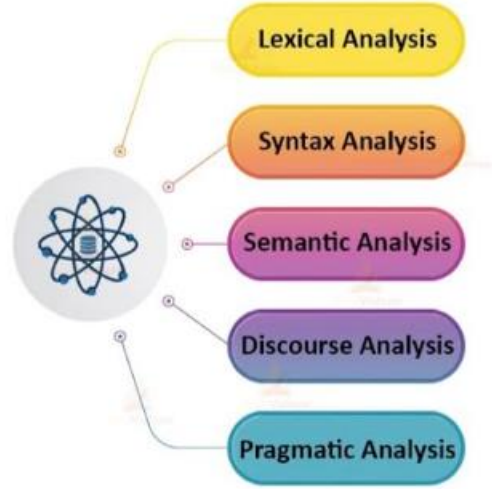


Dođal Dil İşleme Bilgi Düzeyleri-3



Prof.Dr. Banu Diri

Anlamsal (Semantik) Analiz Nedir?

Bir metinden anlamı doğru bir şekilde çıkarma sürecidir.

- ❖ **Gramer Analizi:** Metindeki cümlelerin gramatik yapısını inceler ve doğru gramer kurallarına uygunluğu değerlendirir.
- ❖ **Bağlamın Tanınması:** Belirli bir bağlamdaki kelimeler arasındaki ilişkilerin ve bağlantıların fark edilmesi
- ❖ **Bilgisayarların Anlama ve Yorumlama Yeteneği:** Bilgisayarların cümleleri, paragrafları ve hatta tüm belgeleri anlama ve yorumlama yeteneğinin sağlanması

Bilgisayarların insan dilini anlama ve işleme kabiliyetini geliştirmeye yardımcı olur.

Semantik Analize dayalı araçlar; e-postalar, destek talepleri ve tüketici yorumları gibi [yapılandırılmamış verilerden](#) faydalı bilgilerin otomatik olarak çıkarılmasına yardımcı olabilir.

Anlamsal (Semantik) Analiz çeşitli yöntemler kullanır. Bunlar, metnin ana fikrini, temasını ve bunların bağlantılarını bulmayı gerektirebilir.

- ❖ Popüler bir Anlamsal Analiz yöntemi, metnin ana fikirlerini ve bağlantılarını bulmak için **Makine Öğrenmesi ve Doğal Dil İşlemeyi** birleştirir. Bu, yeni metni analiz etmek ve temel fikirlerini ve ilişkilerini çıkarmak için büyük bir derlem üzerinde eğitilmiş bir makine öğrenimi modelinin kullanılmasını gerektirir.
- ❖ Bir başka strateji de, **belirli bir konudaki kavram ve ilişkilerin önceden oluşturulmuş ontolojilerini ve yapılandırılmış veri tabanlarını kullanmaktır**. Anlamsal Analiz algoritmaları, bu ontolojileri kullanarak metinden ilgili bilgileri daha hızlı bir şekilde bulabilir ve çıkarabilir.

Kelime

Anlamları

(Word Semantics)

Kelime Anlam Belirsizliđi

Bir kelimenin ne anlama geldiđini nasıl biliriz?

Kalem



kalem¹

yazma, çizme vb. işlerde kullanılan araç

Kalem masanın üstünde.

ACABA ?

- Basit bir kelime için bile tek bir anlam yoktur
- Hangi anlamını seçeceğimize karar vermek için kelimenin içinde bulunduğu bağlamı bilmemiz gerekir
- Bağlamda olan ve kelimenin gerçek anlamıyla olan bağıntısını belirleyen kavrama **sense** denir

”Kalem kağıda ihtiyacım var.”

“İki kalem erzak aldım.”

kalem ¹	yazma, çizme vb. işlerde kullanılan araç	Kalem masanın üstünde.
kalem ²	Resmî kuruluşlarda yazı işlerinin görüldüğü yer.	Kalemimiz bugün hizmet vermeyecektir.
kalem ³	yontma işlerinde kullanılan ucu sivri veya keskin araç.	Oymacı kalemi.
kalem ⁴	çeşit, tür	Beş kalem ilaç.
kalem ⁵	bazı deyimlerde yazı	Kaleme almak.
kalem ⁶	yazar	Peyami Safa, edebiyatımızın usta kalemlerindedir.

Anlam belirsizliğini gidermek üzerine yarışmalar yapılmaktadır

Senseval/Semeval 1998 yılında başlamıştır

Kelime Anlam Belirsizliđini gidermek için iki yaklaşım vardır:

- ❑ Sözcük Seçim
 - Seçilmiş Sözcük Yaklaşımı
 - Tüm Sözcükler Yaklaşımı

Kelime Anlam Belirsizliđini gidermek için kullanılan yöntemler:

- ❑ Yararlanılan kaynaklara göre
 - Bilgi Tabanlı Yöntemler
 - Derlem Tabanlı Yöntemler
 - Melez Yöntemler

- ❑ Denetimli ve Denetimsiz Yöntemler

❑ Sözcük Seçim Yöntemi (Lexical Sample) Seçilmiş Sözcük Yaklaşımı

- Anlamlarıyla birlikte küçük bir kelime grubu seçilir
- Bu kelimeler, içerisinde geçtiği cümleler göz önünde bulundurularak bir insan tarafından etiketlenir
- Avantajı, kelimelerin mümkün olan bütün anlamları tablodaki gibi listelenerek (kalem örneği) etiketleme yapacak kişilerin kelimelerin bütün anlamlarını atlamadan etiketleme yapabilmesi sağlanır
- Böylece, elde sınırlı sayıdaki kelimelerin tüm anlamları veya seçilmiş sayıdaki anlamları için etiketleme gerçekleştirilmiş olur

❑ Sözcük Seçim Yöntemi (Lexical Sample) Tüm Sözcükler Yaklaşımı

- Verideki bütün kelimeler aynı anda sisteme yüklenir
- İşaretleyiciler cümlelerdeki bütün kelimeleri anlamlarına göre işaretler
- Kelimenin kökü, büyük harfle yazılıp yazılmadığı, cümlenin hangi ögesi olarak kullanıldığı gibi özellikler kelime anlamlarını birbirinden ayırmak için kullanılır
- Kelimeler anlamlarına göre sınıflandırıldığında karar ağacı gibi sınıflandırma tabanlı yöntemler sistemi, yeni gelen kelimenin anlamını otomatik olarak etiketlemek için eğitilir

❑ **Belirsizlik Giderme Yöntemleri**
Bilgi Tabanlı Belirsizlik Giderme Yöntemleri

