BOUN Treebank *(v2.11)*

Contributors: Büşra Marşan, Furkan Akkurt, Suzan Üsküdarlı, Tunga Güngör, Balkız Öztürk, Arzucan Özgür, Onur Güngör, Muhammet Şen, Merve Gürbüz, Utku Türk, Talha Bedir, Şaziye Betül Özateş, *and others*

Statistics

- Data source: **Turkish National Corpus**
- **9,761** sentences
- **121,214** tokens
- Covering **five registers**: Broadsheet national newspapers, biographical texts, essays, popular culture articles, and instructional texts
- Variation in **sentence length** and **word order**.

Statistics

Order	Count	Percentage (%)
OV	5744	37.21
SV	5416	35.09
SOV	1456	9.43
VS	1116	7.23
VO	714	4.63
OVS	549	3.56
VSO	165	1.07
SVO	144	0.93
OSV	109	0.71
VOS	23	0.15

Order	Count	Percentage (%)
SOV	1456	59.53
OVS	549	22.44
VSO	165	6.75
SVO	144	5.89
OSV	109	4.46
VOS	23	0.94

Statistics
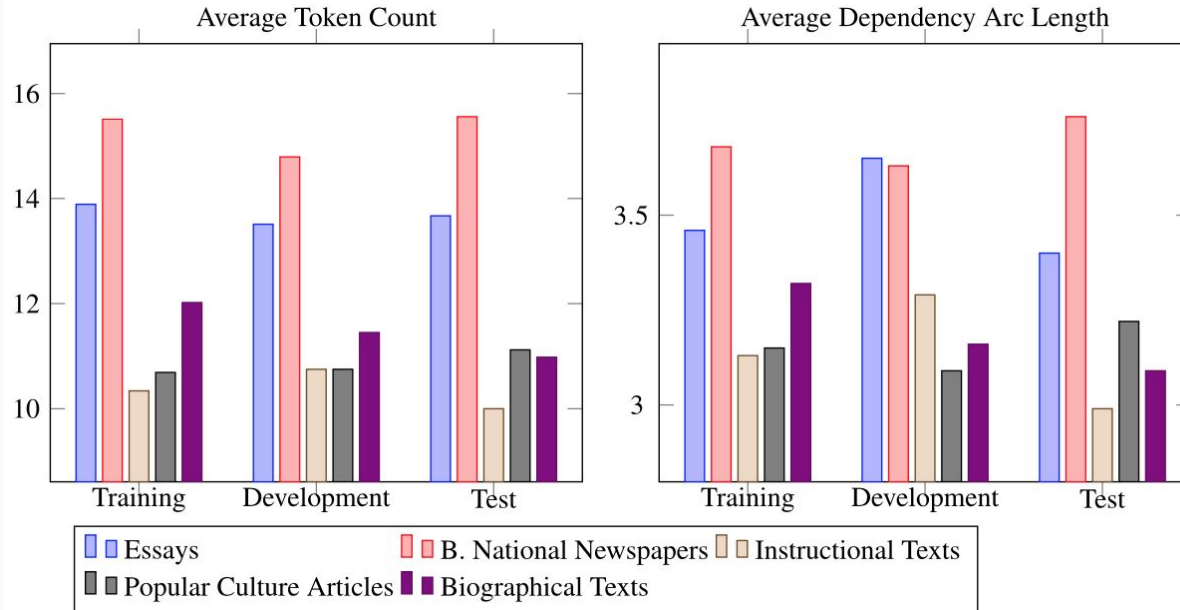
Papers

Türk, U., Atmaca, F., Özateş, Ş. B., Berk, G., Bedir, S. T., Köksal, A., ... & Özgür, A. (2022). **Resources for Turkish dependency parsing: Introducing the BOUN treebank and the BoAT annotation tool.** *Language Resources and Evaluation*, 1-49.

Marşan, B., Akkurt, S. F., Şen, M., Gürbüz, M., Güngör, O., Özateş, Ş. B., ... & Öztürk, B. (2022). **Enhancements to the BOUN Treebank Reflecting the Agglutinative Nature of Turkish.** *arXiv preprint arXiv:2207.11782*.

Akkurt, S. F., Marşan, B., & Uskudarli, S. (2022). **BoAT v2--A Web-Based Dependency Annotation Tool with Focus on Agglutinative Languages.** *arXiv preprint arXiv:2207.01327*.

