magyarlanc 3.0 UD Morphological Tagset

Contents

1. Main Properties

Universal Dependencies[1] (UD) is a framework for cross-linguistically consistent grammatical annotation for over 70 languages including Hungarian. The morphological specification of a word in the UD scheme consists of three levels of representation: lemma, part-of-speech (POS) tag, and feature–value pairs representing morphosyntactic properties of the word. The latter attributes are in a linearized format, in alphabetical order. Every feature has the form `Name=Value`, and features are separated by a vertical bar, such as in `Case=Nom|Number=Sing`. Non-applicable features must not be present in the list of feature–value pairs. The UD framework focuses on syntax, therefore its morphological representation encodes only those phenomena which are important for the syntax, which typically are the inflectional codes.

At the time of writing, the second version of UD is out. However, the development of UD2 for Hungarian and the conversion of Hungarian resources in UD1 are not yet available. Therefore, here we use 'UD' referring to UD1. This document is based on the documentation of UD1 applied for Hungarian[2]. The Szeged Dependency Treebank has a version converted to the format of UD, and magyarlanc 3.0[3] also outputs UD morphological annotation.

2. POS tags

Having a POS tag is obligatory. The tagset contains universal POS tags applied to the Hungarian language. Adpositions (`ADP`), coordinating (`CONJ`) and subordinating (`SCONJ`) conjunctions, interjections (`INTJ`), particles (`PART`), punctuation marks (`PUNCT`), symbols (`SYM`) and foreign words (`X`) do not have any features. Adposition means only postpositions in Hungarian. The POS tag `PART` is only applied for the preverb 'meg'.

3. Features

There are POS categories having obligatory features, which means that they must have a value. These are highlighted with boldface, while their default values are marked with italics throughout the whole document.

3.1. Nominals

Nominals (nouns, adjectives, numerals, and pronouns) share similar features. There are obligatory features applied for all of them: `Case` and `Number`. They and their possible values are listed in Table 2. The nominative case and the singular number are unmarked in Hungarian, therefore these are the default values here.

[1] `http://universaldependencies.org`
[2] `https://universaldependencies.org/docsv1/index.html#hu`
[3] `http://rgai.inf.u-szeged.hu/index.php?lang=en&page=magyarlanc`

POS	description
NOUN	common noun
PROPN	proper noun
ADJ	adjective
NUM	numeral
PRON	pronoun
VERB	main verb
AUX	auxiliary
ADV	adverb
DET	determiner
ADP	adposition
CONJ	coordinating conjunction
SCONJ	subordinating conjunction
INTJ	interjection
PART	particle
PUNCT	punctuation
SYM	symbol
X	other

Table 1: All possible POS tags.

There are 21 possible cases for nominals, however not all of them applicable for all nominals. For example, the locative case (`Loc`) can only be applied for nouns with locative meaning, while the temporal case (`Tem`) can only be used with nominals bearing temporal meaning.

feature	value	description	example
Case	*Nom*	nominative	ember
	Acc	accusative	embert
	Dat	dative	embernek
	Gen	genitive	embernek
	Loc	locative	Győrött
	Ins	instrumental	emberrel
	Dis	distributive	emberenként
	Ess	essive	emberként
	Tra	translative	emberré
	Ine	inessive	emberben
	Ill	illative	emberbe
	Ela	elative	emberből
	Ade	adessive	embernél
	All	allative	emberhez
	Abl	ablative	embertől
	Sup	superessive	emberen
	Sub	sublative	emberre
	Del	delative	emberről
	Tem	temporal	éjfélkor
	Ter	terminative	emberig
	Cau	causative	emberért
Number	*Sing*	singular	ember
	Plur	plural	emberek

Table 2: Obligatory features and their possible values for nominals.

