**This page pertains to UD version 2.**

**UD for Romanian**

**Tokenization and Word Segmentation**

* In general, words are delimited by whitespace characters. Description of exceptions follows.
* According to typographical rules, many punctuation marks are attached to a neighboring word. We always tokenize them as separate tokens (words); that holds even for hyphenated words pronounced joining two syllables in a single one, such as  ne-am, v-ați.

**Morphology**

**Tags**

This is an overview only. For more detailed discussion and examples, see the list of [Romanian POS tags](http://universaldependencies.org/cs/pos/index.html) and [Romanian features](http://universaldependencies.org/cs/feat/index.html).

* Romanian uses all 17 universal POS categories,.
* The pronoun ([PRON](http://universaldependencies.org/cs/pos/PRON.html" \o "cs PRON)) vs. determiner ([DET](http://universaldependencies.org/cs/pos/DET.html" \o "cs DET)) distinction is based on the syntactic position; PRON is independent, DET is subordinated to a NOUN or PRONOUN.
* CONTEMPORARY ROMANIAN has three auxiliary verbs ([AUX](http://universaldependencies.org/cs/pos/AUX_.html)), fi (“to be”), avea (to have) vrea (to want). In Nonstandard Romanian, the auxiliaries putea (can), trebui (must), ști (to know) were annotated.
  + The copula with non-verbal predicates was annotated as AUX.
  + Periphrastic future tense (vrea + infinitive of the main verb; vrea + fi+ participle of the main verb).
  + Periphrastic past tense (avea + participle of the main verb).
  + Periphrastic conditional (conditional form of avea + future of the main verb; conditional form of avea+ fi + participle of the main verb).
  + Periphrastic passive (any form of fi, including periphrastic forms, + participle of the main verb).
* Verbs with modal meaning are not considered auxiliary in Czech.
* There are four main (de)verbal forms, distinguished by the UPOS tag and the value of the [VerbForm](http://universaldependencies.org/cs/feat/VerbForm.html" \o "cs VerbForm) feature:
  + Infinitive Inf, tagged [VERB](http://universaldependencies.org/cs/pos/VERB.html) or [AUX](http://universaldependencies.org/cs/pos/AUX_.html).
  + Finite verb Fin, tagged [VERB](http://universaldependencies.org/cs/pos/VERB.html) or [AUX](http://universaldependencies.org/cs/pos/AUX_.html).
  + Participle Part, tagged [VERB](http://universaldependencies.org/cs/pos/VERB.html) or [AUX](http://universaldependencies.org/cs/pos/AUX_.html) (the so-called l-participle) or [ADJ](http://universaldependencies.org/cs/pos/ADJ.html" \o "cs ADJ) (all other participle types).
  + Gerundive form.

