

TÜRKÇE İÇİN GÖZETİMSİZ SÖZDİZİMSEL BELİRSİZLİK GİDERME^[1]

Araş. Gör. Dr. Özkan ASLAN¹

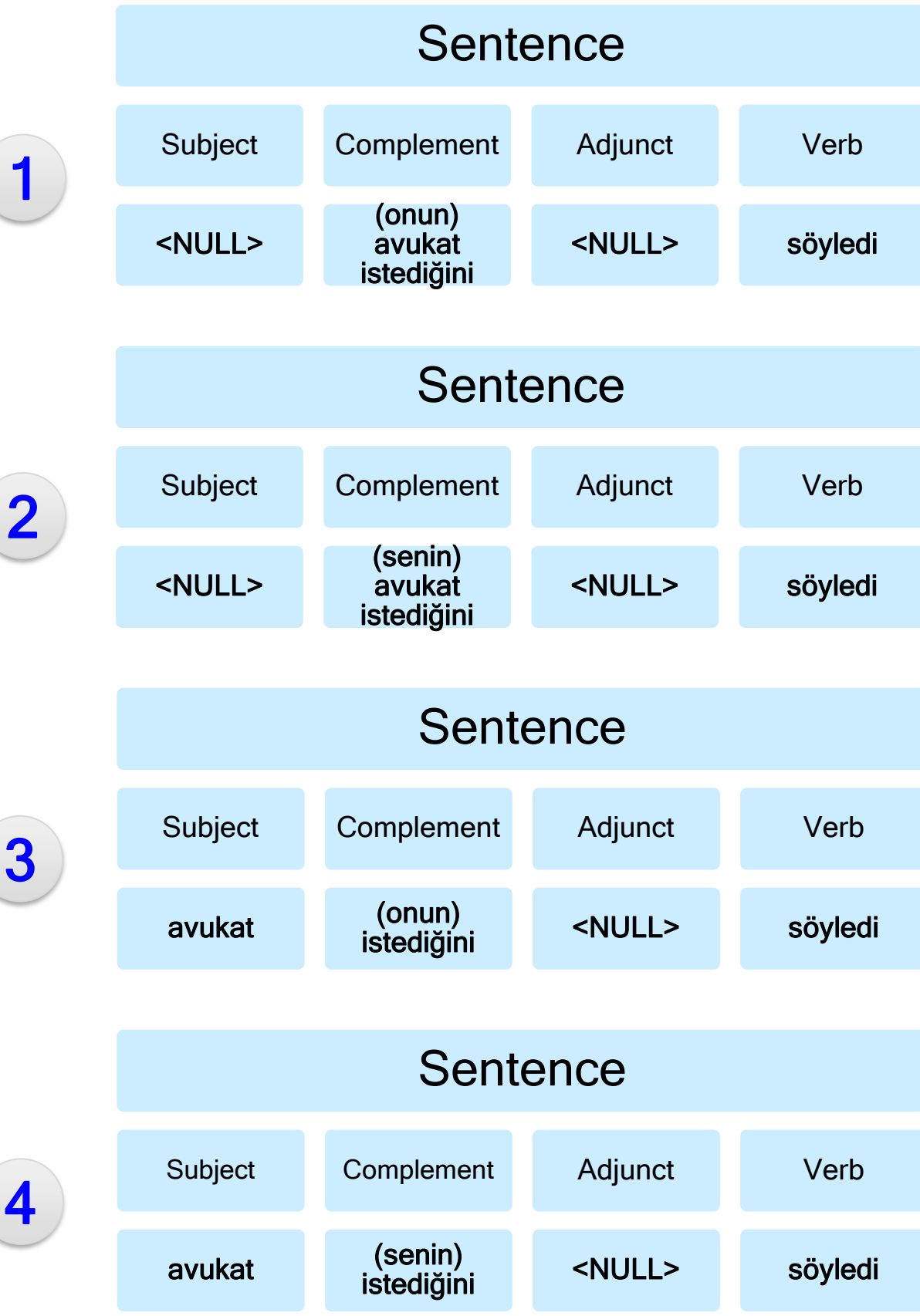
Doktora Tezi, Anadolu Üniversitesi, 2017

Danışmanlar: Doç. Dr. Serkan GÜNAL, Doç. Dr. B. Taner DİNÇER

Giriş

Bir tümce birden çok sözdizim ağacıyla temsil edilebilir.

