

Collaborative Annotation Tool

Furkan Akkurt, Büşra Marşan, Suzan Üsküdarlı
Student Conference on Linguistics (SCOL) 2023
May 5-6, Istanbul

Introduction

- Treebanks for NLP
- High effort
- Annotation tools
- Presenting a collaborative grammatical annotation tool (BoAT)

Grammatical annotation

- POS, features and dependencies
- Universal Dependencies (UD)

```

# text = Maksimum miktar kişi başı 5.000 dolardır.
# text_en = $5,000 per person, the maximum allowed.
1  Maksimum      Maksimum      ADJ JJ  Number=Sing 2  amod  _  _
2  miktar miktar  NOUN    NN  Case=Nom|Number=Sing 6  nsubj  _  _
3  kişi  kişi  NOUN    NN  Number=Sing 4  nmod:poss  _  _
4  başı  baş  NOUN    NN  Number=Sing|Number[psor]=Sing|Person[psor]=3 6  amod  _  _
5  5.000  5.000  NUM CD  Number=Sing 6  nummod  _  _
6-7 dolardır  _  _  _  _  _  _  SpaceAfter=No
6  dolar  do  NOUN    NN  Number=Sing 0  root  _  _
7  dır i  AUX AUX  Aspect=Perf|Mood=Gen|Number=Sing|Person=3|Tense=Pres 6  cop  _  _
8  .  .  PUNCT  .  _  6  punct  _  _

```

UD example with ID, form, lemma, POS, features, dependency relations

```

# text = Maksimum miktar kişi başı 5.000 dolardır.
# text_en = $5,000 per person, the maximum allowed.
1  Maksimum      Maksimum      ADJ JJ  Number=Sing 2  amod      _      _
2  miktar miktar  NOUN    NN  Case=Nom|Number=Sing 6  nsubj    _      _
3  kişi   kişi   NOUN    NN  Number=Sing 4  nmod:poss  _      _
4  başı   baş  NOUN    NN  Number=Sing|Number[psor]=Sing|Person[psor]=3 6  amod      _      _
5  5.000  5.000  NUM CD  Number=Sing 6  nummod    _      _
6-7 dolardır _ _ _ _ _ _ _ SpaceAfter=No
6  dolar  do  NOUN    NN  Number=Sing 0  root      _      _
7  dır i  AUX AUX  Aspect=Perf|Mood=Gen|Number=Sing|Person=3|Tense=Pres 6  cop      _      _
8  . .  PUNCT  .  _ 6  punct    _      _

# text = $5,000 per person, the maximum allowed.
1  $  $  SYM $  _ 0  root 0:root SpaceAfter=No
2  5,000 5,000 NUM CD NumType=Card 1 nummod 1:nummod _
3  per per ADP IN _ 4 case 4:case _
4  person person NOUN NN Number=Sing 1 nmod 1:nmod:per SpaceAfter=No
5  , , PUNCT , _ 1 punct 1:punct _
6  the the DET DT Definite=Def|PronType=Art 7 det 7:det _
7  maximum maximum NOUN NN Number=Sing 1 appos 1:appos _
8  allowed allow VERB VBN Tense=Past|VerbForm=Part 7 acl 7:acl SpaceAfter=No
9  . . PUNCT . _ 1 punct 1:punct _

```

Difference in PUD

Agglutinative challenges

- Stacking morphemes
- Extensive features
- Free word order increasing arc length

Stacked morphemes example

Masadaki kitaplarını görmüştüm. (EN: *I had seen your books that were on the table.*)

Masa-da-ki	kitap-lar-ın-ı	gör-müş-tü-m.
table-LOC-DER	book-PL-POSS.2SG-ACC	see-ANT-PST-1SG

Existent tools

- Categorization
 - Accessibility
 - Collaboration
 - Automation
- brat
- ITU Treebank Annotation Tool
- WebAnno

BoAT v1

- Keyboard-centric
- Used for UD Turkish BOUN treebank (~10k)
- Lemma splitting
- Improving further

