

TÜRKÇE DOĞAL DİL İŞLEME

Kemal Oflazer

Carnegie Mellon Üniversitesi - Katar

ko@cs.cmu.edu

Özet

- Doğal Dil İşleme
- Türkçe
- Türkçe ve Doğal Dil İşleme
- Türkçe Doğal Dil İşleme Çalışmaları
- Önümüzdeki Dönemde Yapılabilecekler

Doğal Dil İşleme

- Doğal Dil İşleme

- İnsan – Bilgisayar etkileşimini doğal dil ile yapabilen
- İnsan – İnsan etkileşimini zenginleştiren/kolaylaştıran

bilgisayar sistemleri geliştiren bilgisayar mühendisliği alt-alanıdır.

Doğal Dil İşleme

- Doğal Dil İşleme

- İnsan – Bilgisayar etkileşimini doğal dil ile yapabilen
- İnsan – İnsan etkileşimini zenginleştiren/kolaylaştıran

bilgisayar sistemleri geliştiren bilgisayar mühendisliği alt-alanıdır.

- Doğal Dil İşleme

- Yapay Zeka
- Makine öğrenmesi
- Biçimsel diller ve Otomatlar kuramı
- İstatistik
- Dilbilim
- Yazılım Mühendisliği

alanlarından kavram ve teknikler kullanır.

Doğal Dil İşleme Uygulamaları

- Soru cevaplama

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme
- Bir toplantıda konuşulanları takip edip özetleme

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme
- Bir toplantıda konuşulanları takip edip özetleme
- E-postalara otomatik yanıtlar gönderme

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme
- Bir toplantıda konuşulanları takip edip özetleme
- E-postalara otomatik yanıtlar gönderme
- Verilen komutları yerine getirme

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme
- Bir toplantıda konuşulanları takip edip özetleme
- E-postalara otomatik yanıtlar gönderme
- Verilen komutları yerine getirme
- Yazım veya sözdizim hatalarını düzeltme

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme
- Bir toplantıda konuşulanları takip edip özetleme
- E-postalara otomatik yanıtlar gönderme
- Verilen komutları yerine getirme
- Yazım veya sözdizim hatalarını düzeltme
- Kompozisyon notlama (SAT)

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme
- Bir toplantıda konuşulanları takip edip özetleme
- E-postalara otomatik yanıtlar gönderme
- Verilen komutları yerine getirme
- Yazım veya sözdizim hatalarını düzeltme
- Kompozisyon notlama (SAT)
- Okuduğunu anlama sınavları için soru üretme

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme
- Bir toplantıda konuşulanları takip edip özetleme
- E-postalara otomatik yanıtlar gönderme
- Verilen komutları yerine getirme
- Yazım veya sözdizim hatalarını düzeltme
- Kompozisyon notlama (SAT)
- Okuduğunu anlama sınavları için soru üretme
- Devamlı olarak yazı okuyup bilgi tabanı güncelleme

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme
- Bir toplantıda konuşulanları takip edip özetleme
- E-postalara otomatik yanıtlar gönderme
- Verilen komutları yerine getirme
- Yazım veya sözdizim hatalarını düzeltme
- Kompozisyon notlama (SAT)
- Okuduğunu anlama sınavları için soru üretme
- Devamlı olarak yazı okuyup bilgi tabanı güncelleme
- Etkileşimli olarak dil öğretme

Doğal Dil İşleme Uygulamaları

- Soru cevaplama
- Yazı ve konuşma çevirme
- Bir dizi yazıyı özetleme
- Bir toplantıda konuşulanları takip edip özetleme
- E-postalara otomatik yanıtlar gönderme
- Verilen komutları yerine getirme
- Yazım veya sözdizim hatalarını düzeltme
- Kompozisyon notlama (SAT)
- Okuduğunu anlama sınavları için soru üretme
- Devamlı olarak yazı okuyup bilgi tabanı güncelleme
- Etkileşimli olarak dil öğretme
- Konuşma/duyma/görme kısıtlı kişilere yardımcı olma

Doğal Dil İşleme Nedir? Detaylı Yanıt

- Dil çözümlemesini, üretimini ve edinimini otomatik olarak yapabilme

Doğal Dil İşleme Nedir? Detaylı Yanıt

- Dil çözümlemesini, üretimini ve edinimini otomatik olarak yapabilme
- **Çözümleme** (“anlama” veya “işleme” ...):
 - **Girdi**: Dil
 - **Çıktı**: Gerekli işlemeyi destekleyen bir “gösterim”

Doğal Dil İşleme Nedir? Detaylı Yanıt

- Dil çözümlemesini, üretimini ve edinimini otomatik olarak yapabilme
- **Çözümleme** (“anlama” veya “işleme” ...):
 - Girdi: Dil
 - Çıktı: Gerekli işlemeyi destekleyen bir “gösterim”
- **Üretim**:
 - Girdi: “Gösterim”
 - Çıktı: Dil

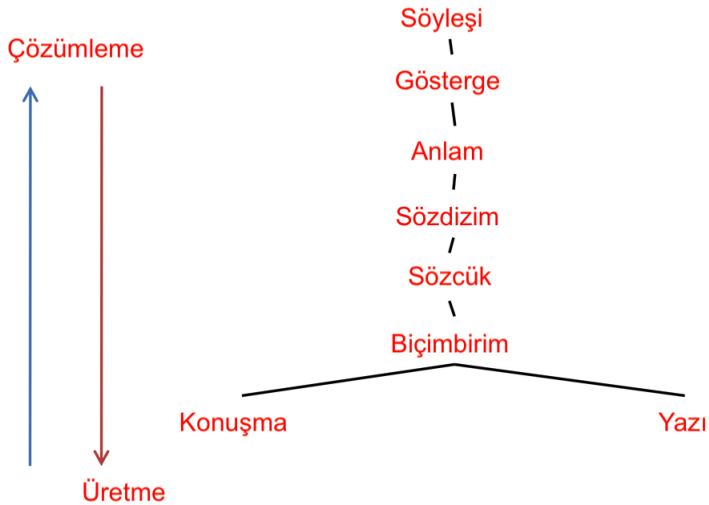
Doğal Dil İşleme Nedir? Detaylı Yanıt

- Dil çözümlemesini, üretimini ve edinimini otomatik olarak yapabilme
- **Çözümleme** (“anlama” veya “işleme” ...):
 - Girdi: Dil
 - Çıktı: Gerekli işlemeyi destekleyen bir “gösterim”
- **Üretim**:
 - Girdi: “Gösterim”
 - Çıktı: Dil
- **Edinim**: “Gösterim” ve işleme algoritmalarını örnek verilerden öğrenme

Doğal Dil İşleme Nedir? Detaylı Yanıt

- Dil çözümlemesini, üretimini ve edinimini otomatik olarak yapabilme
- **Çözümleme** (“anlama” veya “işleme” ...):
 - Girdi: Dil
 - Çıktı: Gerekli işlemeyi destekleyen bir “gösterim”
- **Üretim**:
 - Girdi: “Gösterim”
 - Çıktı: Dil
- **Edinim**: “Gösterim” ve işleme algoritmalarını örnek verilerden öğrenme
- **Gösterim?**

Dilbilimsel Gösterim Düzeyleri



Doğal Dil İşleme Neden Zor?

Çünkü



Doğal Dil İşleme Neden Zor?

- Düzeyler arasındaki ilişki ve aktarım çok karmaşık.

Doğal Dil İşleme Neden Zor?

- Düzeyler arasındaki ilişki ve aktarım çok karmaşık.
- Gösterimlerin detayları ve yeterliliği uygulamanın ne olduğuna bağlı

Doğal Dil İşleme Neden Zor?

- Düzeyler arasındaki ilişki ve aktarım çok karmaşık.
- Gösterimlerin detayları ve yeterliliği uygulamanın ne olduğuna bağlı
- Gösterimler ideal ve kuramsal; **hiç bir zaman gösterimleri doğrudan gözleyemiyoruz!**

Doğal Dil İşleme Neden Zor?

- Düzeyler arasındaki ilişki ve aktarım çok karmaşık.
- Gösterimlerin detayları ve yeterliliği uygulamanın ne olduğuna bağlı
- Gösterimler ideal ve kuramsal; **hiç bir zaman gösterimleri doğrudan gözleyemiyoruz!**
- Girdi genelde “gürültülü”

Çok Yapılılık/Çok Anlamlılık

- Her düzeydeki girdilerin genelde çok sayıda yorumu olabilir!

Çok Yapılılık/Çok Anlamlılık

- Her düzeydeki girdilerin genelde çok sayıda yorumu olabilir!
 - Bir konuşma ses dalgası çok sayıda olası tümceye dönüştürülebilir.

Çok Yapılılık/Çok Anlamlılık

- Her düzeydeki girdilerin genelde çok sayıda yorumu olabilir!
 - Bir konuşma ses dalgası çok sayıda olası tümceye dönüştürülebilir.
 - Bir tümcedeki her sözcüğün çok sayıda biçimbirimsel yorumu olabilir.

Çok Yapılılık/Çok Anlamlılık

- Her düzeydeki girdilerin genelde çok sayıda yorumu olabilir!
 - Bir konuşma ses dalgası çok sayıda olası tümceye dönüştürülebilir.
 - Bir tümcedeki her sözcüğün çok sayıda biçimbirimsel yorumu olabilir.
 - Her kök sözcüğün birden fazla anlamı olabilir.

Çok Yapılılık/Çok Anlamlılık

- Her düzeydeki girdilerin genelde çok sayıda yorumu olabilir!
 - Bir konuşma ses dalgası çok sayıda olası tümceye dönüştürülebilir.
 - Bir tümcedeki her sözcüğün çok sayıda biçimbirimsel yorumu olabilir.
 - Her kök sözcüğün birden fazla anlamı olabilir.
 - Bir tümcenin çok sayıda yapısal çözümlemesi olabilir.

Çok Yapılılık/Çok Anlamlılık

- Her düzeydeki girdilerin genelde çok sayıda yorumu olabilir!
 - Bir konuşma ses dalgası çok sayıda olası tümceye dönüştürülebilir.
 - Bir tümcedeki her sözcüğün çok sayıda biçimbirimsel yorumu olabilir.
 - Her kök sözcüğün birden fazla anlamı olabilir.
 - Bir tümcenin çok sayıda yapısal çözümlemesi olabilir.
 - Bir yapısal çözümlemenin birden fazla mantıksal anlamı olabilir.

