

POS TAGGING GUIDELINES

Representation Information

The capital letters in suffixes correspond to sets of letters. For instance the plural suffix is represented as -lAr. The value of A is one of {a,e}, that is, there will be two surface forms for the plural suffix: -lar (e.g. atlar) and -ler (e.g. kediler).

All sets are as follows:

A = {a,e}

H = {ı, i, u , ü}

C = {c, ç}

D = {d, t}

Tag Set

We use Universal Dependencies POS tag set <http://universaldependencies.org/u/pos/index.html> for all tokens (Turkish, German, 3rd language, emoticons...)

ADJ: adjective

INTJ: interjection

PUNCT: punctuation

ADP: adposition

NOUN: noun

SCONJ: subord. conjunction

ADV: adverb

NUM: numeral

SYM: symbol

AUX: auxiliary verb

PART: particle

VERB: verb

CONJ: coord. conjunction

PRON: pronoun

X: other

DET: determiner

PROPN: proper noun

Guidelines

When deciding the annotation, first follow the tagging rules given below. Then follow these resources in the given order. The order is important because there could be conflicts between them. Universal Dependencies is an ongoing project and how things are handled might change over time.

The links from their webpage is most up-to-date, the treebanks that you can search might not have been updated.

1. General POS rules

<http://universaldependencies.org/u/pos/index.html>

2. German POS rules

<http://universaldependencies.org/de/pos/index.html>

3. English POS rules

<http://universaldependencies.org/en/pos/index.html>

4. German Treebank Search

http://bionlp-www.utu.fi/dep_search/?db=German

Tagging Rules

*** Verbs that are derived into adjectives, nouns, adverbs are still annotated as VERB.**

| | | | |
|--------|---------------|------|--|
| -yAcAk | uyuyacak | VERB | adjectival future participle, e.g. uyuyacak kedi |
| -yAn | uyuyan | VERB | adjectival present participle, e.g. uyuyan kedi |
| -mHş | uyumuş | VERB | adjectival past participle, e.g. uyumuş kedi |
| -yAcAk | uyuyacak | VERB | nominal future participle, e.g. uyuyacaklar |
| -yAn | uyuyan | VERB | nominal present participle, e.g. uyuyanlar |
| -mHş | uyumuş | VERB | nominal past participle, e.g. uyumuşlar |
| -yHncA | uyuyunca | VERB | when |
| -yken | uyurken | VERB | while |
| -yHp | uyuyup | VERB | after doing so |
| -yAlH | uyuyalı | VERB | since doing so |
| -yArAk | uyuyarak | VERB | by doing so |
| -dHkÇA | uyudukça | VERB | as long as |
| -mAdAn | uyumadan | VERB | without having done so |
| -mAk | uyumak | VERB | e.g. uyumak istiyorum |
| -DHğH | uyuduğunu | VERB | e.g. uyuduğunu gördüm |
| -yAcAk | uyuyacağımızı | VERB | e.g. uyuyacağımızı söyledim |

*** Verbs derived from verbs are annotated as VERB.**

| | | | |
|--------|------------|------|-----------|
| -DHr | kestirmek | VERB | causative |
| -t | uyutmak | VERB | causative |
| -Hl | kesilmek | VERB | passive |
| -Hş | tanışmak | VERB | reflexive |
| -yAbil | yapabilmek | VERB | ability |
| -lAn | süslenmek | VERB | acquire |

*** Suffixes that derive adverbs, adjectives, nouns and proper nouns.**

| | | | |
|---------|-------------------------------|-----|--------------------------|
| -DHr | süredir | ADV | since |
| -CA | yavaşça | ADV | ly (as in slow - slowly) |
| -CA | esmerce kız | ADJ | as if |
| -yAsH | kör olası adam | ADJ | feel like |
| -sH | insansı robot | ADJ | just like |
| -HmsH | evimsi, kedimsi, tatlımsı | ADJ | just like |
| -Hmtrak | evimtrak, kedimtrak, acımtrak | ADJ | just like |
| -sAl | yapısal, bilimsel | ADJ | related (lexicalised) |
| -k | ideolojik, psikolojik | ADJ | related (lexicalised) |

| | | | |
|--------|--|-------|--------------------------|
| -mAçA | yuttur maca , aldat maca | NOUN | act of (lexicalised) |
| -cHk | damlac ık | NOUN | diminutive (lexicalised) |
| -cHk | Zehrac ık | PROPN | diminutive |
| -cHğHm | Zehracı ğım | PROPN | |
| | teyzecim | NOUN | diminutive |
| -cAğHz | kızca ğız | NOUN | diminutive |