Sözcük anlamlarının belirtildiği kaynakları kullanır

- ❖ Sözlük
- ❖ Kelime Ağı (**WordNet**)

❖ **Sözlük Anlamının Bağlam ile Örtüşmesini Temel Alan Yöntemler**

- LESK Algoritması
- Sözcükler ve anlamlarını, yakın anlamlı sözcükleri kullanır
- *Oxford Advanced Learner's Dictionary* %50-70 arası başarı

LESK Algoritması

1. Verilen tümce içerisindeki belirsizliği giderilecek olan sözcük seçilir
2. Sözcüğün, sözlükteki açıklaması ve verilen örnek tümceler incelenir
3. Eşleşen sözcükler tespit edilir
4. En çok örtüşme hangi anlamdan elde edilmiş ise, belirsizliği belirlenmek istenen sözcüğe o anlam atanır
5. Tümcedeki anlam taşıyan kelimeler kullanılır, işlevsel kelimeler çıkarılır
6. Türkçe için çalışılıyorsa sözcüklerin kök veya gövde biçimleri kullanılır

Örnek : Eren'nin *konak* gibi büyük evi var.

konak

1) Büyük ve gösterişli ev. Örnek: <i>Konaktan okula dönerken çok yoruldum.</i>	çakışan 3 sözcük	← En doğru tahmin
2) Vali, kaymakam gibi yüksek dereceli devlet görevlilerinin resmî konutu. Örnek: <i>Vali konağının önünde korumalar vardı.</i>	çakışan 2 sözcük	
3) Konakçı. Örnek: <i>Yağmur nedeniyle konaklar yola çıkamadı.</i>	çakışan 1 sözcük	
4) Araba veya hayvanla bir günde alınan yol. Örnek: <i>Buradan orası beş konaktır.</i>		
5) Yolculukta geceyi geçirmek için inilen, konaklanan yer. Örnek: <i>Kervan konağa bu akşam ulaşacak.</i>		
6) Kundak çocuklarının başlarında görülen kepek tabakası. Örnek: <i>Çağla'nın başında konak oluştu.</i>		
7) Gözde oluşan ince tabaka. Örnek: <i>Erdem'in gözünde konak var.</i>		

LESK Algoritmasının **zayıf** yönleri:

1. Sözlükteki açıklamalarda, örnek tümcedeki sözcükler geçmeyebilir
2. Birden fazla açıklama ve örnek tümcede aynı sayıda örtüşme olabilir

Lesk → **Corpus Lesk** (sözcük örtüşmelerine ağırlık değeri verir) (Kilgarriff and Rosenzweig)

Lesk → **Uyarlanmış Lesk** (sadece seçilen sözcük değil, o sözcüğün açıklama ve örnek cümlesindeki diğer sözcüklerin tanımları da kontrol edilir) (Banerjee and Pedersen)

❑ Belirsizlik Giderme Yöntemleri Derlem Tabanlı Yöntemler

❖ Denetimli Yöntemler

- ✓ Makine öğrenmesi yöntemlerine (Karar Ağaçları, Naive Bayes, Destek Vektör Makineleri, ...) anlamları etiketlenmiş bir eğitim kümesi verilerek sistem modellenir
- ✓ Etiketli veri kümesi üzerinden çıkarılan özellikler ve etiket bilgisini kullanır

❖ Yarı Denetimli Yöntemler

❖ Denetimsiz Yöntemler

- ✓ Bir sözcüğü anlamı ile etiketlemeden önce, sözcüğün hangi anlam kümesine üye yapılacağına karar verir
- ✓ *Bağlam kümeleme* kullanılan yöntemlerden biridir
- ✓ Sözcükler bir vektör olarak gösterilir. Aynı ve yakın anlamdaki sözcük vektörleri aynı bağlam kümesi içerisinde yer alır
- ✓ İki sözcük vektörünün benzerliği hesaplanır
- ✓ Başarı düşüktür

Denetimli Yöntemler için WSD

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w|c) = \frac{\text{count}(w,c) + 1}{\text{count}(c) + |V|}$$

	Doc	Words (context of "bass")	Class
Training	1	fish smoked fish	f
	2	fish line	f
	3	fish haul smoked	f
	4	guitar jazz line	g
Test	5	line guitar jazz jazz	?

Priors:

$$P(f) = \frac{3}{4}$$

$$P(g) = \frac{1}{4}$$

$V = \{\text{fish, smoked, line, haul, guitar, jazz}\}$

Conditional Probabilities:

$$P(\text{line} | f) = \frac{(1+1)}{(8+6)} = \frac{2}{14}$$

$$P(\text{guitar} | f) = \frac{(0+1)}{(8+6)} = \frac{1}{14}$$

$$P(\text{jazz} | f) = \frac{(0+1)}{(8+6)} = \frac{1}{14}$$

$$P(\text{line} | g) = \frac{(1+1)}{(3+6)} = \frac{2}{9}$$

$$P(\text{guitar} | g) = \frac{(1+1)}{(3+6)} = \frac{2}{9}$$

$$P(\text{jazz} | g) = \frac{(1+1)}{(3+6)} = \frac{2}{9}$$

Choosing a class:

$$P(f | d5) \propto \frac{3}{4} * \frac{2}{14} * \left(\frac{1}{14}\right)^2 * \frac{1}{14} \\ \approx 0.00003$$

$$P(g | d5) \propto \frac{1}{4} * \frac{2}{9} * \left(\frac{2}{9}\right)^2 * \frac{2}{9} \\ \approx 0.0006$$

Yöntemlerin Karşılaştırılması

Yöntem	Üstünlük	Eksiklik
Bilgi Tabanlı	Daha yüksek doğrulukta sonuç üretir	Algoritmalar sözcük örtüşmesine dayalı, örtüşme seyrekliği sorunu gözlemlenebilir. Başarım sözcük tanımamalarına bağlıdır
Denetimli	Gerçekleştirme açısından diğer yöntemlere üstündür	Kaynakları kısıtlı olan dillerde elde edilen sonuçlar yetersizdir
Denetimsiz	İşaretlenmiş derlem gerekmiyor	Gerçekleştirilmesi zor ve daha düşük başarımlı sonuç verir

WORDNET

- WordNet (kelime ađı), bir dilin tüm kelimelerini, bu kelimelerin tanımlarını ve aralarındaki anlambilimsel ilişkileri içeren bir veritabanıdır
- İlişkiler, eş kümeler (synset) arasında tanımlanmıştır
- 1985 yılında, Princeton Üniversitesinde elle 10 yıla yakın bir sürede **İNGİLİZCE** için oluşturulmuştur
- Günümüzde 200'den fazla dili kapsar

WordNet Hangi Alanlarda Kullanılır

- Doğal Dil İşleme (metin analizi, makine çevirisi, anlam belirsizliği giderme)

WordNet ile Sözlük Arasındaki Fark Nedir?