Data Viewer

Prev Reset Add Row Delete Row 167 Go Write your note here... Next

ID FORM LEMMA UPOS XPOS FEATS HEAD DEPREL DEPS MISC
 Abbr Animacy Aspect Case Clusivity Definite Degree Echo Evident Foreign Gender Mood NounClass Number
 Number[psor] NumType Person Person[psor] Polarity Polite Poss PronType Reflex Register Tense VerbForm Voice

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS	MISC
+ 1	Sel	sel	NOUN	_	Case=Nom Number=Sing Person=3	2	nmod:poss	_	_
+ 2	sularında	su	NOUN	_	Case=Loc Number=Plur Number[psor]=Sing Person=3 Person[psor]=3	4	obl	_	_
+ 3	neler	ne	PRON	Ques	Case=Nom Number=Plur Person=3 PronType=Int	4	nsubj	_	_
- 4-5	yoktu	_	-	-	-	-	-	-	-
+ 4	yok	yok	NOUN	Exist	Number=Sing Person=3 Polarity=Neg	0	root	_	_
+ 5	tu	y	AUX	Zero	Aspect=Perf Evident=Fh Number=Sing Person=3 Polarity=Pos Tense=Past	4	cop	_	_
+ 6	ki	ki	PART	Emph	_	4	advmod:emph	_	SpaceAfter=No
+ 7	PUNCT	TDots	_	4	punct	_	SpacesAfter=\n

Sel(1) sularında(2) neler(3) yoktu(+5) yok(4) tu(5) ki(6) ...(7)

Sel sularında neler yoktu ki...

```

graph TD
    Root["yok  
NOUN  
root"]
    Sularında["sularında  
NOUN  
obl"]
    Neler["neler  
PRON  
nsubj"]
    Tu["tu  
AUX  
cop"]
    Ki["ki  
PART  
advmod:emph"]
    Dots["...  
PUNCT  
punct"]
    Sel["Sel  
NOUN  
nmod:poss"]

    Root --- Sularında
    Root --- Neler
    Root --- Tu
    Root --- Ki
    Root --- Dots
    Sularında --- Sel
  
```

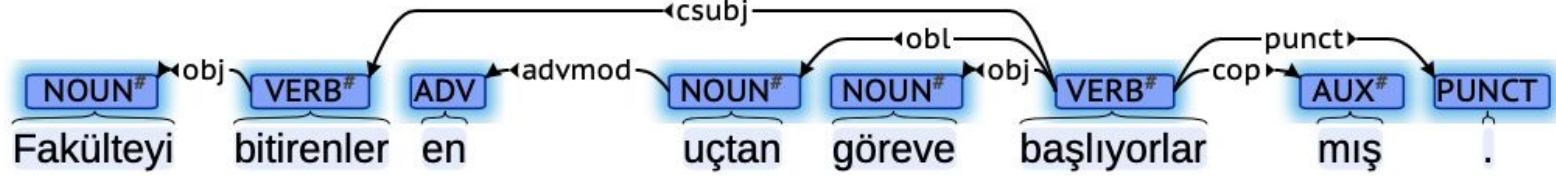
Annotating on BoAT v1

Requirements & Design

- Support for sentence annotation
 - Autocompletion
 - Splitting lemmas
- Improve user experience
- Screen use, customizations
- Keyboard
- Collaboration
- Search

Features

- UD-compliant
- Graph visualization
- Multiple treebanks
- Validation
- Annotation status
- Keyboard shortcuts
- Advanced search
- Inter-annotator agreement



Dependency graph

Errors

ID 7: 'PUNCT' must be 'punct' but it is 'obl'

Errors

Usage

Profile



Go to sentence



Columns



i

Fakülteyi bitirenler en uçtan göreve başlıyorlar mış .

