

Tokenizácia, lematizácia a morfológická anotácia Slovenského národného korpusu

Aktuálna verzia zo 4. 5. 2004

1. Tokenizácia

1.1. Definícia tokenu

Token je arbitrárna jednotka textu, ktorá rozširuje lingvistický význam pojmu slovo. Za token sa v automatickej segmentácii textu považuje akýkoľvek reťazec znakov medzi dvoma medzerami (whitespace), aj jednotlivé znaky interpunkcie, ktoré nemusia byť oddelené medzerou od predchádzajúceho alebo nasledujúceho tokenu (pozri aj 1.2.). Text sa teda z formálneho hľadiska skladá z tokenov a medzier (whitespace).

1.2. Tokenizácia v SNK

V tokenizácii textov SNK princíp identifikácie tokenov ako reťazcov medzi dvoma medzerami neplatí absolútne. Tokeny by ideálne mohli obsahovať vo svojom vnútri whitespace (napr.: *1 234 567 890*), ale v prvej fáze budeme za token považovať všetko medzi whitespace, na ďalších úrovniach spracovania sa dá uvažovať o spojení takýchto tokenov do jednej logickej, resp. lingvistickej jednotky. Na druhej strane, sú prípady, v ktorých znakový reťazec ohraničený medzerami môže byť a často aj je rozložený na dva a viaceré tokeny. Napr.: *Neviem.* (2 tokeny); *Čo?* (2 tokeny);

Pri segmentácii textu rozlišujeme nasledujúce kategórie znakov:

1. **Alfanumerické** – sem patria písmená, číslice a podobné znaky. Z technického hľadiska sme si situáciu uľahčili tým, že do tejto kategórie patria všetky znaky s unicode kategóriou L (písmená), M (špeciálne značky, napr. kombinované diakritické značky) alebo N (čísllice). Tento prístup je v poriadku, ak uvažujeme o bežne rozšírených jazykoch. Pri tokenizácii zlučujeme alfanumerické znaky medzi dvoma medzerami do jedného celku (tokenu).
2. **Interpunkcia** (dvojbodka, bodka, otáznik, výkričník na konci vety, úvodzovky, hviezdičky, matematické symboly a iné) – Každý znak tvorí samostatný token. Napr. vo vete „*Win98 mi nefunguje!!!*“ bude 8 tokenov (úvodzovky dolné, *Win98*, *mi*, *nefunguje*, výkričník, výkričník, výkričník, úvodzovky horné).

Segmentácia je dôležitou etapou v automatickom spracovaní textu, pretože od jej výsledkov je priamo závislá morfológická analýza a dezambiguácia. Navrhnuté zásady tokenizácie môžu vyvolávať otázky pri zložených tvaroch *v rámci, na bielo, a teda*, pri zápisoch so spojovníkom či pomlčkou (často aj chybné použitými) *kde-kto, čím – tým, 8 – krát, Košice-Bratislava*, pri analytickom stupňovaní typu *menej lukratívny*, združených pomenovaniach *Spišská Nová Ves*, číslovkách ako *1 984* a na druhej strane pri aglutinovaných podobách ako *oňho, akoby* a i. Pri týchto slovách, rozdelených na viac tokenov (napriek tomu, že tvoria jednu jazykovú jednotku) alebo zlúčených do jedného tokenu (napriek tomu, že ide o dve pôvodné jednotky), by tým v morfológickom (a určite aj v syntaktickom) spracovaní mohlo dôjsť k ich dezinterpretácii. Súčasná riešenie však nevyklučuje možnosť zapojenia logického modulu do spracovania textu, ktorý by sa neskôr uplatnil ako vhodnejší základ morfológickej lematizácie a značkovania.

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

2. Lematizácia

2.1. Definícia lemy

Lema sa zvyčajne definuje ako “slovníkový” tvar tokenu. Súbor jazykových vlastností, ktoré by mala lema istého slovného tvaru zachytávať, tvoria: tzv. základné hodnoty morfológických kategórií a dištinkcia písania veľkého a malého písmena. Pri lematizácii textov SNK sa podobne ako pri tokenizácii pojem lemy neaplikuje absolútne len v jeho významovom rozsahu.

2.2. Zásady lematizácie v SNK

2.2.1. Lema sa uvádza vždy s malým začiatočným písmenom. Napr. *Alexander* = *alexander*. Je možné namietat', že sa zrušením dištinkcie veľkého a malého písmena strácajú isté sémantické príznaky slov, na druhej strane je však rozsah výsledkov pri vyhľadávaní podľa lemy s malým písmenom podstatne väčší. Navyše, je predpoklad, že sa výskytu slov písaných primárne s veľkými písmenami dajú jednoducho vyhľadať pomocou presne zadaných vyhľadávacích otázok v korpusovom manažéri alebo odfiltrovaním zo zoznamu konkordancií s veľkými a malými písmenami. Informácia o propriálnosti lemy sa uvádza na úrovni tagu (pozri 3.1.).