- Sözlüğe benzer, isimler, sıfatlar, fiiller ve zarflar ayrı olarak gruplanır
- Eş anlamlı kelimeler *synset* (117 bin adet) adı verilen kümeler olarak tutulur
- WordNet içerisinde kelimeler hiyerarşik bir yapı altında dizilir
- *Synset*'ler eş anlamlı öbekler oluşturulduktan sonra anlamsal kapsamlarına göre sıralanırlar
- Anlamsal kapsamı daha geniş olan terimler *hypernym*, daha spesifik olanlar ise *hyponym* olarak adlandırılır
- Bir sözcük, anlamı daha geniş olan bir terimin *hyponym*'i iken, aynı zamanda daha dar anlamlı bir kelimenin *hypernym*'i olabilir

Örnek

“hayvan” sözcüğü “canlı” teriminin altında listelenirken, kendi altında “kedi” sözcüğünü içersin.

“kedi” sözcüğü, “hayvan” sözcüğünün **hyponym**’i, “canlı” sözcüğü de “hayvan” sözcüğünün **hypernym**’i olur.

“canlı” terimi, altında “insan” ve “bitki” grupları da olabilir

Türkçe WordNet Çalışmaları Nelerdir?

Türkçe üzerine yapılmış en büyük çalışmalarından biri KeNet*’tir.

KeNet 80 binden fazla *synset* içerir

Sabancı Üniversitesinin yürüttüğü bir başka proje de **BalkaNet**

** Starlang Yazılım olarak Işık üniversitesi bünyesinde geliştirilmiş Türkçe WordNet KeNet’e açık kaynak ve ücretsiz olarak ulaşılabilir*

WordNet deki ilişkiler

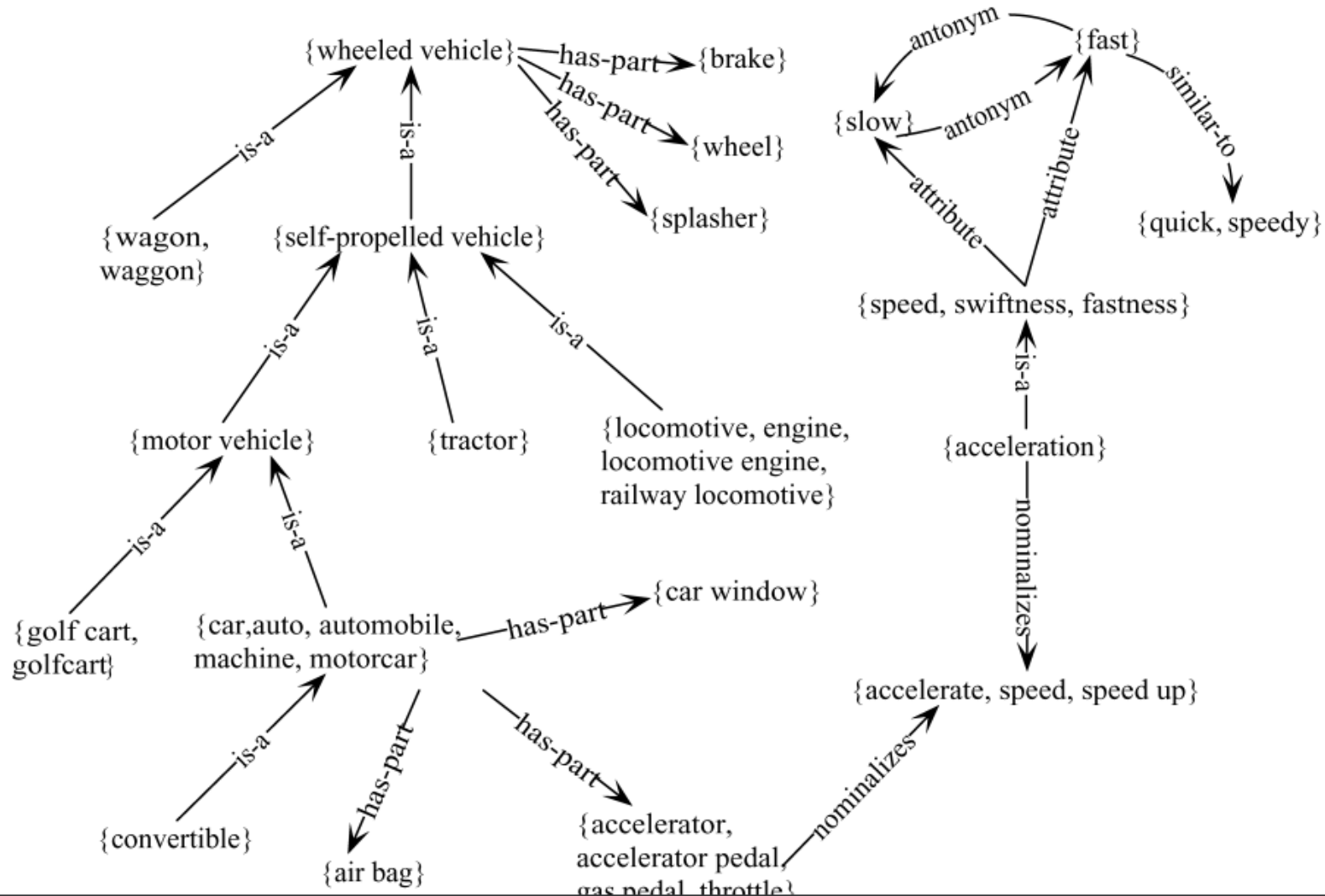
Synsets arasında en sık kodlanan ilişki süper-bağımlı ilişkidir
(hyperonymy, hyponymy veya ISA ilişkisi olarak geçer)

{Furniture}, {piece_of_furniture} en genel hali
{bed} ve {bunkbed} daha spesifik

Böylece,

WordNet, mobilyaların ranza, yatak içerdiğini;
yatak ve ranza gibi kavramların mobilya kategorisini oluşturduğunu anlar

WordNet: Viewed as a graph



Kavram		Kavram		Kavram	
Türkçe	İngilizce	Türkçe	İngilizce	Türkçe	İngilizce
Üst Kavram	Hyperonym	Parçanın Bütünü	Holo Portion	Sonucudur	Is Caused By
Alt Kavram	Hyponym	Yaklaşık Karşıt Anlam	Near Antonym	Durumundadır	Be in State
Bölümün Bütünü	Holo Part	Bütünün Parçası	Mero Portion	Durumudur	State of
Bütünün Bölümü	Mero Part	Alt Olay	Sub Event	Yaklaşık Eş Anlamlı	Near Synonym
Üyenin Bütünü	Holo Member	Olaydır	Is Event of		
Bütünün üyesi	Mero Member	Nedenidir	Causes		

Alt Kavram (Alt Anlamlılık) - (Hyponym)

❖ Sözcükler birbirleriyle aynı cins ve türden olma açısından benzerlikler taşırlar

Örnek

'kuş' sözcüğü canlılar ve hayvanlar üst kavramlarının bir alt türüdür.

