatalarıdır. Özellikle sesçil olmayan İngilizce gibi dillerde yazım hatalarının sayısı azımsanmayacak kadar çoktur. Yazım hatalarını gidermeye yönelik çalışmalar öncelikle sözcüklerin doğru yazılmış olup olmadıklarını denetler ve düzeltmeye çalışır. Yanlış yazılmış sözcükleri düzeltmede,

- Yanlış yazıldığı saptanan sözcüğe en yakın benzer sözcük bulunmaya çalışılır. Bu kapsamda harf sıralarında şaşırma olup olmadığı, eksik harf veya yanlış harf olup olmadığı araştırılır.
- Türkçe gibi kurallı dillerde sözcük kurallara göre düzeltilmeye çalışılır.

Konuya açıklık getirmek amacıyla aşağıdaki örnekleri incelemekte yarar vardır.

Yazılmak istenen sözcük "kelebek" olsun. Ancak yazılan sözcükler şunlardan biri olabilir:

Kelebeklar
Kelebkler
Keleebler

Birinci yanlış sözcükte "e harfi yerine "a" yazıldığı görülmektedir. Yedi harften oluşan bu yanlış sözcüğe, sözlükteki en yakın sözcük "kelebek"tir. Bu iki sözcüğün benzerlik oranı 6/7 dir. Sözcükteki harf sayısı fazla ise bu yöntem olumlu sonuç

vermektedir. Benzerlik yöntemiyle hata giderme her zaman olumlu sonuç vermeyebilir. Örneğin, "liman" ve "limon" sözcükleri 4/5 oranında benzemektedir ancak her iki sözcük de sözlükte vardır. Dolayısıyla her ikisi de doğru olabilir veya biri yanlış yazılmış olabilir.

İkinci örnekte bir harf eksik yazılmıştır. Bu sözcüğün en yakın karşılığı sözlük veri tabanında araştırıldığında "kelebek" sözcüğüne erişilecektir. Çünkü ilk dört ve son üç harfi aynıdır. Bu sonuç bir harfin eksik olduğunu göstermektedir.

Daktilo ve bilgisayar ile yazı yazılırken sıkça yapılan bir yanlış, iki harfin ters sırada basılmasıdır. Üçüncü örnek bu tür yanlışlara bir örnektir. Sıra şaşırması adını verebileceğimiz bu yanlış bulabilmek için yine sözlük veri tabanında en yakın benzer sözlük aranır ve en yakın sözcük ile harf şaşırması olup olmadığına bakılır.

Türkçe gibi kurallı diller için sözlük kullanmaya gerek duymayan yöntemler geliştirilebilir. Örneğin yanlış yazılan ilk sözcükteki "a" harfinin yanlış yazıldığı, doğrusunun "e" olması gerektiği ses bilimi kurallarından çıkarılabilir.

İkinci ve üçüncü yanlış sözcükler incelendiğinde Türkçenin hece kurallarını sağlamadıkları görülür. Dolayısıyla öncelikle yanlış yazılmış olduklarına karar verilir. Verilen karara uygun olarak düzeltme sürecine geçilir.

3.3 Konuşmayı Dilimleme

Sıradan insanların konuşmalarında sözcüklerin nerede başlayıp nerede sonlandığını anlamak kolay değildir. Benzer şekilde tümcenin nerede başlayıp nerede sonlandığı da kolayca anlaşılamaz. Bazı konuşmacılar bir tümceyi tamamlamadan bir başka tümceye başlar.

Konuşmayı metne çevirme sürecinde karşılaşılan bu zorluk konuşmacının soluk alması, ara vermesi ve dilbilgisi kuralları yardımıyla çözümlenmeye çalışılmaktadır.

3.4 Metin Dilimleme

Bazı dillerde yazılmış metinlere baktığımızda her bir sözcüğü ve tümceyi kesin biçimde ayırt edebiliriz. Buna karşın bazı dillerde sözcükleri birbirinden ayırt etmek çok zordur. Çin, Japon, Tayland dillerinde sözcüklerin nerede başlayıp nerede bittiğini anlamak, dolayısıyla sözcükleri ayırt etmek zordur.

Almancada birden fazla sözcük birleştirilerek tek sözcük hâlinde kullanılmaktadır.

Bazı dillerde, tümceler çok uzun kurulmaktadır. Bir tümcenin beş on satır kapsadığı, bir paragraf kadar uzun olduğu görülmektedir. Arap metinleri bu özelliktedir. Benzer duruma XIX. yüzyıl Osmanlı metinlerinde de rastlanmaktadır. O devirde uzun tümce kurmak bir hüner olarak değerlendirilmiştir. Bu yazım alışkanlığının hukuk alanında hâlâ sürdürüldüğü görülmektedir.

Uzun tümcelemin taşıdığı anlamı ortaya çıkarmak için önce tümceyi dilimlemek ve her bir dilimi ayrı anlamlandırmak gerekir.