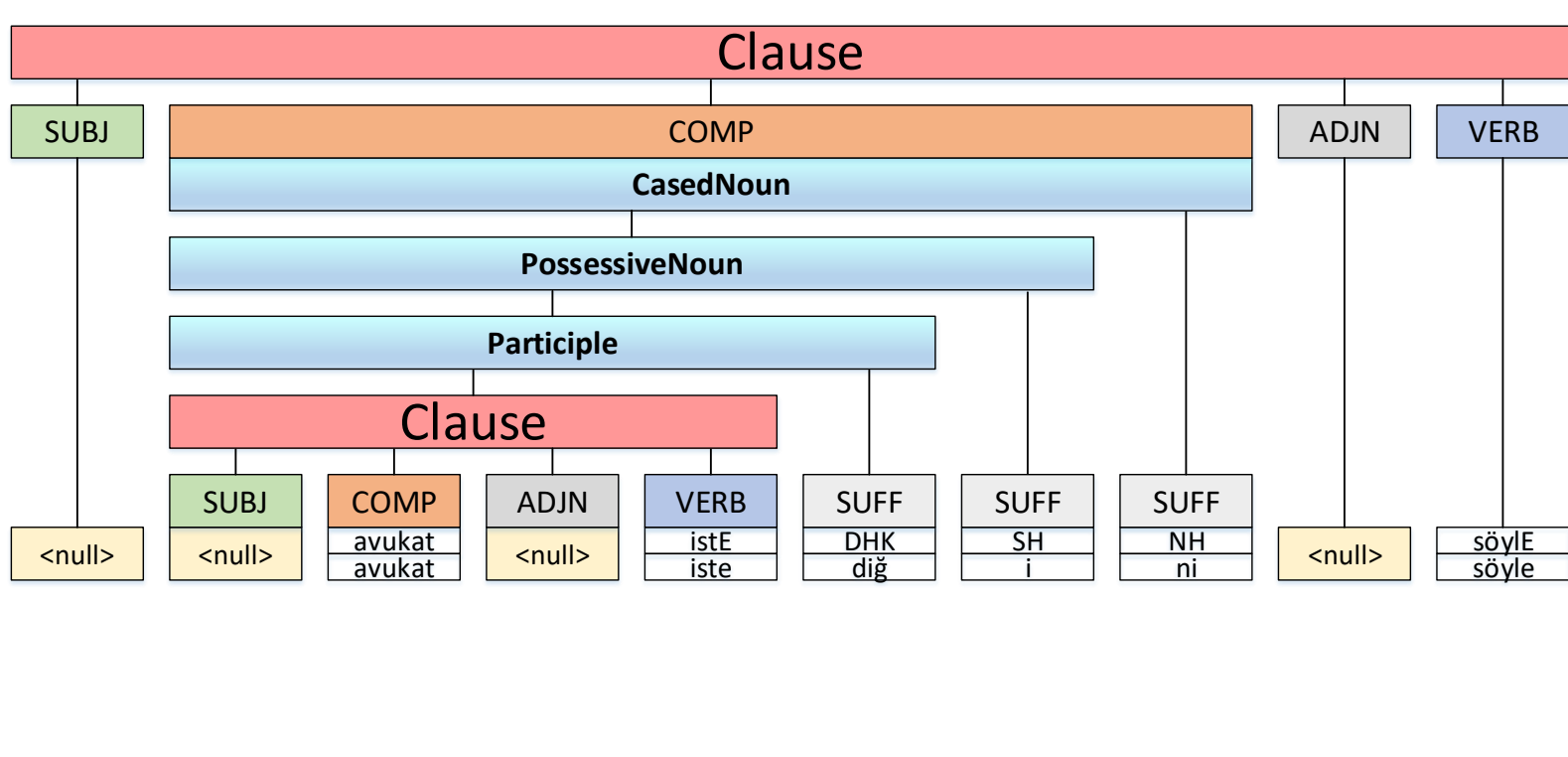
«Avukat istediğini söyledi.»



