

Návrh morfológického tagsetu SNK¹

Martina Forróová – Radovan Garabík – Lucia Gianitsová – Alexander Horák – Mária Šimková:

Abstract: The article brings definitions of rules for automatic text segmentation and semi-automatic lemmatization and POS tagging of the texts in the Slovak National Corpus. The basic unit on the level of text segmentation is token, which is traditionally defined as a sequence of alphanumeric characters constrained by whitespace. On the level of morphological processing each token is morphologically interpreted by the attributes of the basic word form (lemma) and a set of relevant morphological categories (tag). Assuming that the part of the morphological analysis will be done manually the concepts of lemma and morphological tagset are designed in a way to pre-disambiguate as possible ambiguities given by the nature of Slovak language.

1 Tokenizácia

1.1 Definícia tokenu

Token je arbitrárna jednotka textu, ktorá rozširuje lingvistický význam pojmu slovo. Za token sa v morfológickej anotácii korpusu považuje akýkoľvek reťazec znakov medzi dvoma medzerami (teda aj jednotlivé znaky interpunkcie).

Text sa skladá z tokenov a medzier. Tokeny by mohli obsahovať vo svojom vnútri whitespace (príklad: *1 234 567 890*), ale v prvej fáze budeme za token považovať všetko medzi whitespace.

Pri tokenizácii vstupného textu má význam rozdelenie znakov podľa vzťahu k tokenu na:

- znaky, ktoré sú vždy súčasťou tokenu. Postupnosť takýchto znakov vždy tvorí token.
Príklad: písmená abecedy (až na malé výnimky).
- znaky, ktoré sú vždy samy osebe tokenom bez ohľadu na susedné znaky.
Príklad: dvojbodka, matematické symboly a iné.
- znaky, ktoré sú v niektorých prípadoch súčasťou tokenu, v niektorých nie.
Príklad: v tokene *k*k*t* je *** časťou tokenu, v zápise $2 \times 3 - 5 = 2$ je každý znak tokenom, v texte **toto je zvýraznený text** sú obidve hviezdičky samostatné tokeny.

Tokenu sa v procese morfológickej anotácie priradujú atribúty lema a tag.

2 Lematizácia

2.1 Definícia lemy

Lema je základný, "slovníkový" tvar tokenu, napr.: *pekná = pekný, varené = varený* a pod. Nerozlišujeme veľké a malé písmená, teda vlastné a všeobecné podstatné mená, za-

¹ V príspevku sa predstavuje prvá verzia, resp. návrh tagsetu SNK z r. 2003, ktorý bol neskôr prepracovaný a v máji 2004 zverejnený na stránke Slovenského národného korpusu. V súčasnosti sa anotuje podľa najnovšej verzie aktualizovanej koncom r. 2005 (<http://korpus.juls.savba.sk>).

čiatky viet atď. – lema je vždy s malým začiatočným písmenom (*Alexander* = *alexander*).

Pri pomnožných substantívach bude lema v sg. len v relevantných prípadoch, napr. *nohavice* = *nohavica*, ale *Alpy* = *alpy* (nie *alpa*).

Pri určovaní lemy venujeme pozornosť:

1. tvarom

- typu *chlapovi*, *chlapov*, *otcovi*, *otcov* – môže ísť o substantíva s lemov *chlap*, *otec* (*otcovi* – D, L sg., *otcov* – G, A pl.) alebo o privlastňovacie prídavné meno s lemov *chlapov*, *otcov* (*otcov* – N, A sg., *otcovi* – N pl. mask.). Pri určení lemy sa riadime kontextom. Napr.: *Otcovi to nepovedz.* (lema – *otec*), *Otcovi priatelia neprišli.* (lema – *otcov*).
- *nevojak*, *nemocný*, *nevedomosť*, *nejeden*, *nemálo*, *neveľa*, *neraz*. Pri týchto tvaroch ide o samostatné lemy, nie negované tvary lem *vojak*, **mocný*, *vedomosť*, *jeden*, *málo*, *veľa*, *raz*. Kategória negácie sa určuje len pri slovesách a aj tu sa za lemu považujú negované podoby. Sú teda lemy napr. *vidieť*, *mať*, *chcieť*, ale aj lemy *nevidieť*, *nemať*, *nechcieť*...

2. niektorým zámenám ako

- zraštené tvary (*oňho*, *preňho*, *naňho*, *oň*, *preň*, *zaň*) považujeme za tvary zámen *on*, *ono*, predložková časť sa dá chápať ako „gramatická“ (v tomto prípade relačná – pádová) morféma. Teda: *oňho*, *preňho*, *naňho* = lema *on*; *oň*, *preň*, *zaň* = lema *ono*.
- zložené tvary zámen (podľa KSSJ: *ten istý*, *tá istá*, *to isté*, *taký istý*, *tak isto*, *koľko ráz*, *-koľko ráz*, *koľký raz*, *-koľký raz*, *tol'ko ráz*, *tol'ký raz*, *tamto ten*, *tamto tá*, *tamto to*) a tvary, ktoré sa v korpusových textoch vyskytnú s pomlčkou, hoci by sa mali písať so spojovníkom (podľa KSSJ: *čo-to*, *koľký-tol'ký*, *koľko-tol'ko*, *aký-taký*, *ako-tak*, *kde-tu*, *kade-tade*, *kedy-tedy*, *kdesi-čosi*, *čosi-kdesi*, *čosi-kamsi*, *ten-ktorý*, *tá-ktorá*, *to-ktoré*), označíme ako dva tokeny – dve lemy, aj keď rešpektujeme, že ide o jednu lexému. Premietne sa to v morfológickej anotácii. Zloženiny zapísané správne so spojovníkom budú tokenizované ako jeden token a následne anotované ako jedna lema.²
- *koľkí*, *tol'kí*, *koľkých*, *tol'kých* – môžu to byť zámená odkazujúce na základné číslovky (lema – *koľko*, *tol'ko*) aj radové číslovky (lema – *koľký*, *tol'ký*). Napr.: *Koľkí žiaci prišli?* (lema – *koľko*), *Koľkí v poradí boli naši pretekári na súťaži?* (lema – *koľký*).
- *samého*, *samému*, *samého*, *o samom*, *so samým*, *samých*, *samým*, *samých*, *o samých*, *so samými* – nepriame pády zámen *sám* a *samý* sú homonymné, pri určovaní lemy vychádzame z kontextu. Napr.: *Mrzí to jeho samého.* (lema – *sám*), *Ušili šaty zo samého zlata.* (lema – *samý*).
- *jeho*, *jej*, *ich* – môže to byť osobné základné zámeno (*jeho* – G, A sg. s lemov *on*; *jej* – G, D sg. s lemov *ona*; *ich* – G, A pl. s lemov *oni*, *ony*) alebo osobné privlastňovacie zámeno s lemov *jeho*, *jej*, *ich*. Rozlíšime ich na základe postavenia a kontextu. Napr.: *Jeho to mrzí.* (lema – *on*), *Dám to jeho priateľom.* (lema – *jeho*).

² Aj v tokenizácii sa urobili čiastkové modifikácie – napr. všetky takéto zloženiny sa tokenizujú ako osobitné reťazce znakov (*tá*, *-*, *ktorá*).