Serçe, güvercin, çalığıkuşu, bülbül gibi sözcükler de 'kuş' üst kavramının alt örnekleridir.



Eş Anlamlılık - (Hyponym)

- ❖ Farklı okunan ve yazılan ancak aynı kavramsal içeriği ve anlamsal özellikleri olan iki sözcük arasındaki ilişkidir

Örnek

'ak ve beyaz'

'yürek ve kalp'

Sözcükler belli bir bağlamda birbirlerinin yerine kullanılabilirler mi?

- ❖ **Eşdizimsellik** anlam eş anlamlılığını etkileyen ve sözcüklerin birbiri yerine geçmesini engelleyen özelliklerdendir

Örnek

'anamın ak sütü' 'anamın beyaz sütü'

'kalpsiz adam'

Gerçek Eş Anlamlılık Çok Azdır

Eş anlamlılığın üç derecesi vardır:

❖ **Tam eş anlamlılık:** Anlamların tam benzerliğidir
Sözcükler birbirlerinin yerine kullanılabilir
Üzüntü, elem, keder, dert

❖ **Önermesel eşanlamlılık:** Gerektirme ile tanımlanır.
Eşanlamlı sözcükler her zaman birbirinin yerine kullanılamaz

Ali kaplumbağayı gördü
Ali tosbağayı gördü
Ali araştırmacıları kaplumbağaları incelerken gördü
Ali araştırmacıları tosbağaları incelerken gördü

Hepatit ve sarılık birbirinin yerine kullanılabilir.
“Hepatit” daha özel ve dar bir çevrede kullanılır aynı zamanda daha profesyonel bir kullanıma sahiptir.
“Sarılık” daha yaygındır ve günlük bir terimdir.

❖ **Yarı eşanlamlılık:** Anlamaları birbirleriyle tamamen aynı değildir
yüzünden
-den dolayı
nedeniyle
sebebiyle
dolayısıyla

Eş anlamlılık

İlgili sözcük çiftlerinin farklı yönleri olabilir

- Dilin farklı lehçelerine ait olabilirler
günebakan, ayçiçeği, dövranber
- Bazı sözcükler benzer durumlarda farklı biçimlerde kullanılırlar
adam, herif
- Bazı sözcüklerin duygusal anlamları vardır
politikacı, devlet adamı
- Bazı sözcükler sadece belli bir bağlamda kullanılırlar
(alnım) ak, beyaz ten

Karşıt Anlamlılık- (Antonym)

- ❖ İki sözcük arasında anlamsal olarak birbirinin karşıtı olma durumudur
 - kısa-uzun
 - ters-düz
 - üst-alt
- ❖ Temel olarak iki tür karşıtlık vardır
 - Derecelendirilebilen/derecelendirilemeyen
 - Ters/bakışımli

Derecelendirilebilen karşıtlık İki sözcük arasında belli bir ölçütün iki uç noktasında olma ilişkisi vardır

Uzun/kısa, genç/yaşlı (“daha ” ve “çok” sözcükleriyle kullanılabilir)

- Olumsuzları karşıtlık taşımazlar
 - o «kısa değil», «uzun»
 - o «genç değil», «yaşlı» anlamına gelmez

Derecelendirilemeyen karşıtlık İki kavram arasındaki karşıtlığın, bölgesel bir sınır gibi, tam olduğu durumda ortaya çıkar

ölü/canlı, bekar/evli

- Olumsuz biçimleri birbirleri ile eşit anlam taşır
«ölü değil», «canlı»
«bekar değil», «evli» anlamına gelir

Ters karşıtlık (yön gösteren) Sözcüklerden biri bir yönde diğeri ise onun karşıtı yönünde bir hareketi gösteriyorsa
sağ/sol, itmek/çekmek

Bakışimli karşıtlık İki sözcük arasında birbirinin karşıtını içerme ilişkisi vardır
öğretmen/öğrenci, doktor/hasta
Ahmet Ali'ye borç verdi . Ali Ahmet'ten borç aldı

Eş Sosluluk - (Homophony)

- ❖ Kimi sözcükler aynı ses ve yazı biçimde olmalarına rağmen birbirlerinden tümüyle bağımsız kavramsal içerik ve anlamsal özelliklere sahiptirler

Örnek

'yüz'

yüz (I) a. 1. Doksan dokuzdan sonra gelen sayının adı

yüz (II) a. 1. Başta, alın, göz, burun, ağız, yanak ve çenenin bulunduğu ön bölüm, sima, çehre, surat

yüz (III) e. 1. Kol, bacak, yüzgeç vb. organların özel hareketleriyle su yüzeyinde veya su içinde ilerlemek, durmak

yüz (IV) e. (-i) 1. Derisini çıkarmak, derisini soymak

Eş Anlamlılık - (Polysemy)

- ❖ Bir göstergenin yansıttığı temel anlamın yanı sıra yeni yeni kavramları da anlatır durumda olması
- ❖ Eğer bir sözcükten benzetme yoluyla başka anlamlar türetilirse ortaya çıkan sözcükler arasındaki ilişki çok anlamlılık ilişkisidir

Örnek

YÜZ insan yüzü, yorgan yüzü

Baş

- ✓ İnsan ve hayvanlarda beyin, göz, kulak, burun, ağız gibi organları kapsayan vücudun üst bölümünde bulunan organ (Salona girerken başının üzerine çiçekleri attı)
- ✓ Bir topluluğu yöneten kimse (devletin başı)
- ✓ Başlangıç (satır başı)
- ✓ Temel (her şeyin başı sağlıktır)
- ✓ Arazide en yüksek nokta, tepe (yokuşun başı)
- ✓ Yiyeceklerde tane (üç baş soğan)