- Sözdizimsel belirsizlik giderme, bu tezde ağaçları sıralama problemi olarak yorumlandı.
- Bu sıralamada biçimbilimsel ve sözdizimsel açıdan en olası ağaç en üstte yer alır.

Yapı

- Tezdeki ağaç yapıları Öbek Yapı Dilbilgisine (Phrase Structure Grammar) dayanır.
- Fakat özne, tümlec (complement), eklenti (adjunct) gibi işlevsel tanımlamaları da içerir.



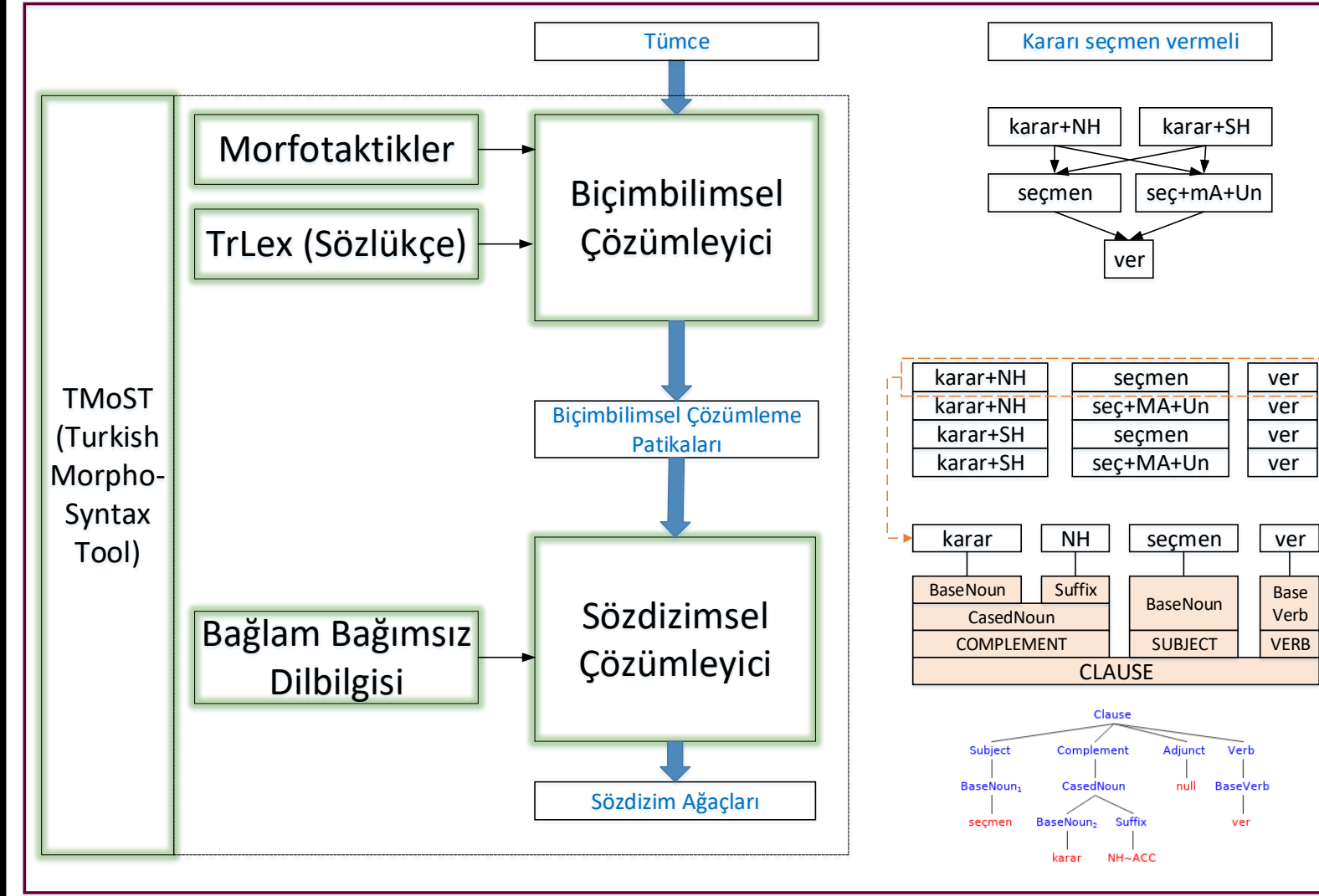
Kısıtlılıklar

- Dil: Türkçe
- Bağlam: Tümce
- Noktalama işaretleri hariç
- Özel adlar hariç
- Yalnızca eylem sonlu dizilimler (SOV/OSV)
- Anlamsal işaretleme yok

