

Doğal Dil İşlemeye Giriş

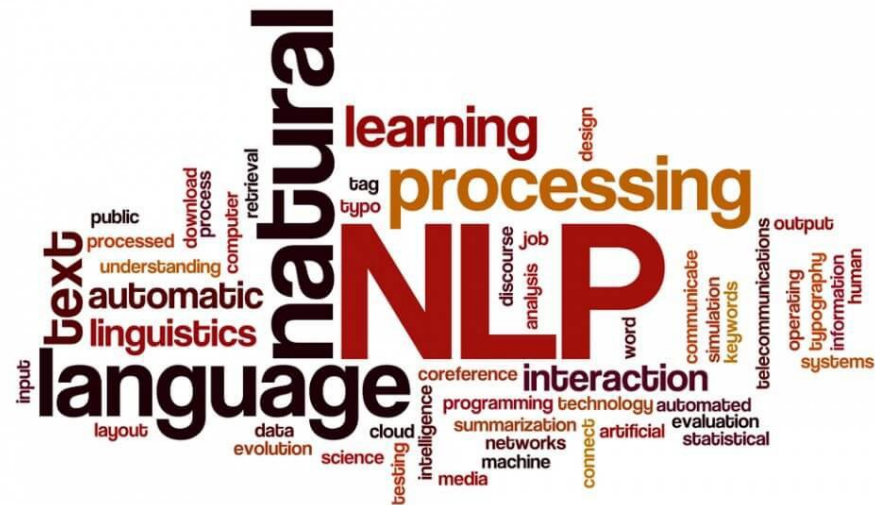
(Introduction to Natural Language Processing)

Necmettin Erbakan Üniversitesi
Bilgisayar Mühendisliği Bölümü
22-23 Güz Dönemi

Doğal Dil İşlemeye Giriş

- Dersin temel amacı;
 - Doğal dilin (natural language) anlaşılması
 - DDİ tekniklerinin öğrenilmesi
 - Uygulama alanları ile tanışılması
 - Mümkün olabilen uygulamaların gerçekleştirilmesi

- Yarıyıl: 7
- Tür: Seçmeli
- T/U: 3+0
- Kredi: 3
- AKTS: 6



Kaynaklar

- Daniel Jurafsky and James H. Martin, «Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition»
- James Allen, «Natural Language Understanding»
- Christopher D. Manning and Hinrich Schutze, «Foundations of Statistical Natural Language Processing»
- Eugene Charniak, «Statistical Language Learning»

Dersin İşlenişi

- **Eğitmen**
 - Arş. Gör. Dr. Sinem ÇINAROĞLU
 - soren@erbakan.edu.tr
- **Bilgi Kaynakları Zinciri**
 - Ders Kitabı → Çevrimiçi Kaynaklar → Arkadaşlar → Dersin Asistanı 😊
→ Öğretmen
- Kaynaklar ve duyurular Google Classroom üzerinden yayınlanacaktır.

Dersin İşlenişı

- Classroom kaydı: Kaynaklar ve duyurular için.

Özelleştir

22-23_Doğal Dil İşlemeye Giriş

Bilgisayar Mühendisliği

Sınıf kodu:

jq6xdxr



Sınıfınıza duyuru yapın



Teslim tarihi yaklaşan ödevler

Teslim tarihi yaklaşan ödev/soru yok

[Tümünü görüntüle](#)