- *si* – môže byť tvarom slovesa *byť* (2. os. sg. prít. času) alebo reflexívnym *si*.
3. číslovkám a číselným výrazom
 - pri číselných výrazoch vyjadrených číslom alebo symbolom lema bude číslo alebo symbol,
 - pri rímskych čísliciach lema bude rímska číslica,
 - číselné výrazy vyjadrené slovom sa lematizujú slovom,
 - číslovky *veľa, viac, najviac; málo, menej, najmenej* sú samostatnými lemmami, keďže nejde o vyjadrenie kategórie stupňa, ale o lexikalizované tvary.
 4. nejednoznačným tokenom ako *XDR, XEU, XML, až100, až-60, ko(z)mický, (Vysoké) Tá3*. Lematizujú sa podľa tvaru, ktorý sa vyskytol v texte, teda *xdr, xeu, xml, až100, až-60, ko(z)mický, ta3*, aj napriek tomu, že pri niektorých ide o zjavné chyby.

3 Morfológické značkovanie

3.1 Forma zápisu atribútov a ich hodnôt v tagsete

Hodnoty jednotlivých kategórií sa zapisujú jednomiestnou značkou. Značky sú tvorené písmenami latinskej abecedy, číslicami a matematickými symbolmi. Súbor týchto značiek tvorí jeden tag k jednému tokenu. Počet značiek je variabilný, ale ich poradie v tagu je záväzné.

3.2 Slovné triedy

V morfológickom tagsete SNK sa množina slovných foriem používaných v slovenčine rozdeľuje do 16 tried, z ktorých 10 v zásade zodpovedá tradične vydeľovaným slovným druhom a 6 obsahuje rôzne špecifické jazykové prvky³:

1. S Substantíva
2. A Adjektíva
3. P Pronominá
4. N Numeráliá
5. V Verbá
6. D Adverbiá
7. E Prepozície
8. O Konjunkcie
9. T Partikuly
10. J Interjekcie
11. R Reflexívne *sa/si*
12. Y Kondicionálová morféma *by* (+ všetky tvary s elementom *by*)
13. Z Interpunkcia
14. W Abreviácie a značky
15. Q Neurčiteľný slovný druh
16. # Neslovný element

³ V ďalších verziách tagsetu sa počet tried zvýšil na 19: pribudli participiá (G), citátové výrazy (%) a číslice (0).

Pri klasických slovných druhoch sa v podstate rešpektuje slovnodruhové zaradenie podľa súčasných slovenských kodifikačných príručiek (KSSJ, PSP). Hraničné prípady predstavujú:

1. Slovesné podstatné mená, ktoré zaraďujeme k substantívam (**S**) – *písanie, hovorenie, rešpektovanie*.
2. Činné a trpné prídavné, ktoré zaraďujeme k adjektívam (**A**) – *písaný, otvorený, obutý, píšuci, hrajúci, stojaci*.⁴
3. Substantíva (apelatíva aj propriá) s adjektívnou paradigmou, ktoré zaraďujeme k substantívam (**S**) – *Chladný, Mináriková, cestovné, vrchný, vedúci, vedúca*.
4. Zrastené tvary (*oňho, preňho, naňho, oň, preň, zaň*) označujeme ako zámená (**P**), predložková časť sa dá chápať ako „gramatická“ (v tomto prípade pádová) morféma.
5. Nerozlišujeme medzi zvratným zámenom *sa, si* a *sa, si* ako súčasťou slovesa. Všetky tvary *sa, si* majú priradenú osobitnú značku **R**. Ich spracovanie ponechávame na osobitné štúdie vypracované na základe korpusu, prípadne na syntaktickú anotáciu.
6. Morféma *by*, ktorá je súčasťou kondicionálneho slovesného tvaru, a ďalšie tvary *s by (keby, aby, žeby, akoby, sľaby)* majú samostatnú značku **Y**.
7. Druhotné predložky, ktoré sú zložené z dvoch a viacerých samostatných tokenov (*s ohľadom na, v závislosti od, na rozdiel od*), sa anotujú ako jednotlivé (dve alebo viaceré lemy) a každej leme priradíme značku **E** (prepozície).⁵
8. Slovné hračky typu *ko(z)mický, 100dola, Vysoké Ta3* zaraďujeme do slovného druhu podľa klasického delenia. Teda *ko(z)mický* ako adjektívum, *100dola, ta3* substantívum.

3.3 Morfológické kategórie

Pri morfológickom značkovani sme vychádzali z teórie gramatických kategórií, ako ich podáva akademická Morfológia slovenského jazyka (1966), príp. iné morfológické práce. V SNK sa však stretáme aj s kategóriami a ich hodnotami, o ktorých sa v tradičnej morfológii ako o kategóriách neuvažovalo (paradigma, slovesná forma). Ide o formálno--morfológické charakteristiky, ktoré sú dôležité na zjednotenie tokenu. Kategórie v našom chápaní nemajú binárnu, ale graduálnu povahu – pre isté slovné formy platia vo väčšej miere ako pre iné.

Pri morfológicky netransparentných aj homonymných tvaroch, ktoré nie sú jednoznačne určiteľné z kontextu, uvádzame pre rod, číslo alebo pád hodnotu neurčený so značkou **o** (*Mužov, žien a detí je päť. Kúpili kanoe. Mesto!*).⁶

3.3.1 Pri ohybných slovných druhoch sa tag začína značkou slovného druhu, ako sa vymedzuje podľa klasických komplexných kritérií a uvádza aj v príslušných lexikografic-

⁴ Tieto tvary, ako uvádzame v pozn. 2, sme postupne začali vymedzovať samostatne ako formálne participiá (**G**).

⁵ Jednoslovné druhotné predložky sa naďalej označujú ako **E**, no substantívne súčasti pri dvoj- a trojslovných predložkách ako **S**.

⁶ Táto značka bola v poslednej verzii tagsetu v podstate zrušená, ostala len vo výnimočných prípadoch (porov. <http://korpus.juls.savba.sk>).

kých príručkách. Napr.: *mama, vedúci, kanoé, pani* – **S**, *pekny, super* – **A**, *taky, on, žiaden, tam* – **P**, *sto, štvrtý, tri, jeden, päťoro* – **N**, *písať, píšem, piš, pišuc, písal, budem* – **V**.

3.3.2 Pri menných slovných druhoch sa na druhom mieste uvádza značka pre typ **paradigmy** s hodnotami substantívna, adjektívna, číslovková, zmiešaná, neúplná, bez paradigmy.