Annotation Details

- **2 versions** of BOUN Treebank are available on the UD repository: 2.8 and 2.11
  - Improvements on **morphological features,** abstraction, lemmatization, **segmentation,** and dependency relations.
- Annotated by: **Trained linguists** who are native speakers of Turkish
- Inter annotator agreement scores of 2.11:
  - Unlabelled attachment and labelled attachment scores were calculated using Cohen's Kappa measure. Their respective values are 98.61 and 97.81.

Annotation Details // Issues

v2.11 targeted a set of issues:

- Derivation
- Null morpheme
- Copula

Annotation Details // Issues

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS	MISC	Case	Number	Person
1	1936	1936	NUM	Year	Case=Nom|Number=Sing|Person=3	2	nmod:poss	_	_	Nom	Sing	3
2-3	yılındayız	_	_	_	_	_	_	_	SpaceAfter=No	_	_	_
2	yılında	yıl	NOUN	_	Case=Loc|Number=Sing|Number[psor]=Sing|Person=3|Person[psor]=3	0	root	_	_	Loc	Sing	3
3	yız	null	AUX	Zero	Case=Nom|Number=Plur|Person=1|Polarity=Pos|Tense=Pres	2	cop	_	_	Nom	Plur	1
4	.	.	PUNCT	Stop	_	2	punct	_	SpacesAfter=\n	_	_	_

Annotation Details // Issues

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS	MISC	Case	Number	Person
1	Ahmed	Ahmed	PROPN	_	Case=Nom|Number=Sing|Person=3	9	nsubj	_	_	Nom	Sing	3
2	Rasim	Rasim	PROPN	_	Case=Nom|Number=Sing|Person=3	1	flat	_	SpaceAfter=No	Nom	Sing	3
3	,	,	PUNCT	Comma	_	9	punct	_	_	_	_	_
4	Büyükada'ya	Büyükada	PROPN	Place	Case=Dat|Number=Sing|Person=3	5	obl	_	_	Dat	Sing	3
5	gidip	git	VERB	Conv	Polarity=Pos|VerbForm=Conv	8	advcl	_	_	_	_	_
6	birkaç	birkaç	DET	Det	_	7	det	_	_	_	_	_
7	gün	gün	NOUN	_	Case=Nom|Number=Sing|Person=3	8	obl:tmod	_	_	Nom	Sing	3
8	kalmayı	kal	VERB	Vnoun	Case=Acc|Number=Sing|Person=3|Polarity=Pos|VerbForm=Vnoun	9	xcomp	_	_	Acc	Sing	3
9-10	tasarlamıştı	_	_	_	_	_	_	_	SpaceAfter=No	_	_	_
9	tasarlamış	tasarla	VERB	Ptcp	Aspect=Perf|Polarity=Pos|VerbForm=Part	0	root	_	_	_	_	_
10	tı	y	AUX	Zero	Evident=Fh|Number=Sing|Person=3|Tense=Past	9	cop	_	_	_	Sing	3
11	.	.	PUNCT	Stop	_	9	punct	_	SpacesAfter=\n	_	_	_