Çok Yapılılık/Çok Anlamlılık

- Her düzeydeki girdilerin genelde çok sayıda yorumu olabilir!
 - Bir konuşma ses dalgası çok sayıda olası tümceye dönüştürülebilir.
 - Bir tümcedeki her sözcüğün çok sayıda biçimbirimsel yorumu olabilir.
 - Her kök sözcüğün birden fazla anlamı olabilir.
 - Bir tümcenin çok sayıda yapısal çözümlemesi olabilir.
 - Bir yapısal çözümlemenin birden fazla mantıksal anlamı olabilir.
- Esas zorluk olası yorumların hangisinin doğru olduğunu çevrimden kestirmek

Çok Yapılılık/Çok Anlamlılık

- Her düzeydeki girdilerin genelde çok sayıda yorumu olabilir!
 - Bir konuşma ses dalgası çok sayıda olası tümceye dönüştürülebilir.
 - Bir tümcedeki her sözcüğün çok sayıda biçimbirimsel yorumu olabilir.
 - Her kök sözcüğün birden fazla anlamı olabilir.
 - Bir tümcenin çok sayıda yapısal çözümlemesi olabilir.
 - Bir yapısal çözümlemenin birden fazla mantıksal anlamı olabilir.
- Esas zorluk olası yorumların hangisinin doğru olduğunu çevrimden kestirmek
- İnsanlar bunu çok kolay bir şekilde yapıyorlar

Çok Yapılılık/Çok Anlamlılık

- Her düzeydeki girdilerin genelde çok sayıda yorumu olabilir!
 - Bir konuşma ses dalgası çok sayıda olası tümceye dönüştürülebilir.
 - Bir tümcedeki her sözcüğün çok sayıda biçimbirimsel yorumu olabilir.
 - Her kök sözcüğün birden fazla anlamı olabilir.
 - Bir tümcenin çok sayıda yapısal çözümlemesi olabilir.
 - Bir yapısal çözümlemenin birden fazla mantıksal anlamı olabilir.
- Esas zorluk olası yorumların hangisinin doğru olduğunu çevrimden kestirmek
- İnsanlar bunu çok kolay bir şekilde yapıyorlar
- Bilgisayarlar henüz yapamıyorlar

Doğal Dil İşleme

- Daha fazla bilgi için bknz. Jurafsky ve Martin **Speech and Language Processing**

(<https://web.stanford.edu/~jurafsky/slp3/>)

Türkçe

- Türkçe Altay dil grubunda 70+ milyon kişi tarafından anadili olarak konuşulan bir dildir.

Türk dilleri	
Coğrafi dağılım:	Özgün olarak Batı Çin 'den Sibirya ve Doğu Avrupa 'ya kadar
Sınıflandırma:	Altay ^[1]
Türk dilleri	
Alt bölümler:	Güneybatı (Oğuz grubu) Kuzeybatı (Kıpçak grubu) Güneydoğu (Uygur grubu) Kuzeydoğu (Sibirya grubu) Ogur grubu Argu grubu
	
Türk dillerinin resmî dil olarak kullanıldığı ülkeler ve özerk bölgeler	

Türkçe

- Türkçe Altay dil grubunda 70+ milyon kişi tarafından anadili olarak konuşulan bir dildir.
- Diğer Türk dilleri de göz önüne alındığında bu rakam 220+ milyona çıkar (Türkçe Vikipedi).

Türk dilleri	
Coğrafi dağılım:	Özgün olarak Batı Çin 'den Sibiryaya ve Doğu Avrupa 'ya kadar
Sınıflandırma:	Altay ^[1]
Türk dilleri	
Alt bölümler:	Güneybatı (Oğuz grubu) Kuzeybatı (Kıpçak grubu) Güneydoğu (Uygur grubu) Kuzeydoğu (Sibiryaya grubu) Ogur grubu Argu grubu
	
Türk dillerinin resmî dil olarak kullanıldığı ülkeler ve özerk bölgeler	

Türkçe

- Türk dilleri ailesinde 41 dil listelenmektedir.

- | | | |
|--------------------|-----------------------------|----------------|
| ● Türkiye Türkçesi | ● Kazakça | ● Başkurtça |
| ● Gagavuzca | ● Karaçay-Balkarca | ● Kumukça |
| ● Karay Türkçesi | ● Nogayca | ● Hakasça |
| ● Azerice (Kuzey) | ● Karakalpakça | ● Tuvaca |
| ● Azerice (Güney) | ● Kırgızca | ● Altayca |
| ● Irak Türkmençesi | ● Tatarca | ● Şor Türkçesi |
| ● Türkmençe | ● Kırım Tatarcası | ● Sahaca |
| ● Özbekçe | ● Moldova-Romanya Tatarcası | ● Çuvaşça |
| ● Uygurca | | |

Türk Dillerinin Temel Ortak Özellikleri

Sözcük dağarcığı

- İsimler sınıflara/cinslere ayrılmıyor (örneğin Almanca, Fransızca'da olduğu gibi)

Türk Dillerinin Temel Ortak Özellikleri

Sözcük dağarcığı

- İsimler sınıflara/cinslere ayrılmıyor (örneğin Almanca, Fransızca'da olduğu gibi)

Sözcük yapısı

- Biçimbirimler kök sözcüğe “tespih taneleri” gibi ekleniyor.

Türk Dillerinin Temel Ortak Özellikleri

Sözcük dağarcığı

- İsimler sınıflara/cinslere ayrılmıyor (örneğin Almanca, Fransızca'da olduğu gibi)

Sözcük yapısı

- Biçimbirimler kök sözcüğe “tespih taneleri” gibi ekleniyor.

Öge Sırası

- Doğal sıra ”Özne – Nesne – Yüklem
- Ama çevrim gerektirdiğinde diğer beş sıra da kullanılabilir.

Türk dillerinden Örnek Sözcükler

sekiz

 Türkiye Türkçesi	 Azəri Türkçesi	 Başkurt Türkçesi	 Kazak Türkçesi	 Kırgız Türkçesi
sekiz	säkkiz	higiz	segiz	segiz
 Özbek Türkçesi	 Tatar Türkçesi	 Türkmen Türkçesi	 Uygur Türkçesi	 Rusça
säkkiz	sigiz	sekiz	säkkiz	vosem'

okumak

 Türkiye Türkçesi	 Azəri Türkçesi	 Başkurt Türkçesi	 Kazak Türkçesi	 Kırgız Türkçesi
okumak	oqumag	ukıv	okuv	okū
 Özbek Türkçesi	 Tatar Türkçesi	 Türkmen Türkçesi	 Uygur Türkçesi	 Rusça
okimäk	uku	okamak	okimak	çitat'

cumhuriyet

 Türkiye Türkçesi	 Azəri Türkçesi	 Başkurt Türkçesi	 Kazak Türkçesi	 Kırgız Türkçesi
cumhuriyet	respublika	respublika	respublika	respublika
 Özbek Türkçesi	 Tatar Türkçesi	 Türkmen Türkçesi	 Uygur Türkçesi	 Rusça
cümhüriyät	respublika	respublika	respublika	respublika
	cömhöriyät		cumhuriyät	

Türkçe

- Kök sözcük dağarcığı Arapça, Farsça, Yunanca, Ermenice, Fransızca, İtalyanca, Almanca ve İngilizce'den etkilenmiştir.

Türkçe

- Kök sözcük dağarcığı Arapça, Farsça, Yunanca, Ermenice, Fransızca, İtalyanca, Almanca ve İngilizce'den etkilenmiştir.
- Cumhuriyet sonrası sadeleştirme çalışmaları

Türkçe

- Kök sözcük dağarcığı Arapça, Farsça, Yunanca, Ermenice, Fransızca, İtalyanca, Almanca ve İngilizce'den etkilenmiştir.
- Cumhuriyet sonrası sadeleştirme çalışmaları

Anne ve Babamın Kuşağı

Bir müsellesin mesahı sathiyesi zemini ile irtifaının zarbının nıfsına müsavidir.

Türkçe

- Kök sözcük dağarcığı Arapça, Farsça, Yunanca, Ermenice, Fransızca, İtalyanca, Almanca ve İngilizce'den etkilenmiştir.
- Cumhuriyet sonrası sadeleştirme çalışmaları

Anne ve Babamın Kuşağı

Bir müsellesin mesahı sathiyesi zemini ile irtifaının zarbının nıfsına müsavidir.

Benim kuşağım+

Bir üçgenin yüzey alanı tabanı ile yüksekliğinin çarpımının yarısına eşittir.

Türkçe ve Doğal Dil İşleme

Sözcük Yapısı

Türkçe bir çok açıdan çok “ilginç” bir dil

- Bitişken sözcük yapısı

Türkçe ve Doğal Dil İşleme

Sözcük Yapısı

Türkçe bir çok açıdan çok “ilginç” bir dil

- Bitişken sözcük yapısı
- Ünlü /Ünsüz uyumu

Türkçe ve Doğal Dil İşleme

Sözcük Yapısı

Türkçe bir çok açıdan çok “ilginç” bir dil

- Bitişken sözcük yapısı
- Ünlü /Ünsüz uyumu
- Çok sayıda üretken türetme biçimbirimi

Türkçe ve Doğal Dil İşleme

Sözcük Yapısı

Türkçe bir çok açıdan çok “ilginç” bir dil

- Bitişken sözcük yapısı
- Ünlü /Ünsüz uyumu
- Çok sayıda üretken türetme biçimbirimi
 - pratikte sonsuz kabul edilebilecek sözcük dağarcığı

Türkçe ve Doğal Dil İşleme

Sözcük Yapısı

Türkçe bir çok açıdan çok “ilginç” bir dil

- Bitişken sözcük yapısı
- Ünlü /Ünsüz uyumu
- Çok sayıda üretken türetme biçimbirimi
 - pratikte sonsuz kabul edilebilecek sözcük dağarcığı
- Türetme ve sözdizimsel ilişkiler

Türkçe ve Doğal Dil İşleme

Sözcük Yapısı

Türkçe bir çok açıdan çok “ilginç” bir dil

- Bitişken sözcük yapısı
- Ünlü /Ünsüz uyumu
- Çok sayıda üretken türetme biçimbirimi
 - pratikte sonsuz kabul edilebilecek sözcük dağarcığı
- Türetme ve sözdizimsel ilişkiler
- Karmaşık bir sözcük söyleyiş vurgusu

Türkçe ve Doğal Dil İşleme

Sözcük Yapısı

Türkçe bir çok açıdan çok “ilginç” bir dil

- Bitişken sözcük yapısı
- Ünlü /Ünsüz uyumu
- Çok sayıda üretken türetme biçimbirimi
 - pratikte sonsuz kabul edilebilecek sözcük dağarcığı
- Türetme ve sözdizimsel ilişkiler
- Karmaşık bir sözcük söyleyiş vurgusu
- Sesbilimsel kurallar – yazım ilişkisi ve evrimi