Formálny atribút paradigma chápeme ako vymedzenie špecifickej formy konkrétneho člena slovného druhu (napr. *taky* je zámeno, no má formu adjektíva). Pre niektoré typy paradigmy (substantívnu, adjektívnu, zámennú a číslovkovú) sa používajú rovnaké značky ako pre slovný druh (**S, A, P, N**).⁷

Zmiešaná paradigma (**F**) platí pre slová, ktoré majú sčasti adjektívnu paradigmu. Patria sem slová typu:

- *pani* (N sg., G, D, L, I pl. nie je adj.)
- *gazdiná* (G, D, L, I pl. nie je adj.) a všetky rovnako sa skloňujúce substantíva
- *sám, jeden* (N sg. nie je adj.) – pri *jeden* len v spojení so životnými mask. (*jeden chlap, hrdina*)
- *onen, žiaden, jeden* v spojení s muž. neživotnými substantívami, napr. *jeden dub, jeden stroj, jedna, jedno, všetok, všetka, všetko, sama, samo* (N, A sg. nie je adj.)
- *všetci, sami, jedni, viacero* (N pl. nie je adj.)
- *všetky, samy, jedny* (N, A pl. nie je adj.)
- *koľko, toľko* a číslovky podľa vzoru *päť*, ktoré síce bývajú použité ako nesklonné (tvary (*od*) *koľko/toľko/päť ľudí, koľko/toľko/päť ľudom*), ale je tu možnosť použiť aj sklonné podoby (*koľki/toľki/piati ľudia, od koľkých/toľkých/piatich ľudí, koľkým/toľkým/piatim ľudom*). Pri skloňovaní mask. neživ., femin. a neutier sa používajú v N a A iba tvary *koľko, toľko, päť*.

Neúplná paradigma (**U**) bude priradená tým leмам substantív, adjektív, zámen a čísloviek, ktoré sa v tradičnej gramatike označujú ako nesklonné (*kupé, super, jeho, sto*), ale majú tendenciu k sklonnosti. Napr. *kanoé*, G pl. *od kanoí*.

Bez paradigmy (**X**) sú také zámena a číslovky, ktoré sú nesklonné a v KSSJ bývajú označené kvalifikátorom *neskl.* alebo *prislov.* (*koľkonásobne, koľkorako, tam, tu, vtedy, vždy, viacnásobne, dvojako*). V týchto prípadoch sa tag určením paradigmy končí. Ďalšie kategórie ako rod, číslo, pád alebo osoba sa neuvádzajú.

Prvé dve značky v tagu čítame takto:

- **SS** – substantívum so substantívnu paradigmou (*mama*)
- **SA** – substantívum s adjektívnu paradigmou (*vedúci*)
- **SF** – substantívum so zmiešanou paradigmou (*pani, gazdiná*)
- **SU** – substantívum s neúplnou (neviditeľnou, nerealizovanou) paradigmou

⁷ Rovnako aj v iných prípadoch sa na označenie tej istej kategórie v rôznych slovných druhoch používa tá istá značka (napr. stupeň má pri **A** aj pri **D** rovnaké označenie **x, y, z** – porov. 3.3.6).

(*kanoe*)

- **AA** – adjektívum s adjektívnou paradigmou (*pekný*)
- **AU** – adjektívum s neúplnou paradigmou (*super*)
- **PS** – zámeno so substantívnou paradigmou (*koľkátko*)
- **PA** – zámeno s adjektívnou paradigmou (*taký*)
- **PP** – zámeno so samostatnou zámennou paradigmou (*on*)
- **PF** – zámeno so zmiešanou paradigmou (*žiaden*)
- **PU** – zámeno s neúplnou paradigmou (*jeho, jej, ich*)
- **PX** – zámeno bez skloňovacej paradigmy (*tam*)
- **NS** – číslovka so substantívnou paradigmou (*sto*)
- **NA** – číslovka s adjektívnou paradigmou (*štvrtý*)
- **NN** – číslovka so samostatnou číslovkovou paradigmou (*tri*)
- **NF** – číslovka so zmiešanou paradigmou (*jeden*)
- **NU** – číslovka s neúplnou paradigmou (*päť, sto, tisíc*)
- **NX** – číslovka bez skloňovacej paradigmy (*päťoro*)

3.3.3 Na treťom mieste sa pri menných slovných druhoch, ktoré majú paradigmu, uvádza **rod** s hodnotami mužský životný, mužský neživotný, ženský, stredný, neurčený.

Rod uvádzame, ak je jednoznačný alebo určiteľný z kontextu, napr. na základe kongruencie s určovaným substantívom (*pekná mama, on, ten, nejaký svet, druhé dieťa*). Ak ani z kontextu nie je možné rod zjednotiť (*Ženy a deti, ony neprišli.*), uplatníme značku **o** (**neurčený**).⁸

Aj pri tzv. bezrodových zámenách (*ja, ty, my, vy, seba, kto, čo, nikto, nič, niekto, niečo...*) sa uvádza konkrétny rod, ak je zistiteľný z kontextu. Diskutabilné prípady, kde by bolo zistenie z kontextu náročné alebo nemožné, budú mať takisto značku **o**.

3.3.4 Číslo s hodnotami singulár, plurál a neurčené uvádzame pri ohybných slovných druhoch na štvrtom mieste. Hodnotu neurčené číslo (**o**) zapíšeme, ak sa pri tvare nedá z kontextu jednoznačne určiť jeho číslo (*Kúpili kanoe.*).

Okrem číslovky *jeden* vyjadrujú základné číslovky (*dva, tri, štyri*) plurálnosť, gramatický protiklad sg./pl. sa tu formálne nevyjadruje. Určujeme ho na základe zhody s nadradeným substantívom ako pri adjektívach alebo zámenách a číslovkách s povahou adjektív (*Tri domy.* – pl.).

3.3.5 Pád uvádzame pri menných slovných druhoch na piatom mieste. Má hodnoty nominatív, genitív, datív, akuzatív, vokatív, lokál, inštrumentál, neurčený. Hodnotu neurčený pád (**o**) priradíme tvarom, pri ktorých ani z kontextu nie je možné pád jednoznačne určiť (*Mesto!* – N alebo A sg.).

3.3.6 Stupeň s hodnotami pozitív, komparatív, superlatív platí pre adjektíva a adverbiá. Kategóriu stupňa určujeme formálne, riadime sa prítomnosťou alebo neprítomnosťou modifikačných morférov *naj-, -ší, -ejší*. Stupeň je určený aj pri vzťahových adjektívach,

⁸ Zmeny v používaní značky **o** porov. podľa Garabík – Gianitsová – Karčová – Šimková (<http://korpus.juls.savba.sk>).

keďže aj tieto typy adjektív sa môžu potenciálne stupňovať (**najdrevenejší stôl* ale *najdrevenejší tanečník*).

Katégoria stupňa sa neurčuje pri číslovkách *veľa*, *viac*, *najviac*; *málo*, *menej*, *najmenej*, keďže tu nejde o gramatickú katégoriu ako pri adjektívach, ale o lexikalizované tvary, ktoré na morfologickej rovine považujeme za samostatné lexémy.

Príklady anotovania menných slovných druhov:

<i>Lingvista</i> anotoval texty z korpusu.	l = lingvista	t = SSms1
Nechcem cestovať v tom kupé .	l = kupé	t = SUns6
Poznám vedúceho oddelenia.	l = vedúci	t = SAmS4
Asi bol lepším človekom, ako vyzeral.	l = dobrý	t = AAmS3y
Pozdravil sa jeho sestru.	l = jeho	t = PUfs3
Pre koľko (koľkých) ľudí to má pripraviť?	l = koľko	t = PUns4
Potešil sa päťoro /piatim nohavičiam.	l = päťoro	t = NUfp3
Z milióna neostalo nič.	l = milión	t = NSis2

3.3.7 Pri slovesách sa na druhom mieste určuje **slovesná forma** s hodnotami: infinitív, prítomný čas (indikatív), imperatív, prechodník, *l*-ové prítomné, futúrum slovesa *byť*.