*** The following suffixes derive adjectives or nouns from nouns. The POS tag is determined based on use.**

| | | | |
|------|--|-----------------|-------------|
| -ki | evde ki kedi | evde | NOUN |
| | | -ki | ADJ |
| | evde kiler | evde | NOUN |
| | | -kiler | NOUN |
| | evin kine | evin | NOUN |
| | | -kine | NOUN |
| -lH | 1.8 motor lu taşıt | motor | NOUN |
| | | -lu | ADJ |
| | 1.8 motor luyla gittim | motor | NOUN |
| | | -luyla | NOUN |
| -sHz | belges iz kanıt | belge | NOUN |
| | | -siz | ADJ |
| | kayıt belges izlere ceza kesilecek | belge | NOUN |
| | | -sizlere | NOUN |
| -lHk | 2 metre lik kumaş | metre | NOUN |
| | | -lik | ADJ |
| | 2 metre likleri adamdan saymıyorlar | metre | NOUN |
| | | -likleri | NOUN |
| -CH | barış çı halk | barış | NOUN |
| | | -çı | ADJ |
| | kitap çı | kitap | NOUN |
| | | -çı | NOUN |

*** lAş could derive verbs from nouns or adjectives.**

| | | | |
|------|--------------------|---------------|-------------|
| -lAş | pembe leşti | pembe | ADJ |
| | | -leşti | VERB |
| | aptal laştı | aptal | ADJ |
| | | -laştı | VERB |

*** copulars are annotated as VERB**

| | | | |
|------|--|---------|------|
| -DHr | güzeld dir | güzel | ADJ |
| | | -dir | VERB |
| -DH | güzeld im | güzel | ADJ |
| | | -dim | VERB |
| -mHş | güzelm işsin | güzel | ADJ |
| | | -mişsin | VERB |
| -yHm | öğrenci yim | öğrenci | NOUN |
| | | -yim | VERB |
| -lAr | öğrenci ler (Onlar öğrenci ler) | öğrenci | NOUN |
| | | -ler | VERB |

| | |
|-------------|-------------|
| Sieben | NUM |
| çok | ADV |
| güzel | ADJ |
| sayı | NOUN |
| -dır | VERB |

(In German, Sieben **ist**... ist is **VERB**)

| | |
|-------------|-------------|
| kezban | NOUN |
| -sın | VERB |

(In German, du **bist**... bist is **VERB**)

*** Similarly, when used as a copular, olmak is also a VERB**

| | |
|---------------|-------------|
| kedi | NOUN |
| var/yok/mutlu | ADJ |
| oldu | VERB |
| mutlu | ADJ |
| olarak | VERB |
| uyudu | VERB |

*** But when it comes after a verb, it is an auxiliary**

| | |
|-----------------|------------|
| eve | NOUN |
| geliyor | VERB |
| olacağım | AUX |

And since we take geçmiş as a verb, olsun below is AUX

| | |
|--------------|------------|
| geçmiş | VERB |
| olsun | AUX |

*** Adjectives, nouns derived into nouns have their original POS tag.**

| | |
|--------------|------------|
| ikisi | NUM |
| geldi | VERB |

| | |
|---------------|------------|
| maviyi | ADJ |
| gördüm | VERB |

Exception to this is vocatives, they are annotated as NOUN.

| | |
|---------------|-------------|
| tatlim | NOUN |
| canim | NOUN |

*** Each word forming a named entity has its original POS tag.**

| | |
|-------------|------|
| 99 | NUM |
| Luftballons | NOUN |

| | |
|-------------|-------|
| Aufbruch | NOUN |
| Neuköln | PROPN |
| Derneği'nde | NOUN |

