

Corpus

| | | | |
|-----------------|--------------|-----------------|----------------------|
| Name | tr.bianet.tr | Analytics date | 24/ 8/ 2023 17:22:25 |
| Source language | Turkish (tr) | Target language | n/a |

Volumes

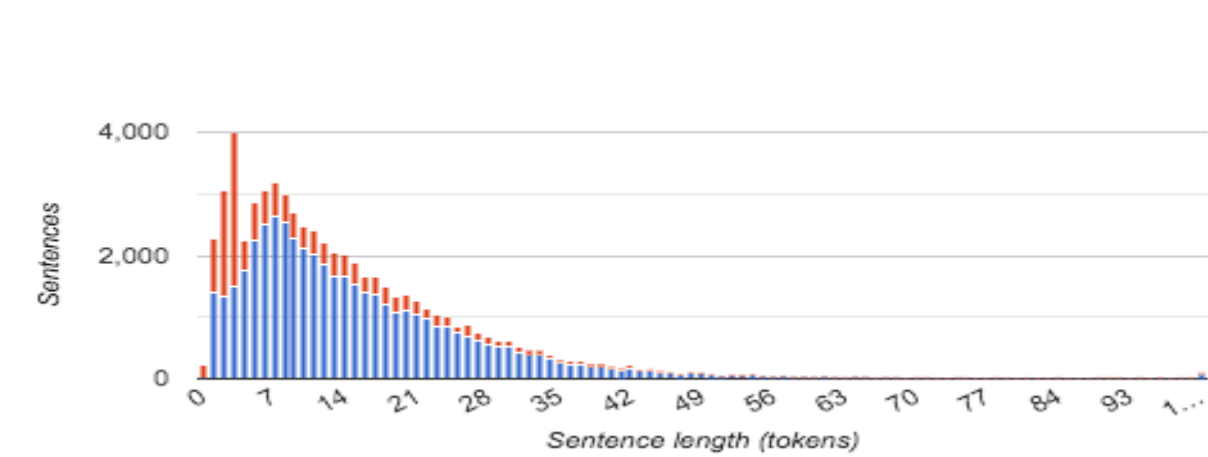
| | | | | | |
|-----------------------|--------|---------------|---------|---------------|-----|
| Sentence pairs | 61.472 | Source size | 5.74 MB | Target size | n/a |
| Unique sentence pairs | 47.155 | Source tokens | 887.812 | Target tokens | n/a |

Type-Token Ratio

| | | | |
|--------|------|--------|-----|
| Source | 0.11 | Target | n/a |
|--------|------|--------|-----|

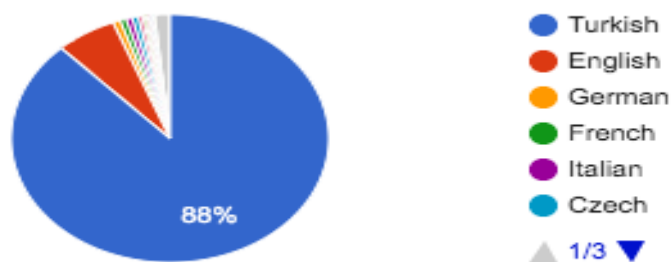
Sentence length distribution

Source

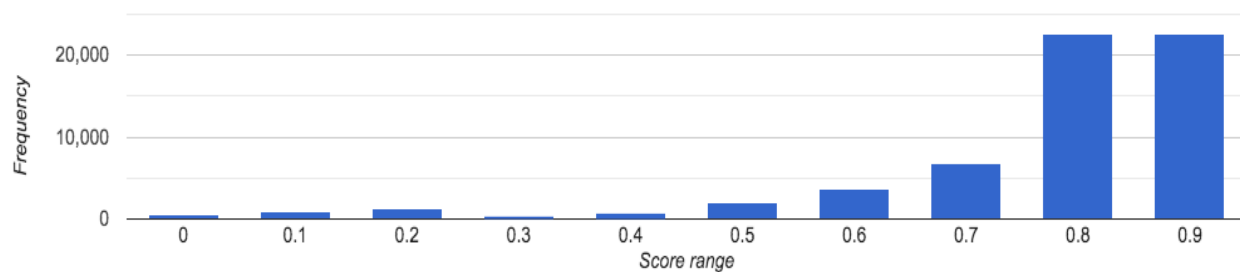


Language distribution

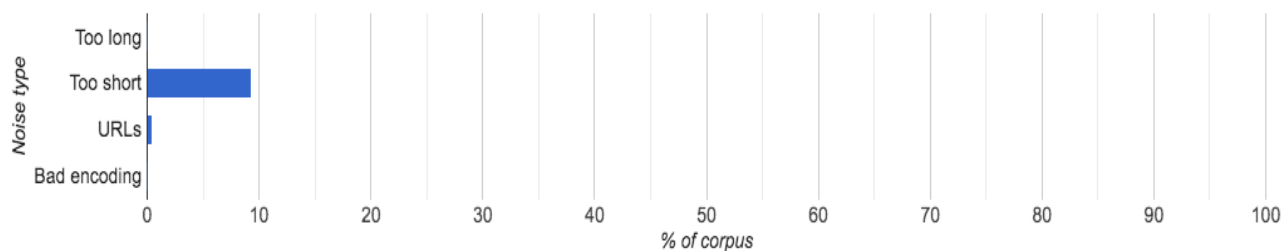
Source



Bicleaner score distribution



Noise distribution



Common n-grams

Source

Length: 1 tokens

bir (Freq: 8265)
türkiye (Freq: 2152)
olarak (Freq: 1972)
i s t a n b u l (Freq: 1944)
g ö z a l t ı n a (Freq: 1762)

Length: 2 tokens

soka a ç1kma (Freq: 561)
prof. dr. (Freq: 559)
gözüaltına alındı (Freq: 508)
a 1r ceza (Freq: 460)
insan hakları (Freq: 431)

Length: 3 tokens

halkların demokratik partisi (Freq: 387)
soka a ç1kma yasa 1 (Freq: 326)
a 1r ceza mahkemesi (Freq: 318)
üniversitesi prof. dr. (Freq: 233)
recep tayyip erdoğan (Freq: 226)

Length: 4 tokens

cumhurbaşkanı recep tayyip erdoğan (Freq: 189)
soka a ç1kma yasa 1 ilan (Freq: 157)
adli kontrol _artıyla serbest (Freq: 107)
yer alan habere göre (Freq: 102)
nöbetçi genel yayın yönetmenliği (Freq: 87)

Length: 5 tokens

e_ genel başkan selahattin demirtaş (Freq: 80)
başkanları selahattin demirtaş ve figen (Freq: 63)
selahattin demirtaş ve figen yüksekda (Freq: 63)
e_ genel başkanları selahattin demirtaş (Freq: 61)
soka a ç1kma yasa 1 ilan edildi (Freq: 56)