Eğretileme - (Metaphor)

❖ Bir kavramı diğer bir kavram üzerinden anlamak/anlatmak için kullanılır

- Eğer bir kavram diğeri üzerinden benzetme ile ifade edilirse iki kavram arasındaki anlamsal ilişki eğretilemedir

Örnek

Borç içinde yüzmek

Borç => deniz

Borçlu olmanın miktarı => denizin derinliği

Denizin içinde yüzmek => çok fazla borcun içinde olmak

Eğretileme

Kan beynine sıçramak

İçinde öfke biriktirmek

Öfke kusmak

Burnundan ateşler saçmak

Tepesinin taşı atmak

Öfkeyle dolmak

Parça-bütün İlişkisi (Part-whole)

- ❖ Bir bütünün parçası ile bütün arasındaki ilişkiyi anlatır
 - Sözcükler bir bütün ile onun parçalarını gösteriyorsa, kurdukları anlamsal ilişkidir

Örnek

masa-çekmece

kitap-sayfa

ev-çatı

Rakamlarla Wordnet

Category	Unique Forms	# of Senses
Noun	117,097	145,104
Verb	11,488	24,890
Adjective	22,141	31,302
Adverb	4,601	5,720

WordNet Noun Relations

Relation	Definition	Example
Hypernym	From concepts to superordinates	<i>breakfast</i> → <i>meal</i>
Hyponym	From concepts to subtypes	<i>meal</i> → <i>lunch</i>
Has-Member	From groups to their members	<i>faculty</i> → <i>professor</i>
Member-Of	From members to their groups	<i>copilot</i> → <i>crew</i>
Has-Part	From wholes to parts	<i>table</i> → <i>leg</i>
Part-Of	From parts to wholes	<i>course</i> → <i>meal</i>
Antonym	Opposites	<i>leader</i> → <i>follower</i>

Hyponym (alt kavram)/Hypernym (üst kavram) (Is-A relationships)
Bütün ilişkilerin %60'ı

Verb Relations in WordNet

Relation	Definition	Example
Hypernym	From events to superordinate events	<i>fly</i> → <i>travel</i>
Troponym	From events to their subtypes	<i>walk</i> → <i>stroll</i>
Entails	From events to the events they entail	<i>snore</i> → <i>sleep</i>
Antonym	Opposites	<i>increase</i> ↔ <i>decrease</i>

Adjective & Adverb Relations in WordNet

Relation	Definition	Example
Antonym	Opposite	<i>heavy</i> ↔ <i>light</i>
Adverb	Opposite	<i>quickly</i> ↔ <i>slowly</i>

Troponym: bir şeyin nasıl yapıldığı çeşidini gösterir

Entails: bir şeyin nasıl yapıldığını gösterir

❖ Anlamsal Kelime Ağı Üzerinde Benzerlik Ölçen Yöntemler

İki sözcüğün anlam açısından birbirine ne kadar benzediği ve aralarındaki ilişkinin çıkarılmasına dayanır

- **Sözcük benzerliği** : İki sözcüğün eş veya yakın anlamlı olmasının ölçüsüdür
- **Sözcük ilişkisi** : Sözcükler arasındaki ilişkiyi gösterir. Zıt anlamlı iki sözcük arasında ilişki derecesi (akıllı-aptal) yüksektir ancak, aralarında benzerlik olmayabilir. Buna karşılık iki sözcük birbiri ile ilişkili olabilir ama benzerlik olmayabilir (benzin-araba)

Wordnet'le anlamsal benzerlik bulma

➤ İki temel yaklaşım

❖ Bağ sayma

- Taksonomi yeterli

❖ Ortak / Müşterek Bilgi (Mutual Information)

- Taksonomi ve corpus kullanır

Leacock & Chodorow (1998)

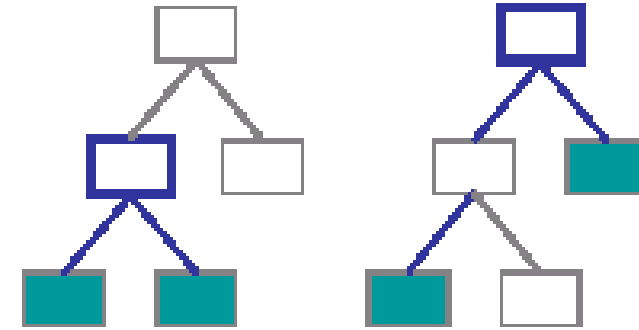
$$sim_{LC}(c_1, c_2) = -\log \frac{len(c_1, c_2)}{2L}$$

- ***len(c1,c2)*** iki synset arasındaki en kısa yolun uzunluğu (*benzerlik değeriyle ters orantılı*)
- ***L***, tüm taksonominin derinliği

Wu & Palmer (1994)

$$sim_{Wu\&Palmer}(c_1, c_2) = \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3}$$

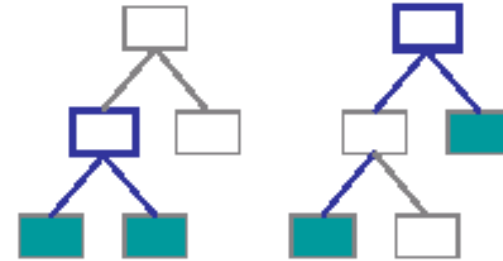
- N_1 ve N_2 , en yakın ortak üst synset'lerine iki synsetin IS-A bağlarıyla uzaklıkları (*benzerlik değeriyle ters orantılı*)
- N_3 , en yakın ortak üst synset'in kök synset'e IS-A bağlarıyla uzaklığı (*büyüklüğü ortak synset'in spesifikliğini gösterir*)



Jiang-Conrath (1997)- Lin (1998)

$$\begin{aligned} \text{sim}_{\mathcal{JC}}(c_1, c_2) &= \frac{2 \log(p(\text{lcs}(c_1, c_2)))}{\log(p(c_1)) + \log(p(c_2))} \\ \text{sim}_L(c_1, c_2) &= \frac{2 * \log(p(\text{lcs}(c_1, c_2)))}{\log(p(c_1)) + \log(p(c_2))} \end{aligned}$$

$\text{lcs}(c_1, c_2)$ en yakın ortak üst synset



- A, her iki kavramı da içeren en spesifik kavramı kullanır
(İki kavramın beraber geçtiği doküman sayısına benzer)
- B, iki kavramdan herhangi birini içeren doküman sayısına benzer