2.2.2. Pri tokenoch patriacich do ohybných slovných tried má lema v zásade tieto hodnoty morfológických kategórií:

Substantíva Zámená Číslovky	Adjektíva Adjektívaliá	Slovesá
príslušný rod	maskulínium	infinitív
singulár (ak existuje)	singulár	
nominatív (ak existuje)	nominatív	
	pozitív	

2.2.3.1. Substantíva:

a) Substantíva sa lematizujú tvarom N sg. príslušného rodu (*domom* = *dom*). Pri tvaroch *nevojak*, *nemocný*, *nevedomosť*, ako aj pri slovách z iných slovných tried (*nejeden*, *nemálo*, *neveľa*, *neraz*) ide o samostatné lemy, nie o negované tvary lem *vojak*, **mocný*, *vedomosť* (*jeden*, *málo*, *veľa*, *raz*).

b) Pri pluráliách tantum môže byť lema v sg. v relevantných prípadoch, inak sa lematizujú tvarom N pl. príslušného rodu. Napr.: *nohavice* = *nohavica*, ale *Alpy* = *alpy* (nie *alpa*).

c) Substantíva s adjektívnou paradigmou (*vedúci*, *hlavná*) si zachovávajú rod, v akom sa nachádzajú v texte. Napr.: (z) *vreckového* = *vreckové*, nie *vreckový*; (*od*) *vedúcej* = *vedúca*, nie *vedúci*.

2.2.3.2. Adjektíva:

a) Adjektíva a adjektívaliá (**adjektívne zámená, číslovky a prídavné mená**) sa lematizujú tvarom N sg. mask. pozitívu. Napr.: *peknej* = *pekný*; aj *ťarchavá* = *ťarchavý*; *ktoré* = *ktorý*; *každou* = *každý*, *prvého* = *prvý*.

b) Privlastňovacie adjektíva sa lematizujú samostatne adjektívnym tvarom N pl. mask., nie ako substantíva (Napr. *s otcovou sestrou* = *s, otcov, sestra*). Pritom pri tvaroch typu *otcovi*, *otcov* – môže ísť o substantívum s lemov *otec* (*otcovi* – D, L sg., *otcov* – G, A pl.) alebo o privlastňovacie prídavné meno s lemov *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*).

c) Prvé časti zložených adjektív, ktoré sa zvyčajne píšú so spojovníkom (alebo nesprávne s pomlčkou) sú samostatné tokeny. Lematizujeme ich tvarom, ktorý sa nachádza v texte (*česko – slovenský*, 1 = *česko*; *bielo – červený*, 1 = *bielo*).

2.2.3.3. Zámená:

a) Tvary zámen *ja*, *ty*, *my*, *vy*, *seba*, *kto*, *čo*, *nikto*, *nič* a pod. sa lematizujú vždy samostatne,

príslušným tvarom N, teda *my, nás, nami* nemá lemu *ja*, ale *my*.

b) Plurálové tvary zámen *oni, ony, tí, tie* sa lematizujú N sg. podľa rodu ako *on, ona, ono, ten, tá, to*, ak sa rod dá z kontextu určiť.

c) Tvary *jeho, jej, ich* lematizujeme N sg. príslušného rodu ako osobné základné zámeno (*jeho* – G, A sg. s lemovou *on*; *jej* – G, D sg. s lemovou *ona*; *ich* – G, A pl. s lemovou *on, ona, ono*) alebo lemovou *jeho, jej, ich*, ak ide o osobné prívlastňovacie zámeno. Rozlíšime *ich* na základe postavenia a kontextu. Napr.: *Jeho to mrzí.* (lema – *on*), *Dám to jeho priateľom.* (lema – *jeho*).

d) Prívlastňovacie zámená ako *môj, moja, moje, moji, tvoj, tvoja, tvoje, tvoji, náš, naša, naše, naši...* sa lematizujú N sg. mask. ako *môj, tvoj, náš...* Takisto pre vymedzovacie zámeno *sám, sama, samo* platí len lema *sám*. Pritom tvary *samého, samému, samého, o samom, so samým, samých, samým, samých, o samých, so samými* sú homonymné nepriame pády zámen *sám* a *samý*. Aj tu pri určovaní lemy vychádzame z kontextu. Napr.: *Mrzí to jeho samého.* (lema – *sám*), *Ušili šaty zo samého zlata.* (lema – *samý*).

e) Tvary zámena *koľko* (*koľkí, koľké...*) a *tol'ko* (*tol'kí, tol'ké*), ktoré odkazujú na základnú číslovku, sa lematizujú N sg. spoločným pre všetky rody (*koľko, tol'ko*). Tvary *koľkí, tol'kí, koľkých, tol'kých* môžu byť popri tom aj zámená odkazujúce na radové číslovky, lema bude N sg. mask. (*koľký, tol'ký*). Napr.: *Koľkí žiaci prišli?* (lema – *koľko*), *Koľkí v poradí boli naši pretekári na súťaži?* (lema – *koľký*).