Bitişken Sözcük Yapısı

- Finlandiyalılaştıramadıklarımızdanmışsınızcasına

Bitişken Sözcük Yapısı

- Finlandiyalılaştıramadıklarımızdanmışsınızcasına
- Finlandiya+lı+laş+tır+ama+dık+lar+ımız+dan+mış+sınız+casına

Bitişken Sözcük Yapısı

- Finlandiyalılaştıramadıklarımızdanmışsınızcasına
- Finlandiya+lı+laş+tır+ama+dık+lar+ımız+dan+mış+sınız+casına
- yapabileceksek

Bitişken Sözcük Yapısı

- Finlandiyalılaştıramadıklarımızdanmışsınızcasına
- Finlandiya+lı+laş+tır+ama+dık+lar+ımız+dan+mış+sınız+casına
- yapabileceksek
 - yap+abil+ecek+se+k

Bitişken Sözcük Yapısı

- Finlandiyalılaştıramadıklarımızdanmışsınızcasına
- Finlandiya+lı+laş+tır+ama+dık+lar+ımız+dan+mış+sınız+casına
- yapabileceksek
 - yap+abil+ecek+se+k
 - if we will be able to do (something)

Bitişken Sözcük Yapısı

- Finlandiyalılaştıramadıklarımızdanmışsınızcasına
- Finlandiya+lı+laş+tır+ama+dık+lar+ımız+dan+mış+sınız+casına
- yapabileceksek
 - yap+abil+ecek+se+k
 - if we will be able to do (something)
- Ancak, genelde ortalama ≈ 3 biçimbirim/sözcük (kök dahil)

Bitişken Sözcük Yapısı

- Finlandiyalılaştıramadıklarımızdanmışsınızcasına
- Finlandiya+lı+laş+tır+ama+dık+lar+ımız+dan+mış+sınız+casına
- yapabileceksek
 - yap+abil+ecek+se+k
 - if we will be able to do (something)
- Ancak, genelde ortalama ≈ 3 biçimbirim/sözcük (kök dahil)
 - Ama yüksek frekanslı sözcüklerin genelde bir biçimbirimi var!

Yüksek Frekanslı Sözcükler

	Sözcük	Biçimbirim	Yorum		Sözcük	Biçimbirim	Yorum
1	bir	1	4	11	kadar	1	2
2	bu	1	2	12	ama	1	3
3	da	1	1	13	gibi	1	1
4	için	1	4	14	ol+an	2	1
5	de	1	2	15	var	1	2
6	çok	1	1	16	ne	1	2
7	ile	1	2	17	sonra	1	2
8	en	1	2	18	i+se	2	2
9	daha	1	1	19	o	1	2
10	ol+arak	2	1	20	ilk	1	1

Bitişken Sözcük Yapısı

Ancak beterin beteri de var (:-)

Bitişken Sözcük Yapısı

Ancak beterin beteri de var (:-)

İnuktikut

- Parismunngaujumaniralausimanngittunga
- Paris+mut+nngau+juma+niraq+lauq+si+ma+nngit+jun
- Paris'e gitmek istediğimi hiç bir zaman söylemedim.

Bitişken Sözcük Yapısı

Ancak beterin beteri de var (:-)

İnuktikut

- Parismunngaujumaniralausimanngittunga
- Paris+mut+nngau+juma+niraq+lauq+si+ma+nngit+jun
- Paris'e gitmek istediğimi hiç bir zaman söylemedim.

Fince

- kahdensienkymmenensienkahdeksansien
- kahde+ns+i+en+kymmene+ns+i+en+kahdeksa+ns+i+en
- ikincilerinonuncularınsekizincilerin
- yirmi sekizincilerin

Sözcük Yapısı

Sesbilimsel Kurallar

- Ünlü uyumu
 - ev+ler, av+lar
 - al+dı, ev+di, bul+du, gül+dü

Sözcük Yapısı

Sesbilimsel Kurallar

- Ünlü uyumu
 - ev+ler, av+lar
 - al+dı, ev+di, bul+du, gül+dü
- Ünsüz uyumu
 - al+dı, ak+tı, ...

Sözcük Yapısı

Sesbilimsel Kurallar

- Ünlü uyumu
 - ev+ler, av+lar
 - al+dı, ev+di, bul+du, gül+dü
- Ünsüz uyumu
 - al+dı, ak+tı, ...
- Ünlü düşmesi
 - sağl[a]+ıyor, oda+[ı]m (sakal+ım) ...

Sözcük Yapısı

Sesbilimsel Kurallar

- Ünlü uyumu
 - ev+ler, av+lar
 - al+dı, ev+di, bul+du, gül+dü
- Ünsüz uyumu
 - al+dı, ak+tı, ...
- Ünlü düşmesi
 - sağl[a]+ıyor, oda+[ı]m (sakal+ım) ...
- Ünsüz düşmesi
 - kalem+[n]in, oda+nın ...

Sözcük Yapısı

Sesbilimsel Kurallar

- Ünlü uyumu
 - ev+ler, av+lar
 - al+dı, ev+di, bul+du, gül+dü
- Ünsüz uyumu
 - al+dı, ak+tı, ...
- Ünlü düşmesi
 - sağl[a]+ıyor, oda+[ı]m (sakal+ım) ...
- Ünsüz düşmesi
 - kalem+[n]in, oda+nın ...
- Başka
 - tıb[b]+ı, hukuk+u/hukuğ+u, cami+si, cami+[s]i ...

Sorunlar, Sorunlar!

- Yazılı dilde kuralların denetlenmesine engel teşkil eden yazımlar var.

Sorunlar, Sorunlar!

- Yazılı dilde kuralların denetlenmesine engel teşkil eden yazımlar var.
- 2012'ye / 2011'e / 2/3'si / 2/3'ü, 15:00'te / 15:00'da

Sorunlar, Sorunlar!

- Yazılı dilde kuralların denetlenmesine engel teşkil eden yazımlar var.
- 2012'ye / 2011'e / 2/3'si / 2/3'ü, 15:00'te / 15:00'da
 - Yazıldığı hali ile köklerde ünlü yok!

Sorunlar, Sorunlar!

- Yazılı dilde kuralların denetlenmesine engel teşkil eden yazımlar var.
- 2012'ye / 2011'e / 2/3'si / 2/3'ü, 15:00'te / 15:00'da
 - Yazıldığı hali ile köklerde ünlü yok!
- BAB'a / AB'ye / BBC'ye, / BM'ye / BM'e

Sorunlar, Sorunlar!

- Yazılı dilde kuralların denetlenmesine engel teşkil eden yazımlar var.
- 2012'ye / 2011'e / 2/3'si / 2/3'ü, 15:00'te / 15:00'da
 - Yazıldığı hali ile köklerde ünlü yok!
- BAB'a / AB'ye / BBC'ye, / BM'ye / BM'e
- “Kök sözcüğün” okunuşunun nasıl bittiğini bilmek durumundayız.

Sorunlar, Sorunlar!

Okunuş – Yazım İlişkisinin Evrimi

- Yazanlar okuyanların İngilizce veya Fransızca bildiğini düşünüyorlar!

Sorunlar, Sorunlar!

Okunuş – Yazım İlişkisinin Evrimi

- Yazarlar okuyanların İngilizce veya Fransızca bildiğini düşünüyorlar!
- Kök sözcükler özgün dillerindeki yazımları ile yazılırken

Sorunlar, Sorunlar!

Okunuş – Yazım İlişkisinin Evrimi

- Yazanlar okuyanların İngilizce veya Fransızca bildiğini düşünüyorlar!
- Kök sözcükler özgün dillerindeki yazımları ile yazılırken
- Eklenen Türkçe ekler sözcüğün özgün dilindeki söylenişine göre ekleniyor!

Sorunlar, Sorunlar!

Okunuş – Yazım İlişkisinin Evrimi

- Yazanlar okuyanların İngilizce veya Fransızca bildiğini düşünüyorlar!
- Kök sözcükler özgün dillerindeki yazımları ile yazılırken
- Eklenen Türkçe ekler sözcüğün özgün dilindeki söylenişine göre ekleniyor!
 - Godot'yu, Bordeaux'ya ...

Sorunlar, Sorunlar!

Okunuş – Yazım İlişkisinin Evrimi

- Yazanlar okuyanların İngilizce veya Fransızca bildiğini düşünüyorlar!
- Kök sözcükler özgün dillerindeki yazımları ile yazılırken
- Eklenen Türkçe ekler sözcüğün özgün dilindeki söylenişine göre ekleniyor!
 - Godot'yu, Bordeaux'ya ...
 - serverlar ve clientlar

Sorunlar, Sorunlar!

Okunuş – Yazım İlişkisinin Evrimi

- Yazanlar okuyanların İngilizce veya Fransızca bildiğini düşünüyorlar!
- Kök sözcükler özgün dillerindeki yazımları ile yazılırken
- Eklenen Türkçe ekler sözcüğün özgün dilindeki söylenişine göre ekleniyor!
 - Godot'yu, Bordeaux'ya ...
 - serverlar ve clientlar
 - Daha da kötüsü: server'lar ve client'lar

Sorunlar, Sorunlar!

Okunuş – Yazım İlişkisinin Evrimi

- Yazanlar okuyanların İngilizce veya Fransızca bildiğini düşünüyorlar!
- Kök sözcükler özgün dillerindeki yazımları ile yazılırken
- Eklenen Türkçe ekler sözcüğün özgün dilindeki söylenişine göre ekleniyor!
 - Godot'yu, Bordeaux'ya ...
 - serverlar ve clientlar
 - Daha da kötüsü: server'lar ve client'lar
- Twitter'da yazılanlara hiç girmeyelim (-:-)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklere bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

okuma

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

okuma

- oku+ma (reading)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

okuma

- oku+ma (reading)
- oku+ma (don't read!)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

okuma

- oku+ma (reading)
- oku+ma (don't read!)
- ok+um+a (to my arrow)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

okuma

- oku+ma (reading)
- oku+ma (don't read!)
- ok+um+a (to my arrow)

okuna

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

okuma

- oku+ma (reading)
- oku+ma (don't read!)
- ok+um+a (to my arrow)

okuna

- oku+na (let it be read!)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

okuma

- oku+ma (reading)
- oku+ma (don't read!)
- ok+um+a (to my arrow)

okuna

- oku+na (let it be read!)
- ok+un+a (to your arrow)

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

okuma

- oku+ma (reading)
- oku+ma (don't read!)
- ok+um+a (to my arrow)

okuna

- oku+na (let it be read!)
- ok+un+a (to your arrow)

- Ortalama ≈ 2 çözümleme/sözcük.