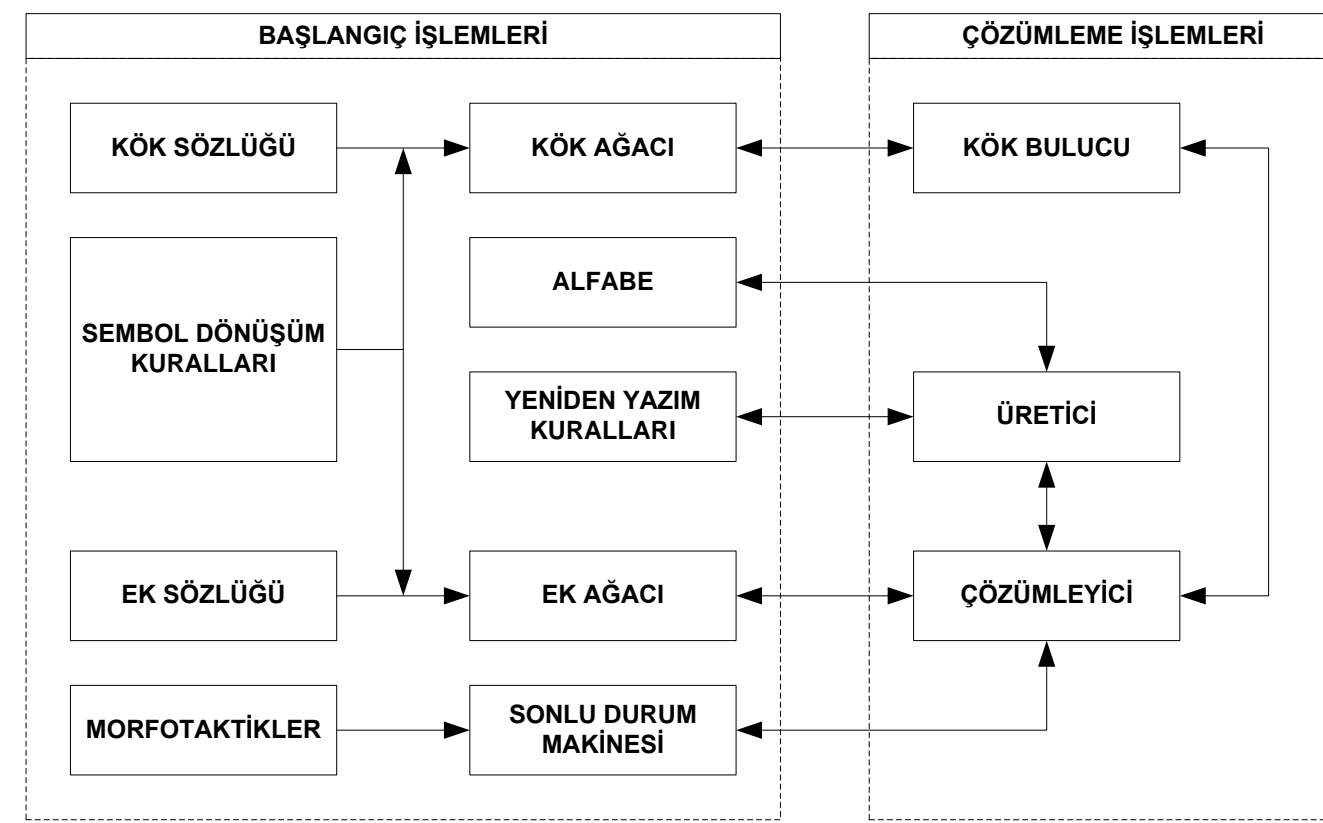
Kaynaklar

- Aslan, Ö., 2017. Türkçe için gözetimsiz sözdizimsel belirsizlik giderme (Basılmamış Doktora Tezi, Anadolu Üniversitesi).
- Aslan, O., Gunal, S., & Dincer, B. T. (2018). A computational morphological lexicon for Turkish: TrLex. *Lingua*, 206, 21-34.