f) Tvary zámen *onen* (*oná, ono*), *žaden* (*žiadna, žiadne*), *všetok* (*všetka, všetko*) sa lematizujú N sg. mask. (*onen, žaden, všetok*).

g) Zrastené tvary (*oňho, preňho, naňho, oň, preň, zaň*) považujeme za tvary zámen *on, ono*. Lema sa teda skladá z osamostatnenej predložky a N sg. príslušného rodu (*oňho* = *o_on*, *preňho* = *pre_on*, *naňho* = *na_on*). Tieto tvary považujeme za aglutinované a na úrovni tagu túto informáciu značíme (pozri 3.3.13).

h) Zložené tvary zámen (podľa KSSJ: *ten istý, tá istá, to isté, taký istý, tak isto, koľko rás, -koľko rás, koľký rás, -koľký rás, tol'ko rás, tol'ký rás, tamto ten, tamto tá, tamto to*) aj tvary, ktoré sa v korpusových textoch môžu vyskytnúť so spojovníkom *i* (nesprávnou) pomlčkou (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*; číslovky typu *3-krát* a pod.), označíme ako dva (resp. tri) tokeny a každému tokenu priradíme lemu, aj keď rešpektujeme, že ide o jednu lexému.

i) Zvratné zámeno *si* lematizujeme tvarom *si* (*dávať si darčeky*) a odlišujeme od tvaru slovesa *byť* (*si* – 2. os. sg. prít. času).

2.2.3.4. Číslovky:

a) Tvary základných čísloviek *jeden* sa lematizujú N sg. mask. Napr.: *jednej (žene) = jeden*.

b) Tvary základných čísloviek *dva, tri, štyri, päť...* sa lematizujú N pl. mask. než. Napr.: *dve (deti) = dva, traja (draci) = tri*.

c) Tvary čísloviek, ktoré sa skloňujú ako substantíva (*sto, tisíc, milión, miliarda...*) sa lematizujú N sg. príslušného rodu.

d) Číslovky vyjadrené číslom alebo symbolom sa lematizujú ako číslo alebo symbol. Rímske číslice majú lemu v tvare rímskej číslice. Číselné výrazy vyjadrené slovom sa lematizujú slovom.

e) Číslovky *veľa, viac, najviac; málo, menej, najmenej* sú samostatnými lemmami, keďže nejde o vyjadrenie kategórie stupňa, ale o lexikalizované tvary.

f) Zloženým číslovkám *dva razy, päť rás* priradíme dve lemy. Druhý token má vždy lemu *raz*. V prípade nesprávneho textového zápisu násobnej číslovky s nesamostatným *-krát* (napr. *dva krát*) uvádzame aj samostatnú lemu *krát*.

2.2.3.5. Slovesá:

a) Lemou je infinitív, pri ktorého určovaní prihliadame aj na vid. Napr. *stretnem = stretnúť, stretávam = stretávať*. Aj gramatické morfémy ako *som, si, sme, ste, budem* a *i*. lematizujeme infinitívom formálneho *byť*.

b) Negované tvary sa lematizujú negovaným tvarom infinitívu. Napr.: popri leme *vidieť, mať, chcieť* určujeme aj lemu *nevidieť, nemať, nechcieť*.

Zvláštnu pozornosť si vyžaduje lematizácia negovaného tvaru slovesa *byť*. Zatiaľ čo pri tvaroch minulého i budúceho času je negácia vyjadrená synteticky (*nebol, nebude*), v prítomnom

čase sa negácia vyjadruje pomocou častice *nie* (*nie je*). Prítomnosť tejto častice ovplyvňuje lematizáciu slovesa (*nie je*; *l = nie, nebyť*) a zároveň naznačuje negáciu (pozri aj 3.3.11.).

c) Pri prídavných, ktoré vyčleňujeme ako samostatnú slovnú triedu (pozri 3.2.), sa ako lema uvádza tvar N sg. mask. Napr.: *vyhodených = vyhodený, čítajúcich = čítajúci*.

d) Pri slovesných podstatných menách, ktoré zaraďujeme k substantívam, sa ako lema uvádza N sg. neutr. Napr.: *čítaním = čítanie*.

2.2.4. Lemou slov z neohybných slovných tried je slovníkový tvar, teda v zásade taký tvar, v akom sa nachádzajú v texte, napr. častice ako *tuším, hádam, prosím*, predložky *končiac, začiatkom, koncom* a i. neohybné slová sa lematizujú svojím zmeraveným tvarom (*tuším, hádam, prosím, končiac, začiatkom, koncom* a i.). Výnimkou sú prípady:

a) vystupňované **príslovky** sa lematizujú pozitívom. Napr.: *lepšie = dobre*.

b) pri vokalizovaných **predložkách** je lema nevokalizovaná podoba. Napr.: *vo (vreci) = v*.