3.5 Sözcük Niteliklerindeki Belirsizlik

Doğal dil işlemenin en temel konularının başında sözcüklerin dilbilgisi açısından niteliklerinin belirlenmesi ve bu nitelikler ile sözcükleri etiketlemek gelir. Bir tümce içindeki bir sözcüğün niteliğini belirlemek dilbilgisi bilen biri için kolay olmakla beraber bu işlemin bilgisayar tarafından yapılması oldukça zordur. Zorluğun nedeni her dilde bir sözcüğün birden fazla niteliğinin bulunmasından kaynaklanmaktadır.

En temel dilbilgisi kaynakları sözcükleri aşağıdaki gibi sınıflandırmaktadır:

Ad soylular

- Ad
- Sıfat
- Adıl
- Belirteç
- İlgeç
- Bağlaç
- Ünlem

Eylem soylular

- Eylem

Bu sınıflar ayrıca kendi içlerinde alt kümelere, alt kümeler de kendi altında yeni niteliklere ayrılmaktadır. Örneğin sıfatlar aşağıdaki gibi iki aşamalı kümelenebilirler:

- Niteleme Sıfatları
- Karşılaştırma sıfatları
- Pekiştirme sıfatları
- Küçültme sıfatları
- Unvan sıfatları
- Belirtme Sıfatları
- İşaret sıfatları

- Sayı sıfatları
- Belgisiz sıfatlar
- Soru sıfatları

Bu temel özelliklerin yanı sıra ad soylu sözcüklerin çoğul olma özellikleri de vardır. Eylem soylu sözcükler kişi ve zaman bilgisi içerirler.

Türkçe gibi bitişken dillerde kök sözcüğe ulanan ekler son derece etkindir. Köke eklenen ekler, sözcüğün anlamını değiştirebildiği gibi sözcüğün niteliğini de değiştirir.

Göz + lük + çü + ler

Bu örnekte kök sözcük "göz" dür ve görme organını belirtir, ad soyludur.

Kök sözcüğe eklenen "lük " yapım eki görmeye yardımcı olan aracın adıdır; ad soyludur ve gözlük artık bir gövdedir. Gözlük gövde sözcüğüne eklenen "çü" yapım eki, gözlük ticareti veya da üretimi ile ilgilenen kişi veya kuruluşları belirtir; ad soyludur; gövdedir. Gözlükçü gövdesine eklenen "ler" çekim eki, bu sözcüğü çoğul yapmaktadır. Göz sözcüğüne farklı ekler eklenerek sözcüğün niteliği de değiştirilebilir. Örneğin;

Göz+le+mek : Eylem

Göz + de : Sıfat

Yukarıda verilen örnekte "göz" sözcüğünün görme organı anlamı kullanılmıştır. Aslında "göz" sözcüğü farklı anlamlarda da kullanılmaktadır, örneğin; çekmece gözü, iğne gözü, bellek gözü, köprü gözü, iki gözlü ev (oda anlamında), suyun gözü (suyun çıktığı yer), çubuklarda filiz veren nokta, göze gelmek vb.

Yukarıda verilen örneklerden görüldüğü gibi tek başına "göz" sözcüğünü dilbilgisi açısından etiketlemeye kalktığımızda ortaya çok sayıda belirsiz durum çıkacaktır. Çok kaba bir araştırmayla Türkçedeki sözcüklerin yarısının en az iki anlamı bulunduğu söylenebilir. Biçimbilimi belirsizliği diyeceğimiz bu belirsizlikleri gidermek için dillere bağlı ve dillerden bağımsız yöntemler geliştirilmiştir.

Komşu sözcüklere (önceki ve sonraki ya da önceki iki veya daha fazla) bakarak sözcüğün niteliğini bulmaya çalışan yöntemler bulunmaktadır. Bu yöntemler çoğunlukla istatistiksel temellidir ve "n-gram" adını alırlar (n komşu sözcük sayısını

belirler). İstatistiksel yöntemle sözcük niteliğini belirlemek için derlemlerden yararlanılır.

Komşu sözcüklere bakarak kural tabanlı karar verme yöntemleri de geliştirilmiştir. Örneğin "ela göz" ikilisinde, "ela" sözcüğünün niteleme sıfatı olduğu ve bir ad soyluyu niteleyeceği bilindiği için "göz" sözcüğünün ad soylu olduğu söylenebilir.

Tek başına sözcüğe bakarak niteliğini ortaya koymak için bitişken dillerde köke ulanan yapım ve çekim eklerine bakmak bir yöntem olarak geliştirilmiştir. Bu ilkeyi temel alarak Türkçe için geliştirilmiş yöntemlerin nitelik belirlemede %90'dan daha iyi sonuç ürettikleri söylenmektedir. Sözcük niteliklerindeki belirsizliklerin giderilmesi konusunda, H. Sak [8], D. Yüret [9] ve R. Ehsani [10] nin çalışmaları incelenebilir.

3.6 Anlam Belirsizliği

Her dilde anlamları farklı ancak yazılışları aynı olan sözcükler bulunmaktadır. Sesteş denilen bu sözcükler, bir yazının anlaşılmasında zaman zaman anlam belirsizliklerine neden olmaktadır.