Annotation Details // Issues

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS	MISC	Case	Number	Person
1	ABD	ABD	PROPN	Place	Abbr=Yes|Case=Nom|Number=Sing|Person=3	2	nmod:poss	_	_	Nom	Sing	3
2	Başkanı	başkan	NOUN	_	Case=Nom|Number=Sing|Number[psor]=Sing|Person=3|Person[psor]=3	3	nmod	_	_	Nom	Sing	3
3	George	George	PROPN	_	Case=Nom|Number=Sing|Person=3	18	nsubj	_	_	Nom	Sing	3
4	Bush	Bush	PROPN	_	Case=Nom|Number=Sing|Person=3	3	flat	_	SpaceAfter=No	Nom	Sing	3
5	,	,	PUNCT	Comma	_	18	punct	_	_	_	_	_
6-7	Irak'taki	_	_	_	_	_	_	_	_	_	_	_
6	Irak'ta	Irak	PROPN	Place	Case=Loc|Number=Sing|Person=3	9	nmod	_	_	Loc	Sing	3
7	ki	ki	PART	Attr	_	6	dep:der	_	_	_	_	_
8	Amerikan	Amerikan	PROPN	_	Case=Nom|Number=Sing|Person=3	9	nmod:poss	_	_	Nom	Sing	3
9	birliklerinin	birlik	NOUN	_	Case=Gen|Number=Plur|Number[psor]=Sing|Person=3|Person[psor]=3	10	nmod:poss	_	_	Gen	Plur	3
10-11	omuzlarındaki	_	_	_	_	_	_	_	_	_	_	_
10	omuzlarında	omuz	NOUN	_	Case=Loc|Number=Plur|Number[psor]=Sing|Person=3|Person[psor]=3	12	nmod	_	_	Loc	Plur	3
11	ki	ki	PART	Attr	_	10	dep:der	_	_	_	_	_
12	yükün	yük	NOUN	_	Case=Gen|Number=Sing|Person=3	17	nsubj	_	SpaceAfter=No	Gen	Sing	3
13	,	,	PUNCT	Comma	_	17	punct	_	_	_	_	_
14	sonbaharda	sonbahar	NOUN	_	Case=Loc|Number=Sing|Person=3	16	obl	_	_	Loc	Sing	3
15	müttefiklerin	müttefik	NOUN	_	Case=Gen|Number=Plur|Person=3	16	nsubj	_	_	Gen	Plur	3
16	gelişiyle	gel	VERB	Vnoun	Case=Ins|Number=Sing|Number[psor]=Sing|Person=3|Person[psor]=3|Polarity=Pos|VerbForm=Vnoun	17	obl	_	_	Ins	Sing	3
17	hafifleyeceğini	hafifle	VERB	Ptcp	Aspect=Prosp|Case=Acc|Number=Sing|Number[psor]=Sing|Person=3|Person[psor]=3|Polarity=Pos|Tense=Fut|VerbForm=Part	18	ccomp	_	DerivedFrom=hafif	Acc	Sing	3
18	söyledi	söyle	VERB	_	Aspect=Perf|Evident=Fh|Number=Sing|Person=3|Polarity=Pos|Tense=Past	0	root	_	SpaceAfter=No	_	Sing	3
19	.	.	PUNCT	Stop	_	18	punct	_	SpacesAfter=\n	_	_	_

Graphs ∨
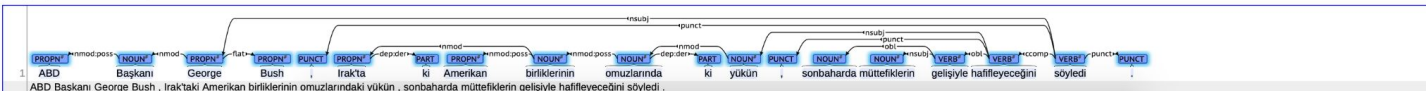
ABD Başkanı George Bush , Irak'taki Amerikan birliklerinin omuzlarındaki yükün , sonbaharda müttefiklerin gelişiyle hafifleyeceğini söyledi .

Annotation Details // Issues

ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS	MISC	Case	Number	Person
1	Bu	bu	DET	Det	_	2	det	_	_	_	_	_
2	kararla	karar	NOUN	_	Case=Ins|Number=Sing|Person=3	12	obl	_	_	Ins	Sing	3
3	polise	polis	NOUN	_	Case=Dat|Number=Sing|Person=3	8	obl	_	_	Dat	Sing	3
4	ve	ve	CCONJ	_	_	5	cc	_	_	_	_	_
5	savcılıklara	savcılık	NOUN	_	Case=Dat|Number=Plur|Person=3	3	conj	_	DerivedFrom=Savcı	Dat	Plur	3
6	bu	bu	DET	Demons	_	7	det	_	_	_	_	_
7	yönde	yön	NOUN	_	Case=Loc|Number=Sing|Person=3	8	obl	_	_	Loc	Sing	3
8	yapılan	yap	VERB	Ptcp	Polarity=Pos|Tense=Pres|VerbForm=Part|Voice=Pass	9	acl	_	_	_	_	_
9	şikâyetlerin	şikâyet	NOUN	_	Case=Gen|Number=Plur|Person=3	11	nsubj	_	_	Gen	Plur	3
10	de	de	PART	Emph	_	9	advmod:emph	_	_	_	_	_
11	azalacağı	azal	VERB	Ptcp	Aspect=Prosp|Case=Nom|Number=Sing|Number[psor]=Sing|Person=3|Person[psor]=3|Polarity=Pos|Tense=Fut|VerbForm=Part	12	csubj	_	_	Nom	Sing	3
12	belirtildi	belir	VERB	_	Aspect=Perf|Evident=Fh|Number=Sing|Person=3|Polarity=Pos|Tense=Past|Voice=CauPass	0	root	_	SpaceAfter=No	_	Sing	3
13	.	.	PUNCT	Stop	_	12	punct	_	SpacesAfter=\n	_	_	_

Annotation Interface