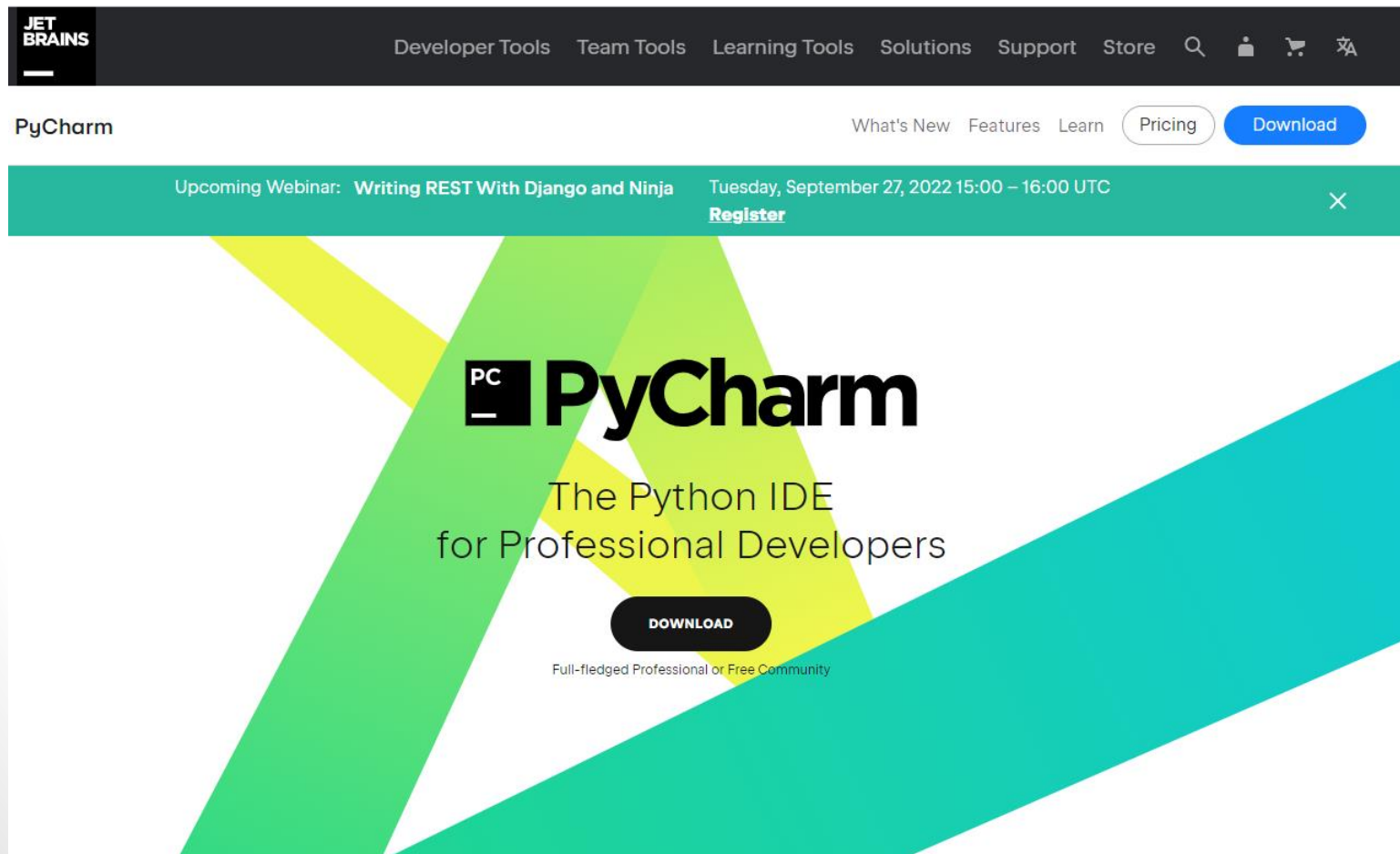
Burası sınıfınızla konuşabileceğiniz yerdir

Duyuru paylaşmak, ödev yayınlamak ve öğrenci sorularına yanıt vermek için akışı kullanın

 Akış ayarları

Dersin İşlenişi

- Uygulamalar için PyCharm Community Ed. 2019:
<https://www.jetbrains.com/pycharm/>



The screenshot shows the PyCharm website homepage. At the top is a dark navigation bar with the 'JET BRAINS' logo on the left and links for 'Developer Tools', 'Team Tools', 'Learning Tools', 'Solutions', 'Support', and 'Store' on the right. Below this is a white header with the 'PyCharm' name on the left and links for 'What's New', 'Features', 'Learn', 'Pricing', and a blue 'Download' button on the right. A teal banner below the header announces an 'Upcoming Webinar: Writing REST With Django and Ninja' on 'Tuesday, September 27, 2022 15:00 – 16:00 UTC' with a 'Register' link and a close button. The main content area features a large graphic with green and yellow diagonal stripes. In the center, the 'PyCharm' logo is displayed above the text 'The Python IDE for Professional Developers'. Below this is a black 'DOWNLOAD' button and the text 'Full-fledged Professional or Free Community'.