Katégorie času a spôsobu sa neurčujú osobitne, keďže sú obsiahnuté v jednotlivých určitých tvaroch sloves, ktoré značíme takto:

Indikatív

- tvary **prítomného času** (*pišem*, *pišeš*, *piše*, *pišeme*, *pišete*, *pišu*) – sloveso *pišať* v prítomnej forme (**VK**);
- tvary **minulého času** (*pisal som*, *pisal si*, *pisal*, *pisali sme*, *pisali ste*, *pisali*) – sloveso *pišať* vo forme *l*-ového prítomného (**VL**) + sloveso *byť* v prítomnej forme (**VK**), ak je prítomné;
- tvary **budúceho času nedokonavých sloves** (*budem písať*, *budeš p.*, *bude p.*, *budeme p.*, *budete p.*, *budú p.*) – sloveso *byť* vo futúre (**VB**) + sloveso *pišať* v infinitíve (**VI**);
- tvary **budúceho času dokonavých sloves** (*napíšem*, *napíšeš*, *napíše*, *napíšeme*, *napíšete*, *napíšu*) – sloveso *napísať* v prítomnej forme (**VK**).

Imperatív

- len tvary **prítomného času** (*piš!*, *pišme!*, *pište!*) – sloveso *pišať* v imperatívnej forme (**VM**).

Kondicionál

- tvary **prítomného času** (*pisal/-la/-lo by som*; *pisal/-la/-lo by si*; *pisal/-la/-lo by*; *pisali by sme*; *pisali by ste*; *pisali by*) – sloveso *pišať* vo forme *l*-ového prítomného (**VL**) + *by* (**Y**) + sloveso *byť* v prítomnej forme (**VK**), ak je prítomné;
- tvary **minulého času** (*bol/-la/-lo by som* *pisal/-la/-lo*; *bol/-la/-lo by si* *pisal/-la/-lo*; *bol/-la/-lo by* *pisal/-la/-lo*; *boli by sme* *pisali*; *boli by ste* *pisali*; *boli by* *pisali*) – sloveso *bol* vo forme *l*-ového prítomného (**VL**) + *by* (**Y**) + sloveso *byť* v prítomnej forme (**VK**), ak je prítomné + *pišať* vo forme *l*-ového prítomného (**VL**).

3.3.8 Slovesný vid s hodnotami dokonavý, nedokonavý alebo obojvidový sa nachádza pri slovesách na treťom mieste.

3.3.9 Osoba s hodnotami prvá, druhá, tretia je kategória relevantná pre verbá (uvádza sa pri nich na piatom mieste).

V préterite a kondicionáli sloviess je gramatická morféma pre 3. os. nulová, čo nás vedie k nutnosti označovať osobu aj pri *l*-ovom prídastí, vždy podľa kontextu. Ak sa vo vete nachádza tvar *písal som/si/sme/ste*, značíme osobu aj pri *l*-ovom prídastí, aj pri gramatickej morféme; ak tvar *písal, písali*, osobu značíme ako prítomnú pri *l*-ovom prídastí. Pri kondicionáli (prítomnom i minulom) platí to isté.

Pri zložených tvaroch sloviess v budúcom čase sa osoba určuje len pri futúre slovesa *byť*, nie pri infinitíve.

3.3.10 Kongruencia v rode s hodnotami v mužskom rode živ. a neživ., v ženskom rode, v strednom rode, neurčená. Kategória kongruencie v rode je relevantná len pre verbá so slovesnou formou *l*-ového prídastia a uvádza sa na šiestom mieste.

Hodnotu neurčená kongruencia (**o**) zapíšeme, ak sa pri tvare nedá z kontextu jednoznačne určiť, s ktorým rodom je sloveso v kongruencii (*Ženy a deti neprišli.*).

3.3.11 Negáciu s hodnotami afirmácia alebo negácia označujeme iba pri slovesách, v tagu je vždy na poslednom mieste.

Príklady tagov slovesných tvarov:

<i>Študent nepriđe.</i>	l = neprísť	t = VKdsc-
<i>Študent neprišiel .</i>	l = neprísť	t = VLdscm-
<i>Vy ste neprišli.</i>	l = neprísť	t = VLdpbh-
<i>Ženy a deti neprišli.</i>	l = neprísť	t = VLdpc-
<i>Tuším vo vzduchu búrkku.</i>	l = tušiť	t = VKesa+
<i>Napísali by sme to rukou.</i>	l = napísať	t = VLdpam+
	l = by	t = Y
	l = byť	t = VKepa+

3.3.12 Okrem týchto kategórií rozoznávame v tagsete aj atribút, ktorý vyjadruje istú formálnu vlastnosť niektorých tokenov, ale nemá význam kategoriálnosti v lingvistickom zmysle – **zápis číslovky** s hodnotami: slovo, číslica alebo symbol, číselno-slovný tvar. Tento atribút označujeme iba pri číslovkách.

3.4 Všeobecné zásady značkovania

Značku uvádzame vtedy, keď je kategória pre daný tvar relevantná. Napr. pre zámeno *ako* je nerelevantná kategória paradigmy, rodu, čísla, pádu, osoby, určíme teda iba **slovnú triedu**; pri slovesách v neurčitku (*vniknúť*) nemožno určiť kongruenciu v rode, číslo a osobu, určíme teda iba **slovnú triedu** (sloveso), **slovesnú formu** (infinitív), **slovesný vid** (dokonavý), **afirmáciu** (afirmatív).

Značky sa priradujú hodnotám (morfologických) kategórií, ktoré sú pre danú slovnú formu relevantné aj v prípade, ak nie sú z formy slova „viditeľné“, teda nie sú dostatočne formovo transparentné. V niektorých prípadoch ich môžeme určiť z kontextu, ktorý je pre potreby ručného značkovania neobmedzený. Prihliadame najmä na kongruenciu v rámci syntagmy alebo na valenčnú väzbu. Špecifické je aj určovanie osoby pri *l*-ovom

príčasť na základe prítomnosti, resp. neprítomnosti gramatickej morfémy (*som, si, sme, ste*). Napr.:

Nechcem cestovať v tom kupé. – neutr., sg., L
Pre super dievča to nebol problém. – neutr., sg., A
Pozdravil sa jeho sestru. – femin., sg., D
Nevedel, pre koľko/koľkých ľudí to má pripraviť. – mask., pl., A
Dozvedel sa to od päť/piatich chlapov. – mask. živ., pl., G
Potešil sa päťoro/piatim nohavičiam. – femin., pl., D
Nenašiel som ani kúsok. – 1. os., sg., kongr. s muž. živ. rodom

Ide v podstate o absolútnu morfológickú homonymiu, keďže tieto slová majú len jednu formu, ktorou sa zapájajú do syntaktických vzťahov a väzieb. Tým sú však jasne definovateľné z kontextu. Niektoré tvary, uvedené v odbornej literatúre ako nesklonné, sa v bežnej praxi začínajú skloňovať. Vyskloňované a nevyskloňované tvary jednej lemy sa môžu v istom páde vyskytovať popri sebe a používateľ má možnosť nájsť všetky prípady a zistiť pokročilosť flektivizácie. Napr. pri substantíve *Philips* (lema *philips*) zachytíme takúto tvary G sg.: *Cieľom transakcie je transformácia Phillips..., hovorí O. Š. z Phillips Slovakia, ...s kapitálovou pomocou Phillipsu, ...veľkých spotrebičov od Phillipsa*; pri zámene *koľko* zachytíme A pl. v podobách: *pre koľko ľudí – pre koľkých ľudí*. Frekvenčná analýza potom ukáže pomer v ich výskyte.