*** Each word forming a multiword expression has its original POS tag.**

| | |
|----------|------|
| o | PRON |
| kadar | ADP |
| zamandır | ADV |

| | |
|----------|------|
| zu | ADP |
| dem | DET |
| Beispiel | NOUN |

*** Exclamations and feedback markers are interjections.**

| | |
|----|------|
| Oh | INTJ |
|----|------|

| | |
|-----|------|
| Ach | INTJ |
| ja | INTJ |

Note that **ja** is not used as interjection all the time.

| | |
|-----------|------------|
| aber | CONJ |
| man | PRON |
| kann | AUX |
| ja | ADV |
| nicht | PART |
| alles | PRON |
| essen | VERB |

*** Words primarily belonging to another POS keep their original POS when used in exclamations.**

| | |
|----------|-------|
| Aman | INTJ |
| Allah'ım | PROPN |

*** Question particles are annotated as AUX <http://coltekin.github.io/ud-docs-tr/tr/pos/all.html>**

| | |
|-----------|------------|
| geldi | VERB |
| mi | AUX |

| | |
|--------------|------------|
| güzel | ADJ |
| miyim | AUX |

They can also be used in coordinations and subordinating clauses.

| | |
|-----------|-------------|
| yeşil | ADJ |
| mi | CONJ |
| mavi | ADJ |
| mi | CONJ |
| bilemedim | VERB |

| | |
|------------|--------------|
| Oraya | PRON |
| vardın | VERB |
| mi | SCONJ |
| denizi | NOUN |
| göreceksin | VERB |

*** Words that are used both as pronouns and as determiners, are annotated according to use.**

| | |
|-------|------|
| Hangi | DET |
| ev | NOUN |

| | |
|------------|------|
| Hangisini | PRON |
| seviyorsun | VERB |

| | | |
|-------|------|--------------------|
| Kendi | DET | (in English “own”) |
| evi | NOUN | |

| | |
|-------|------|
| Kendi | PRON |
| geldi | VERB |

| | |
|--------|------|
| Bu | DET |
| kedi | NOUN |
| uyumuş | VERB |

| | |
|--------|------|
| Bu | PRON |
| uyumuş | VERB |

| | |
|--------|-----|
| Hiçbir | DET |
|--------|-----|

| | |
|---------|------|
| anlamı | NOUN |
| yok | ADJ |
| Hiçbiri | PRON |
| yok | ADJ |

*** The clitic dA could be used as an ADV, a CONJ, or a SCONJ**

| | | |
|-----------|------------|------------------------------------|
| Ali | PROPN | |
| de | ADV | (to test, replace with bile, dahi) |
| geldi | VERB | |

| | |
|-----------|-------------|
| Ali | PROPN |
| de | CONJ |
| Ahmet | PROPN |
| de | CONJ |

| | |
|-----------|--------------|
| Ali | PROPN |
| geldi | VERB |
| de | SCONJ |
| sofraya | NOUN |
| oturduk | VERB |

*** Değil when used as a negation particle is annotated as a PART. It can also be used as a CONJ or SCONJ.**

| | |
|--------------|-------------|
| Ali | PROPN |
| iyi | ADJ |
| değil | PART |
| -dir | VERB |

| | |
|--------------|--------------|
| Ali | PROPN |
| değil | CONJ |
| Ahmet | PROPN |
| geldi | VERB |
| Değil | SCONJ |
| kitap | NOUN |
| okumak | VERB |
| sinemaya | NOUN |
| bile | ADV |
| gitmez | VERB |

*** ise could be standalone or attached to a verb.**

| | |
|-------------|-------------|
| güzel | ADJ |
| isen | VERB |

| | |
|-------------|-------------|
| güzel | ADJ |
| -sen | VERB |

| | |
|-------|------|
| geldi | VERB |
| isen | AUX |

| | |
|-----------|------|
| geldiysen | VERB |
|-----------|------|

It can also be used as an adverbial

| | |
|---------|-------|
| Ali | PRONP |
| gelmedi | VERB |
| Ahmet | PRONP |
| ise | ADV |
| geldi | VERB |