Sesteş sözcükler ad soylu olabilecekleri gibi hem ad hem de eylem soylu olabilirler. Türkçede hem ad ve hem de eylem soylu sesteş sözcüklerin sayısının az olduğu söylenmektedir. Anlam belirsizliği özellikle diller arası bilgisayarlı çeviride önemli bir sorun olarak karşımıza çıkmaktadır. Konuyu açmak amacıyla aşağıdaki gazete başlığını okumakta yarar vardır.

Köprücüler İstanbul'da toplanıyor

Çok satan ve saygın bir gazetemizde çıkan bu başlığı ilk okuyan bir okur, köprü inşaatı ile ilgilenen kuruluş yetkililerinin İstanbul'da bir toplantıda bir araya geleceklerini düşünür. Yazıyı okumaya başlayınca şaşırır. Çünkü toplantı sonunda birinci geleceklere ödülleri verileceğinden söz edilmektedir. Okuyucu biraz kafasını yorduğunda toplantıya katılanların köprü yapımcıları olmadığını, birç oyuncularını olduğunu anlayacaktır. Dış kaynaklı bu haberi dilimize çeviren kişi, İngilizcede (bridge) sesteş olan birç oyunu ve köprü sözcüklerini karıştırmıştır. Aslında İngilizcede "bridge" sözcüğünün başka anlamları da vardır.

Yukarıda verdiğimiz örnekte de anlaşılacağı gibi anlam belirsizliği diller arası çevirilerde önemli bir sorun olarak karşımıza çıkmaktadır. Bu tür

belirsizlikleri gidermek için tek bir sözcüğe bakmak yerine kullanıldığı tümceye hatta içinde yer aldığı metnin tamamına bakmak gerekebilir.

1958 Jacques Tati yapımı ve ülkemizde "Amcam" adıyla gösterilmiş olan Fransız yapımı filmin adının (mon oncle) yanlış çevrildiğini dikkatli izleyiciler kolayca fark edebilmişlerdir. Çünkü film, kadın oyuncunun dayısı ile olan ilişkileri üzerine kurulmuştu. Bu örnek, bir dilde akrabalık ile ilgili sözcüklerin zengin olmasına karşın diğer dilin bu alanda fakir olmasının neden olduğu belirsizlik için güzel bir örnektir. Fransızcada "oncle" sözcüğü amca ve dayı için kullanılmaktadır. Bir bilgisayarın filmi izleyip Türkçe çevirisinde "oncle" sözcüğüne karşılık olarak amca mı yoksa dayı mı deme konusunda karar vermesi beklenmemelidir.

3.7 Söz Dizimsel Belirsizlik

DDİ ile ilgilenen bilişimciler doğal dillerin bilgisayar programlama dilleri kadar kesin olmasını beklerler. Ancak doğal diller, bu konuda DDİ alanında çalışanları hayal kırıklığına uğrattılar.

Bazı diller söz dizim kuralları açısından diğer dillere oranla daha katı özelliklere sahiptir iken bazıları son derece esneklerdir. Konuya açıklık getirmek için kısa ve basit bir tümceyi ele alalım. Örneğin İngilizcede şu iki tümceyi ele alalım;

"Ali gave the bears the quins because they were hungry"

"Ali gave the bears the quins because they were sweet"

Dilbilgisi yapısı açısından değerlendirildiğinde bu iki tümcenin aynı yapıda olduğu kolayca söylenebilir. Ancak anlatmak istedikleri farklıdır. Birinci tümce, ayılar aç olduğu için Ali'nin ayvaları ayılara verdiğini söylemektedir. İkinci tümce ise ayılar tatlı olduğu için Ali'nin ayvaları ayılara verdiği söylenmektedir. Bu örneklerden görüldüğü gibi, tümce içinde sözcüklerin sıralanışı aynı olsa bile tümcelerin anlatmak istediği şey farklı olabilir. Benzer söz dizimsel belirsizlikler Türkçe tümceler için de geçerlidir:

Ali ayvaları ayılara aç oldukları için verdi.

Ali ayvaları ayılara tatlı oldukları için verdi.

Bu örneklerdeki belirsizlikler ancak sözcüklerin ilişkileri ile çözülebilmektedir. Sözcükler arasındaki ilişkiler, "sözcük ağı veya sözcük ağacı" denilen bir

yapıya oturtulmaktadır. Bu ağaç üzerinde dilde olan her sözcük yer bulmaktadır. Ağaç üzerindeki bir sözcüğe bakıldığında üst ve alt ilişki kümeleri görülebilmektedir.

Yukarıdaki ilk tümcede "açlık" ile "canlı" sözcükleri ilişkilidir. Ayı bir canlı olduğuna göre "açlık" sözcüğü "ayı" ile ilişkilendirilmektedir. İkinci tümcede "tatlı" ile "ayva" sözcüğü ilişkilendirilmektedir. Çünkü meyve gibi nesnelere tatlılığından söz edilebilir. Ayıların tatlı olması söz konusu değildir.

Söz dizimi kuralları açısından Türkçe çok esnek sayılabilir. Aşağıdaki örnekler Türkçenin bu özelliği için çarpıcı örneklerdir.