Sözcük Yapısı – Çokyapılılık

- Sözcükler farklı şekilde eklerle bölünebilir veya ekler farklı işlevlere sahip olabilir.

koyun

- koyun (sheep / bosom)
- koy+un (your bay)
- koy+[n]un (of the bay)
- koy+un (put!)
- koyu+[u]n (your dark (one))

okuma

- oku+ma (reading)
- oku+ma (don't read!)
- ok+um+a (to my arrow)

okuna

- oku+na (let it be read!)
- ok+un+a (to your arrow)

- Ortalama ≈ 2 çözümleme/sözcük.
 - Ama sözcüklerin ≈ 65 'inin tek çözümü var.

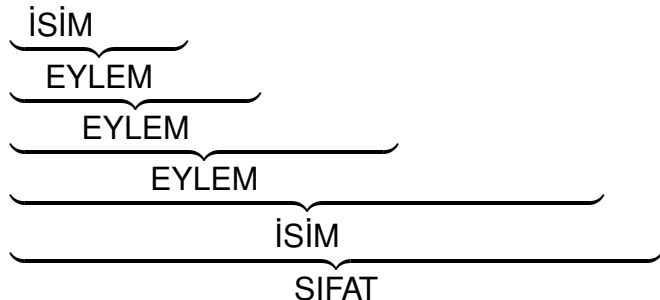
Sözcük Yapısı

- **ruhsatlandırılmamasındaki** - 9 biçimbirim

Sözcük Yapısı

- **ruhsatlandırılmamasındaki** - 9 biçimbirim

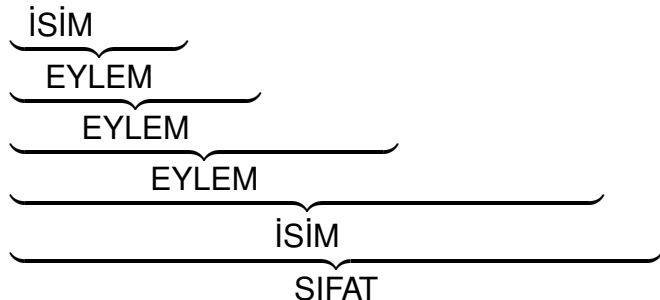
- ruhsat + lan + dır + ıl + ama + ma + sı + nda + ki



Sözcük Yapısı

- **ruhsatlandırılmamasındaki** - 9 biçimbirim

- ruhsat + lan + dır + ıl + ama + ma + sı + nda + ki



- İsim kök ile başlayıp bir dizi türetmeden sonra tümcede sıfat rolü oynayacak bir sözcük oluşturuyoruz.

Sözcük Yapısı

Üretken Türetme

- İsimlerden: isim (kitapçık), sıfat (renkli, renksiz), eylem (taşlaş, arabalan)

Sözcük Yapısı

Üretken Türetme

- İsimlerden: isim (kitapçık), sıfat (renkli, renksiz), eylem (taşlaş, arabalan)
- Sıfatlardan: isim (büyüklük), eylem (büyükleş)

Sözcük Yapısı

Üretken Türetme

- İsimlerden: isim (kitapçık), sıfat (renkli, renksiz), eylem (taşlaş, arabalan)
- Sıfatlardan: isim (büyüklük), eylem (büyükleş)
- Eylemlerden: eylem (yaptır), “isim” (gitme, okuduğu), “sıfat” (gideceği), zarf (gidip, giderken, gidince)

Sözcük Yapısı – Üretken Türetme

Kök	Türetme Sayısı	Sözcük Sayısı	Toplam
masa (isim)	0	112	112
	1	4,663	4,775
	2	49,640	54,415
	3	493,975	548,390
oku (eylem)	0	702	702
	1	11,366	12,068
	2	112,877	124,945
	3	1,336,266	1,461,211

Sözcük Yapısı – Üretken Türetme

- Boğaziçi haber derlemi

Sözcük Yapısı – Üretken Türetme

- Boğaziçi haber derlemi
 - 4.1M değişik sözcük

Sözcük Yapısı – Üretken Türetme

- Boğaziçi haber derlemi
 - 4.1M değişik sözcük
 - 490M'dan 491M'a çıkınca 5000 yeni sözcük ekleniyor.

Sözcük Yapısı – Üretken Türetme

- Boğaziçi haber derlemi
 - 4.1M değişik sözcük
 - 490M'dan 491M'a çıkınca 5000 yeni sözcük ekleniyor.
 - En sık 50K sözcük tüm derlemin %89'unu kapsıyor

Sözcük Yapısı – Üretken Türetme

- Boğaziçi haber derlemi
 - 4.1M değişik sözcük
 - 490M'dan 491M'a çıkınca 5000 yeni sözcük ekleniyor.
 - En sık 50K sözcük tüm derlemin %89'unu kapsıyor
 - En sık 300K sözcük tüm derlemin %97'sini kapsıyor

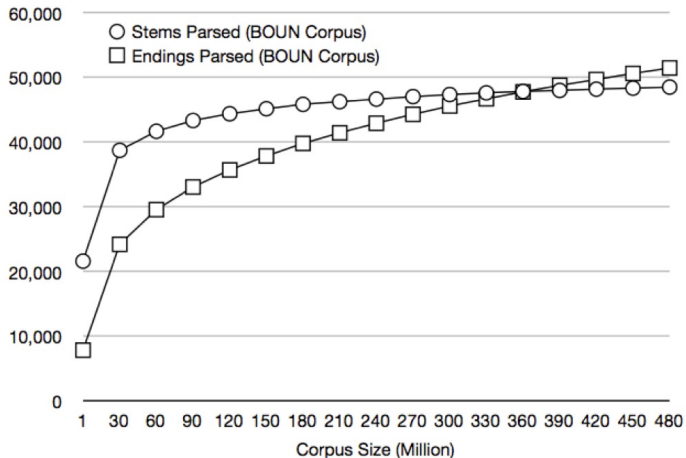
Sözcük Yapısı – Üretken Türetme

- Boğaziçi haber derlemi
 - 4.1M değişik sözcük
 - 490M'dan 491M'a çıkınca 5000 yeni sözcük ekleniyor.
 - En sık 50K sözcük tüm derlemin %89'unu kapsıyor
 - En sık 300K sözcük tüm derlemin %97'sini kapsıyor
 - 3.4M sözcük 10 kereden az geçiyor

Sözcük Yapısı – Üretken Türetme

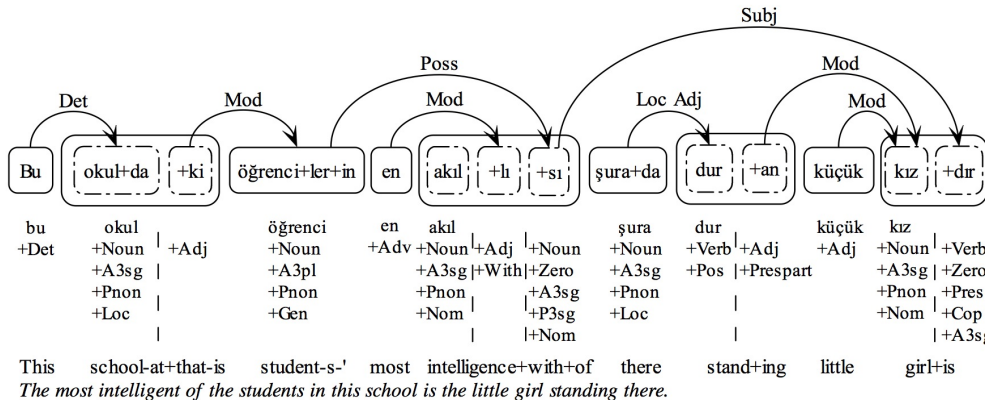
- Boğaziçi haber derlemi
 - 4.1M değişik sözcük
 - 490M'dan 491M'a çıkınca 5000 yeni sözcük ekleniyor.
 - En sık 50K sözcük tüm derlemin %89'unu kapsıyor
 - En sık 300K sözcük tüm derlemin %97'sini kapsıyor
 - 3.4M sözcük 10 kereden az geçiyor
 - 2.0M sözcük bir kere geçiyor

Sözcük Yapısı – Üretken Türetme



Sözcük Yapısı – Türetme ve sözdizimsel ilişkiler

- Sözdizimsel ilişkiler, türetmelerin ayırdığı çekim grupları arasındadır:



Türkçe ve Doğal Dil İşleme – Sözdizim

- (Nerdeyse) Tamamen serbest bir öge dizimi
- Her ne kadar **Özne**, **Nesne**, **Yüklem** normal öge sırası olarak belirtilse de
 - Diğer 5 sıra da kullanılabiliyor
 - Hatta, belirsiz nesnenin yüklemle beraber hareket edeceği kuralı da ihlal edilebiliyor.¹
 - Yapayım sana bir yemek.
 - Ben de içtim sigara.

¹Örnekler: Sarah Kennely

Sözdizim

- Ekin Çağla'yı gördü.

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)
- Gördü Ekin Çağla'yı.

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)
- Gördü Ekin Çağla'yı.
 - Ama görmesini beklemiyorduk.

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)
- Gördü Ekin Çağla'yı.
 - Ama görmesini beklemiyorduk.
- Gördü Çağla'yı Ekin.

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)
- Gördü Ekin Çağla'yı.
 - Ama görmesini beklemiyorduk.
- Gördü Çağla'yı Ekin.
 - Zaten görmesini bekliyorduk (?)

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)
- Gördü Ekin Çağla'yı.
 - Ama görmesini beklemiyorduk.
- Gördü Çağla'yı Ekin.
 - Zaten görmesini bekliyorduk (?)
- Ekin gördü Çağla'yı.

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)
- Gördü Ekin Çağla'yı.
 - Ama görmesini beklemiyorduk.
- Gördü Çağla'yı Ekin.
 - Zaten görmesini bekliyorduk (?)
- Ekin gördü Çağla'yı.
 - Başkası da görebilirdi.

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)
- Gördü Ekin Çağla'yı.
 - Ama görmesini beklemiyorduk.
- Gördü Çağla'yı Ekin.
 - Zaten görmesini bekliyorduk (?)
- Ekin gördü Çağla'yı.
 - Başkası da görebilirdi.
- Çağla'yı gördü Ekin.

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)
- Gördü Ekin Çağla'yı.
 - Ama görmesini beklemiyorduk.
- Gördü Çağla'yı Ekin.
 - Zaten görmesini bekliyorduk (?)
- Ekin gördü Çağla'yı.
 - Başkası da görebilirdi.
- Çağla'yı gördü Ekin.
 - Başkasını da görebilirdi.