2.2.5. Skratky, nejednoznačné tokeny, chybné alebo variantné zápisy ako *XDR, XEU, XML, ko(z)mický, (Vysoké) Ta3; až100, pavieho, veľký, Wintston* (správne *Winston*) sa lematizujú podľa tvaru, aký sa vyskytol v texte, teda: *xdr, xeu, xml, ko(z)mický, ta3; až100, pavi, veľký, wintston*, aj napriek tomu, že pri niektorých ide o zjavné chyby. Pri chybných tvaroch ako *Goldsteinová (tvár), prosím ta* lematizujeme podľa pravidiel (*Goldsteinov, ty*).

3. Morfológické značkovanie

3.1. Forma zápisu atribútov a ich hodnôt v tagsete

Hodnoty jednotlivých kategórií sa zapisujú jednomiestnym znakom – písmenom latinskej abecedy, číslicou, matematickým symbolom. Súbor týchto znakov tvorí jeden **tag (značku)** k jednému tokenu a leme. Tag je teda súborom znakov vyjadrujúcim hodnoty formálnych kategórií, ktoré považujeme pri danom tvare za relevantné. Počet znakov je variabilný, ale ich poradie v tagu je záväzné.

Každý tag sa skladá z dvoch častí. Prvá časť určuje morfológické a gramatické kategórie tokenu, druhá (nepovinná) časť zaraďuje token do určitých špeciálnych skupín (vlastné mená, defektné zápisy). Vo väčšine prípadov token nepatrí ani do jednej zo špeciálnych skupín, druhá časť tagu vtedy úplne chýba. V prípade vlastného mena uvedieme za prvou časťou tagu znak : (dvojbodka – U+003A COLON) a špeciálny znak **r**; v prípade chybného, resp. nesprávneho zápisu slova bude za dvojbodkou nasledovať znak **q**.

Napr.: *od Minárika* *l = Minárik* *t = SSms2:r*
 velmi *l = velmi* *t = Dx:q*
 Goldsteinovú tvár *l = Goldsteinov* *t = AFfs4x:q*

Pri určovaní triedy nesprávnych zápisov sa riadime aj frekvenciou výskytu v korpuse, resp. typom "chyby". Ak tvar slova nie je síce spisovný, ale používaný v štylisticky relevantných prípadoch (*neni, do Košicoch, za prvé, postavím sa do rady, prádlo*), neoznačujeme ho ako chybný. Za defektný zápis považujeme preklepy a zjavné pravopisné chyby.

Prvá časť tagu sa začína vždy znakom pre slovnú triedu a nasledujú znaky pre ostatné kategórie. Pri menných slovných druhoch nasleduje kategória paradigmy (bližšie pozri 3.3.2.) a za ňou zvyčajne základné morfológické charakteristiky (rod, číslo, pád – pozri 3.3.3. – 3.3.5., pri adjektívach aj stupeň – 3.3.6.). Pri verbách je na druhom mieste slovesná forma (pozri 3.3.7.), ďalej sú vyjadrené hodnoty vidu, čísla, osoby, kongruencie v rode (ak sú relevantné) a negácia (pozri 3.3.8. – 3.3.11.). Pri pronomínach, numeráliách, participiách, prepozíciách a tvaroch *keby, aby* značíme ďalšie atribúty, zväčša formálne vlastnosti ako aglutinovanosť, samostatné použitie číslovky, druh participia, forma predložky, kondicionálnosť (pozri 3.3.12. – 3.3.16.).

Napr.: *Lingvista anotoval texty z korpusu*. Vo vete je 6 tokenov (*lingvista, anotoval, texty, z, korpusu, .* (bodka)), každému priradíme lemu (l) a tag (t):

<i>Lingvista</i>	l = <i>lingvista</i>	t = SSms1
<i>anotoval</i>	l = <i>anotovať</i>	t = VLescm+
<i>texty</i>	l = <i>text</i>	t = SSip4
<i>z</i>	l = <i>z</i>	t = Eu2
<i>korpusu</i>	l = <i>korpus</i>	t = SSis2
.	l = .	t = Z

3.2. Slovné triedy

V morfológickom tagsete SNK sa množina slovných foriem používaných v slovenčine rozdeľuje do 19 tried, z ktorých 10 v zásade zodpovedá tradične vydeľovaným slovným druhom a 9 obsahuje rôzne špecifické jazykové prvky:

1. **S** **Substantíva**
2. **A** **Adjektíva**
3. **P** **Pronomína**
4. **N** **Numerálie**
5. **V** **Verbá**
6. **D** **Adverbiá**
7. **G** **Formálne participiá**
8. **E** **Prepozície**
9. **O** **Konjunkcie**
10. **T** **Partikuly**
11. **J** **Interjekcie**
12. **R** **Reflexívne sa/si**
13. **Y** **Kondicionálová morféma by**
14. **Z** **Interpunkcia**
15. **W** **Abreviácie a značky**
16. **Q** **Neurčiteľný slovný druh**
17. **#** **Neslovný element**
18. **%** **Citátový výraz**
19. **0** **Číslice**