Podobne postupujeme aj pri homonymii len niektorých tvarov. Napr. *pekné* – môže byť N, A pl. fem., N, A sg. neutra, N, A pl. neutra, N, A pl. mask. neživ. Tvarovú homonymiu zjednodzňujeme z kontextu: (*Dievčatá*) *sú pekné.* – N pl. neutra.

4 Značky a ich použitie v rámci jednotlivých slovných druhov

4.1 Substantíva

Pozícia 1: Substantíva – S

Pozícia 2: Paradigma

Značka	Hodnota	Príklad
S	substantívna	<i>chlap, hrdina, kuli, dub, stroj, žena, ulica, dlaň, kosť, mesto, srdce</i>
A	adjektívna	<i>hlavný, príbuzní, Borský, starká, hradská, vedúci, prepitné, vstupné</i>
F	zmiešaná	<i>pani, gazdiná</i>
U	neúplná	<i>kanoe, kupé</i>

Pozícia 3: Rod

Značka	Hodnota	Príklad
m	mužský životný	<i>chlap, hrdina, kuli, vedúci, hlavný, Borský</i>
i	muž. neživotný	<i>dub, stroj</i>
f	ženský	<i>žena, ulica, dlaň, kosť, pani, gazdiná, vedúca, Hlavatá, hradská</i>
n	stredný	<i>mesto, srdce, vysvedčenie, dievča, kanoe, kupé</i>

Pozícia 4: Číslo

Značka	Hodnota	Príklad
s	singulár	<i>chlap, pán, kuli, vedúci, hlavný, Borský, vedúca, Hlavatá, hradská</i>
p	plurál	<i>chlapí, páni, vedúci, hlavní, Borskí, vedúce, Hlavaté, hradské</i>
o	neurčené	nie je možné zjednotiť na základe kontextu: <i>Kúpili kanoe.</i>

Pozícia 5: Pád

Značka	Hodnota	Príklad
1	nominatív	<i>chlap, pán, kuli, vedúci, hlavný, Borský, vedúca, Hlavatá, hradská</i>
2	genitív	<i>chlapa, pána, kuliho, vedúceho, hlavného, Borského, vedúcej</i>
3	datív	<i>chlapovi, pánu, kulimu, vedúcemu, hlavnému, Borskému, vedúcej</i>
4	akuzatív	<i>chlapa, pána, kuliho, vedúceho, hlavného, Borského, vedúcu</i>
5	vokatív	<i>pane, mamo, mami, tati, Táni, Zuzi</i>
6	lokál	<i>o chlapovi, pánovi, kulim, vedúcom, hlavnom, Borskom, vedúcej</i>
7	inštrumentál	<i>s chlapom, pánom, kulim, vedúcim, hlavným, Borským, vedúcou</i>
o	neurčený	<i>Mesto!</i> (N alebo A sg.)

4.2 Adjektíva

Pozícia 1: Adjektíva – A

Pozícia 2: Paradigma

Značka	Hodnota	Príklad
A	adjektívna	<i>pekný, pekná, pekné, cudzí, cudzia, cudzie, pávi, otcov, matkin</i>
U	neúplná	<i>super, fajn, nanič</i>

Pozícia 3: Rod

Značka	Hodnota	Príklad
m	muž. životný	<i>pracovitý (chlap), besný (pes), besní (psi), otcov (priateľ)</i>
i	muž. neživot.	<i>staré (duby), besné (psy), nový (stroj), super (nápad)</i>
f	ženský	<i>emancipované (ženy), široké (ulice), otvorená (dlaň), biela (kosť)</i>
n	stredný	<i>veľké (mestá), dobré (vysvedčenie), fajn (dievča)</i>
o	neurčený	nie je možné zjednotiť na základe kontextu: <i>upravené (ženy a deti)</i>

Pozícia 4: Číslo

Značka	Hodnota	Príklad
s	singulár	<i>pracovitý (chlap), besný (pes), korpusový (manažér), super (nápad)</i>
p	plurál	<i>pracovití (chlapí), korpusoví (lingvisti), fajn (ľudia), matkini</i>

Značka	Hodnota	Príklad
		<i>(známi)</i>
o	neurčené	nie je možné zjednotzniť na základe kontextu: <i>(Kúpili) super (kanoe.)</i>

Pozícia 5: Pád

Značka	Hodnota	Príklad
1	nominatív	<i>pracovitý (chlap), besný (pes), korpusový (manažér), super (nápad)</i>
2	genitív	<i>pracovitého (chlapa), besného (psa), super (nápadu)</i>
3	datív	<i>pracovitému (chlapovi), besnému (psovi), super (nápadu)</i>
4	akuzatív	<i>pracovitého (chlapa), besného (psa), super (nápad)</i>
5	vokatív	<i>vážený (pane), milá (mamo, mami), drahý (tati, oci, Ďuri)</i>
6	lokál	<i>o pracovitom (chlapovi), besnom (psovi), super (nápade)</i>
7	inštrumentál	<i>s pracovitým (chlapom), besným (psom), super (nápadom)</i>
0	neurčený	<i>Pekné/Super (mesto)! (N alebo A sg.)</i>

Pozícia 6: Stupeň

Značka	Hodnota	Príklad
x	pozitív	<i>pracovitý (chlap), besný (pes), korpusový (manažér), super (nápad)</i>
y	komparatív	<i>pracovitejší (chlap), besnejší (pes)</i>
z	superlatív	<i>najpracovitejší (chlap)</i>