Sözdizim

- Ekin Çağla'yı gördü.
- Çağla'yı Ekin gördü.
 - Çağla'yı gören Ekin idi (başkası değil!)
- Gördü Ekin Çağla'yı.
 - Ama görmesini beklemiyorduk.
- Gördü Çağla'yı Ekin.
 - Zaten görmesini bekliyorduk (?)
- Ekin gördü Çağla'yı.
 - Başkası da görebilirdi.
- Çağla'yı gördü Ekin.
 - Başkasını da görebilirdi.
- Sözdizimi modelleyen gramerlerin bu serbestliği çok “doğal” bir şekilde tanımlayabilmesi gerekir.

Sözdizim– Mantıksal Anlambilim İlişkisi

- Her hastaya bir doktor baktı \rightarrow
 $\forall x, \text{hasta}(x) \{ \exists(y), \text{doktor}(y), \text{baktı}(y, x) \}$

Sözdizim– Mantıksal Anlambilim İlişkisi

- Her hastaya bir doktor baktı \rightarrow
 $\forall x, \text{hasta}(x) \{ \exists(y), \text{doktor}(y), \text{baktı}(y, x) \}$
- Bir doktor her hastaya baktı \rightarrow
 $\exists y, \text{doktor}(y) \{ \forall(x), \text{hasta}(x), \text{baktı}(y, x) \}$

Sözdizim– Mantıksal Anlambilim İlişkisi

- Her hastaya bir doktor baktı \rightarrow
 $\forall x, \text{hasta}(x) \{ \exists(y), \text{doktor}(y), \text{baktı}(y, x) \}$
- Bir doktor her hastaya baktı \rightarrow
 $\exists y, \text{doktor}(y) \{ \forall(x), \text{hasta}(x), \text{baktı}(y, x) \}$
- Keza eğer bu mantıksal çıktıyı üretecekse, gramer öge sırası bilgisini kullanmalı.

Türkçe Doğal Dil İşleme Çalışmaları

- Güney Gönenç, Unique decipherability of codes with constraints with application to syllabification of Turkish words, COLING 1973
- Aydın Köksal, A first approach to a computerized model for the automatic morphological analysis of Turkish, Doktora Tezi, Hacettepe Üniversitesi, 1975
- Zeki Sagay, A computer translation from English to Turkish, Yüksek Lisans Tezi, ODTÜ, 1981
- Jorge Hankamer, Finite state morphology and left to right phonology, West Coast Conference on Formal Linguistics, 1986
- Albert Stoop, TRANSIT in the world of Machine Translation: Towards an automatic translator for Dutch and Turkish, Third Conference on Turkish Linguistics, 1987

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- NATO-Science for Stability Programı tarafından 1994–1999 arası desteklenen altyapı projesi

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- NATO-Science for Stability Programı tarafından 1994–1999 arası desteklenen altyapı projesi
 - Bilkent Üniversitesi (Kemal Oflazer), ODTÜ (Cem Bozşahin), Halıcı Bilgisayar

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- NATO-Science for Stability Programı tarafından 1994–1999 arası desteklenen altyapı projesi
 - Bilkent Üniversitesi (Kemal Oflazer), ODTÜ (Cem Bozşahin), Halıcı Bilgisayar
 - Ciddi bir öğrenci birikimi ve altyapı geliştirme süreci

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- NATO-Science for Stability Programı tarafından 1994–1999 arası desteklenen altyapı projesi
 - Bilkent Üniversitesi (Kemal Oflazer), ODTÜ (Cem Bozşahin), Halıcı Bilgisayar
 - Ciddi bir öğrenci birikimi ve altyapı geliştirme süreci
 - Uluslararası çok sayıda yayın ve görünürlük

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- NATO-Science for Stability Programı tarafından 1994–1999 arası desteklenen altyapı projesi
 - Bilkent Üniversitesi (Kemal Oflazer), ODTÜ (Cem Bozşahin), Halıcı Bilgisayar
 - Ciddi bir öğrenci birikimi ve altyapı geliştirme süreci
 - Uluslararası çok sayıda yayın ve görünürlük
- Bilkent Üniversitesi, ODTÜ, Boğaziçi Üniversitesi'nde çok sayıda yüksek lisans ve doktora tezi

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- NATO-Science for Stability Programı tarafından 1994–1999 arası desteklenen altyapı projesi
 - Bilkent Üniversitesi (Kemal Oflazer), ODTÜ (Cem Bozşahin), Halıcı Bilgisayar
 - Ciddi bir öğrenci birikimi ve altyapı geliştirme süreci
 - Uluslararası çok sayıda yayın ve görünürlük
- Bilkent Üniversitesi, ODTÜ, Boğaziçi Üniversitesi'nde çok sayıda yüksek lisans ve doktora tezi
- Çoğunlukla kural tabanlı çalışmalar

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- Sözcük çözümlemesi

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- Sözcük çözümlemesi
- Doğru sözcük çözümlerinin seçilmesi

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- Sözcük çözümlemesi
- Doğru sözcük çözümlerinin seçilmesi
- LFG/HPSG/ATN formalizmaları ile gramer modelleri geliştirilmesi

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- Sözcük çözümlemesi
- Doğru sözcük çözümlerinin seçilmesi
- LFG/HPSG/ATN formalizmaları ile gramer modelleri geliştirilmesi
- Kural tabanlı kısıtlı İngilizce - Türkçe çeviri sistemleri geliştirilmesi

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- Sözcük çözümlemesi
- Doğru sözcük çözümlerinin seçilmesi
- LFG/HPSG/ATN formalizmaları ile gramer modelleri geliştirilmesi
- Kural tabanlı kısıtlı İngilizce - Türkçe çeviri sistemleri geliştirilmesi
- İlk istatistik temelli çalışmalar

Türkçe Doğal Dil İşleme Çalışmaları

1990-2000

- Sözcük çözümlemesi
- Doğru sözcük çözümlerinin seçilmesi
- LFG/HPSG/ATN formalizmaları ile gramer modelleri geliştirilmesi
- Kural tabanlı kısıtlı İngilizce - Türkçe çeviri sistemleri geliştirilmesi
- İlk istatistik temelli çalışmalar
- İlk konuşma işleme çalışmaları

Türkçe Doğal Dil İşleme Çalışmaları

2000 sonrası

- Kaynak oluşturma çalışmaları

Türkçe Doğal Dil İşleme Çalışmaları

2000 sonrası

- Kaynak oluşturma çalışmaları
- Konuşma işleme çalışmaları

Türkçe Doğal Dil İşleme Çalışmaları

2000 sonrası

- Kaynak oluşturma çalışmaları
- Konuşma işleme çalışmaları
- Çoğunlukla istatistik / makina öğrenmesi yöntemlerinin kullanılmaya başlanması

Kaynak Oluşturma Çalışmaları

- ODTÜ Türkçe Dengeli Derlem
 - Bilge Say, ODTÜ

Kaynak Oluşturma Çalışmaları

- ODTÜ Türkçe Dengeli Derlem
 - Bilge Say, ODTÜ
- ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem
 - Bilge Say ve Kemal Oflazer (ODTÜ ve Sabancı) + Sabine Buchholz (Toshiba), Gülşen Eryiğit (İTÜ) ve Ruken Çakıcı'nın (Edinburgh/ODTÜ) düzeltmeleri, Gülşen Eryiğit'in ek deneme kümesi ve Evrensel Bağımlılık Ağaç Yapılı Derlemi.

Kaynak Oluşturma Çalışmaları

- ODTÜ Türkçe Dengeli Derlem
 - Bilge Say, ODTÜ
- ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem
 - Bilge Say ve Kemal Oflazer (ODTÜ ve Sabancı) + Sabine Buchholz (Toshiba), Gülşen Eryiğit (İTÜ) ve Ruken Çakıcı'nın (Edinburgh/ODTÜ) düzeltmeleri, Gülşen Eryiğit'in ek deneme kümesi ve Evrensel Bağımlılık Ağaç Yapılı Derlemi.
- Türkçe Kavramsal Sözlük (WordNet)
 - Özlem Çetinoğlu, Orhan Bilgin, Sabancı

Kaynak Oluşturma Çalışmaları

- ODTÜ Türkçe Dengeli Derlem
 - Bilge Say, ODTÜ
- ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem
 - Bilge Say ve Kemal Oflazer (ODTÜ ve Sabancı) + Sabine Buchholz (Toshiba), Gülşen Eryiğit (İTÜ) ve Ruken Çakıcı'nın (Edinburgh/ODTÜ) düzeltmeleri, Gülşen Eryiğit'in ek deneme kümesi ve Evrensel Bağımlılık Ağaç Yapılı Derlemi.
- Türkçe Kavramsal Sözlük (WordNet)
 - Özlem Çetinoğlu, Orhan Bilgin, Sabancı
- Konuşma tanıma ve üretme uygulamaları için kaynaklar
 - TÜBİTAK, ODTÜ, Sabancı, Boğaziçi

Kaynak Oluşturma Çalışmaları

- ODTÜ Türkçe Dengeli Derlem
 - Bilge Say, ODTÜ
- ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem
 - Bilge Say ve Kemal Oflazer (ODTÜ ve Sabancı) + Sabine Buchholz (Toshiba), Gülşen Eryiğit (İTÜ) ve Ruken Çakıcı'nın (Edinburgh/ODTÜ) düzeltmeleri, Gülşen Eryiğit'in ek deneme kümesi ve Evrensel Bağımlılık Ağaç Yapılı Derlemi.
- Türkçe Kavramsal Sözlük (WordNet)
 - Özlem Çetinoğlu, Orhan Bilgin, Sabancı
- Konuşma tanıma ve üretme uygulamaları için kaynaklar
 - TÜBİTAK, ODTÜ, Sabancı, Boğaziçi
- Çok büyük boyutta (genelde haber) yazı derlemleri
 - Boğaziçi, Sabancı

Kaynak Oluşturma Çalışmaları

- Paralel Çeviri Metin Derlemleri
 - Sabancı, TÜBİTAK, Koç, Uppsala
- ODTÜ Metin Düzeyinde İşaretlenmiş Derlem
 - Deniz Zeyrek, ODTÜ
- ODTÜ Sözlü Türkçe Derlemi
 - Şükriye Ruhi, ODTÜ
- Türkçe Ulusal Derlemi
 - Yeşim Aksan, Mustafa Aksan (Mersin Üniversitesi)

ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem

- ODTÜ Türkçe Dengeli Derlem'den seçilmiş olan 5600 tümceden oluşur.
- Her tümcenin sözcükleri
 - biçimbilimsel olarak çekim gruplarına ayrılarak işaretlenmiş, ve
 - çekim gruplarının arasındaki bağımlılık (dependency) ilişkileri kodlanmıştır.

ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem

Ana fikir:

hızlı arabanızdaydı
(it was in your **fast/speedy** car)

hız lı arabanızda ydı

Modifier

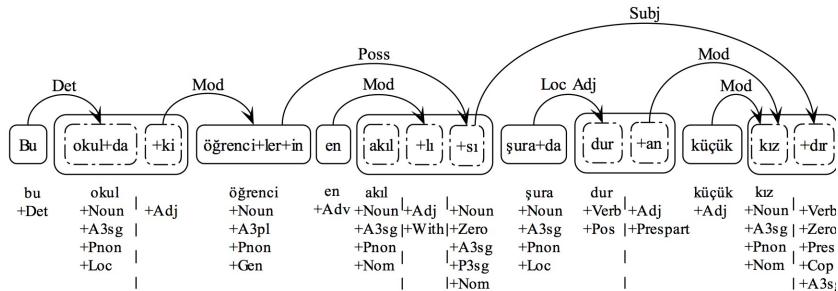
hız+Noun+A3sg+Pnon+Nom

^DB +Adj+With

araba+Noun+A3sg+P3pl+Loc

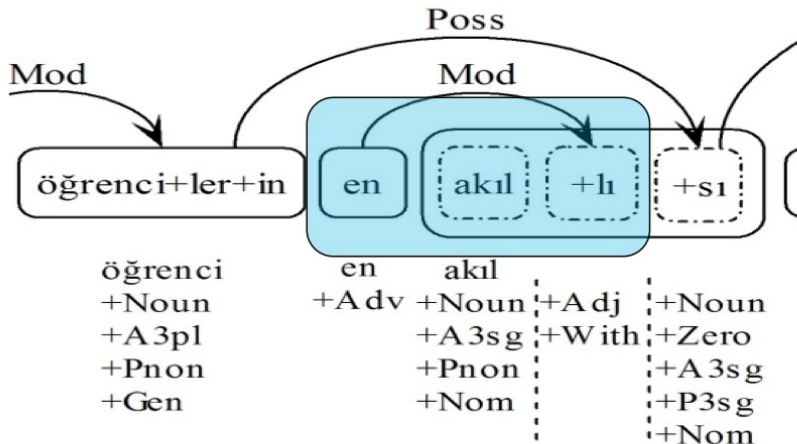
^DB +Verb+Past+A3sg

ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem



This school-at+that-is student-s-' most intelligence+with+of there
The most intelligent of the students in this school is the little girl standing there.

ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem



ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem

```
<?xml version="1.0" encoding="windows-1254" ?>
<Set sentences="1">
<S No="1">
<W IX="1" LEM="" MORPH="" IG="[(1,"Brecht+Noun+Prop+A3sg+Pnon+Abl")]' REL="[2,1,(ABLATIVE.ADJUNCT)]"> Brecht'ten </W>
<W IX="2" LEM="" MORPH="" IG="[(1,"yap+Verb+Pos")(2,"Adj+PastPart+P1sg")]' REL="[5,1,(MODIFIER)]"> yaptığım </W>
<W IX="3" LEM="" MORPH="" IG="[(1,"bu+Det")]' REL="[5,1,(DETERMINER)]"> bu </W>
<W IX="4" LEM="" MORPH="" IG="[(1,"uzun+Adj")]' REL="[5,1,(MODIFIER)]"> uzun </W>
<W IX="5" LEM="" MORPH="" IG="[(1,"alıntı+Noun+A3sg+Pnon+Abl")]' REL="[6,1,(OBJECT)]"> alıntıdan </W>
<W IX="6" LEM="" MORPH="" IG="[(1,"sonra+Postp+PCAbI")]' REL="[12,2,(MODIFIER)]"> sonra </W>
<W IX="7" LEM="" MORPH="" IG="[(1,".+Punc")]' REL="[,( )]"> , </W>
<W IX="8" LEM="" MORPH="" IG="[(1,"ev+Noun+A3sg+Pnon+Dat")]' REL="[9,1,(OBJECT)]"> eve </W>
<W IX="9" LEM="" MORPH="" IG="[(1,"ilişkin+Postp+PCDat")]' REL="[11,1,(MODIFIER)]"> ilişkin </W>
<W IX="10" LEM="" MORPH="" IG="[(1,"ben+Pron+PersP+A1sg+Pnon+Gen")]' REL="[11,1,(POSSESSOR)]"> benim </W>
<W IX="11" LEM="" MORPH="" IG="[(1,"ütopya+Noun+A3sg+P1sg+Dat")]' REL="[12,1,(DATIVE.ADJUNCT)]"> ütopyama </W>
<W IX="12" LEM="" MORPH="" IG="[(1,"gel+Verb+Pos")(2,"Verb+Able+Aor+A1pl")]' REL="[13,1,(SENTENCE)]"> gelebiliriz </W>
<W IX="13" LEM="" MORPH="" IG="[(1,".+Punc")]' REL="[,( )]"> . </W>
</S>
</Set>
```

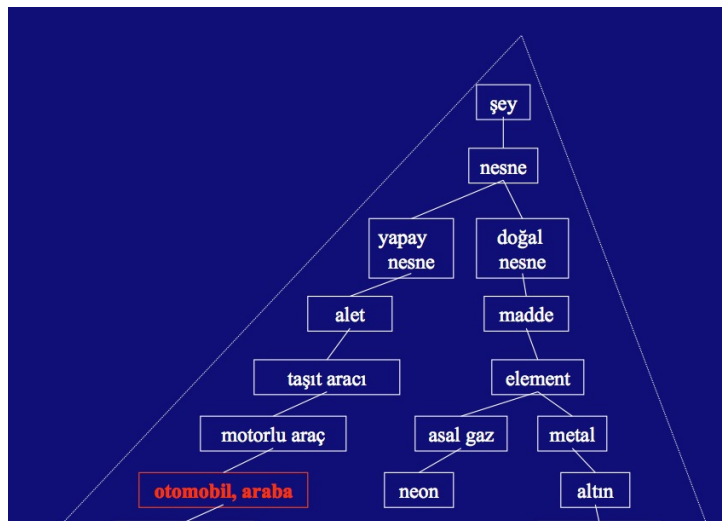
ODTÜ–Sabancı Türkçe Ağaç Yapılı Derlem

- CoNLL-X Shared Task: Multi-lingual Dependency Parsing
2006 ve 2007’de çok sayıda araştırmacı tarafından bağımlılık çözümlemesi yarışmalarında kullanıldı.
- Gülşen Eryiğit ve Ruken Çakıcı tarafından doktora tezlerinde kullanıldı.
- Boğaziçi’nde **Natural language watermarking via morphosyntactic alterations** adlı çalışmada kullanıldı (Meral ve diğerleri, Computer Speech and Language, 2009).

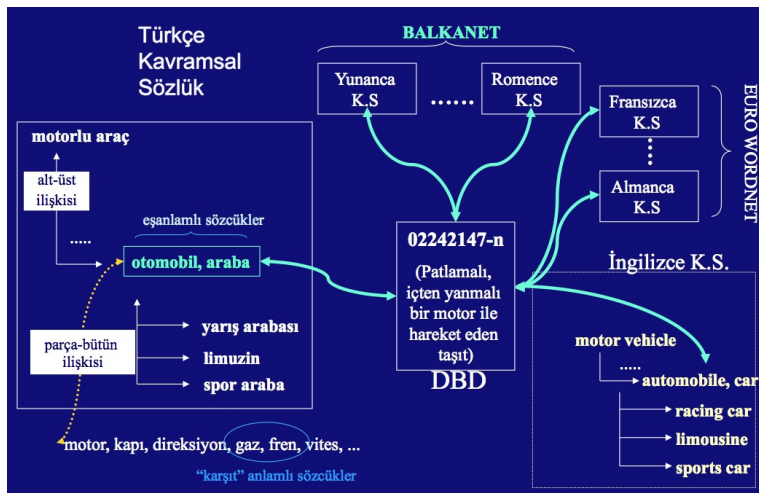
Türkçe Kavramsal Sözlük – WordNet

- Kavram – sözcük ilişkileri
- Kavramlararası ilişkiler
 - Altkavram, üstkavram
 - Parça, bütün
- Sözcüklerarası ilişkiler
 - Eş anlamlılık,
 - Karşıt anlamlılık
- $\approx 12,000$ kavram kümesi (synset)
- $\approx 16,000$ sözcük
- $\approx 17,500$ ilişki

Türkçe Kavramsal Sözlük – WordNet



Türkçe Kavramsal Sözlük – WordNet



Metin Derlemleri

- Bu derlemler genelde konuşma tanıma ve bazı başka doğal dil işleme uygulamaları için n-gram istatistiksel dil modeli oluşturmak için kullanılmakta.
 - Sabancı – \approx 125M sözcük
 - Boğaziçi – \approx 500M sözcük

Parallel Metin Derlemleri

- Bu derlemler istatistiksel çeviri sistemlerinin eğitimi amaçlı
 - $\approx 53K$ paralel tümce: İngilizce-Türkçe AB, BM vb. dokümanları (Sabancı)
 - $\approx 20K$ paralel tümce: İngilizce-Türkçe BTEC seyahat derlemi (TÜBİTAK)
 - İsveççe (150K sözcük) – Türkçe (126K sözcük) paralel metinler (Uppsala)
 - South East Balkan Times gazetesi
 - Opus Projesi

Ağaç Yapılı Paralel Derlemeler

- **The English-Swedish-Turkish Parallel Treebank**, (Megyesi ve diğerleri, LREC 2010)
 - İsveççe (160K sözcük) – Türkçe (145K sözcük) paralel metinler
 - Her iki taraftaki tümceler **biçimbilimsel ve sözdizimsel** olarak işaretlenmiş
 - Sözdizimsel çözümlemeler Eryiğit ve Nivre'nin çözümleyicileri ile
- **Işık Üniversitesi İngilizce - Türkçe Ağaç Yapılı Derlem** (Yıldız ve diğerleri, ACL 2014)
 - Penn Treebank çevirisi
 - Ağaç yapıları Penn Treebank'ten yarı otomatik olarak dönüştürülmüş