Pri klasických slovných druhoch sa v podstate rešpektuje slovnodruhové zaradenie podľa súčasných slovenských kodifikačných príručiek (KSSJ, 2003; PSP, 2000). Hraničné prípady predstavujú:

1. Substantíva (apelatíva aj propriá) s adjektívnou paradigmou, ktoré zaraďujeme k substantívam (**S**) – *Chladný, Mináriková, cestovné, vrchný, vedúci, vedúca*.
2. Slovesné podstatné mená, ktoré zaraďujeme k substantívam (**S**) – *písanie, hovorenie, rešpektovanie*.
3. Činné a trpné prídavné, ale aj adjektíva, ktoré vznikli slovnodruhovým prechodom od sloves. Vydeľujeme ich zvlášť ako triedu formálnych prídavných (**G**) – *písaný, otvorený, obutý, píšuci, hrajúci, stojaci*. Od adjektív ich odlišujeme na základe posúdenia ich formy i pôvodu. Za (formálne) trpné prídavné považujeme tie adjektíválie, ktoré vznikli z infinitívneho kmeňa slovesa pridaním participiálnej morfémy *-n+y*, *-t+y* (*sklad-a:t* → *sklad-a:n-y*), za (formálne) činné prídavné tie, ktoré vznikli z prítomného kmeňa (resp. od jeho variantu v 3. os. pl.) pridaním participiálnej morfémy *-úc+i/-uc+i*, *-iac+i/-ac+i* (*sklad-a:j-ú* → *sklad-a:j-úc-i*). Napr.: *píšuci, písaný, žnúci, žatý, bijúci, bitý, spiaci, šijúci, šitý, sejúci, siaty* (**G**). Medzi participiá nezaraďujeme adjektívne deverbatíva (adjektíva), ktoré vznikli deriváciou z neurčitkového kmeňa, napr. *písaci, skladaci, žaci, bici, spaci, šijaci, sejací* (**A**).
4. Na tejto úrovni anotácie nerozlišujeme medzi *sa, si* ako zvratným zámenom a *sa, si* ako zvratnou súčasťou slovesa. Všetky tvary *sa, si* majú priradenú osobitnú značku **R**. Ich rozlíšenie ponechávame na osobitné štúdie vypracované na základe korpusu, prípadne na syntaktickú anotáciu. V tomto prípade zaniká možnosť určovania morfológických kategórií pri *sa/si* ako zvratnom zámenne. Zjednotňuje sa len *si* ako zvratný element (*pospať si*,

zaspievať **si**, zaspieval si **si** – **R**) a ako súčasť tvaru 2. os. prítomného času indikatívu a prítomného a minulého času kondicionálu slovesa *byť* (*ty si kľáčal, zaspieval si si, kľáčal by si, bol by si si zaspieval* – **VK**).

5. 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.
6. Morfému *by*, ktorá je súčasťou kondicionálového slovesného tvaru, značíme samostatne znakom **Y**. Ďalšie tvary s *by* (napr. spojky *keby, aby, žeby, akoby, sťaby*, častice *aby, keby*) značíme podľa funkcie ako spojky (**O**) alebo častice (**T**), ale prihliadame aj na to, že v týchto tvaroch je aglutinovaná aj morféma *by*. Zaznamenávame to pridaním znaku pre kondicionál (**OY, TY**). Napr.: **Keby sme to vedeli!** l = *keby* t = **TY**
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**).
7. Zrastené tvary (*oňho, preňho, naňho, oň, preň, zaň*) označujeme ako zámená (**P**), predložková časť sa zachytáva ako súčasť lemy (pozri 2.2.3.3), ich aglutinovanosť je vyjadrená pomocou formálnej kategórie a jej hodnoty aglutinované (**g**).
Napr.: **Starám sa oňho.** l = *o_on* t = **PPms4g**
8. Medzi abreviácie a značky (**W**) patria značky ako *l, km, H2O, X569847*; skratky typu *atd., tzv., t. j., pod., kt., i., XML, SND*; nie však skratkové slová typu *Satur, Slovnaft, Rempo*, ktoré zaraďujeme podľa funkcie i významu k substantívam.
9. Viacslovné lexikálne jednotky sa v SNK skladajú z niekoľkých tokenov. Prvé, resp. nesamostatné časti (slová) väčšieho lexikálneho celku často nie je možné určiť. Priradíme im znak **Q**, neurčiteľný slovný druh: *po (Q) slovensky (Dx), fast (Q) food (S), Los (Q) Angeles (S)*; *slovensko-český, 2-krát* sa zobrazí ako tri tokeny, ktoré je potrebné značiť samostatne: *slovensko (Q), - (Z), český (AA)*; *2 (0), - (Z), krát (Q)*. Podobne postupujeme aj pri zlúčených slovách typu *až60*, ktoré tiež označíme ako neurčiteľný slovný druh (**Q**).
10. Druhotné predložky typu *s ohľadom na, v závislosti od, na rozdiel od, v prípade* značujeme ako spojenie predložiek a substantíva. Napr.: *s (E), ohľadom (l = ohľad, S), na (E)*;
11. K citátovým výrazom (**%**) nezaraďujeme jednotlivé slová cudzieho pôvodu (*kuskus, ska, sitar, djembe, česnečka, květák* – **S**), ktoré sa v texte vyskytnú ako súčasť slovenskej vety a je prispôbené slovenskej gramatike, ale skôr celé reťazce tokenov – cudzie viacslovné výrazy, frázy a vety, ktoré nie sú v procese zdomáčňovania, ale fungujú ako rečové jednotky pôvodného jazyka, resp. sú ako citátové použité v danom texte: *Take it easy!*; *Šaj pes dovakeras*; „*správne vychlazená dvanáctka*“; *Ta naše povaha česká*; *náměstí (%)*. Tieto tokeny nemusia byť vyčlenené z textu úvodzovkami.
12. Čísllice (rímske i arabské), resp. číselné znaky a kombinácie číselných znakov vyčleňujeme ako samostatnú triedu (**0**).

