

Checking rules

Adj parent

- na adjektivu (tag A*) nevisí uzel s afunem *Atr*. (Výjimky: elipsa, lemma rodičovského uzlu je: *další, někteří, každý, jiný, jediný, ostatní, mnohý*, název barvy *_o; příjmení typu Novákov (AU* a v lemmatu je _Y).

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář)

Afun not preposition

- uzel má na morfologické rovině tag pro předložku (R na první pozici), ale na a-rovině není jako předložka anotován (tj. nemá afun *AuxP*).

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář)

<afun1> undef <afun2>

- *AuxP* a *AuxC* nemohou viset na sobě. Výjimky.

Ampersand parent not previous

- ampersand &, pokud má *AuxG*, visí na slově předcházejícím (pokud není prvním znakem ve větě)

Apostrophe parent

- apostrof, pokud má *AuxG*, visí na slově předcházejícím (pokud je součástí výrazů jako *Moody's*). Pokud má funkci uvozovky, visí otevírací uvozovky na slově následujícím a zavírací uvozovky na slově předcházejícím.

Atv parent

- *Atv* (doplňěk) visí na slovese nebo adjektivu

AuxP without preposition

- *AuxP* mohou být jen morfologické předložky (tj. v tagu je na první pozici R) nebo víceslovné předložky (tj. na uzlu visí uzel s afunem *AuxY*)

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář)

AuxZ under AuxZ

- na *AuxZ* nemůže viset jiné *AuxZ*; dvě částice zachycujeme jako sestry

Cannot be extra dep

- *Pred* a *Denom* nemohou být *is_extra_dependency* (tj. *Pred_E* a *Denom_E* je nedovolená kombinace)

Cannot have children

- *AuxY* nemá děti kromě *AuxG*

- *AuxT*, *AuxR*, *AuxV* nemá děti kromě *AuxZ*, *AuxG*

- *AuxK*, *AuxG*, *AuxX* nemohou mít děti (bez výjimky)

Dash parent

- pomlčka/spojovník může mít *Coord/Apos*; pokud má *AuxG*, pak je prvním nebo posledním slovem ve větě, visí na uzlu s *Coord/Apos/AuxC*, na spojce *-li*, na uzlu s příponou *_P* nebo na slově (nikoli znaku; % a \$ se považují za slova) předcházejícím (*C-dur*; *5%-ní*) nebo následujícím (*Praha-Holešovice*). Další výjimky: rodičem je *Pred*, *AuxC*.

Deprecated afun

- *ExD*, *AtvV* kombinované funkce *AtrAtr*, *AtrAdv*, *AdvAtr*, *AtrObj*, *ObjAtr* jsou zrušené afuny

Different afuns in Coord/Apos

- členy v koordinaci nebo v apozici by měly mít stejné funkce; výjimkou jsou kombinace *Atv – Adv*, *Pred – Denom*, *Pred – Partl* a *Denom – Partl*.

Dot parent not previous

- tečka, pokud má *AuxG*, visí na slově předcházejícím

Invalid afun for lemma

- hlídá povolené afuny pro interpunkci:

'> => Coord|Apos|AuxX

'> => Coord|Apos|AuxG

'(' => Apos|AuxG

'.' => Pred|Coord|Apos|AuxG

'-' => Pred|Coord|Apos|AuxG

'/' => Coord|Apos|AuxG

'&' => Coord|AuxG

Invalid AuxK

- *AuxK* mohou dostat jen tečka, vykřičník a otazník (a trojtečka).

Invalid AuxV parent

- *AuxV* závisí na slovese (tag má na první pozici *V*)

Kontrola nehlásí chybu, pokud uzel má *is_extra_dependency: AuxV_E*

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář)

Invalid _Co _Ap _P _E

- afuny *AuxK*, *AuxG*, *AuxX*, *AuxS* nemohou mít žádnou příponu (tj. např. *AuxG_E* je nedovolená kombinace).

Invalid Denom case

- Uzel s afunem *Denom* by měl být v 1. pádě (Pád je v morfologickém tagu na 5. pozici).

Výjimka: názvy měsíců/datum, např. *28. ledna.Denom*

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář)

Invalid Sb case

- Podmět v jiném než 1. pádě (2. pád se ukazuje jen jako "Warning")

(Pád je v morfologickém tagu na 5. pozici)

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář).