- Say, B., Zeyrek, D., Oflazer, K. ve Özge, U. (2002). Development of a Corpus and a Treebank for Present day Written Turkish. Proceedings of the Eleventh International Conference of Turkish Linguistics, pp. 183-192. Eastern Mediterranean University, Cyprus.

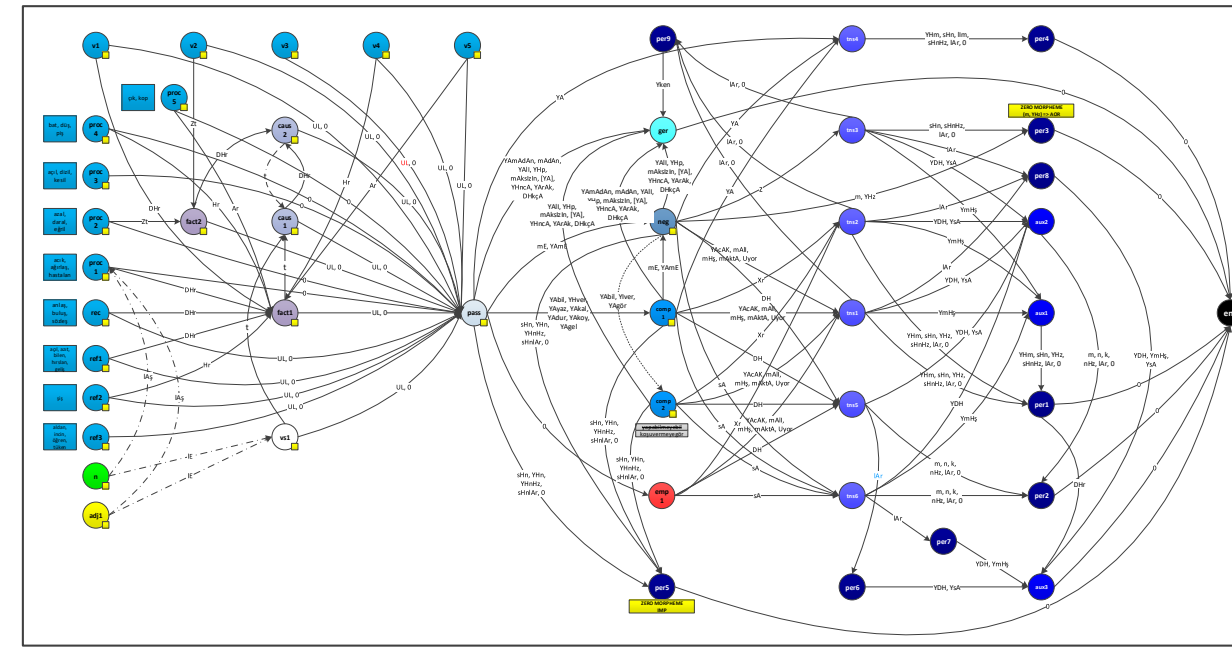
Dizge (TMOst)