Problematika analytických tvarov, ktoré môžu spôsobovať problémy najmä vzhľadom na nevyhnutnosť uchopiť a pretransformovať klasický systém gramatických tvarov pre potreby počítačového spracovania, sa rieši v kapitole 3.5., po podrobnom predstavení systému kategórií a ich hodnôt.

3.3. Kategórie a ich hodnoty

Pri morfológickom značovaní sme vychádzali z teórie gramatických kategórií, ako ich podáva akademická Morfológia slovenského jazyka (1966), príp. iné morfológické práce. V tagsete SNK sa však stretáme aj s kategóriami a ich hodnotami, o ktorých sa v tradičnej morfológii ako o kategóriách neuvažovalo (paradigma, slovesná forma, aglutinovanosť, kondicionálnosť). 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 znakom **o** (*Mužov, žien a detí je päť. Kúpili kanoe. Mesto!*).

3.3.1. Pri ohybných slovných druhoch sa tag začína znakom pre slovný druh, ako sa vymedzuje podľa klasických komplexných kritérií a uvádza aj v KSSJ. Napr.: *mama, vedúci, kanoe, pani* – **S**, *pekný, otcov, super* – **A**, *taký, on, žiaden, tam* – **P**, *sto, štvrtý, tri, jeden, päť* – **N**, *písať, píšem, piš, píšuc, písal, budem* – **V**. Ako samostatná slovná trieda boli vyčlenené formálne participiá (*prestretý, zívajúci* – **G**) pre ich dynamiku a často nejasné postavenie na pomedzí slovies a prídavných mien.

3.3.2. Pri menných slovných druhoch sa na druhom mieste uvádza značka pre typ **paradigmy** s hodnotami substantívna, adjektívna, zámenná (pronominálna), číslovková (numerálna), zmiešaná, neúplná a príslovková (adverbiálna).

Formálny atribút paradigma chápeme ako vymedzenie špecifickej formy konkrétneho člena slovného druhu (napr. *taký* je zámeno, no má formu adjektíva – **PA**). Pre typy paradigmy substantívna, adjektívna, zámenná, číslovková a príslovková sa ponechávajú rovnaké značky ako pre slovný druh (**S, A, P, N, D**).

Zmiešaná paradigma (F) platí pre slová, ktoré majú len sčasti kongruentnú paradigmu. Vývoj ich deklinácie prešiel zložitými procesmi a nie sú jednoznačne zaraditeľné k čistým deklináčnym typom. Patria sem slová typu: *kuli* + všetky rovnako sa skloňujúce subst., *gazdiná* + všetky rovnako sa skloňujúce subst. (vzor: *pekná* okrem G, D, L, I pl.), *otcov, matkin* (všetky individuálne privlastňovacie príd. mená), *on, ona, ono, kto, čo, nikto, nič...*, *môj (tvoj, náš, váš), ten (tá, to), sám, onen, žiaden, všetok, jeden*.

Príklad	prevládajúca paradigma	odlišné pády
<i>sám, onen, žiaden, jeden, všetok</i> (mask. živ.)	pekný (sg.)	N sg.
<i>sám, onen, žiaden, jeden, všetok</i> (mask. neživ.)	pekný (sg.)	N, A sg.
<i>sama, samo, jedna, jedno, všetka, všetko</i> (fem., neutr.)	pekná, pekné (sg.)	N, A sg.
<i>všetci, sami, jedni</i> (mask. životné)	pekni (pl.)	N pl.
<i>všetky, samy, jedny</i> (mask. neživ., fem., neutr.)	pekné (pl.)	N, A pl.