Telefonla bağlan dünyaya

Bağlan telefonla dünyaya

Dünyaya telefonla bağlan

Telefon, bağlanmak ve dünya gibi üç sözcük ancak yukarıda gösterildiği gibi üç biçimde sıralanabilir. Bu üç tümcenin anlatmak istediği şey aynıdır. Diğer bir deyişle anlamları aynıdır. Bu üç tümcenin aynı anlamı vermelerinin nedeni sözcüklerin aldığı eklerdir. Benzer bir örnek bitişken olmayan bir dilde görülemez.

4. Dillerin Karşılaştırılması

Daha önce söylendiği gibi dünyada 4000'den fazla konuşulan dil olmasına karşın yaygın olarak kullanılan dillerin sayısı sınırlıdır. Yaygın kullanım sıralamasında Türkçe beşinci sırada yer almaktadır. Dillerin bilimsel olarak karşılaştırılmasını yapmak üzere geliştirilmiş ölçütlere rastlanmamakla beraber bazı ölçütleri üretmek için aklımızı kullanabiliriz.

Duygu, düşünce ve eylemlerimizi anlatmaya yarayan bir araç olarak tanımlanabilir. Bu tanımdan da anlaşılacağı gibi önemli olan çevremizdekilerle iletişim kurabilmektir. Kolaycı bir yaklaşımla bir dili diğer dillere karşı güçlü gösteren özelliğin söz varlığı olduğu söylenebilir.

Dilleri karşılaştırırken söz varlığının genişliği bir ölçü olabilir ancak tek bu ölçü ile dilleri karşılaştırmanın doğru olmayacağını söyleyebiliriz. Kanımızca ölçü olabilecek başlıklar aşağıda sıralanmıştır:

- Anlaşılabilirlik
- Söyleyiş kolaylığı

- Söz varlığı
- İşlevlik
- Belirginlik
- Kurallı olmak

Bu kısımda yapacağımız değerlendirmelerin dil bilimi açısından değil, mühendis gözüyle yapıldığını vurgulamak isterim.

4.1 Anlaşılrlık

Telefon hatlarının iletişim kalitesini ölçmek için kullanılan bir yöntem, bir dilin anlaşılrlığını ölçmek için kullanılabilir. Deneyin yapılış biçimi şöyledir: Birbirini görmeyen iki kişi telefon aracılığı ile konuşmaya başlarlar. Söz gelimi telefonun bir ucunda Eren diğer ucunda Gözde bulunmaktadır. Bu kişilerin düzgün konuştukları varsayılacaktır. Eren kendisine verilmiş olan bir metni belli bir hızla okur; diğer taraftaki Gözde duyduklarını yazar. Eren'in okuması bittiğinde Gözde aynı boyda ancak farklı bir metni okumaya başlar. Bu kez Eren duyduklarını yazar. Deney bittiğinde iki tarafın yazdıkları gerçek metinler ile karşılaştırılır.

Doğal olarak her iki dinleyicinin yanlış duydukları dolayısıyla yanlış yazdıkları sözcükler olacaktır. Anlaşılrlık derecesi aşağıdaki gibi yazılabilir:

$$AD = 1 - \frac{YS_1 + YS_2}{M_1 + M_2}$$

Bu bağıntıda,

AD : Anlaşılrlık derecesi

YS : Yanlış sayısı, YS_1 : birinci dinleyenin yanlış sayısı, YS_2 : ikinci dinleyenin yanlış sayısı.

M : Okunan metindeki sözcük sayısı, M_1 : Birinci okuyucunun okuduğu metnin sözcük sayısı, M_2 : İkinci okuyucunun okuduğu metnin sözcük sayısı.

Bu deney çok sayıda farklı iki kişi arasında yapılır ve her deney sonunda bulunan anlaşılrlık derecelerinin ortalaması alınırsa bir dile ilişkin anlaşılrlık derecesi ölçülmüş olur. Aynı deney farklı diller için benzer şekilde gerçekleştirildiğinde diğer diller için de anlaşılrlık derecesi hesaplanmış olacaktır. Deneyin bilimsel olabilmesi için aynı telefon bağlantısının kullanılmasına seçilen kişilerin eğitim düzeylerinin ve konuşma düzgünlüklerinin yaklaşık aynı olmasına dikkat edilmelidir.

Türk dilinin iletişim ortamlarında anlaşılrlık deneyleri yapılmaktadır. Bu deneyler, Türkçenin anlaşılrlığının yüksek olduğunu göstermektedir. Bu konu ile ilgili deneysel sonuçlar H. Palaz, Y. Bıcıl, A. Kanak ve M. U. Doğan'ın çalışmalarında yer almaktadır [11]

4.2 Söyleyiş Kolaylığı

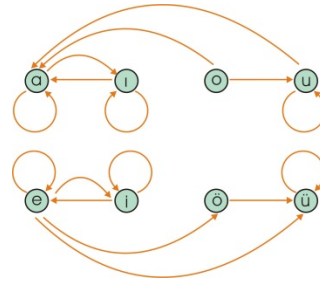
Dillerin söyleyiş kolaylığı ve kulağa geliş biçimleri hakkında değişik görüşler dil bilimcileri tarafından söylenmektedir. Dil bilimciler Türkçe için doğunun İtalyancası nitelemesini yaparlar. Bu değerlendirmelerin öznel olduğu bir gerçektir. Ancak;

- Türkçede ses uyumunun bulunması,
- Sözcük ve tümce içindeki vurguların kurallı olması ve
- Sesli harflerinin zenginliği

Türkçeyi kulağa hoş gelen bir dil yapmaktadır ifadesi daha teknik bir değerlendirme sayılabilir.