Benzerlik Metotlarının Karşılaştırılması

- ❑ Bütün metotlar İngilizce 38 kelime çiftine uygulanmış
- ❑ Bulunan benzerlik değerlerinin, insan yargılarıyla olan korelasyonları yandaki tabloda verilmiştir

<i>Method</i>	<i>Type</i>	<i>Correlation</i>
Wu & Palmer 1994	Edge Counting	0.74
Li 2003	Edge Counting	0.82
Leacock & Chodorow 1998	Edge Counting	0.82
Resnik 1999	Info. Content	0.79
Lin 1998	Info. Content	0.82
Lord 2003	Info. Content	0.79
Jiang & Conrath 1998	Info. Content	0.83
Tversky 1977	Feature Based	0.73
Adapted Lesk 2002	Feature Based	0.37*
Rodriguez 2003	Hybrid	0.71

Kaynaklar

- Steve Vincent, Text Extraction, Similarity and WordNet
- http://www.ise.gmu.edu/~carlotta/teaching/INFS-795-s06/readings/Similarity_in_WordNet.pdf
- <http://www.cs.utah.edu/~sidd/documents/msthesis03ppt.pdf>
- <https://starlangyazilim.com/kelime-anlam-belirsizligi-giderme-nedir/>
- ‘Türkçe Doğal Dil İşleme’, Eşref Adalı, Akçağ, 2020
- <https://acikders.ankara.edu.tr/>

Söylem (Discourse) Analiz



- Söylem Analizinde analiz edilecek olan metin içerisinde birden fazla cümle yer alır.
- Metin: haber makalesi, hikaye veya kurgusal bir içerik olabilir.

ANAPHORA ?

I went to see my grandfather at the hospital. The old man has been there for weeks. He had surgery a few days ago.

- Referans veren ifadeleri ve öncül isimleri (*antecedents*) bulmak

I went to see **my grandfather** at the hospital. *The old man* has been there for weeks. *He* had surgery a few days ago.

- Hem tek cümleli hem de çok cümleli metinlerle ilgili sorunlar
- Nasıl bir modele ihtiyaç vardır?

Coreference

- Anaphore kullanımına bir örnek
 - John saw Mary in the park. As every morning, she was walking her dog.
- «She» neye referans eder? (she → isim, canlı)

- Adaylar

- John
 - Mary
 - The park
 - Every morning
 - Her dog
- } isim (John → canlı ama dişi değil)
- } isim (the park → cansız)
- } isim (every morning → cansız)
- } isim (her dog → canlı, she referans edebilir.)

Ancak, cümlenin yapısı gereği olamaz.

She, Mary referans etmektedir.

MUC-7 Coreference Task

<COREF ID="6" TYPE="IDENT" REF="5" MIN="Aeroflot">The Russian airline Aeroflot</COREF> has been hit with <COREF ID="19">a writ</COREF> for loss and damages, filed in <COREF ID="15">Hong Kong</COREF> by <COREF ID="7" TYPE="IDENT" REF="4" MIN="families">the families of <COREF ID="22" MIN="passengers">seven passengers killed in <COREF ID="8" TYPE="IDENT" REF="9" MIN="crash">an air crash</COREF></COREF></COREF>.

All 75 people on board <COREF ID="12" MIN="Airbus">the <COREF ID="10" TYPE="IDENT" REF="6">Aeroflot</COREF> <COREF ID="25">Airbus</COREF> </COREF> died when <COREF ID="11" TYPE="IDENT" REF="12">it</COREF> ploughed into a Siberian mountain in March 1994.

MUC Yarışmasında coreference çıkarılması için bir görev verilmişti.

Bir uçak kazasında hayatını kaybeden yedi yolcunun aileleri tarafından Hong Kong'da açılan tazminat davası Rus şirketini vurdu. Mart 1994'te Sibiry'a'da bir dağa çarpan Aeroflot Airbus uçağındaki 75 kişinin tamamı hayatını kaybetmişti.

MUC-7 Coreference Task

<COREF ID="6" TYPE="IDENT" REF="5" MIN="Aeroflot">The Russian airline Aeroflot</COREF> has been hit with <COREF ID="19">a writ</COREF> for loss and damages, filed in <COREF ID="15">Hong Kong</COREF> by <COREF ID="7" TYPE="IDENT" REF="4" MIN="families">the families of <COREF ID="22" MIN="passengers">seven passengers killed in <COREF ID="8" TYPE="IDENT" REF="9" MIN="crash">an air crash</COREF></COREF></COREF>.

All 75 people on board <COREF ID="12" MIN="Airbus">the <COREF ID="10" TYPE="IDENT" REF="6">Aeroflot</COREF> <COREF ID="25">Airbus</COREF> </COREF> died when <COREF ID="11" TYPE="IDENT" REF="12">it</COREF> ploughed into a Siberian mountain in March 1994.

Coreference Resolution (Zamir Çözümleme)

- **Uyum kısıtlamaları (Agreement constraints)**
 - Cinsiyet, tekil/çoğul, canlı/cansız
- **Sözdizimsel kısıtlamalar (Syntactic constraints)**
 - Aynı yapıdaki cümleler-paraellik
- **Cümle Sıralaması (Sentence ordering)**
 - Tekrarlılık

RAP Algoritması_1 – Resolution of Anaphora Procedure

Sentence recency	100
Subject emphasis	80
Existential emphasis (“there is ...”)	70
Accusative (direct object) emphasis	50
Indirect object emphasis	40
Non-adverbial emphasis	50
Head noun emphasis	80

Sürekliliğin işlenmesi

- her cümle işlendikten sonra tekrarlılık yarıya indirilir.

- **Examples:**

- An Acura Integra is parked in the lot. (subject)
- There is an Acura Integra parked in the lot. (existential predicate nominal)
- John parked an Acura Integra in the lot. (object)
- John gave Susan an Acura Integra. (indirect object)
- In his Acura Integra, John showed Susan his new CD player. (demarcated adverbial PP)

RAP Algoritması_2 – Resolution of Anaphora Procedure

- Potansiyel referansları belirleyin (en fazla dört cümle geriye)
- Zamirle, sayı veya cinsiyet açısından uyuşmayan potansiyel referansları kaldırın.
- Rol paralelliği (+35) veya cataphor (-175) için geçerli herhangi bir değeri ekleyerek referansın toplam puan değerini hesaplayın.
- En yüksek belirginlik değerine sahip referansı seçin.
- Eşitlik durumunda, sıralama konumu açısından en yakın referansı seçin.
- Yeni bir cümleye geçerken, listedeki mevcut varlıklar için tüm puanları yarıya indirin.