Neúplná paradigma (U) bude priradená tým substantívam, adjektívam, zámenám a číslovkám, ktoré sa v tradičnej gramatike označujú ako nesklonné (*kupé, super, jeho*), s tendenciou k sklonnosti (*kanoe* – G pl. *od kanoi*), alebo ako obyčajne nesklonné (*pani, kolko, tolko, viacero, päť, sto, tisíc...*). V týchto prípadoch je istá distribúcia sklonných a nesklonných podôb závislá od mnohých činiteľov (*sto, tisíc* v zložených číslovkách). Za substantívum s neúplnou paradigmou môže byť považovaný aj tvar takeého substantíva cudzieho pôvodu, ktoré je v zásade sklonné (*Phillips, Tesco*), no pisateľ uprednostnil v danom páde nesklonnú podobu (*od Phillips* – **SU**), hoci sa vyskytuje zväčša sklonná podoba (*od Phillipsu* – **SS**).

Slová typu *kolko, tolko, viacero* sa síce skloňujú v pluráli podľa vzoru *pekný (kolki, kolké, tolki, tolké, viacerí, viaceré)*, ale často sa používa ich nesklonná podoba. Podobne pri číslovkách, ktoré sa skloňujú ako *päť*. Zároveň pri týchto číslovkách sú v N a A pl. pre mask. neživ., femin. a neutrá možné iba nesklonné tvary. Napr.: *Vidím päť žien*. (Riešenie značkovania v týchto prípadoch pozri 3.3.3.)

Adverbiálnu – príslovkovú paradigmu (D) majú také zámená a číslovky, ktoré sú nesklonné a v KSSJ bývajú označené kvalifikátorom *neskl.* alebo *príslov.*, resp. majú adverbiálnu funkciu (*kolkonásobne, kolkorako, tam, tu, vtedy, vždy, viacnásobne, dvojako*). V týchto prípadoch sa tag určením paradigmy končí. Ďalšie kategórie ako rod, číslo alebo pád sa neuvádzajú.

Prvé dve značky v tagu čítame takto:

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

- **AA** – adjektívum s adjektívnou paradigmou (*pekný*)
- **AF** – adjektívum so zmiešanou paradigmou (*otcova*)
- **AU** – adjektívum s neúplnou paradigmou (*super*)

- **PS** – zámeno so substantívnou paradigmou (*koľkátko*)
- **PA** – zámeno s adjektívnou paradigmou (*taký*)
- **PP** – zámeno so samostatnou zámenou paradigmou (*ja*)
- **PF** – zámeno so zmiešanou paradigmou (*on, sám, žiaden*)
- **PU** – zámeno s neúplnou paradigmou (*jeho, jej, ich, koľko, tolko*)
- **PD** – zámeno s adverbialnou paradigmou (*tam, niekedy*)
-
- **NS** – číslovka so substantívnou paradigmou (*milión, raz*)
- **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, päťoro, veľa*)
- **ND** – číslovka s adverbialnou paradigmou (*dvakrát, mnohonásobne*)

3.3.3. Na treťom mieste sa pri menných slovných druhoch, ktoré majú inú ako príslovkovú paradigmu, uvádza **rod** s hodnotami mužský životný (**m**); mužský neživotný (**i**); ženský (**f**); stredný (**n**); neurčený (**o**); pri zámenách a slovesách aj všeobecný (**h**).

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

Pri tzv. bezrodových zámenách (*ja, ty, my, vy, seba*) sa uvádza len všeobecný rod (**h**) a z kontextu sa ďalej nezisťuje, či ho možno určiť presnejšie. S tým súvisí aj určenie rodu pre sloveso, ktoré je so zámenom v zhode.

Napr.: *Vy ste neprišli!* l = vy t = PPhp1 l = neprišť t = VLdpbh-

Pri číslovkách (*tisíc, sto, veľa, málo, viac* a pod.) i zámenách (*koľko, tolko*), ktoré sa používajú ako nesklonný tvar *päť* odlišujeme N a A od ostatných pádov. Pri spojení s genitívom substantíva sa číslovka správa ako neutrum v singulári. Teda rod a číslo určujeme aj podľa zhody so slovesom:

N:	<i>piati chlapci prišli</i>	l = päť	t = NUmp1
	<i>päť chlapov, žien, detí prišlo</i>	l = päť	t = NUns1
A:	<i>videl piatich chlapov</i>	l = päť	t = NUmp4
	<i>videl päť chlapov, žien, detí</i>	l = päť	t = NUns4

G:	<i>od päť/piatich chlapov, žien, detí</i>	l = päť	t = NUm(i/f/n)p2
D:	<i>k päť/piatim chlapom, ženám, deťom</i>	l = päť	t = NUm(i/f/n)p3
L:	<i>o päť/piatich chlapoch, ženách, deťoch</i>	l = päť	t = NUm(i/f/n)p6
I:	<i>s päť/piatimi chlapmi, ženami, deťmi</i>	l = päť	t = NUm(i/f/n)p7

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

Napr.: *Kúpili kanoe.* l = kanoe t = SUno4

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

Napr.: *Tri domy.* l = tri t = NNip1

O určovaní čísla pri číslovkách (*tisíc, sto, veľa, málo, viac* a pod.) i zámenách (*koľko, tolko*), ktoré sa používajú ako nesklonný tvar *päť* pozri 3.3.3.