JET BRAINS

Developer Tools Team Tools Learning Tools Solutions Support Store

PyCharm

What's New Features Learn Pricing Download

Upcoming Webinar: Writing REST With Django and Ninja Tuesday, September 27, 2022 15:00 – 16:00 UTC Register

PC PyCharm

The Python IDE
for Professional Developers

DOWNLOAD

Full-fledged Professional or Free Community

Performans Değerlendirme ve Notlandırma

- ▶ Ara Sınav Notu ~ %40
- ▶ Final Notu ~ %60
 - ▶ Final Sınavı ~ %60 ?
 - ▶ Proje ~ %40 ?
 - ▶ Ara Sınav tarihine kadar proje önerisi verilmelidir.
 - ▶ Projeler bireysel olacaktır.
 - ▶ Son iki hafta proje sunumu yapılacaktır.
- ▶ Ekstralar
 - ▶ Proje özgünlüğü

Performans Değerlendirme ve Notlandırma

- ▶ Birbirinize yardım edebilirsiniz, ancak projelerinizi kopyalamayın.
- ▶ *Kopya çekip yakalanmasanız bile, sınavlarda başarısız olursunuz!!!*
- ▶ *Kopya çekip yakalanırsanız, kredinizi ve güvenilirliğinizi kaybedersiniz, ve sınavlarda yine başarısız olursunuz!!!*
- ▶ *En kolay çözüm:*
 - ▶ *Yeterince sıkı çalışın, böylece hile yapmanıza, başkalarının çalışmalarını kopyalamanıza gerek kalmaz.*

İçerik

- Doğal Dil İşlemeye Genel Bakış
- Dilbiliminin Esasları
- Gramer ve Diller
- Düzenli Diller
- Dil Modelleri N-Grams
- Biçimbilimsel Analiz (Morphological Analysis)
- Sözdizimsel Analiz (Syntactic Analysis)- Sözcüksel Analiz (Lexical Analysis-Part of Speech Tagging/POS)
- Anlam Bilgisi (Semantic)- Söylem (Discourse) Bilgisi-Edim (Pragmatic) Bilgisi

İçerik

- Makine Öğrenmesi (Machine Learning)
- Metin Sınıflandırma (Text Classification)
- Bilgiye Erişim Sistemleri (Information Retrieval)
- Bilgi Çıkarımı (Information Extraction)
- Kelime Anlamları (Word Semantic)
- Kelime Gömmeleri (Word Embedding)
- Duygu Analizi (Sentiment Analizi)
- Soru Cevaplama Sistemleri (Question Answering)
- Proje ve Proje Sunumları

Dil Nedir?

- Chomsky'e göre:

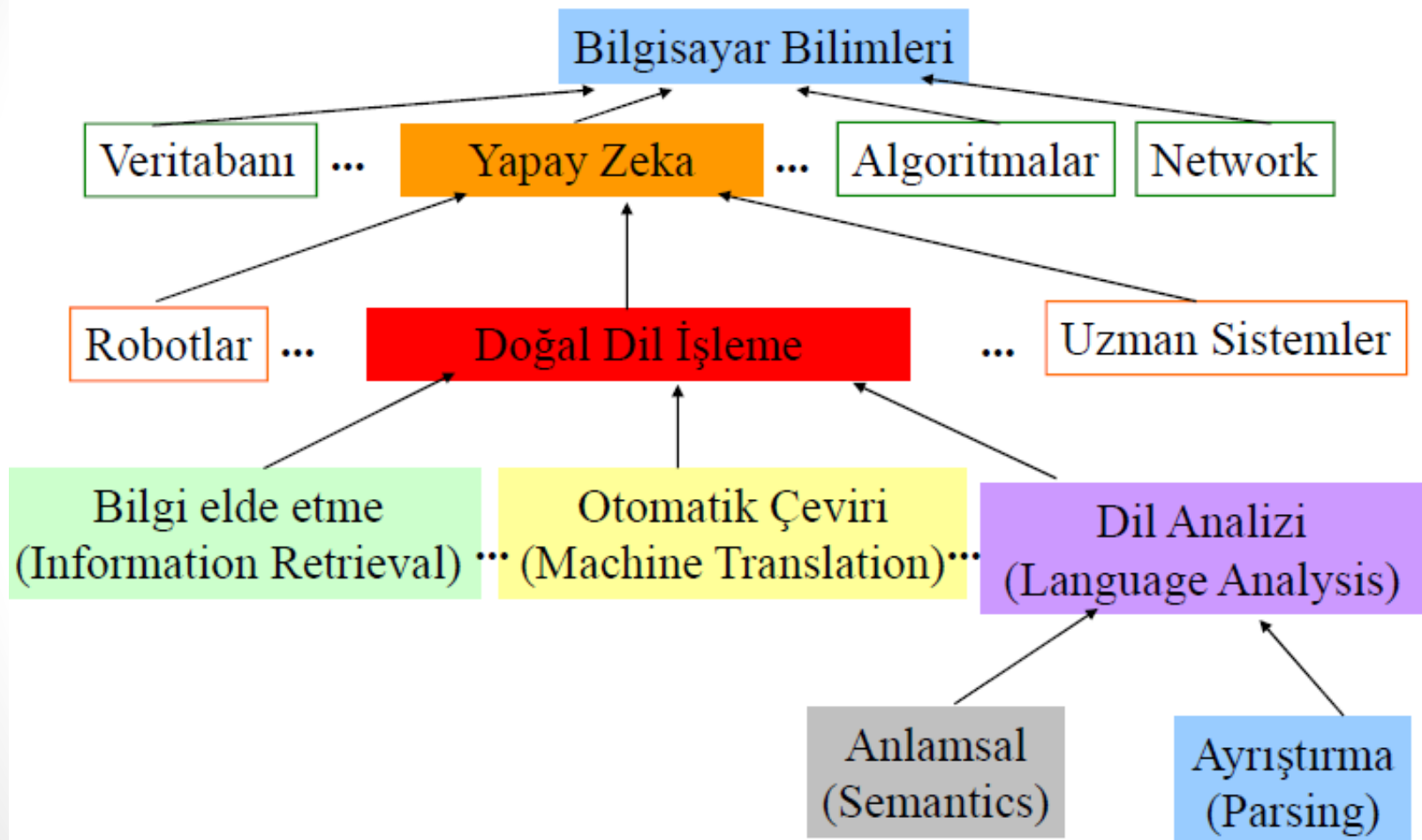
“Sözcük ve cümle birimleri aracılığıyla, düşünceyi konuşmayla ilişkilendiren çok seviyeli bir sistemdir”

- İnsanlar arasında bir iletişim aracıdır.
- Dilin bilgisayar ortamında modeli oluşturulursa iletişim için önemli bir araç elde edilmiş olur.

Doğal Dil İşleme Nedir?

- Temel amacı;
 - Doğal dilin **derinlemesine** anlaşılması
 - Karakter dizisi veya anahtar kelime eşleşme olarak değil
 - İnsan-makine iletişimi
 - İnsan-insan iletişimi
 - Metin veya konuşmanın işlenmesi
- Yapay Zeka & Dil Bilimi

DDİ'nin Bilgisayar Bilimindeki Yeri



Doğal Dil İşleme

- Doğal dil alanındaki temel araştırmalar;
 - İşlev ve yapısının daha iyi anlaşılması
 - Bilgisayar ve insanlar arasındaki iletişimi kolaylaştırmak
 - Bir dilden diğerine çeviri yapmak
- Almanya, ABD, Japonya gibi ülkeler
- İngilizce
- Türkçe'de çalışmalar

Doğal Dilin Formları

- Doğal Dil İşleme sisteminin girdisi/çıktısı:
 - Yazılı metin
 - Konuşma
- Yazılı metinlerin işlenmesi için:
 - Sözcüksel (lexical), sözdizimsel (syntactic) ve anlamsal (semantic) bilgi
 - Söylem (discourse) bilgisi, gerçek dünya bilgisi
- Konuşma dilinin işlenmesi için:
 - Yazılı metni işlemek için gerekli her şey
 - Konuşma tanıma ve konuşma sentezi

Doğal Dil İşlemenin Görevleri

- Kelime sayımı (Word count),
 - kelimenin ne olduğunu bilmemizi gerektiren bir DDİ uygulaması
- Satır veya bayt sayısı
- DDİ görevleri;
 - Sohbet Ajanları (Conversational Agents)
 - Makine Çevirisi (Machine Translation)
 - Soru Cevaplama (Question Answering)
 - Bilgi Çıkarımı (Information Extraction)
 - ...