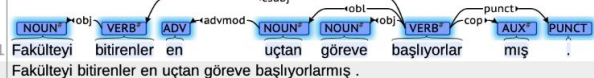
1 2 3 4 5 6 7 8

Status: New

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS	MISC
1	Fakülteyi	fakülte	NOUN	_	Case=Acc Number=Sing Person=3	2	obj	_	_
2	bitirenler	bitir	VERB	Ptcp	Polarity=Pos VerbForm=Part	6	csubj	_	_
3	en	en	ADV	_	_	4	advmod	_	_
4	uçtan	uç	NOUN	_	Case=Ab Number=Sing Person=3	6	obl	_	_
5	göreve	görev	NOUN	_	Case=Dat Number=Sing Person=3	6	obj	_	_
6-7	başlıyorlarmış	_	_	_	_	_	_	_	DerivedFrom=baş SpaceAfter=No
6	başlıyorlar	başla	VERB	Ptcp	Aspect=Prog Evident=Nfh Number=Plur Person=3 Polarity=Pos VerbForm=Part	0	root	_	_
7	mış	y	AUX	Zero	Evident=Nfh Number=Sing Person=3 Polarity=Pos Tense=Past	6	cop	_	_
8	.	.	PUNCT	Stop	_	6	punct	_	SpacesAfter=\n



Graphs





















Errors

Without error!

Annotating

9 results found.

Query was lemma: başka, upos: ADJ in the treebank *UD_Turkish-BOUN*

sent_id	text	
news_1874	Başka bir krize neden olan MGK toplantısı ise sadece 30 dakika sürmüştü.	 
news_373	başka biri olarak döneceksindir yolculuğundan	 
ess_1212	Nietzsche'nin dediği gibi, fiziğin bile bir dünya açıklaması olmayıp yalnızca bir dünya yorumu ve düzenlemesi olduğunu anlamak için başka bir kaynağa değil, bilim tarihine bakmak yeter.	 
pop_189	Sonra başka bir şey duyamaz oldum.	 
pop_233	Yoksa yalnız başkaları için değil, kendisi için de çekilmez derecede inceleyici, karar almaktan yoksun duruma düşecektir.	 
ess_1107	Bir durumdan bir başka duruma geçiş, çağrışımlar.	 
bio_803	Bir insanın bir başka insana, o çocuklara yaptığını, yapabileceğini benim kafam almıyor.	 
bio_1493	Bu kadar ara verip aynı çizgiyi tutturan başka tiyatro bilmiyorum ben.	 
news_928	Ardından başka oyuncuların rol seçimi için gönderdiği kasetlerin birinde kısacık bir görüntüsüyle karşılaşan Martin Brest, onu tercih etmiş.	 

Searching

Text: Bir durumdan bir başka duruma geçiş, çağrışımlar.

Annotator: Upload version

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS	MISC
1	Bir	bir	DET	Indef	_	2	det	_	_
2	durumdan	durum	NOUN	_	Case=Abl Number=Sing Person=3	6	nmod	_	_
3	bir	bir	DET	Indef	_	5	det	_	_
4	başka	başka	ADJ	Adj	_	5	amod	_	_
5	duruma	durum	NOUN	_	Case=Dat Number=Sing Person=3	6	nmod	_	_
6	geçiş	geç	VERB	Vnoun	Case=Nom Number=Sing Person=3 Polarity=Pos VerbForm=Vnoun	0	root	_	SpaceAfter=No
7	,	,	PUNCT	Comma	_	8	punct	_	_
8	çağrışımlar	çağrışım	NOUN	_	Case=Nom Number=Plur Person=3	6	conj	_	SpaceAfter=No
9	.	.	PUNCT	Stop	_	6	punct	_	SpacesAfter= \n

Searching *cont*

Evaluation

- Version comparison (Jun 22)
 - ~30% speedup
- User testing (Jul 22)
 - Functionality
 - 3 experienced linguists
 - Valuable feedback
- Automated testing (Sep 22)
 - Basic usage
- Learnability (Dec 22)
 - 2 inexperienced linguists
- Currently testing
 - New treebank for *ota*

Discussion, conclusion

- Interdisciplinary collaboration
- Open-source
- Further improvements
 - Automation
 - Adjustments for *ota*

Thank you.

References

- Universal Dependencies: universaldependencies.org
- UD_Turkish-BOUN:
universaldependencies.org/treebanks/tr_boun/index.html
- BoAT: github.com/BOUN-TABILab-TULAP/Web-Annotation-Tool

Contact

- Email: furkanakkurt9285@gmail.com
- GitHub: furkanakkurt1335
- BM 36, North Campus, Boğaziçi University