**Nominal Features**

* Nominal words ([NOUN](http://universaldependencies.org/cs/pos/NOUN.html), [PROPN](http://universaldependencies.org/cs/pos/PROPN.html" \o "cs PROPN) and [PRON](http://universaldependencies.org/cs/pos/PRON.html" \o "cs PRON)) have an inherent [Gender](http://universaldependencies.org/cs/feat/Gender.html) feature with one of two values: Masc or Fem. The neuter is in Romanian classified as masc singular and feminine plural.
  + The following parts of speech inflect for Gender because they must agree with nouns: [ADJ](http://universaldependencies.org/cs/pos/ADJ.html" \o "cs ADJ), [DET](http://universaldependencies.org/cs/pos/DET.html" \o "cs DET), [NUM](http://universaldependencies.org/cs/pos/NUM.html" \o "cs NUM), [VERB](http://universaldependencies.org/cs/pos/VERB.html), [AUX](http://universaldependencies.org/cs/pos/AUX_.html). For verbs (including auxiliaries), only participles for gender.
* The two main values of the [Number](http://universaldependencies.org/cs/feat/Number.html) feature are Sing and Plur. The following parts of speech inflect for number: [NOUN](http://universaldependencies.org/cs/pos/NOUN.html), [PROPN](http://universaldependencies.org/cs/pos/PROPN.html" \o "cs PROPN), [PRON](http://universaldependencies.org/cs/pos/PRON.html" \o "cs PRON), [ADJ](http://universaldependencies.org/cs/pos/ADJ.html" \o "cs ADJ), [DET](http://universaldependencies.org/cs/pos/DET.html" \o "cs DET), [VERB](http://universaldependencies.org/cs/pos/VERB.html), [AUX](http://universaldependencies.org/cs/pos/AUX_.html) (finite, participles and converbs), marginally [NUM](http://universaldependencies.org/cs/pos/NUM.html).
* [Case](http://universaldependencies.org/cs/feat/Case.html) has 5 possible values: Nom, Gen, Dat, Acc, Voc. It occurs with the nominal words, i.e., [NOUN](http://universaldependencies.org/cs/pos/NOUN.html), [PROPN](http://universaldependencies.org/cs/pos/PROPN.html" \o "cs PROPN), [PRON](http://universaldependencies.org/cs/pos/PRON.html" \o "cs PRON), [ADJ](http://universaldependencies.org/cs/pos/ADJ.html" \o "cs ADJ), [DET](http://universaldependencies.org/cs/pos/DET.html" \o "cs DET), [NUM](http://universaldependencies.org/cs/pos/NUM.html). It can occur with participles but only with those tagged as ADJ. To nouns there is no distinction of form, they are annotated as nom.,-acc. and gen.,-dat. Distinct forms have only some pronouns.
  + The Case feature also occurs with prepositions ([ADP](http://universaldependencies.org/cs/pos/ADP.html)). Here it is a lexical feature. Prepositions do not inflect for case but they subcategorize for the case of their noun phrase.

**Degree and Polarity**

* [Degree](http://universaldependencies.org/cs/feat/Degree.html) applies to adjectives ([ADJ](http://universaldependencies.org/cs/pos/ADJ.html" \o "cs ADJ)) and adverbs ([ADV](http://universaldependencies.org/cs/pos/ADV.html" \o "cs ADV)) and has rarely one of three possible values: Pos, Cmp, Sup.
* [Polarity](http://universaldependencies.org/cs/feat/Polarity.html) has two values, Pos and Neg, and applies primarily to verbs ([VERB](http://universaldependencies.org/cs/pos/VERB.html), [AUX](http://universaldependencies.org/cs/pos/AUX_.html)), adjectives ([ADJ](http://universaldependencies.org/cs/pos/ADJ.html" \o "cs ADJ)) and adverbs ([ADV](http://universaldependencies.org/cs/pos/ADV.html" \o "cs ADV)) that can be negated using the bound morpheme *ne-*.
  + The Polarity feature is not used with pronouns and determiners, although there is a subset of negative pronouns and determiners. The PronType=Neg feature is used there instead.

**Verbal Features**

* Finite verbs always have one of three values of [Mood](http://universaldependencies.org/cs/feat/Mood.html): Ind, Imp or Subj. The conditional mood is only used with conditional auxiliaries *(avea, avea + fi)*
* Verbs in the indicative mood always have one of four values of [Tense](http://universaldependencies.org/cs/feat/Tense.html): Past simple, Pres , Imperfect, Pluperfect.
* The future is only used with auxiliaries (vrea, vrea+fi).
  + Imperative forms do not have the Tense feature (note that past and present conditionals are distinguished analytically; also the past subjunctive and the perfect future).