Biçimbilimsel Çözümleyici (Morfolog)



Morfotaktikler



Sözlükçe (TrLex^[2])

	Anlam	Sözlükbiçim
Toplam	113.069	83.763
Tekil Sözcük	78.458	53.499
Çoklu Sözcük	34.611	30.301

Eylemler:

Yanulam	Örnek
NULL	11.856 avare ol, enayitç, mirasa kon-, sürç-
ABL	478 bahis aç, havalan-, nasibini al-, olu-
ACC	3.893 başar-, ertele-, görev say-, konuştur-
DAT	2.338 çanak tut-, hazırlan-, seslen-, uğur getir-
ILE	430 boy ölçüş-, evlen-, ilgilen-, konuş-
LOC	181 dinlen-, ikamet et-, sahne al-, rol al-
NOM	217 aç-, dök-, kazan-, say-, taşı-, yap-
<Karma>	3.419
Toplam	22.812

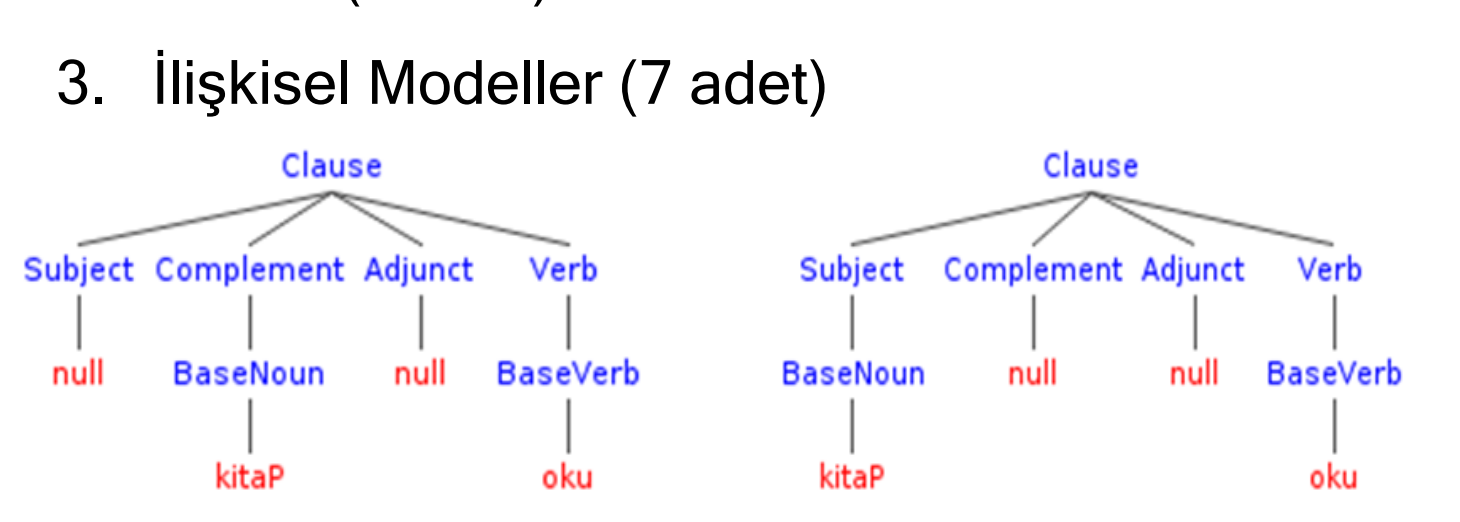
Veri

- (Eğitim) Metin Koleksiyonu
 - İnternet kaynaklı 42.630.365 tümce
- (Sınama) Ağaç Yapılı Derlem (AUT)
 - METU Corpus'tan^[3] 510 tümce