Beside the obligatory features, nominals can have features marking the number and person of the possessor and the number of the possessee. Table 3 lists these features and their possible values.

feature	value	description	example
Number[psor]	Sing	singular	kutyája
	Plur	plural	kutyájuk
Number[psed]	Sing	singular	kutyáé
	Plur	plural	kutyáéi
Person[psor]	1	1st person	kutyám
	2	2nd person	kutyád
	3	3rd person	kutyája

Table 3: Features and values marking possession.

Adjectives Beside the obligatory features and the possession-marking features, adjectives have three specific features. The value of the feature `Degree` marks the degree of comparison. Ordinal numbers are treated as adjectives, and they get an `Ord` value of the feature `NumType`. Ordinal numbers are not comparable, therefore they do not have a `Degree` feature. If the adjective is derived from a verb, it gets a participle feature (`VerbForm`). Table 4 lists the specific features of adjectives with their possible values. Common features are marked with dots.

feature	value	description	example
Case	*Nom*	nominative	szép
	...	...	...
Degree	Pos	positive	szép
	Cmp	comparative	szebb
	Sup	superlative	legszebb
	Abs	absolute superlative	legeslegszebb
NumType	Ord	ordinal	első
Number	*Sing*	singular	szép
	...	...	...
Number[psor]	...	...	...
Number[psed]	...	...	...
Person[psor]	...	...	...
VerbForm	PartPres	present participle	épülő
	PartPast	past participle	épült
	PartFut	future participle	épülendő

Table 4: The features of adjectives with their possible values.

Numerals Beside the common features of nominals, numerals have two extra features relating to their type and semantics. The feature `NumType` indicates if the word is a cardinal (`Card`), a fractal (`Frac`) or a distributive (`Dist`) numeral. This is an obligatory feature, whose defult value is `Card`. Ordinal numbers are treated as adjectives. The feature `NumType[sem]` is a lexical feature indicating the format of the numeral. Table 5 lists the extra features of numerals and their possible values.

Pronouns Beside the common features of nominals, pronouns must have one of these features: `Poss` or `PronType` or `Reflex`. Both of these features indicate the type of the pronoun. Table 6 lists the possible values of these features.

3.2. Verbs

Main verbs (VERB) and auxiliaries (AUX) share a common feature set, which is listed in Table 7. Obligatory features are `VerbForm` and `Voice`. The default value of `Voice` is active (`Act`) which means that it lacks derivational suffix of causative voice (`Cau`). The feature `VerbForm` divides verbs into two groups: finite verbs (`Fin`) and infinitives (`Inf`). Since auxiliaries can only be finite verbs, they have the same features and values as main finite verbs do. While the presence and the exact stock of auxiliaries in the Hungarian language are constantly disputed, there are four auxiliaries distinguished in UD1: *volna*, *fog*, *talál* and *szokott*.

feature	value	description
Case	*Nom*	nominative
	...	...
NumType	*Card*	cardinal
	Frac	fractal
	Dist	distributive
NumType[sem]	Dot	dotted number
	Formula	formula
	Quotient	quotient
	Result	result
	Signed	number with sign
	Time	temporal expression
	Percent	number with percent sign
	Grade	number with grade
	Measure	measurement
	Other	other
Number	*Sing*	singular
Number[psor]	...	...
Number[psed]	...	...
Person[psor]	...	...

Table 5: The features of numerals with their possible values.

feature	value	description	example
Case	*Nom*	nominative	
	...	...	
Number	*Sing*	singular	
	...	...	
Number[psor]	...	...	
Number[psed]	...	...	
Person	...	...	
Person[psor]	...	...	
Poss	Yes	possessive	tiéd
PronType	Prs	personal	te
	Rcp	reciprocal	egymás
	Int	interrogative	ki
	Rel	relative	aki
	Dem	demonstrative	ez
	Tot	total	minden
	Ind	indefinite	valaki
Reflex	Yes	reflexive	magad

Table 6: The features of pronouns with their possible values.