Jedná-li se o pravidelnou změnu (*přišlo deset lidí -> kolem deseti lidí*), je třeba do komentáře poznámky *Other* napsat „sb“.

Invalid Sb parent

- Rodičem uzlu s afunem *Sb* je sloveso (tj. uzel s tagem začínajícím na *V*) nebo *- : =*.

Kontrola nehlásí chybu, pokud rodič má *is_extra_dependency*.

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář).

Invalid tag/aux combination

- hlídá odpovídající tagy pro afuny *AuxC*, *AuxX*, *AuxG*, *AuxK*, *AuxV*, *AuxT*, *AuxR*

AuxC => J,

AuxX => Z :

AuxG => Z :

AuxK => Z :

AuxV => V

AuxT => P7--[34]

AuxR => P7--[34]

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář).

Invalid top node

- jenom Coord, Apos, Pred, Denom, Vocat, a fun_E mohou být pod AuxS

Multiple Pnom/Sb

Pod slovesem nemůže viset dvakrát Pnom nebo Sb.

Must be top node

- Pred a Denom nemohou být hlouběji ve stromě (pouze v případě "obráceného syntaktického vztahu" (př. *jak je řekl Klaus, je to nesmysl*) visí Pred pod AuxC_P). Hlouběji zanořené Pred nebo Denom musí mít příponu _P (*is_parenthesis root*)

Noun parent

- pod substantivem (tag NN) by neměl viset Adv, Obj, Sb, Atv (rozvití substantiva je Attr).

Not enough members

- koordinace/apozice by měla mít alespoň dva členy (s příponou _Co/Ap). Výjimkou jsou jednočlenné koordinace typu *rajčata apod.*

parent not být/bývat

- Pnom by neměl viset na ničem jiném než na *být* a jeho vidových derivátech (rozeznává i "*bývávat*")

Parent not _P

- závorka, pokud má AuxG, visí na uzlu s příponou _P (Povolen je také uzel s Coord/Apos a AuxZ)

Parent too far

- mezi uvozovku a jejího rodiče se v povrchovém pořadí nemůže dostat nic kromě interpunkce (počáteční uvozovky visí na slově bezprostředně následujícím, koncové na slově bezprostředně předcházejícím).

Preposition case not matching

- kontrola pro chyby na M-rovině. Pád předložky (AuxP) a synovského uzlu by měl být ve shodě, tj. na předložce se 6. pádem visí jméno v 6. pádě.

Několik výjimek:

-- rodičovský/synovský uzel má poznámku Other: překlep

-- rodičovský/synovský uzel má poznámku M-layer

-- výjimka synovský uzel má _E

A pak je jedna datumová výjimka: měsíc po předložce na-1 RR--4----- může mít genitiv (*přesunuli výuku na 4. prosince / 21. listopadu ...*). Podobnou výjimku máme i pro Denom.

S2 not AuxY

- uzel s tagem S2 má a fun AuxY.

Slash parent

- lomítko, pokud má AuxG visí na slově předcházejícím (*1/8*) nebo následujícím (*km/h, 5/8*) nebo na slově s _P (pokud odděluje vsuvku).

Tag J, – should be AuxC

- morfologická podřadící spojka (tag J,) by měla být AuxC

- (výjimky jsou *aniž, jakož, takže, jako, než-2, li-2*)

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář)

Tag J^ - should be Coord/Apos/AuxZ

- morfologická souřadící spojka (tag J^) by měla být Coord/Apos/AuxZ

Je-li chyba v morfologickém tagu, je třeba dát poznámku *M-layer* (není třeba komentář)

Verb parent

- pod slovesem (tag V) by neměl viset Atr

Warning: no children

- AuxP a AuxC by měly mít nějaké dítě

Warning: No punctuation

- uzel s příponou *_P* (*is_parenthesis_root*), na kterém nevisí interpunkce (na uzlu s příponou *_P* by měla viset interpunkce (závorka, pomlčka apod.) oddělující parentezi od základního textu. Je však celá řada výjimek (nespojité parenteze, parenteze na začátku nebo na konci věty)

Slovnědruhové kategorie v morfologickém tagu:

N noun

A adjectivum

P Pronoun

C číslovka

V verb

D aDverb

R pReposition,

J, podřadící spojka

J^ souřadící spojka

T částice

I citoslovce

F cizí slovo

Q písmeno

B zkratka

S segment