DDİ Görevi: Sohbet ajanları

- «2001: A Space Odyssey» filmi
 - HAL
 - https://youtu.be/oR_e9y-bka0
- Sohbet ajanları ve ya diyalog sistemleri
- Dil bilgisi;
 - Konuşma tanıma ve sentezleme
 - İçerilen İngilizce kelime bilgisi
 - Kelime gruplarının nasıl kümelendiği ve ne anlama geldiği
 - Söylem

DDİ Görevi: Makine Çevirisi

- Tam otomatik
- İnsan çevirmenlere yardım

Enter Source Text:

这不过是一个时间的问题。

Translation from Stanford's *Phrasal*:

This is only a matter of time.

Enter Source Text:

تعرض الرئيس اللبناني اميل لحود لـ حملة عنيفة في مجلس النواب الذي انعقد امس في جلسة تشريعية عادية تحولت الي " محاكمة " لـ رئيس الجمهورية علي موقفه من المحكمة الدولية و " الملاحظات " التي ادلى بها حول هذا الموضوع .

Translate

Clear

Enter Translation:

lebanese

president

suffered

exposed

president emile

before

presented

offer

Done!

DDİ Görevi: Makine Çevirisi

- Google Çeviri



Killing Palestinians and wounding nine in the raids Sector

Nine Palestinians were wounded among civilians in an Israeli air raid in the neighborhood result in the Gaza Strip. This comes immediately after the killing of two prominent Al-Aqsa Martyrs Brigades in the Israeli occupying forces carried out air and infantry forces in the Balata camp in the West Bank.



Bashir meets Fraser, the Security Council will not impose forces Darfur

Is scheduled to meet with Sudanese President Omar al-Bashir Jenday Fraser Assistant Minister for Foreign Affairs of the American attempt to persuade officials in Khartoum, Sudanese Darfur deployment of the nationalities. For his part, US Ambassador to the United Nations that it has no intention of the Security Council to impose its forces in the province.



Rmsfield and Cheney insist on keeping the American forces in Iraq

Called American Defense Minister Donald Rmsfield Americans to show patience on Iraq. I take Vice President Dick Cheney calls Democrats withdrawal of American forces from Iraq link and the possibility of early withdrawal of attacks inside the United States.



Killing civilians and wounding officer suicide attack in Afghanistan

The international force to help establish security (ISAF) killed civilians and the wounding of an officer in an attack against Afghan forces convoy south Atlantic Afghanistan. In the capital Kabul, a hand grenade exploded at the passage of manufacture: French patrol was not reported injuries or damage.



DDİ Görevi: Soru Cevaplama

- IBM Watson
 - 16 Şubat 2011, Jeopardy
 - <https://www.youtube.com/watch?v=P18EdAKuC1U>

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL



Bram Stoker

DDİ Görevi: Bilgi Çıkarımı

Subject: **curriculum meeting**

Date: January 15, 2012

To: Dan Jurafsky

Event: Curriculum mtg
Date: Jan-16-2012
Start: 10:00am
End: 11:30am
Where: Gates 159

Hi Dan, we've now scheduled the curriculum meeting.

It will be in Gates 159 tomorrow from 10:00-11:30.

-Chris

Create new Calendar entry

DDİ Görevi: Bilgi Çıkarımı & Duygu Analizi



Özellikler:

yakınlaştırma

fiyat

boyut ve ağırlık

flash

kullanım kolaylığı

Boyut ve ağırlık

- ✓ nice and compact to carry!
- ✓ since the camera is small and light, I won't mind those heavy, bulky professional cameras either!
- ✗ the camera feels flimsy, is plastic and very light in weight you have to be very delicate in the handling of this camera



Dil Teknolojisi

mostly solved

Spam detection

Let's go to Agra!



Buy V1AGRA ...



Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

making good progress

Sentiment analysis

Best roast chicken in San Francisco!



The waiter ignored us for 20 minutes.



Coreference resolution

Carter told Mubarak he shouldn't run again.

Word sense disambiguation

I need new batteries for my *mouse*.



Parsing

I can see Alcatraz from the window!

Machine translation (MT)

第13届上海国际电影节开幕...



The 13th Shanghai International Film Festival...

Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30



Party
May 27
add

still really hard

Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose



Economy is good

Dialog

Where is Citizen Kane playing in SF?



Castro Theatre at 7:30. Do you want a ticket?