Türkçe ve akraba dillerde ses uyum kurallarının varlığı bilinmektedir. Ses uyum kuralları, aslında ses organlarımızdan kaynaklanan bir sonuçtur. Ses uyum kurallarını hatırlatmak için değişik bir bakış açısıyla Şekil-6 ve Çizelge-1 verilmiştir. Şekil-6 ve Çizelge-1 birlikte incelendiğinde aşağıdaki sonuçlar kolayca çıkarılabilir:

- Dudaklarımız düz ve açık iken ağızımızın gerisinden çıkardığımız ses "a" sesidir. "a" ünlüsünün ardından ağız ve dudamızın şeklini bozmadan, yine bir "a" ve çok az değiştirerek "ı" sesini çıkarabiliriz.



Şekil-6: Türkçenin ünlü uyumu

Çizelge-1 : Türkçenin Ünlü seslerinin Çıkış Yerlerine Göre Kümelenişi

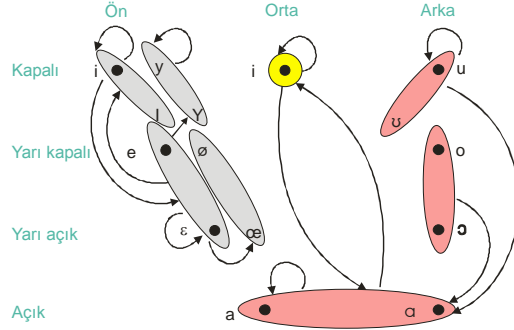
	Ön		Orta		Arka	
	Düz	Yuvarlak	Düz	Yuvarlak	Düz	Yuvarlak
Kapalı	i	ü	ı			u
Açık	e	ö			a	o

- Dudaklarımız düz ve kapalı iken ağzımızın ortasından çıkardığımız ses "ı" sesidir. "ı" ünlüsünün ardından ağız ve dudamızın şeklini bozmadan yine bir "ı" ve çok az değiştirerek "a" sesini çıkarabiliriz.
- Dudaklarımız yuvarlak ve açık iken ağzımızın gerisinden "o" sesini çıkarabiliriz. "o" ünlüsünden sonra tekrar "o" sesini çıkarabiliriz. Ancak kulağa hoş gelmediğinden Türkçe sözcüklerde bu duruma rastlanmaz. Dudak yapısında küçük bir değişiklikte "u" ve "a" sesleri kolayca çıkarılabilir.
- Dudaklarımız yuvarlak ve kapalı iken ağzımızın gerisinden "u" sesini çıkarabiliriz. "u" ünlüsünün ardından yine "u" ünlüsü gelebilir veya ağız biçimimizi biraz değiştirerek "a" sesini çıkarabiliriz.
- Dudaklarımız düz iken ağzımızın ön kısmından çıkardığımız seslerden biri "e" diğeri "i" sesidir. "e" ünlüsünden sonra "e" ünlüsünü çıkarmak için ağız yapımızda bir değişiklik yapmaya gerek yoktur. "e" ünlüsünün ardından ağız ve dudamızın şeklinde küçük bir değişiklik yaparak "i", "ö" ve "ü" seslerinden birini çıkarabiliriz. "i" ünlüsünden sonra hiçbir değişiklik yapmadan "i" ve az değişiklikle "e" ünlüsü de kolay çıkarılabilir. Ancak "ö" ve "ü" ünlülerinden sonra "e" veya "i" ünlüsünü söylemek zordur.
- Dudaklarımız yuvarlak ve açık iken ağzımızın ön kısmından "ö" sesini çıkarabiliriz. "ö" ünlüsünden sonra tekrar "ö" sesini çıkarabiliriz. Ancak kulağa hoş gelmediğinden Türkçe sözcüklerde bu duruma rastlanmaz.
- Dudaklarımız yuvarlak ve kapalı iken ağzımızın ön kısmından "ü" sesini çıkarabiliriz. "ü" ünlüsünün ardından yine "ü" gelebilir.