4.3 Pronominá

Pozícia 1: Pronominá – P

Pozícia 2: Paradigma

Značka	Hodnota	Inventár (podľa KSSJ a doplnené)
S	substantívna	<i>kol'kátko, vsetučko</i>
A	adjektívna	<i>oný, oná, oné, onaký, onakvý, inšie, iný, inakší</i> <i>aký, ký, kýže (bárs-, bár-, bohvie-, čertvie-, daj-, (dáký), hocij-, hoc-, kadej-, kdej-, ktovie-, neviem-, vol'aj-, (vol'áky), všakov-, (všakový), všelij-, nej- aký -kol'vek, -si)</i> <i>nijaký, nijakovský</i> <i>ktorý (bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, neviem-, nie-, poda-, ponie-, sotva-, málo-, vol'a-, zriedka-ktorý -kol'vek, -si)</i> <i>čí, čia, čie, nič (da-, nie-, vol'a- či -kol'vek, -si), taký, takýto, henta-ký</i> <i>kol'ký, kol'ki, kol'ké, kol'koraký, kol'konásobný, kol'káty</i> <i>tol'ký, tol'ki, tol'ké, tol'koraký, tol'konásobný, tol'kýto, toličizný, toli-kánsky</i> <i>každý, každučký, každučičký, žiadny (žiadna, žiadne, žiadni, žiadne), samý</i> <i>bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, leda-, neviem-, nie-, vol'a-, všeli- kol'ký/kol'koraký/kol'konásobný (-kol'vek)</i>
P	zámenná	<i>ja, ty, on, ona, ono, onô, my, vy, seba, sebe</i> <i>kto, nikto, nik (bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, kade-, kde-, ktovie-, málo-, neviem-, nie-, sotva-, vol'a-, všeli-, zriedka- kto -kol'vek, -si)</i> <i>čo, nič, pranič (bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, kade-, kde-, ktovie-, leda-, málo-, neviem-, nie-, sotva-, vol'a-, všeli- čo -kol'vek, -si)</i> <i>ten, tá, to (tento, táto, toto, tamten, tamtá, tamto, henten, hentá, hento, tenže, táže, tože)</i> <i>môj, -a, -e (tvoj, náš, váš, svoj)</i>
F	zmiešaná	<i>onen, všetok (všetka, všetko, všetci, všetky), sám, sama, samo, žiaden</i> <i>kol'ko, kol'kože, tol'ko</i>
U	neúplná	<i>jeho, jej, ich</i>
X	bez paradigmy	<i>kol'konásobne, kol'kokrát, kol'kýkrát, tol'konásobne, tol'kokrát,</i>

Značka	Hodnota	Inventár (podľa KSSJ a doplnené)
		<p><i>tol'koto, toľkýkrát</i> <i>bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, leda-, neviem-, nie-, voľa-, všeli- koľko (-koľvek), -koľkonásobne, -koľkokrát, -koľkýkrát</i> <i>ako, nijako, nijakovsky, tak, takto, onak, ináč, inak, inakšie, takisto, všakovak</i> <i>(bárs-, bár-, bohvie-, čertvie-, daj- (dáko), hocij-, hoc-, ktovie-, neviem-, nej-, voľáj- (voľáko), všelij- ako -koľvek, -si)</i> <i>kde, nikde, inde, všade, odvšade, zovšade (bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, kade-, ktovie-, málo-, neviem-, nie-, sotva-, voľa-, všeli-, zriedka- kde -koľvek, -si)</i> <i>kade, odkade, pokade, skade, skadeže, nikade, odnikade, znikade, inokade (bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, kde-, ktovie-, neviem-, nie-, odnie-, voľa-, všeli-, zda-, znie- kade -koľvek, -si), tade, tadeto, odtade, potade, stade, stadeto</i> <i>kam, nikam, inam, tam, tamto, tamže, sem, ta, tu, tuto, tuná, hen, hentam, (bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, neviem-, niekam -koľvek, -si)</i> <i>kadiaľ, odkadiaľ, skadiaľ, skadiaľže, znikadiaľ, odnikadiaľ, (bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, neviem-, nie-, odnie-, voľa-, všeli-, zda-, znie-, zvolá- kadiaľ -koľvek, -si)</i> <i>tadiaľ, tadiaľto, odtadiaľ, stadiaľ, stadiaľto, všadiaľ, odvšadiaľ, zovšadiaľ</i> <i>dokiaľ, dokiaľže, dokiaľkoľvek, odkiaľ, odniekiaľ, pokiaľ, odnikiaľ, odinakiaľ, (bohvie-, čertvie-, ktovie-, neviem- odkiaľ -koľvek, -si)</i> <i>dotiaľ, zatiaľ, odtiaľ, odtiaľto, potiaľ, potiaľto</i> <i>kedy, dokedy, nakedy, odkedy, inokedy (bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, málo-, neviem-, nie-, sotva-, voľa-, zriedka-, bár-, bár- kedy -si, -koľvek)</i> <i>vtedy, dovtedy, odvtedy, onehdy, nikdy, vždy, vždycky dosiaľ, posiaľ,</i> <i>nakoľko, natoľko, načo, prečo, prečože, zač, začo, nato, preto, zato (bohvie-, ktovie-, čertvie-, neviem- prečo)</i></p>

Pozícia 3: Rod

Značka	Hodnota	Príklad
m	muž. životný	<i>oni, on (brat), môj, ten, aký, taký, každý, žiadny, nejaký chlap</i>
i	muž. neživot.	<i>on, ony (duby), môj, ten, aký, taký, každý, žiadny, nejaký sveter</i>
f	ženský	<i>ona, ony, moja, tá, aká, taká, každá, žiadna, nejaká (žena)</i>
n	stredný	<i>ono, ony, moje, to, aké, také, každé, žiadne, nejaké (dieta)</i>
o	neurčený	<i>nie je možné zjednotiť na základe kontextu: naše (ženy a deti)</i>

Pozícia 4: Číslo

Značka	Hodnota	Príklad
s	singulár	<i>ja, môj, svoj, kto, čo, aký, ktorý, ten, taký, nikto, nič, nijaký, nejaký</i>
p	plurál	<i>my, moji, svoji, akí, ktorí, tí, tie, takí, nijakí, nejakí</i>
o	neurčené	nie je možné zjednotiť na základe kontextu: <i>Aké (kupé)?</i>

Pozícia 5: Pád

Značka	Hodnota	Príklad
1	nominatív	<i>ja, môj, svoj, kto, čo, aký, ktorý, ten, taký, nikto, nič, nijaký, nejaký</i>
2	genitív	<i>mňa, môjho, seba, svojho, koho, čoho, akého, ktorého, toho, takého</i>
3	datív	<i>mne, môjmu, sebe, svojmu, komu, čomu, akému, ktorému, tomu</i>
4	akuzatív	<i>mňa, oňho, môjho, seba, svojho, koho, čo, akého, ktorého, toho</i>
5	vokatív	<i>môj (otče, pane), naša (mami), náš (tati, oci, Ďuri)</i>
6	lokál	<i>o mne, mojom, sebe, svojom, kom, čom, akom, ktorom, tom, takom</i>
7	inštrumentál	<i>so mnou, mojím, sebou, svojou, kým, čím, akým, ktorým, tým, takým</i>
o	neurčený	<i>Môj (kabát)! (N alebo A sg.)</i>

Pozícia 6: Osoba

Značka	Hodnota	Príklad
a	prvá osoba	<i>ja, my, môj, -a, -e, náš, -a, -e</i>
b	druhá osoba	<i>ty, vy, tvoj, -a, -e, váš, -a, -e</i>
c	tretia osoba	<i>on, ona, ono, oni, ony, jeho, jej, ich</i>