Model

1. Dil Modelleri

- Biçimbirim, gövde, son ek, kanal
- 1, 2, 3 gram



- Modeller tek tek ve birleşik olarak ele alındı.
- İki tür model birleştirme yapıldı:
 - Eşit oylama: İleri yönlü öznitelik seçimi
 - Ağırlıklı oylama: Genetik algoritmalar

Deney

4 başarı ölçüsü kullanıldı:

T ₁	T ₂	T ₃
1 Ağaç	1 Ağaç	1 Ağaç
2 Ağaç	2 Ağaç	2 Ağaç
3 Ağaç	3 Ağaç	3 Ağaç
4 Ağaç	4 Ağaç	4 Ağaç
5 Ağaç	5 Ağaç	5 Ağaç
6 Ağaç	6 Ağaç	6 Ağaç
7 Ağaç	7 Ağaç	7 Ağaç
8 Ağaç	8 Ağaç	8 Ağaç

Ortalama: 3

2/5=0,4 6/8=0,75 1/2=0,5 Ortanca 0,5

5 8 2 Ortanca 5

Bağlı Ortanca: 0,5x5=2,5

T ₁	T ₂	T ₃
1 Ağaç	1 Ağaç	1 Ağaç
2 Ağaç	2 Ağaç	2 Ağaç
3 Ağaç	3 Ağaç	3 Ağaç
4 Ağaç	4 Ağaç	4 Ağaç
5 Ağaç	5 Ağaç	5 Ağaç
6 Ağaç	6 Ağaç	6 Ağaç
7 Ağaç	7 Ağaç	7 Ağaç
8 Ağaç	8 Ağaç	8 Ağaç

Ortalama Benzerlik: 0,85

T ₁	T ₂	T ₃
1 Ağaç	1 Ağaç	1 Ağaç
2 Ağaç	2 Ağaç	2 Ağaç
3 Ağaç	3 Ağaç	3 Ağaç
4 Ağaç	4 Ağaç	4 Ağaç
5 Ağaç	5 Ağaç	5 Ağaç
6 Ağaç	6 Ağaç	6 Ağaç
7 Ağaç	7 Ağaç	7 Ağaç
8 Ağaç	8 Ağaç	8 Ağaç

Bağlılık: 0,30 (1,2):0,75 (1,6):0,80 (1,1):1,00

Bağlılık: 0,55 (1,2):0,75 (1,6):0,80 (1,1):1,00

Ortalama Bağlılık: 0,62

Sonuç

Tekil Modeller:

Model	Ortalama	B. Ortanca	Benzerlik	Bağlılık
Biçimbirim_3	4,61	3,33	0,65	0,30
*OBBD	5,13	4,00	0,64	0,29
Biçimbirim_2	4,70	3,56	0,64	0,29
Kanal_1	5,26	4,00	0,61	0,24
Kanal_2	5,59	4,21	0,61	0,23
Biçimbirim_1	5,62	4,00	0,63	0,23
Kanal_3	6,18	5,33	0,60	0,19
Son ek_1	6,86	5,71	0,60	0,17
Son ek_2	6,98	6,00	0,58	0,15
Son ek_3	7,00	6,00	0,58	0,15

Genel Değerlendirme:

	Ortalama	B. Ortanca	Benzerlik	Bağlılık
Taban	8,76	8,36	0,5498	0,0137
Deneyler	4,61	3,33	0,6500	0,3000
Râhin	2,11	1,45	0,7752	0,3289
OB + BBG	1,25	1,14	0,9175	0,5763

Teşekkür

Bu tez çalışması Anadolu Üniversitesi BAP Komisyonunca kabul edilen 1410F415 no.lu proje kapsamında desteklenmiştir.

1ozkanaslan@mu.edu.tr

Doktora Sonrası Araştırmacı, Bilişsel Bilimler, Enformatik Enstitüsü

oz82 github