Yukarıdaki açıklamalara bakarak aşağıdaki genellemeleri yapabiliriz:

- Ağız yapımızda hiç değişiklik yapmadan çıkarabileceğimiz ardışık ünlü, ilk ünlünün aynıdır.
- İlk ünlüyü söylerken dudamızın durumu neyse (açık veya kapalı) ardışık ünlüyü söylerken de aynı kalması ardışık ünlüyü söylemeyi kolaylaştırır.
- İlk ünlüyü söylerken dudamızın durumu neyse (düz ya da yuvarlak) ardışık ünlüyü söylerken de aynı kalması ardışık ünlüyü söylemeyi kolaylaştırır.
- Yuvarlak ünlülerden "o" ve "u" düz ünlülere geçiş ancak arka bölgede kalmak koşuluyla yapılabilir. Bu durumda geçiş sadece "a" ünlüsüne olabilir.
- Düz "e" den sadece, aynı bölgede (ön) kalmak koşulu ile yuvarlak ünlülere ("ü" ve "ö") geçiş olabilir.

Ünlüler arası geçiş olaylarını ünlüler dörtgeninde gösterdiğimizde yukarıda yaptığımız yorumlar açık biçimde görülmektedir, Şekil-7.



Şekil-7: Ünlüler arası geçişler

Şu iki yapay sözcüğü okumaya çalışalım.

tenteredi, tintoridö

Hangisini söylemek daha kolay? Hangisi çenenizi daha az yoruyor? İlk sözcüğü söylemek daha kolay ve çenemizi daha az yoruyor. Bu sözcük Türkçe bir sözcük olmasa da ünlülerin dizilişi Türkçe ses kurallarına uygun olduğu için kolay söylenebilmektedir.

4.3 Söz Varlığı

Bir dilin etkinliği söz varlığının zenginliği ile de ölçülebilir. Bir dilin söz varlığının genişliği ile o dili konuşan insanın düşünme yeteneği doğru orantılıdır. Bir dilin söz varlığının zenginliğini araştırırken sadece sözlükteki sözcük sayısına bakmak doğru sonuç vermez.

İngilizceye veya Fransızcaya ilişkin sözlüklere baktığımızda sözcük zenginliğini görebiliyoruz. Buna karşın Türkçe sözlüklerdeki söz varlığı az gibi görünüyor. Bükümlü dillerde her somut ve soyut nesneye yeni bir ad vermek gerekmektedir. Bu özellik bu tür dillerde söz varlığının geniş olmasını gerekli kılıyor denebilir. Bitişken ve Sami dillerinde kök sözcüğe eklenen ekler ile yeni sözcük türetme olanağı bulunmaktadır. Türetilen sözcüklerin bazıları sözlüklerde yer almasa bile anlamı o dili kullananlar tarafından anlaşılmaktadır. Durumu çok bildik örnekler ile anlatabiliriz. Çizelge-2'de görme organı olan "göz" sözcüğünden sadece yapım ekleri eklenerek türetilmiş Türkçe sözcükler ve bunlara karşılık gelen İngilizce sözcükler gösterilmiştir.

Çizelge-2: Türkçenin Sözcük Türetme Yeteneği

Türkçe	İngilizce
Göz	Eye
Gözlük	Eyeglasses
Gözlükçü	Optician
Gözlükçülük	Opticians
Gözcü	Watchman
Gözcülük	Opthamolgy
Gözlem	Observation
Gözleme	Observing
Gözlemci	Observer
Gözlemcilik	Observation
Gözde	Favourite
Gözü	Mirror
Gözünlük	Mirror stand
Gözücü	Mirror maker
Gözücülük	Mirror makers

Çizelge-2'de görülen ayna anlamındaki "gözü" sözcüğü eskiden kullanılan bir sözcüktür. Anlamını öğrendikten sonra günümüzde kullanılmayan bu sözcükten türetilen dört sözcüğü de kolayca yorumlayabilmemiz Türkçenin sözcük türetme yeteneğinin bir kanıtıdır. Tablo-2'nin söylediği bir başka sonuç; Türkçe sözcüklere karşılık İngilizcede bir sözcük bulunabildiğidir. Ancak bunların önemli

bir kısmı görme organı ile ilişkili değildir. Bu durum sözcüklerin insan tarafından algılanması ve yorumlanmasında zorluğa neden olabilir.

Sami dillerinin kökten yeni sözcük türetme yetenekleri de üstündür. Örneğin "KTB" den oluşan bir kökten, "kitap, kâtip, kâtibe, mekteb" gibi sözcükler türetilmektedir.

Türkçede sözcükler (pekiştirme sıfatları ve yabancı sözcükler dışında) sadece son ek almalarına karşın İngilizcede ön ek de alabilmektedir. Ön ek kullanılmaması Türkçe için bir eksiklik olarak sayılabilir.

Ancak bu eksiklik Çizelge-3'te gösterildiği gibi ön ekin kazandırdığı anlamı veren yeni sözcüklerle veya ek sözcüklerle giderilmektedir.

Çizelge-3: Türkçede Ön Ek Kullanılmaz

Türkçe	İngilizce
Olanaksız	Impossible
Yeniden canlandırma	Reactivate
Tepkin	Reactive
Düzeltilmek	Reform
Mutsuz	Unhappy
Önlem	Precaution
Ön yagılı	Preconceived
Özürlü	Disabled

Bir dilin söz varlığının zenginliği değerlendirilken sadece ad soylu sözcüklere bakmak yeterli olamaz, eylem soylu sözcükleri de incelemek gerekir. Türkçede her eylemin etken, edilgen gibi temel biçiminin dışında dönüşlü, işteş ve ettirgen durumları diğer dillere göre önemli bir üstünlük sağlamaktadır. Çizelge-4'te Türkçenin eylem yapıları gösterilmiştir.