Cataphor, bir ifadeye atıfta bulunmadan önce o ifadenin gelecekte kullanılacağına işaret eder. Başka bir deyişle cataphor, bir ifadenin daha sonra belirtilen bir ifadeye atıfta bulunmasını sağlar.

- **Örneğin:** "Ona güvenebilirim. O, her zaman söylediği gibi yapar" cümlesindeki "o", daha sonraki ifade "her zaman söylediği gibi yapar" ifadesine bir cataphor örneğidir.

- John saw a beautiful Acura Integra at the dealership last week. He showed it to Bill. He bought it.

	Rec	Subj	Exist	Obj	Ind Obj	Non Adv	Head N	Total
John	100	80				50	80	310
Integra	100			50		50	80	280
dealership	100					50	80	230

Sentence recency	100
Subject emphasis	80
Existential emphasis (“there is ...”)	70
Accusative (direct object) emphasis	50
Indirect object emphasis	40
Non-adverbial emphasis	50
Head noun emphasis	80

- Birinci cümlede zamiri referans edecek 4 adet isim var : John, Integra, dealership
- İkinci cümlede «He» için 3 aday var John, Integra, dealership
- «John saw a beautiful Acura Integra at the dealership last week» aynı cümlede olduğu için hepsi 100 puan alır
- «John» özne olduğu için 80 puan alır
- «Integra» nesne olduğu için 50 puan alır
- Üçü de bu cümlede zarf olmadığından 50 puan alır
- Başlık ismi de olmadıklarından 80 puan alırlar
- Toplam puanlar sırası ile 310, 280 ve 230.
- En yüksek puan John olduğu için bu bize «He» için en olası adayın John olduğunu söyler.

John saw a beautiful Acura Integra at the dealership last week. He showed it to Bill. He bought it.

	Rec	Subj	Exist	Obj	Ind Obj	Non Adv	Head N	Total
John	100	80				50	80	310
Integra	100			50		50	80	280
dealership	100					50	80	230

He showed it to Bill.

- «He» için en olası referans **John** olduğu için (John, he ve Integra, it aynı gruba dahil edilerek toplam değer alınır.
 $310 + 155 = 465$ ve $280 + 140 = 420$

Referent	Phrases	Value
John	{John, he ¹ }	465
Integra	{a beautiful Acura Integra}	140
dealership	{the dealership}	115

Referent	Phrases	Value
John	{John, he ¹ }	465
Integra	{a beautiful Acura Integra, it}	420
dealership	{the dealership}	115

Referent	Phrases	Value
John	{John, he ¹ }	465
Integra	{a beautiful Acura Integra, it}	420
Bill	{Bill}	270
dealership	{the dealership}	115

- Bill için puan $\text{Rec} + \text{Ind Obj} + \text{Non Adv} + \text{Head N} = 270$
- En yüksek puan **John** olduğu için bu bize yine «He» için en olası adayın **John** olduğunu söyler.

He bought it.

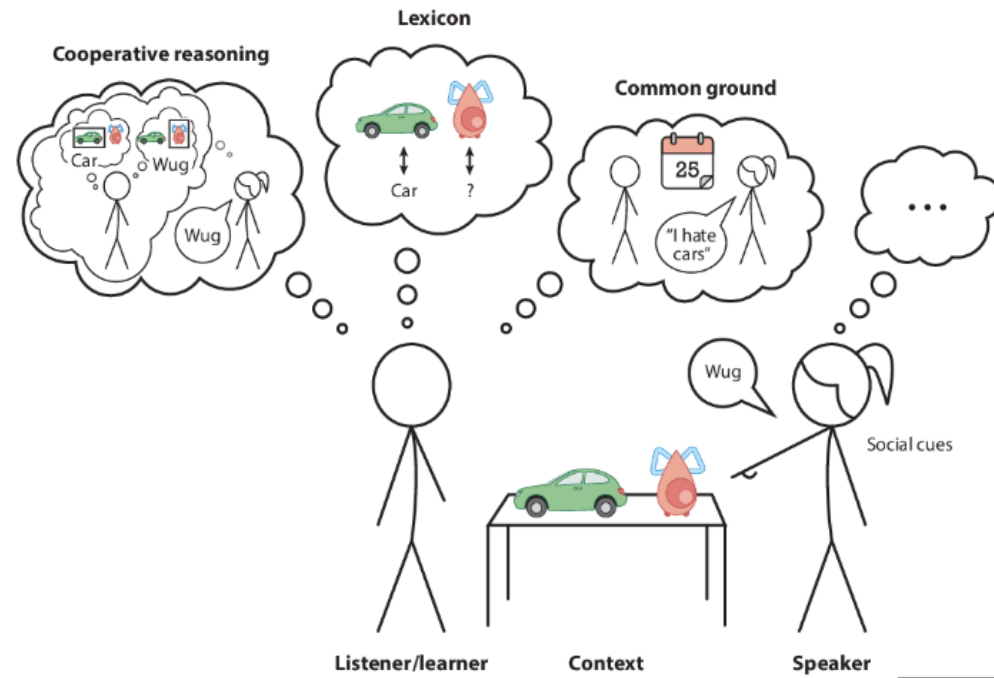
Referent	Phrases	Value
John	{John, he ¹ }	232.5
Integra	{a beautiful Acura Integra, it ¹ }	210
Bill	{Bill}	135
dealership	{the dealership}	57.5

- Üçüncü cümleye geçeceğimiz için her adayın puanı yarıya düşürülür.
- En yüksek puan **John** olduğu için «He» referans olan «John»

- İkinci cümleye geçeceğimiz için her adayın puanı yarıya düşürülür.

Referent	Phrases	Value
John	{John}	155
Integra	{a beautiful Acura Integra}	140
dealership	{the dealership}	115

Kullanım Bilimi Analizi (Pragmatic Analysis)



NLP'de Pragmatik Belirsizlik nedir?

- Aynı cümlenin farklı durumlarda farklı anlamlara sahip olabileceği gerçeğine atıfta bulunmak
- Bir konuşmada konuşan veya yazarın söyledikleri ile dinleyen veya okuyanların anladığını inceleyen bilim dalı
- Aynı cümlenin birden çok yorumunun olması

Pragmatik, Anlam Bilim ile nasıl farklılık gösterir?

NLP'deki Pragmatik, dilin anlamını anlar ve bağlamı da göz önünde bulundurur. Anlam bilimi ise sadece cümledeki kelimelerinin gerçek anlamını düşünür.

Pragmatiğin ana avantajı nedir?