*** Hashtags are annotated according to their normal POS tag without the # symbol**

| | |
|---------|------|
| #keyf | NOUN |
| #happy | ADJ |
| #wünche | VERB |

Hashtags with multiple words are annotated as X

| | |
|-------------------|---|
| #geziparki | X |
| #GiveVoiceToCizre | X |

*** Acronyms are proper names** <http://universaldependencies.org/u/pos/PROPN.html>

| | |
|-----|-------|
| THY | PROPN |
| DW | PROPN |

*** Abbreviations are assigned the POS tag of the full form**

| | |
|-----|------|
| Dr. | NOUN |
|-----|------|

If the full form is a multiword, assign the function POS

| | |
|------|-----|
| z.B. | ADV |
|------|-----|

*** Usernames are annotated as proper nouns**

| | |
|-----------|-------|
| @username | PROPN |
|-----------|-------|

*** URLs are symbols in UD** <http://universaldependencies.org/u/pos/SYM.html>

| | |
|-------|-----|
| [URL] | SYM |
|-------|-----|

* Emoticons are symbols <http://universaldependencies.org/u/pos/SYM.html>

:) SYM

* Twitter-specific tokens are annotated as X

| | |
|------|---|
| RE | X |
| RT | X |
| <NL> | X |

* Country and language names are proper names

| | |
|---------|-------|
| Almanca | PROPN |
| Almanya | PROPN |

| | |
|--------------|------------|
| Alman | ADJ |
| müziği | NOUN |

| | |
|--------------|--------------|
| İki | NUM |
| Alman | PROPN |
| geldi | VERB |

* Bir (ein,eine) is a determiner unless we want to explicitly specify the number of something. Then it is NUM.

| | |
|------------|------------|
| Şimdi | ADV |
| bir | DET |
| konuyu | NOUN |
| tekrar | ADV |
| açacağım | VERB |

| | |
|------------|------------|
| Masada | NOUN |
| bir | NUM |
| değil | CONJ |
| iki | NUM |
| yumurta | NOUN |
| var | ADJ |

* Ya is an interjection, unless it is used to connect two sentences, then it is an SCONJ

| | |
|-----------|-------------|
| Allah | PROPN |
| kahretsin | VERB |
| ya | INTJ |

| | |
|--------|------|
| bu | DET |
| senin | PRON |
| Termin | NOUN |
| var | ADJ |

| | |
|-----------|--------------|
| ya | SCONJ |
| beni | PRON |
| hep | ADV |
| gıcık | NOUN |
| ediyor | VERB |

*** If it is part of 'ya da', then both ya and da are CONJ**

| | |
|------------|-------------|
| Siteyle | NOUN |
| ilgili | ADJ |
| extrem | ADJ |
| öneri | NOUN |
| ya | CONJ |
| da | CONJ |
| eleştirisi | NOUN |
| olanlar | VERB |

*** yok could be used as an interjection (if it means 'no'), unless it is an expletive (there isn't...). The following example has both usages:**

| | |
|------------|-------------|
| Yok | INTJ |
| ya | INTJ |
| bu | DET |
| sene | NOUN |
| izin | NOUN |
| yok | ADJ |

*** diye acts as a subordinate conjunction**

| | |
|-------------|--------------|
| " | PUNCT |
| Fantastisch | INTJ |
| ! | PUNCT |
| " | PUNCT |
| diye | SCONJ |
| geziyorum | VERB |

*** Hadi is an interjection**

| | |
|-------------|-------------|
| ağlayacağım | VERB |
| şimdi | ADV |
| hadi | INTJ |
| kaçalım | VERB |

*** Other frequent words and their tags**

| | |
|-----------------------------------|--|
| Evet, hayır, ja, nein | INTJ (ja could have other uses, see previous explanations above) |
| Bismillah, inşallah, maşallah,... | INTJ |