Çizelge-4: Türkçenin Eylem Yapıları

Eylem türü	Türkçe	İngilizce
Etken	Görmek	To see
Edilgen	Görünmek	To be seem
Edilgen	Görölmek	To be seen
Dönüşlü	Giyinmek	To dress
İşteş	Görüşmek	To see each other or to discuss
Ettirgen	Görüştürmek	To bring someone to see or to discuss each other

İkinci derece ettirgen	Görüşürtmek	To have somebodies to see or to dicuss each other
Ettirgen edilgen	Görüşürülme	To be brought to see or to discuse somebody

Diller, o dili kullanan ulusların yaşam biçimi, alışkanlıkları ve ilgi alanlarına göre gelişim göstermektedir. Felsefi konulara ilgi duyan toplumlarda bu alanda sözcük sayısı artarken aile bağları kuvvetli olan toplumlarda akrabalıkları tanımlayan sözcüklerin sayısı artmaktadır. Bazı diller olayları somut eylem sözcükleri ile anlatmaya çalışırken bazı diller deyimlerle anlatmayı yeğlerler. Bir deyim ile anlatılan bir eylem veya durumu, bir başka dilde bir sayfalık bir yazı ile anlatmak gerekebilir.

Türkçede eksik olarak söz edilen konulardan biri sözcüklerin cinsiyete göre değişmesidir.

Örneğin, melik-melike, rahip-rahibe, memur-memure gibi aynı meslek veya görevi üstlenmiş erkekler ve kadınlara farklı adların verilmesi özelliği Türkçede yoktur. Kişinin cinsiyetini belirtmek gerektiğinde sözcüğün öncesine veya arkasına cinsiyeti belirleyen sözcük konarak bu eksiklik giderilmektedir. Örneğin memur hanım, hanım memur gibi.

Ad soylu sözcüklere dişilik ve erkeklik özelliği kazandırılmış dillerde örneğin Fransızca ve Arapçada sözcüğün başında bulunan ek, sözcüğün cinsiyetini belirtir. Fransızcadan örnek verirse "la mur" (duvar), "le port" (kapı), duvar sözcüğü dişi ve kapı sözcüğü erkektir. Böyle bir ayırımın neden gerektiğini anlamak gerçekten güçtür, belki de anlamsızdır, denebilir.

4.4 İşleklilik

Eklemeli dillerde sözcüklere yapılan eklemeler sözcüklere yeni anlamlar kazandırmaktadır. Bu özellikler dilin söz varlığını zenginleştirmekle kalmayıp dile işleklilik te kazandırmaktadır.

Ad soylu sözcüklere yapılan eklemelerin kazandırdığı yeni sözcüklere örnekler Çizelge-3'te verilmiştir.

Eylem soylu sözcüklerin işlekliliği Çizelge-4'te gösterilmiştir. Eylemlerin etken ve edilgen hâlleri tüm dillerde görülmektedir. Ancak eylemlerin

dönüştürme, işteş, ettirgen, ikinci derece ettirgen, ettirgen edilgen hâllerini oluşturmak Türkçedeki kadar kolay ve düzenli değildir. Hatta bazı dillerde olanaklı değildir.

Bir dilin gelişmişliğinin bir göstergesi de sözcüklere yüklenen anlamlardır. Örneğin "kırmak" eyleminin en temel kullanımı, "sert şeyleri vurarak veya ezerek parçalamak" anlamındadır. Türkçe gibi gelişmiş dillerde kırmak sözcüğünün çok fazla anlamına rastlamak olasıdır.

Örneğin;

- Soğuk hayvanları kırdı (öldürmek)
- Fiyatları kırmak (indirim yapmak)
- Kalbimi kırdı (gücendirmek)
- Pulunu kırmak (tavla oyununda)
- Buğdayı kırdırmak (kaba öğütme)
- Direksiyonu kırmak (Çevirmek, yönünü değiştirmek)
- Soğuğun belini kırmak (Soğuğun etkisini ciddi olarak azaltmak)
- Senet kırmak (senedi paraya çevirmek)
- Dersi kırmak (dersten kaçmak)

4.5 Belirginlik

Tümce içindeki sözcüklerin dizilişinde belli bir sıranın olması beklenir. Her dilde, tümce içindeki sözcüklerin sıralanışında belirli kurallara uyulur. Bazı dillerde bu sıralamada değişiklik yapılırsa, anlam bozulur ya da kaybolur. Türkçe tümcelerde de sözcükler belli bir düzende sıralanır. Ancak sıra değiştirildiğinde anlam kaybı olmaz. Ancak vurgu farkı ortaya çıkar. Örneğin şu üç cümlenin anlamı aynı, vurgu yapılan sözcük farklıdır.