Bilgisayarın, cümlelerin gerçek dünya anlamını anlamasını sağlar. Bağlamı, alaycılığı vb. dikkate alır.

Konuşmalarda gerçekte istenen nedir?

A: Şurada bir çay bahçesi var. (gidelim mi?)

B: Hayır. (Ben oraya gitmek istemiyorum.)

A: Neden? (Neden oraya gitmek istemiyorsun?)

B: Yorgunum. (Gitmek istemiyorum çünkü yorgunum.)

Bağlam 4 başlık altında incelenir.

1- Düz Bağlam

Konuşan ve dinleyenlerin konuşmaya ilişkin temel bilgileri bildikleri, bu nedenle kullanılan sözcük ve tümcelerin anlamlarında belirsizliğin olmadığı bağlam türüdür.

2- Dilsel Bağlam

Bir kelimenin birden fazla anlamı olabilir. Kelimenin anlamı kullanıldığı yere göre yani bağlama göre değişir.

Çekmecenin gözünde mendillerim vardı.

Emel'in gözü yaşardı.

Göze geldi.

3- Fiziksel Bağlam

Konuşmanın bağlamı konuşmanın yapıldığı yere bağlıdır.

A: Özgür'ü gördün mü?

B: Beyaz spor arabayı parkta gördüm.

4- Sosyal Bağlam

Konuşma ve yazının bağlamı bulunulan kültüre bağlıdır.

Türkiye de okumaya gelen birinin garsonluk yapması...

Garson: Bir bardak çay daha ister misiniz?

Türk : Seni durduran yok.

Garson : Beni durdurmak mı istediniz?

Türk : Getir şu çayı.

PRAGMATICS

Pragmatics is the study of how context influences how we interpret and make meaning of communication. It is often described as the study of "language in use".

OVERVIEW

Sometimes, the literal meaning of what is said isn't the implied meaning. Pragmatics goes beyond the literal meaning of words and phrases to understand the intended message based on factors like the social situation, relationship between the speakers, cultural context, situational context, and the way the words are said.

EXAMPLES

- Use of metaphor
- Use of irony
- Use of sarcasm
- Use of deixis
- Use of euphemisms
- Jokes and humor
- Use of silence / mis en scene
- Use of hyperbole
- Use of tautologies

Pragmatik, kelimelerin ve ifadelerin kelime anlamının ötesine geçer ve amaçlanan mesajı anlamak: Sosyal durum, konuşmacılar arasındaki ilişki, kültürel bağlam, durumsal bağlam, kelimelerin söylenme şekli, vd. bağlıdır.

- ❖ **Sarcasm (alay etme)** : Şiddetli yağmur yağıyorsa ve birisi "Hava harika, değil mi? diyorsa, aslında havanın güzel olduğunu kastetmiyordur.
- ❖ **Irony** : "Yarın sınav var, ne güzel!".
- ❖ **Implication (imalı konuşma)** : "Bugün hava öyle güzel ki, neredeyse işe gitmek istemiyorum."
- ❖ **Deictic-Deixis (Anlamları kullanıldıkları bağlama bağlı olan kelimeler)** : "Saat beşte orada buluşuruz"
Bağlam olmadan, "orada "nın neyi ifade ettiğini veya "beş "in sabah mı yoksa akşam beşi mi kastettiğini bilemeyiz. Belirli bir kafeden bahsediyorsak ve genellikle akşamları buluşuyorsak, bu bilginin anlamını doğru bir şekilde yorumlayabiliriz.
- ❖ **Politeness (kibar konuşma)** : Pragmatik anlayış, sosyal bağlama, konuşmacılar arasındaki ilişkiye, kültürel normlara ve daha fazlasına bağlı olarak değişebilen dildeki nezaket derecelerini de içerir. Resmi bir ortamda veya bir amirle konuşurken, "Raporu bana ver" demek yerine, "Lütfen raporu bana uzatır mısınız?" denebilir.
- ❖ **Metaphors (canlı bir imaj yaratmak için bir şeyi başka bir şeyin terimleriyle ifade etmek)** : "Zaman bir hırsızdır"
- ❖ **Indirect Speech Acts (dolaylı konuşma eylemi)** : Bir yemek masasında birisi "Tuzu uzatabilir misin?" derse, gerçekten o kişinin fiziksel olarak tuzu uzatabilecek durumda olup olmadığını sormuyordur, bunu yapmasını talep ediyordur.
- ❖ **Euphemisms (hoş olmayan, hassas veya tabu olabilecek bir şeyi ifade etmenin daha yumuşak yolu)** : "Öldü" yerine "Vefat etti" deniyorsa, ölüm hakkında daha az sert bir şekilde konuşmak isteniyordur.
- ❖ **Jokes And Humor (şaka ve mizah)**: Kültürlere bağlı olarak değişir.

- ❖ **Cultural References** : Konuşmacılar genellikle fikirlerini daha etkili bir şekilde iletmek için kültürel referans kullanır. "O gerçek bir Sherlock Holmes" denirse, o kişinin çok gözlemci olduğunu veya gizemleri çözmede iyi olduğu ima edilmiş olur. Ancak, bunu anlamak için Sherlock Holmes karakteri ve nitelikleri hakkında bilgi sahibi olmak gerekir.
- ❖ **Hedging (bir ifadenin etkisini azaltmak, genellikle yüzleşmekten kaçınmak veya bir eleştiriyi yumuşatmak)** : Birisi "Yanılıyor olabilirim ama bence argümanınız hatalı" derse, "yanılıyor olabilirim" ifadesi bir korunma işlevi görür. Konuşmacının eleştirisinin iddiasını azaltır ve hata olasılığını kabul eder.
- ❖ **Phatic Communication (belirli bilgileri iletmekten ziyade, yakınlık kurmak gibi sosyal görevleri yerine getirmek için kullanılır)** : Tanıştığınızda biri size "Nasılsın?" diye sorarsa, genellikle iyi halinizin ayrıntılı bir açıklamasını istemez. Sosyal bir etkileşim başlatmanın geleneksel bir yoludur. Beklenen yanıt "İyiyim, teşekkür ederim" olmalıdır.
- ❖ **Use Of Silence (söylenmeyi anlamayı içerir)** : Sessizlik, bağlama bağlı olarak çok şey anlatabilir. Bir toplantı sırasında birisi bir fikir önerdiğinde uzun bir duraklama olursa anlaşmazlık, gariplik, şaşkınlık anlamına gelebilir.
- ❖ **Hyperbole (mübalağa)** : "O kadar açım ki bir atı yiyebilirim"