**Pronouns, Determiners, Quantifiers**

* [PronType](http://universaldependencies.org/cs/feat/PronType.html) is used with pronouns ([PRON](http://universaldependencies.org/cs/pos/PRON.html" \o "cs PRON)), determiners ([DET](http://universaldependencies.org/cs/pos/DET.html" \o "cs DET)) and adverbs ([ADV](http://universaldependencies.org/cs/pos/ADV.html" \o "cs ADV)).
* [NumType](http://universaldependencies.org/cs/feat/NumType.html) is used with numerals ([NUM](http://universaldependencies.org/cs/pos/NUM.html" \o "cs NUM)), The [Poss](http://universaldependencies.org/cs/feat/Poss.html" \o "cs Poss) feature marks possessive personal determiners.
* The [Reflex](http://universaldependencies.org/cs/feat/Reflex.html) feature is marked by the language specific features: expl:pv, expl:poss, expl:impers, and expl:pass.
* [Person](http://universaldependencies.org/cs/feat/Person.html) is a lexical feature of personal pronouns ([PRON](http://universaldependencies.org/cs/pos/PRON.html" \o "cs PRON)) and has three values, 1, 2 and 3. With personal possessive determiners ([DET](http://universaldependencies.org/cs/pos/DET.html" \o "cs DET)), the feature actually encodes the person of the possessor. Person is not marked on other types of pronouns and on nouns, although they can be almost always interpreted as the 3rd person.
  + As a cross-reference to subject, person is also marked on finite verbs ([VERB](http://universaldependencies.org/cs/pos/VERB.html), [AUX](http://universaldependencies.org/cs/pos/AUX_.html)).
* There are two [layered features](http://universaldependencies.org/u/overview/feat-layers.html), [Gender[psor]](http://universaldependencies.org/cs/feat/Gender-psor.html) and [Number[psor]](http://universaldependencies.org/cs/feat/Number-psor.html). They appear with certain possessive adjectives and determiners and encode the lexical gender/number of the possessor. The extra layer is needed to distinguish these lexical features from the inflectional gender and number that mark agreement with the modified (possessed) noun.

**Other Features**

* Besides the layered features listed above, there are several other language-specific features:
  + NumType

Definite

Polite

* + [Variant](http://universaldependencies.org/cs/feat/Variant.html) … distinguishes short and long forms of adjectives, a Slavic-wide phenomenon

**Syntax**

**Core Arguments, Oblique Arguments and Adjuncts**

* Nominal subject ([nsubj](http://universaldependencies.org/cs/dep/nsubj.html" \o "cs nsubj)) is a noun phrase in the nominative case, without preposition.
  + An infinitive verb may serve as the subject and is labeled as clausal subject, [csubj](http://universaldependencies.org/cs/dep/csubj.html" \o "cs csubj).
  + A finite subordinate clause may serve as the subject and is labeled csubj.
* For the purpose of UD the objects are divided to core objects, labeled [obj](http://universaldependencies.org/cs/dep/obj.html" \o "cs obj) or [iobj](http://universaldependencies.org/cs/dep/iobj.html" \o "cs iobj), and oblique objects, labeled [obl](http://universaldependencies.org/cs/dep/obl-arg.html).
  + All prepositional objects are considered oblique.
  + Accusative objects of some verbs alternate with finite clausal complements, which are labeled [ccomp](http://universaldependencies.org/cs/dep/ccomp.html" \o "cs ccomp).
  + If a verb subcategorizes for the infinitive (e.g. modal verbs or verbs of control), the infinitival complement is labeled [xcomp](http://universaldependencies.org/cs/dep/xcomp.html" \o "cs xcomp).
  + If a verb subcategorizes for two core objects, one of them accusative (or ccomp) and the other non-accusative, then the non-accusative object is labeled [iobj](http://universaldependencies.org/cs/dep/iobj.html" \o "cs iobj). Core nominal objects in other situations are labeled just [obj](http://universaldependencies.org/cs/dep/obj.html).
* Adjuncts are usually prepositional phrases, but they can be bare noun phrases as well. They are labeled [obl](http://universaldependencies.org/cs/dep/obl.html). For the prepositional objects which cannot be omitted we used the label nmod:pmod.
  + Temporal modifiers are annotated as: nmod:tmod; advmod:tmod or advcl:tcl.
* Extra attention has to be paid to clitic forms of pronouns (accusative) and (dative). They can function as:
  + Core objects with weak form can be ([obj](http://universaldependencies.org/cs/dep/obj.html" \o "cs obj) or [iobj](http://universaldependencies.org/cs/dep/iobj.html" \o "cs iobj)): or [expl](http://universaldependencies.org/cs/dep/expl-pass.html) if the obj. iobj is named also by a noun or another pronoun.
  + The reflexive can be [expl:pv](http://universaldependencies.org/cs/dep/expl-pv.html), expl:impers, expl:poss, expl:pass.
* In passive clauses (both reflexive and periphrastic passive), the subject is labeled with [nsubj:pass](http://universaldependencies.org/cs/dep/nsubj-pass.html" \o "cs nsubj:pass) or [csubj:pass](http://universaldependencies.org/cs/dep/csubj-pass.html" \o "cs csubj:pass), respectively.
  + The auxiliary verb in periphrastic passive is labeled [aux:pass](http://universaldependencies.org/cs/dep/aux-pass.html" \o "cs aux:pass).
  + If the demoted agent is present, it has the form of a bare instrumental phrase and its relation is labeled [nmod:agent](http://universaldependencies.org/cs/dep/obl-agent.html" \o "cs obl:agent).