Tümce Çözümleme

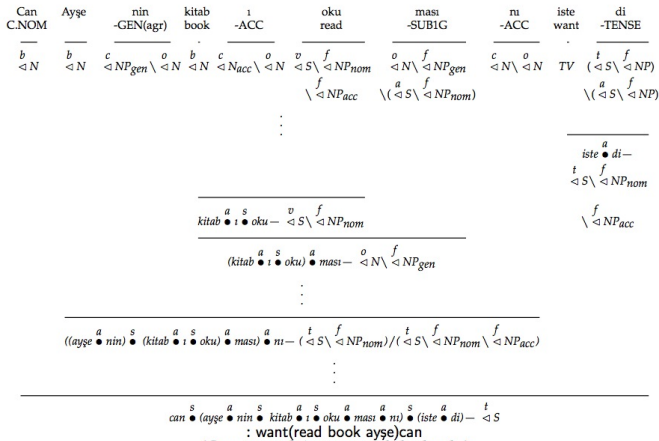
Türkçe Ağaç Yapılı Derlem'den otomatik olarak öğrenilen çözümleyiciler:

- Gülşen Cebiroğlu Eryiğit, *Dependency parsing of Turkish*, Doktora Tezi, İTÜ, 2006
- Ruken Çakıcı, *Wide-Coverage Parsing for Turkish*, Doktora Tezi, University of Edinburgh, 2008

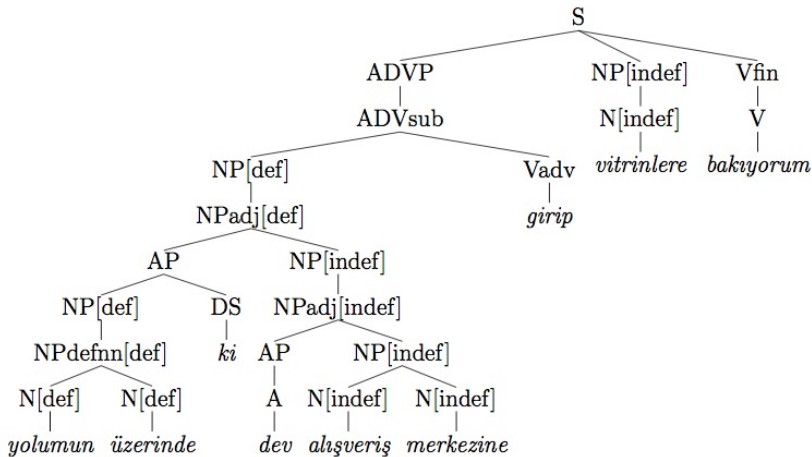
Derin Sözdizimsel Çözümleme

- Cem Bozşahin *The Combinatory Morphemic Lexicon*, Computational Linguistics, 2002
- Özlem Çetinoğlu, *A Large Scale LFG Grammar for Turkish*, Sabancı Üniversitesi, Doktora Tezi, 2009

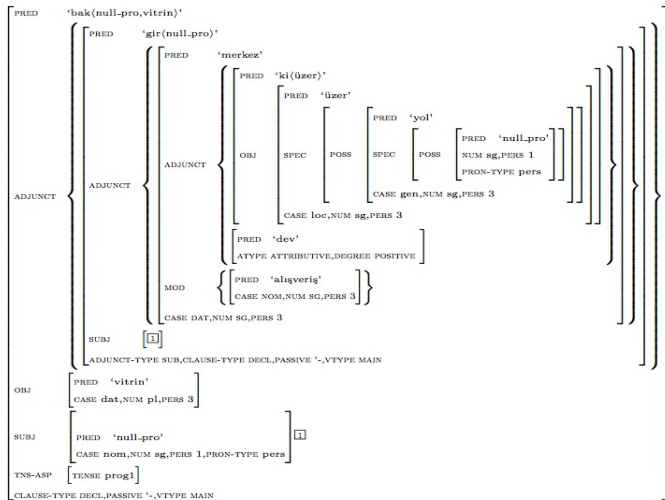
“Combinatory Categorical Grammar” Tabanlı Çözümleyici



Sözcüksel İşlevsel Gramer Tabanlı Çözümleyici

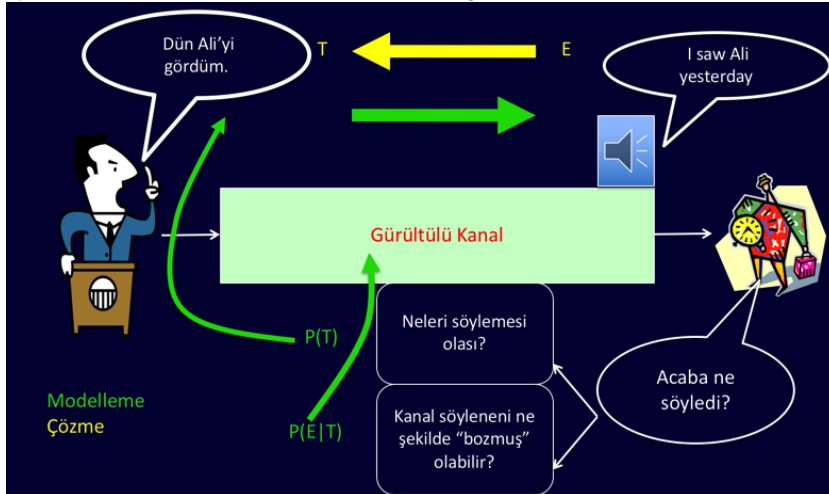


Sözcüksel İşlevsel Gramer Tabanlı Çözümleyici




İstatistiksel Çeviri

- Çeviri bir “Gürültülü Kanal” problemi olarak modellenir



İstatistiksel Çeviri

- Çeviri bir “Gürültülü Kanal” problemi olarak modellenir

I saw Ali yesterday	İyi Türkçe mi? $P(T)$	İyi Çeviri mi? $P(E T)$	Sonuç
Bugün Ali'ye gittim			
Okulda kalmışlar			
Var gelmek ben			
Dün Ali'yi gördüm			
Gördüm ben dün Ali'yi			
Dün Ali'ye gördüm			

İngilizce nasıl Türkçe'ye dönüşür?

İngilizce nasıl Türkçe'ye dönüşür?

if we will be able to make ... become strong

İngilizce nasıl Türkçe'ye dönüşür?

if we will be able to make ... become strong

if we will be able to make ... become strong

İngilizce nasıl Türkçe'ye dönüşür?

if we will be able to make ... become strong

if we will be able to make ... become strong

... strong become to make be able will if we

İngilizce nasıl Türkçe'ye dönüşür?

if we will be able to make ... become strong

if we will be able to make ... become strong

... strong become to make be able will if we

... sağlam +laş +tır +abil +ecek +se +k

İngilizce nasıl Türkçe'ye dönüşür?

if we will be able to make ... become strong

if we will be able to make ... become strong

... strong become to make be able will if we

... sağlam +laş +tır +abil +ecek +se +k



... sağlamlaştıracaksak

İstatistiksel Çeviri Çalışmaları

- Farklı gösterimler üzerinde deneyler
- İleri istatistiksel modellerle deneyler
- Kural ve istatistik tabanlı melez sistemler

İstatistiksel Çeviri Çalışmaları

- Türkçe ve İngilizce sözcükleri biçimbirimlerine bölerek çeviri modelleme.
 - Sistem İngilizce biçimbirim dizisi alıp Türkçe biçimbirim dizisi üretir (veya tersi)
- İlknur Durgar-El Kahlout, *Statistical Machine Translation from English to Turkish*, Doktora Tezi, Sabancı Üniversitesi, 2009
- Coşkun Mermer vd., *The TÜBİTAK–UEKAE Statistical Machine Translation System for IWSLT 2009*, IWSLT 2009
- Coşkun Mermer ve Ahmet Akın, *Unsupervised Search for The Optimal Segmentation for Statistical Machine Translation*, ACL 2010

İstatistiksel Çeviri Çalışmaları

- İngilizce öbek – Türkçe sözcük eşleştirmesi:
 - İngilizce tümce üzerinde sözdizimsel çözümleme
 - İşlevsel sözcüklerin bağlı oldukları sözcüklere sanki biçimbirimmiş gibi eklenince oluşan yapıların Türkçe sözcüklere çevrilmesi (Bir nevi İngilizce'yi Türkçeleştirme!)
- Reyhan Yeniterzi **Syntax-to-Morphology Mapping and Syntactic Reordering in English to Turkish Phrase-based SMT**, Yüksek Lisans Tezi, Sabancı Üniversitesi, 2009
- Yeniterzi ve Oflazer **Syntax-to-Morphology Mapping in Factored Phrase-Based Statistical Machine Translation from English to Turkish**, ACL 2010

İstatistiksel Çeviri Çalışmaları

- İleri istatistiksel modelleme yöntemleri kullanarak daha sağlıklı sözcük eşleştirme ve sonucunda daha iyi çeviri
- Coşkun Mermer, **Unsupervised Learning of Word Alignments for Statistical Machine Translation**, Boğaziçi Üniversitesi, Doktora Tezi, Devam etmekte (yarına kadar (-:-))
- Coşkun Mermer ve Murat Saraçlar, **Bayesian Word Alignment for Statistical Machine Translation**, ACL 2011

Çeviri Çalışmaları – Melez Yöntemler

- Ferhan Türe, Hybrid MT System from Turkish to English, Yüksek Lisans Tezi, Sabancı Üniversitesi, 2008.

Çeviri Çalışmaları – Melez Yöntemler

- Ferhan Türe, Hybrid MT System from Turkish to English, Yüksek Lisans Tezi, Sabancı Üniversitesi, 2008.
 - Türkçe-İngilizce parallel gramer ile kural-tabanlı tümce/öbek çözümleme ve çoklu aktarım

Çeviri Çalışmaları – Melez Yöntemler

- Ferhan Türe, **Hybrid MT System from Turkish to English**, Yüksek Lisans Tezi, Sabancı Üniversitesi, 2008.
 - Türkçe-İngilizce **parallel gramer** ile kural-tabanlı tümce/öbek çözümleme ve çoklu aktarım
 - İngilizce tarafında istatistiksel modellerle en olası tümcenin seçilmesi

Çeviri Çalışmaları – Melez Yöntemler

- Ferhan Türe, **Hybrid MT System from Turkish to English**, Yüksek Lisans Tezi, Sabancı Üniversitesi, 2008.
 - Türkçe-İngilizce **parallel gramer** ile kural-tabanlı tümce/öbek çözümleme ve çoklu aktarım
 - İngilizce tarafında istatistiksel modellerle en olası tümcenin seçilmesi
- Cüneyd Tantuğ, **A hybrid machine translation model between close agglutinative languages**, İTÜ, Doktora Tezi,, 2007

Çeviri Çalışmaları – Melez Yöntemler

- Ferhan Türe, **Hybrid MT System from Turkish to English**, Yüksek Lisans Tezi, Sabancı Üniversitesi, 2008.
 - Türkçe-İngilizce **parallel gramer** ile kural-tabanlı tümce/öbek çözümleme ve çoklu aktarım
 - İngilizce tarafında istatistiksel modellerle en olası tümcenin seçilmesi
- Cüneyd Tantuğ, **A hybrid machine translation model between close agglutinative languages**, İTÜ, Doktora Tezi,, 2007
 - Türkmençe tarafında biçimbilimsel çözümleme ve Türkçe'ye çoklu aktarım

Çeviri Çalışmaları – Melez Yöntemler

- Ferhan Türe, **Hybrid MT System from Turkish to English**, Yüksek Lisans Tezi, Sabancı Üniversitesi, 2008.
 - Türkçe-İngilizce **parallel gramer** ile kural-tabanlı tümce/öbek çözümleme ve çoklu aktarım
 - İngilizce tarafında istatistiksel modellerle en olası tümcenin seçilmesi
- Cüneyd Tantuğ, **A hybrid machine translation model between close agglutinative languages**, İTÜ, Doktora Tezi,, 2007
 - Türkmençe tarafında biçimbilimsel çözümleme ve Türkçe'ye çoklu aktarım
 - Türkçe tarafında istatistiksel modellerle en olası tümcenin seçilmesi

Bazı uygulamalar

- Soru cevaplama sistemleri
- Duygu çözümlemesi
- Haber klipleri arama sistemleri

Reklam arası! Ne durumdayız?