Finite verbs Table 8 lists all possible features and values of finite verbs. The feature `Mood` can take multiple values simultaneously as the feature of potentiality can be added to other features indicating mood. In Hungarian, the suffixes of a conjugated verb form represents five morphosyntactic attributes: tense, mood, person, number, and the definiteness of the object. Therefore, these are the obligatory features of verbs. The active voice, the present tense, the indicative mood, and the 3rd person singular are unmarked in Hungarian, therefore these are the default values of the corresponding features.

Infinitives There are two kinds of infinitives: the impersonal infinitive and the personal infinitive. Table 9 lists possible features and values of impersonal infinitives. The value of the feature `VerbForm` must be `Inf`. Voice, aspect and mood are represented by (derivational) suffixes in Hungarian, thus they can be present on an impersonal

feature	value	description	example
Aspect	Freq	frequentative	ütöget
Definite	Def	definite	üti
	Ind	indefinite	üt
	2	2nd person	ütlek
Mood	Ind	indicative	üt
	Cnd	conditional	ütne
	Imp	imperative	üss
	Pot	modal	üthet
Number	Sing	singular	üt
	Plur	plural	ütnek
Person	1	1st person	ütök
	2	2nd person	ütsz
	3	3rd person	üt
Tense	Pres	present	üt
	Past	past	ütött
VerbForm	Fin	finite	üt
	Inf	infinitive	ütni
Voice	Act	active	üt
	Cau	causative	üttet

Table 7: The features of main verbs and auxiliaries with their possible values.

feature	value	description
Aspect	Freq	frequentative
Definite	Def	definite
	Ind	indefinite
	2	2nd person
Mood	*Ind*	indicative
	Cnd	conditional
	Imp	imperative
	Pot	modal
Number	*Sing*	singular
	Plur	plural
Person	1	1st person
	2	2nd person
	3	3rd person
Tense	*Pres*	present
	Past	past
VerbForm	Fin	finite
Voice	*Act*	active
	Cau	causative

Table 8: The features of finite verbs with their possible values.

infinitive form as well.

Table 10 lists all possible features and values of personal infinitives. Personal infinitives also have a `Tense=Pres` attribute–value pair in the Szeged Dependency Treebank, even though personal infinitives do not have a slot for representing this feature.

3.3. Adverbials

Similarly to adjectives, adverbials can also have the `Degree` feature, unless they are not adverbial participles derived from a verb (`VerbForm=Trans`). Adverbial pronouns have a feature indicating the pronoun type (`PronType`). Table 11 lists the possible features and values of adverbials. There are no obligatory features for adverbials.

feature	value	description	example
Aspect	Freq	frequentative	ütögetni
Mood	Pot	modal	üthetni
VerbForm	Inf	infinitive	ütni
Voice	*Act*	active	ütni
	Cau	causative	üttetni

Table 9: The features of impersonal infinitives with their possible values.

feature	value	description	example
Aspect	Freq	frequentative	ütögetnem
Mood	Pot	modal	üthetnem
Number	Sing	singular	ütnöm
	Plur	plural	ütnünk
Person	1	1st person	ütnöm
	2	2nd person	ütnöd
	3	3rd person	ütnie
VerbForm	Inf	infinitive	ütnöm
Voice	Act	active	ütnöm
	Cau	causative	üttetnem

Table 10: The features of personal infinitives with their possible values.

feature	value	description	example
Degree	Pos	positive	későn
	Cmp	comparative	később
	Sup	superlative	legkésőbb
	Abs	absolute superlative	legeslegkésőbb
PronType	Int	interrogative	mikor
	Rel	relative	amikor
	Dem	demonstrative	itt
	Tot	total	mindenhol
	Neg	negative	sehol
	Ind	indefinite	valahol
VerbForm	Trans	adverbial participle	ütve

Table 11: The features of adverbials with their possible values.

3.4. Determiners

Definite and indefinite article form a separate POS category with two obligatory features, as can be seen in Table 12.

feature	value	description	example
Definite	Def	definite	a
	Ind	indefinite	egy
PronType	Art	article	a

Table 12: The features of determiners with their possible values.