4.4 Numerálie

Pozícia 1: Numerálie – N

Pozícia 2: Paradigma

Značka	Hodnota	Inventár (podľa KSSJ – doplnené a opravené)
S	substantívna	nula, sto, tisíc, milión, bilión, miliarda, stámilióny, státisíce, raz
A	adjektívna	<i>jediný (jedinký), mnohý (nemnohý, premnohý) prvý, druhý, tretí, štvrtý, piaty, šiesty, siedmy, ôsmy, deviaty, desiaty, jedenásty, dvanásty, trinásty, štrnásty, pätnásty, sedemnásty, osemnásty, devätnásty, dvadsiaty, tridsiaty, štyridsiaty, päťdesiaty, šesťdesiaty, sedemdesiaty, osemdesiaty, deväťdesiaty stý, dvojtý, trojtý, štvorstý, päťstý, sedemstý, tisící, dvojtisíci, stotisíci miliónty, miliardty, meruôsmy, nulty entý, ostatný (predostatný), posledný (neposledný, predposledný) dvojnásobný, trojnásobný, štvornásobný, päťnásobný, šesťnásobný, sedemnásobný, osemnásobný, deväťnásobný, desaťnásobný, dvanásťnásobný, dvadsaťnásobný, sedemdesiatnásobný, stonásob-</i>

Značka	Hodnota	Inventár (podľa KSSJ – doplnené a opravené)
		<i>ny, tisíc násobný, viacrázový, mnohonásobný, viacnásobný jednaký, dvojaký, trojaký, štvoraký, päťoraký, šesťoraký, sedmoraký, osmoraký, deväťoraký, desatoraký, dvanásťoraký, dvadsatoraký, storaký, tisícoraký, obojaký mnohoraký, viacerí, viaceré</i>
N	čísloková	<i>dva, dvaja, dve (obidva, oba, obidvaja, obaja, obidve, obe), tri/traja, štyri/štyria</i>
F	zmiešaná	<i>viacero, jeden, jedna, jedno (nejeden, -na, -no, podajeden, -na, -no), päť/piati (šesť/šiesti, sedem/siedmi, osem/ôsmi, deväť/deviati, desať/desiati, jedenásť/jedenásti, dvanásť/dvanásti, trinásť/trinásti, štrnásť/štrnásti, pätnásť/pätnásti, šesťnásť/šesťnásti, sedemnásť/sedemnásti, osemnásť/osemnásti, devätnásť/devätnásti, dvadsať/dvadsať, tridsať/tridsať, štyridsať/štyridsať, päťdesiat/päťdesiat, šesťdesiat/šesťdesiat, sedemdesiat/sedemdesiat, osemdesiat/osemdesiat, deväťdesiat/deväťdesiat, dvadsaťpäť/dvadsaťpiati, päťadvadsať)</i>
U	neúplná	<i>dvesto, tristo, štyristo, päťsto, šesťsto, sedemsto, osemsto, deväťsto dvetisíc, tritisíc, štyritisíc, päťtisíc, šesťtisíc, sedemtisíc, osemtisíc, desaťtisíc, štyridsaťtisíc, sedemdesiatitisíc, stotisíc, dvestotisíc, sedemstotisíc</i>
X	bez paradigmy	<i>byľ/u, byľ/ku, habadej, hodne, veľ'a (neveľ'a, priveľ'a), viac, viacej (najviac, najviacej), mnoho (premnoho, primnoho), moc, plno, nadostač, málo (-ičko, -ičičko, -inko, -lilinko, pramálo, nemálo), (menej), najmenej, pomenej, trocha/trochu (troška/trošku, -ička/u -ička/u, -inka/u, -ilinka/u), pár, zopár, dost', trištvrté, tri štvrté, pol prvý-, druhý-, tretí-, štvrtý-, piaty-, šiesty-, siedmy-, ôsmy-, jeden-, dva-, tri-, štyri-, päť-, šesť-, sedem-, osem-, deväť-, desať-, dvanásť-, dvadsať-, tridsať-, štyridsať-, sto-, dvesto-, tisíc-, milión- krát, raz/ráz dvoj-, troj-, štvor-, päť-, šesť-, sedem-, osem-, deväť-, desať-, dvanásť-, dvadsať-, sedemdesiat-, sto-, tisíc- násobne dvojito, dvojmo, trojito, trojmo, štvormo viac-, mnoho-, veľ'a-, nejeden-, pár-, zopár-, nesčíselne-, nastokrát, -raz/ráz, mnoho-, viac- násobne, neraz jedn-, dvoj-, troj-, štvor-, päťor-, šesťor-, sedmor-, osmor-, deväťor-, desator-, dvanásťor-, dvadsator-, stor-, tisícor-, (oboj-) ako, mnohorako dvoje, troje, štvoro, päťoro, šesťoro, sedmoro, osmoro, deväťoro, desatoro, dvadsatoro, tisícoro, obidvoje, oboje, viacero</i>

Pozícia 3: Rod

Značka	Hodnota	Príklad
m	muž. životný	<i>jeden (pán), jedni, dvaja, traja, štyria, piati (páni), päť (chlapov)</i>
i	muž. neživot.	<i>jeden (stôl), jedny (okuliare), dva, tri, štyri (duby), päť (dubov)</i>
f	ženský	<i>jedna (kniha), jedny (dvere), dve, tri, štyri (priateľky), päť (žien)</i>
n	stredný	<i>jedno (jablko), jedny (plúca), dve, tri, štyri (mestá), päť (miest)</i>
o	neurčený	nie je možné zjednotiť na základe kontextu: <i>tri ženy a deti</i>

Pozícia 4: Číslo

Značka	Hodnota	Príklad
s	singulár	<i>jeden (človek), druhá (osoba), trojaké (víno), štvornásobné (víťazstvo)</i>
p	plurál	<i>jedny (ženy), druhé (osoby), trojaké (vína), štvornásobné (víťazstvá)</i>
o	neurčené	nie je možné zjednotiť na základe kontextu: <i>druhé (kupé)</i>

Pozícia 5: Pád

Značka	Hodnota	Príklad
1	nominatív	<i>jeden (človek), druhá (osoba), trojaké (víno), štvornásobné (víťazstvo)</i>
2	genitív	<i>jedného (človeka), druhej (osoby), trojakého, štvornásobného (víťazstva)</i>
3	datív	<i>jednému (človeku), druhej (osobe), trojakému, štvornásobnému (víťazstvu)</i>
4	akuzatív	<i>jedného (človeka), druhú (osobu), trojaké (víno), štvornásobné (víťazstvo)</i>
5	vokatív	<i>jeden (človeče)</i>
6	lokál	<i>o jednom (človeku), druhej (osobe), trojakom (víne), štvornásobnom víťazstve</i>
7	inštrumentál	<i>s jedným (človekom), druhou (osobou), trojakým (vínom), štvornásobným víťazstvom</i>
o	neurčený	<i>dve ženy, dva stromy (N alebo A sg.)</i>

Pozícia 6: Zápis číslovky

Značka	Hodnota	Príklad
w	slovo	<i>päť, sto, milión</i>
q	číslica, symbol	<i>5, 100, 1 000 000</i>
r	číselno-slovný tvar	<i>5-krát, 100-násobný, 1 000 000-ty</i>

4.5 Verbá

Pozícia 1: Verbá – V

Pozícia 2: Slovesná forma

Značka	Hodnota	Príklad
I	infinitív	<i>písať, napísať, byť, anotovať</i>
K	prézent (indik.)	<i>píše, napíše, je, anotuje, píšeš, napíšeš, si, anotuješ</i>
M	imperatív	<i>píš! napíš! buď! anotuj!</i>
H	prechodník	<i>píšuc, napíšuc, anotujúc</i>
L	l-ové prídavné	<i>písal, napísala, bolo, anotovali</i>
B	futúrum byť	<i>budem, budeš, bude, budeme, budete, budú</i>