Dil İşlemede Bilgi

Arabic text ... يَجِبُ عَلَى الْإِنْسَانِ أَنْ يَكُونَ أَمِينًا وَصَادِقًا مَعَ نَفْسِهِ وَمَعَ أَهْلِهِ وَجِيرَانِهِ وَأَنْ ...
يَبْذُلَ كُلَّ جُهِدٍ فِي إِعْلَاءِ شَأْنِ الْوَطَنِ وَأَنْ يَعْمَلَ عَلَى مَا

Bir bilgisayar bu **dizgeden** nasıl bir **anlam** çıkarabilir?

- Phonology** - Kelimenin hangi sesleri ilişkilidir? (sesler/harfler/telafuz)
- Morphology** - Kelimenin en küçük biriminin anlamı nedir?
- Syntax** - Kelimeler birbirleriyle nasıl ilişkilidir?
- Semantics** - «Kelimelerin birleştirilmesi» ile oluşan anlam nedir?
- Pragmatics** - «Meta-anlamı» nedir? (sözeylem)
- Discourse** - Büyük bir metin parçasının işlenmesi
- Inference** - Her şeyi anlamlandırmak

Doğal Dil İşleme Zordur!

- Doğal dil insan iletişimini verimli kılmak için tasarlanmıştır.
- Neden?
- Sonuç olarak;
 - dinleyicinin/okuyucunun sahip olduğunu varsaydığımız birçok “yaygın anlam” bilgisini atlıyoruz.
 - dinleyicinin/okuyucunun nasıl çözeceğini bildiğini varsaydığımız birçok belirsizliği saklıyoruz.
- **Belirsizlik** DDİ'deki **HER** adımı zorlaştırır.

Belirsizlik (Ambiguity)

- Bu cümlenin farklı anlamlarını bulun:
 - «I made her duck»
- Buradaki belirsizliğin nedenleri ne olabilir?

Belirsizlik

- Bu cümlenin farklı anlamlarını bulun:
 - «I made her duck»
- Onun için ördek pişirdim
- Ona ait olan ördeği pişirdim
- Sahip olduğu oyuncak bir ördek yaptım.
- Sihir kullandım ve onu ördeğe çevirdim.

Belirsizlik

- Bu cümlelerin farklı anlamlarını bulun:
 - «I made her duck»
- **duck** → biçimbilimsel ve sözdizimsel olarak belirsiz
 - isim veya fiil
- **her** → sözdizimsel olarak belirsiz
 - İsmi -e hali (dative) veya iyelik eki (possessive)
- **make** → anlamsal olarak belirsiz
 - pişirmek veya yaratmak
- **make** → sözdizimsel olarak belirsiz
 - Geçişli (transitive) : doğrudan bir nesne alan (2)
 - Çift geçişli (di-transitive): iki nesne alan (4)

Belirsizlik

- Sesbilim (Phonetics)
 - I mate or duck
 - I'm eight or duck
 - Eye maid; her duck
 - Aye mate, her duck
 - I maid her duck
 - I'm aid her duck
 - I mate her duck
 - I'm ate her duck
 - I'm ate or duck
 - I mate or duck

Türkçe'de Belirsizlik

- «Adamı gördüm.»
 - I saw the man.
 - I saw my island.
 - I visited my island.
 - I bribed the man.
- Biçimbilimsel olarak belirsizlik:
 - ada-m-ı → ada + iyelik + durum eki
 - adam-ı → ada + durum eki
- Anlamsal olarak belirsizlik:
 - gör → to see
 - gör → to visit
 - gör → to bribe

Belirsizliklerin Çözümü

- Farklı seviyelerdeki belirsizlikleri çözmek için modeller ve algoritmalar:
 - **Sözcük türü etiketleme (Part-of Speech Tagging/POS)**
 - duck → isim veya fiil
 - **Sözcük anlam belirsizliği giderme (Word-sense disambiguation/WSD)**
 - make → yaratmak veya pişirmek
 - **Sözcüksel anlam belirsizliği giderme (Lexical Disambiguation)**
 - POS ve WSD
 - **Sözdizimsel belirsizlik (Syntactic Ambiguity)**
 - Örn; her duck → olasılıksal ayrıştırma

Belirsizliklerin Çözümü

- Bağımlılık Ayırıştırma (Dependency Parsing)

I made her duck