Reklam arası! Ne durumdayız?



Reklam arası! Ne durumdayız?

1	Turkish and Its Challenges for Language and Speech Processing	1
	Kemal Oflazer and Murat Saraçlar	
2	Morphological Processing for Turkish	21
	Kemal Oflazer	
3	Morphological Disambiguation for Turkish	53
	Dilek Zeynep Hakkani-Tür, Murat Saraçlar, Gökhan Tür, Kemal Oflazer, and Deniz Yuret	
4	Language Modeling for Turkish Text and Speech Processing	69
	Ebru Arısoy and Murat Saraçlar	
5	Turkish Speech Recognition	93
	Ebru Arısoy and Murat Saraçlar	
6	Turkish Named-Entity Recognition	115
	Reyyan Yeniterzi, Gökhan Tür, and Kemal Oflazer	

Reklam arası! Ne durumdayız?

7	Dependency Parsing of Turkish	133
	Gülşen Eryiğit, Joakim Nivre, and Kemal Oflazer	
8	Wide-Coverage Parsing, Semantics and Morphology	153
	Ruket Çakıcı, Mark Steedman, and Cem Bozşahin	
9	Deep Parsing of Turkish with Lexical-Functional Grammar	175
	Özlem Çetinoğlu and Kemal Oflazer	
10	Statistical Machine Translation and Turkish	207
	Kemal Oflazer, Reyyan Yeniterzi, and İlknur Durgar-El Kahlout	
11	Machine Translation Between Turkic Languages	237
	A. Cüneyd Tantuğ and Eşref Adalı	
12	Sentiment Analysis in Turkish	255
	Gizem Gezici and Berrin Yanıkoğlu	

Reklam arası! Ne durumdayız?

13	The Turkish Treebank	273
	Gülşen Eryiğit, Kemal Oflazer, and Umut Sulubacak	
14	Linguistic Corpora: A View from Turkish.....	291
	Mustafa Aksan and Yeşim Aksan	
15	Turkish Wordnet.....	317
	Özlem Çetinoğlu, Orhan Bilgin, and Kemal Oflazer	
16	Turkish Discourse Bank: Connectives and Their Configurations	337
	Deniz Zeyrek, Işın Demirşahin, and Cem Bozşahin	



Önümüzdeki Dönem için Öneriler

- Çok daha büyük bir ağaç yapılı derlem: $> 100K$ tümce

Önümüzdeki Dönem için Öneriler

- Çok daha büyük bir ağaç yapılı derlem: $> 100K$ tümce
 - Daha yüksek başarı oranına sahip sözdizimsel çözümleyicilerin eğitimi, ve

Önümüzdeki Dönem için Öneriler

- Çok daha büyük bir ağaç yapılı derlem: $> 100K$ tümce
 - Daha yüksek başarı oranına sahip sözdizimsel çözümleyicilerin eğitimi, ve
- Büyük boyutta paralel derlem oluşturulması ($> 100M$ paralel tümce)

Önümüzdeki Dönem için Öneriler

- Çok daha büyük bir ağaç yapılı derlem: $> 100K$ tümce
 - Daha yüksek başarı oranına sahip sözdizimsel çözümleyicilerin eğitimi, ve
- Büyük boyutta paralel derlem oluşturulması ($> 100M$ paralel tümce)
 - Tahminim bunun yakın bir zamanda olanaksız olduğu (“doğal” kaynak yok!)

Önümüzdeki Dönem için Öneriler

- Çok daha büyük bir ağaç yapılı derlem: $> 100K$ tümce
 - Daha yüksek başarı oranına sahip sözdizimsel çözümleyicilerin eğitimi, ve
- Büyük boyutta paralel derlem oluşturulması ($> 100M$ paralel tümce)
 - Tahminim bunun yakın bir zamanda olanaksız olduğu (“doğal” kaynak yok!)
- Çok daha büyük bir dil modeli derlemi oluşturulması ($> 5G$ sözcük)

Önümüzdeki Dönem için Öneriler

- Çok daha büyük bir ağaç yapılı derlem: $> 100K$ tümce
 - Daha yüksek başarı oranına sahip sözdizimsel çözümleyicilerin eğitimi, ve
- Büyük boyutta paralel derlem oluşturulması ($> 100M$ paralel tümce)
 - Tahminim bunun yakın bir zamanda olanaksız olduğu (“doğal” kaynak yok!)
- Çok daha büyük bir dil modeli derlemi oluşturulması ($> 5G$ sözcük)
- Wordnet’in büyütülmesi

Önümüzdeki Dönem için Öneriler

Yöntem geliştirme

- Türkçe gibi dillerde standart n-gram modellerin ötesinde, kök sözcük ve biçimbirimleri ayrı ayrı kullanan ayrıştırılabilir dil modelleme tekniklerinin geliştirilmesi.

Önümüzdeki Dönem için Öneriler

Yöntem geliştirme

- Türkçe gibi dillerde standart n-gram modellerin ötesinde, kök sözcük ve biçimbirimleri ayrı ayrı kullanan ayrıştırılabilir dil modelleme tekniklerinin geliştirilmesi.
- “Stanford NLP Tools” benzeri bir Türkçe (ön)işleme silsilesinin geliştirilmesi.

Önümüzdeki Dönem için Öneriler

Yöntem geliştirme

- Türkçe gibi dillerde standart n-gram modellerin ötesinde, kök sözcük ve biçimbirimleri ayrı ayrı kullanan ayrıştırılabilir dil modelleme tekniklerinin geliştirilmesi.
- “Stanford NLP Tools” benzeri bir Türkçe (ön)işleme silsilesinin geliştirilmesi.
 - İTÜ Web Servisi

Önümüzdeki Dönem için Öneriler

Yöntem geliştirme

- Türkçe gibi dillerde standart n-gram modellerin ötesinde, kök sözcük ve biçimbirimleri ayrı ayrı kullanan ayrıştırılabilir dil modelleme tekniklerinin geliştirilmesi.
- “Stanford NLP Tools” benzeri bir Türkçe (ön)işleme silsilesinin geliştirilmesi.
 - İTÜ Web Servisi
 - + Varlık Referans Eşleştirilmesi

Önümüzdeki Dönem için Öneriler

Yöntem geliştirme

- Türkçe gibi dillerde standart n-gram modellerin ötesinde, kök sözcük ve biçimbirimleri ayrı ayrı kullanan ayrıştırılabilir dil modelleme tekniklerinin geliştirilmesi.
- “Stanford NLP Tools” benzeri bir Türkçe (ön)işleme silsilesinin geliştirilmesi.
 - İTÜ Web Servisi
 - + Varlık Referans Eşleştirilmesi
 - + Olay referans Eşleştirilmesi

Önümüzdeki Dönem için Öneriler

Yöntem geliştirme

- Türkçe gibi dillerde standart n-gram modellerin ötesinde, kök sözcük ve biçimbirimleri ayrı ayrı kullanan ayrıştırılabilir dil modelleme tekniklerinin geliştirilmesi.
- “Stanford NLP Tools” benzeri bir Türkçe (ön)işleme silsilesinin geliştirilmesi.
 - İTÜ Web Servisi
 - + Varlık Referans Eşleştirilmesi
 - + Olay referans Eşleştirilmesi
 - + Anlamsal Rollerin Çıkarılması

Önümüzdeki Dönem için Öneriler

Yöntem geliştirme

- Sosyal medya (Facebook, Twitter, bloglar vb.) gibi, çok farklı yazım biçimlerinin kendinden oluştuğu ortamlar için çok daha gürbüz işleme tekniklerinin geliştirilmesi.

Önümüzdeki Dönem için Öneriler

Yöntem geliştirme

- Sosyal medya (Facebook, Twitter, bloglar vb.) gibi, çok farklı yazım biçimlerinin kendinden oluştuğu ortamlar için çok daha gürbüz işleme tekniklerinin geliştirilmesi.
- Tüm bu teknik ve kaynakların diğer Türk dillerine aktarılması

Önümüzdeki Dönem için Öneriler – Uygulamalar

- Açık uçlu soru-cevap sistemlerinin geliştirilmesi

Önümüzdeki Dönem için Öneriler – Uygulamalar

- Açık uçlu soru-cevap sistemlerinin geliştirilmesi
- Siri veya Google Assistant'a tümleşik Türkçe diyalog sistemlerinin geliştirilmesi

Önümüzdeki Dönem için Öneriler – Uygulamalar

- Açık uçlu soru-cevap sistemlerinin geliştirilmesi
- Siri veya Google Assistant'a tümleşik Türkçe diyalog sistemlerinin geliştirilmesi
- Derin öğrenme için, biçimbirimsel yapıyı dikkate alan birleştirici (compositional) sözcük gömme (word embedding) modellerinin kuramsal olarak geliştirilmesi

Önümüzdeki Dönem için Öneriler – Uygulamalar

- Açık uçlu soru-cevap sistemlerinin geliştirilmesi
- Siri veya Google Assistant'a tümleşik Türkçe diyalog sistemlerinin geliştirilmesi
- Derin öğrenme için, biçimbirimsel yapıyı dikkate alan birleştirici (compositional) sözcük gömme (word embedding) modellerinin kuramsal olarak geliştirilmesi
- Bu modellerin tümce çözümleme, derin öğrenme ile çeviri vb. gibi problemlere uygulanması.

Teşekkürler

- Sorularınız?