Pozícia 3: Vid

Značka	Hodnota	Príklad
d	dokonavé	<i>napišem, napísal som, napíšme, napísal by som, bol by som napísal</i>
e	nedokonavé	<i>píšem, písal som, budem písať, píšme, písal by som, bol by som písal</i>
j	obojvidové	<i>aplikovať</i>

Pozícia 4: Číslo

Značka	Hodnota	Príklad
s	singulár	<i>píšem, písal som, budem písať, píš!, písal by som, bol by som písal, napíšem</i>
p	plurál	<i>píšeme, písali sme, budeme písať, píšete!, písali by sme, boli by sme písali</i>

Pozícia 5: Osoba

Značka	Hodnota	Príklad
a	prvá osoba	<i>píšem, písal som, budem písať, píšme!, písal by som, bol by som písal</i>
b	druhá osoba	<i>píšeš, písal si, budeš písať, piš!, píšte!, písal by si, bol by si písal, napíšete</i>
c	tretia osoba	<i>píše, písal, bude písať, písal by, bol by písal, napíšu, napísali, napísal by</i>

Pozícia 6: Kongruencia v rode

Značka	Hodnota	Príklad
m	muž. životný	<i>(lingvista) anotoval, (lingvisti) anotovali</i>
i	muž. neživot.	<i>(stôl) sa rozpadol, (stoly) sa rozpadli</i>
f	ženský	<i>(búrka) prišla, (búrky) prišli</i>
n	stredný	<i>(zrkadlo) viselo, (zrkadlá) viseli</i>
o	neurčený	<i>(Ženy a deti) rástli.</i>

Pozícia 7: Negácia

Značka	Hodnota	Príklad
+	afirmácia	<i>písať, napísať, pracovať, kresliť, behať, vidieť, chcieť, prosiť, vedieť, nosiť</i>
-	negácia	<i>nepísať, nenapísať, nepracovať, nekresliť, nebehať, nevidieť, nechcieť, neprosiť, nevedieť, nenosiť</i>

4.6 Adverbiá

Pozícia 1: Adverbiá – D

Pozícia 2: Stupeň

Značka	Hodnota	Príklad
x	pozitív	<i>vhodne</i>
y	komparatív	<i>vhodnejšie</i>
z	superlatív	<i>najvhodnejšie</i>

4.7 Prepozície

Pozícia 1: E

Napr.: *na, do, v, po, pred, z, za, okolo, popod, poza, ponad, medzi, začiatkom, prostredníctvom, počínajúc*

4.8 Konjunkcie

Pozícia 1: O

Napr.: *a, aj, ako, ale, alebo, ani, avšak, ba, ba aj, ba dokonca, bud', bud' – alebo, či, i, iba, nielen – ale aj, no, však, aj keď, ako, až, kedy, keď, keďže, lebo, pretože, že*

4.9 Partikuly

Pozícia 1: **T**

Napr.: *a, muž, vari, či, teda, no, ale, tak, však, kiez, bodaj, asi, iba, ešte, sotva, samozrejme, áno, nie, vôbec, žiaľbohu*

4.10 Interjekcie

Pozícia 1: **J**

Napr.: *ach, fuj, ohó, nevidali, heš, mé, bé, krá, hav, čľup, cink, juj, hej, pch, tresk, beda, bože, dočerta, ahoj, nazdar, čau, vitaj*

4.11 Reflexívne *sa/si*

Pozícia 1: **R**

Pri značkovani sa nerozlišuje medzi zvratným zámenom *sa, si* a *sa, si* ako súčasťou slovesa. Všetky tvary *sa, si* majú priradenú osobitnú značku **R**. Neurčujú sa ani morfológické kategórie pri *sa/si* ako zvratnom zámene. Zjednodzňuje sa len homonymia *si*: zvratný element (*pospat' si, zaspievať si, zaspieval si si – R*) – 2. os. prítomného času, indikatív slovesa *byť* (*ty si klácal, zaspieval si si – VK*).

4.12 Kondicionálová morféma *by* (+ všetky tvary s elementom *by*)

Pozícia 1: **Y**

Morféma *by* a všetky tvary s elementom *by*, napr. spojky *keby, aby, žeby, akoby, sta-by, častice aby, keby*. Tvary *ak by, že by, ani čo by*, ktoré spoluvytvárajú kondicionál, určujeme jednotlivo. Teda *ak by* je spojením spojky *ak* (**O**) a kondicionálovej morfémy *by* (**Y**).

4.13 Abreviácie a značky

Pozícia 1: **W**

Patria sem značky ako *l, km, H₂O, X569847*, skratky typu *atd., tzv., t. j., pod., kt., i., XML, SND*, nie však skratkové slová typu *Satur, Slovnafť, Rempo*.

4.14 Interpunkcia

Pozícia 1: **Z**

Napr.: *..., ?, :, -, /, (,), +, ;*

4.15 Neurčiteľný slovný druh

Pozícia 1: **Q**

Napr.: *až60, až-100*

4.16 Neslovný element

Pozícia 1: **#**

Napr.: -----, XXXXXXXX, ***

Literatúra

- EAGLES Recommendations for the Morphosyntactic Annotation of Corpora*. Eagles Document– CWG–MAC/R 1996. <<http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html/>>
- Forróová, Martina, Horák, Alexander: *Morfologická anotácia korpusu*. In: Slovenčina na začiatku 21. storočia. Prešov, Fakulta humanitných a prírodných vied PU 2004, s. 174 – 183.
- Hajič, Jan: *Popis morfológických značiek – pozičný systém*. <<http://ucnk.ff.cuni.cz/manual/znacky.html>>
- Hajič, Jan – Krbec, Pavel – Květoň, Pavel – Oliva, Karel – Petkevič, Vladimír: *Serial Combination of Rules and Statistics: A Case Study in Czech Tagging*. Prague Dependency Treebank. CD ROM. V. 1.0. Praha: Ústav formální a aplikované lingvistiky MFF UK 2001.
- Hana, Jiří – Hanová, Hana: *Manual for morphological annotation*. ÚFAL/CKL Technical report TR-2002-14. Praha: Ústav formální a aplikované lingvistiky MFF UK 2002.
- Krátký slovník slovenského jazyka*. 4. vyd. Red.: M. Považaj. Bratislava: VEDA 2003.
- Leech, Geoffrey: *Anotační systémy pro značkování korpusu*. In: Studie z korpusové lingvistiky. Acta Universitatis Carolinae – Philologica 3 – 4. Praha: Karolinum 2000, s. 185 – 197.
- Morfológia slovenského jazyka*. Red.: J. Ružička. Bratislava: Veda 1966.
- Oravec, Ján – Bajzík, Eugénia – Furdík, Juraj: *Súčasný slovenský spisovný jazyk. Morfológia*. Bratislava: Slovenské pedagogické nakladateľstvo 1984.
- Pravidlá slovenského pravopisu*. Bratislava: Veda 2000.
- Slovenčina a čeština v počítačovom spracovaní*. Zborník referátov zo seminára (Bratislava 26. – 27. októbra 2001). Red.: A. Jarošová. Bratislava: Veda 2001.