**Non-verbal Clauses**

* The copula verb fi (be) is used in equational, attributional, possessive and benefactory nonverbal clauses. Purely existential clauses (without indicating location) use fi as well but it is treated as the head of the clause and tagged [VERB](http://universaldependencies.org/cs/pos/VERB.html).

**Relations Overview**

* The following relation subtypes are used in Romanian:
  + [nsubj:pass](http://universaldependencies.org/cs/dep/nsubj-pass.html) for nominal subjects of passive verbs
  + [csubj:pass](http://universaldependencies.org/cs/dep/csubj-pass.html) for clausal subjects of passive verbs
  + [nmod:agent](http://universaldependencies.org/cs/dep/obl-agent.html) for agents of passive verbs
  + [nmod:pmod](http://universaldependencies.org/cs/dep/obl-agent.html) for prepositional objects which can be omitted as determiners of some verbs.
  + [expl:impers](http://universaldependencies.org/cs/dep/obl-arg.html) for impersonal value of the reflexive
  + [expl:pv](http://universaldependencies.org/cs/dep/expl-pv.html) for reflexive clitics of inherently reflexive verbs
  + [expl:pass](http://universaldependencies.org/cs/dep/expl-pass.html) for reflexive clitics in reflexive passives
  + [expl:poss](http://universaldependencies.org/cs/dep/expl-pass.html) for reflexive clitics with possessive value
  + [aux:pass](http://universaldependencies.org/cs/dep/aux-pass.html) for passive auxiliaries
  + [nmod:tmod](http://universaldependencies.org/cs/dep/nummod-gov.html) for prepositional nouns with temporal value
  + [advcl:tcl](http://universaldependencies.org/cs/dep/det-numgov.html" \o "cs det:numgov)  for adverbial clauses with temporal value
  + [advmod:tmod](http://universaldependencies.org/cs/dep/advmod-emph.html) for adverbs with temporal value
* The following main types are also used alone: [expl](http://universaldependencies.org/u/dep/expl.html) if a nsubj, obj or iobj is repeated by a pronoun.

**Treebanks**

There are [two](http://universaldependencies.org/treebanks/cs-comparison.html) UD-Romanian treebanks:

UD\_Romanian-RRT

UD\_Romanian-Nonstandard

© 2014–2017 [Universal Dependencies contributors](http://universaldependencies.org/introduction.html#contributors). Site powered by [Annodoc](http://spyysalo.github.io/annodoc) and [brat](http://brat.nlplab.org/)

.