- Elini sabunla yıka (vurgu el)
- Sabunla elini yıka (vurgu sabun)
- Yıka elini sabunla (vurgu yıkamak)

Hint-Avrupa dillerinde tümce yapısı katıdır. Bu nedenle tümce içindeki sözcüklerin yerleri değiştirilince tümcenin anlamı değişir veya bozulur. Tümce yapısının esnek olması, Türkçenin anlam gücünü artırmakta ve konuşanlara yetenek ve kolaylık sağlamaktadır. Bu özellik doğal dil işleme çalışmalarında ek zorluklara neden olmaktadır.

4.6 Kurallı Olmak

Dillerin, insanlığın gelişimine koşut olarak geliştikleri, bu nedenle katı kurallarının olması beklenmemeli, diye düşünülmektedir. Ancak bu

düşünce Türkçe için geçerli değildir. Sanki 5000 yıl önce bir dil kurultayı yapılmış ve Türkçenin kuralları belirlenmiştir. Bu kurallar günümüze kadar değişmemiştir. Türkçenin bu özelliği, dil bilimcileri hayrete düşürmektedir. Türkçenin değişmeyen kuralları ses biliminde ve biçim biliminde hiç bozulmamıştır. Biçim bilimini kurallarında, bir kural dışı örnek verilebilmektedir. Bilindiği gibi ünlü ile biten bir sözcüğe iyelik ekleri ulanırken araya "n" harfi eklenir. Bu kurala göre "*sunun*" dememiz gerekirken "*suyun*" demekteyiz. Kural bozulması olarak değerlendirebileceğimiz bu durum aslında "su" sözcüğünün geçmiş zamanda "*suy*" biçiminde söylenmesinden kaynaklanmaktadır.

Dilleri, adlara eklenen durum ekleri açısından da inceleyebiliriz. Türkçede ad durum eklerinin kazandırdığı anlam çok kesindir. Bu kesinlik Hint-Avrupa dillerinde aynı düzeyde değildir.

Örneğin;

Müziği dinliyorum. (I am listining to music)
Partililere konuşuyorum.(I am speaking to members of parties)
Seni düşünüyordum. (I think of you)

Türkçede sık kullanılan eylem köklerinin tek hece olduğu bir gerçektir. Ayrıca Türkçe eylemlerin kökleri hiçbir zaman değişmez. Örneğin "gitmek" eylemini tüm zamanlar için yazarsak (**gittim**, **gitmişim**, **gidiyorum**, **gideceğim**, **giderim**) bu özelliği açıkça görürüz.

Hint-Avrupa dil ailesinde bulunan, Farsça, Fransızca ve İngilizce dillerinde gitmek eyleminin şimdiki zamanı ile geçmiş zamanı arasında harf benzerliği bile yoktur. Örneğin, I go - I went (Ing), Je vais - Je suis allé (Fr). hahem reft - reftem (Farsça), (ilk örnekler geniş zaman, ikincisi geçmiş zamana ilişkin örneklerdir). Türkçe eylemlerin çekiminde bu tür kural dışılıklara rastlanmaz.

Kaynaklar

- [1] Delibaş, A. "Doğal Dil İşleme ile Türkçe Yazım Hatalarının Denetlenmesi, Yüksek, Lisans Tezi, İTÜ Fen Bilimleri 2008
- [2] Şentürk, F. "Yetkin Bul ve Değiştir", TBV BBHD, Sayı 3 2010.

- [3] Adalı, Ş, "Türkçe Belgelerden Bilgi Çıkarımı İçin Bir Tümeleşik Mimari", Doktora Tezi, İTÜ Fen Bilimleri 2009
- [4] Şentürk, T. "Türkçe Metin Seslendirme", TBV BBHD, Sayı 4 2011.
- [5] Tantug, A. C. "Akraba ve Bitişken Diller Arasında Bilgisayarlı Çeviri İçin Karma Bir Model", Doktora Tezi, İTÜ Fen Bilimleri 2007
- [6] Orhun, M. "Uygurcadan Türkçeye Bilgisayarlı Çeviri", Doktora Tezi, İTÜ Fen Bilimleri 2010
- [7] Çetinoğlu, Ö. "Building a lexical functional grammar for Turkish".Doktora Tezi, Sabancı Üniversitesi, 2009
- [8] Sak, H., Gungor, T. and Saraclar, M., "Morphological Disambiguation of Turkish Text with Perceptron Algorithm", Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing '07, Springer-Verlag, Berlin, Heidelberg, pp.107-118, 2007
- [9] Yüret, D. and Töre, F., "Learning morphological disambiguation rules for Turkish", Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, HLT-NAACL '06, Association for Computational Linguistics, Stroudsburg, PA, USA, pp.328-334, 2006.
- [10] Razieh Ehsani, "Turkish Morphological Disambiguation using Multiple Conditional Random Fields" Yüksek Lisans Tezi, İTÜ Fen Bilimleri Ens. 2012
- [11] Palaz, H., Bicil, Y., Kanak, A., Doğan, M., U., "New "Turkish Intelligibility Test for Assessing Speech Communication Systems" Speech Communication, 2005