3.3.5. Pád uvádzame pri menných slovných druhoch na piatom mieste. Má hodnoty nominatív (**1**), genitív (**2**), datív (**3**), akuzatív (**4**), vokatív (**5**), lokál (**6**), inštrumentál (**7**), neurčený (**o**). Vokatív vymedzujeme ako "oslovovací" pád, uvádzame ho teda pri osloveniach. Hodnotu neurčený (**o**) priradíme tvarom, pri ktorých ani z kontextu nie je možné pád jednoznačne určiť.

Napr.:	<i>od študenta</i>	l = študent	t = SSms2
	<i>Dievča, pod' sem!</i>	l = dievča	t = SSns5
	<i>Mesto!</i>	l = mesto	t = SSns0 (N, A alebo V sg.)

3.3.6. Stupeň s hodnotami pozitív/irelevantný stupeň (**x**), komparatív (**y**), superlatív (**z**) platí pre adjektíva, adverbíá a participiá. Kategóriu stupňa určujeme formálne, riadime sa prítomnosťou alebo neprítomnosťou modifikačných morféme *naj-*, *-š-*, *-ej:š-*. Stupeň je určený aj pri vzťahových adjektívach, keďže aj tieto typy adjektív sa môžu potenciálne stupňovať (**najdrevenejší stôl*, ale *najdrevenejší tanečník*).

Pri privlastňovacích prídavných menách, príslovkách, ktoré sú nestupňovateľné, a pri prídavných menách uvádzame hodnotu **pozitív/irelevantný stupeň (x)**. Ide teda o základnú a v týchto prípadoch jedinú formu, pri ktorej je hodnota stupňa irelevantná.

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

Napr.:	<i>dobrý človek</i>	l = dobrý	t = AAms1x
	<i>lepší človek</i>	l = dobrý	t = AAms1y
	<i>najdrevenejší tanečník</i>	l = drevený	t = AAms1z
	<i>svieti menej</i>	l = málo	t = Dy
	<i>otcov priateľ</i>	l = otcov	t = AAms1x
	<i>hore</i>	l = hore	t = Dx
	<i>pečený koláč</i>	l = pečený	t = Gtis1x
	<i>viac ľudí</i>	l = viac	t = NUns1

3.3.7. Pri slovesách sa na druhom mieste určuje **slovesná forma** s hodnotami: infinitív (**I**), prítomný čas (**K**), imperatív (**M**), prechodník (**H**), *l*-ové prídavné meno (**L**), futúrum (najčastejšie slovesa *byť*) (**B**).

Nájďme však aj prípady, keď nedokonavé sloveso vyjadruje budúci čas syntetickým tvarom (*letieť* – *budem letieť* aj **poletím**; *nieť* – *budem nieť* aj **ponesiem**...). Tieto tvary označujeme ako futúrum a značíme **B**.

Napr.:	<i>písať</i>	l = písať	t = VIe+
	<i>píšem</i>	l = písať	t = VKesa+
	<i>píš</i>	l = písať	t = VMesb+
	<i>píšuc</i>	l = písať	t = VHe+
	<i>písal</i>	l = písať	t = VLescm+
	<i>budem</i>	l = byť	t = VBesa+
	<i>ponesiem</i>	l = nieť	t = VBesa+

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

Indikatív

- tvary **prítomného času** (*píšem, píšeš, píše, píšeme, píšete, píšu*)
sloveso *písať* v prítomnej forme (**VK**)
- tvary **minulého času** (*písal som, písal si, písal, písali sme, písali ste, písali*)
sloveso *písať* vo forme *l*-ového prídavného mena (**VL**) + formálne 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., ponesiem*)
formálne sloveso *byť* vo futúre (**VB**) + sloveso *písať* v infinitíve (**VI**)
sloveso *nieť* vo futúre (**VB**)
- tvary **budúceho času dokonavých sloves** (*napíšem, napíšeš, napíše, napíšeme, napíšete, napíšu*)
sloveso *písať* v prítomnej indikatívnej forme (**VK**)

Imperatív

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

Kondicionál

- tvary **prítomného času** (*písal/-la/-lo by som*; *písal/-la/-lo by si*; *písal/-la/-lo by*; *písali by sme*; *písali by ste*; *písali by*)
sloveso *písať* vo forme *l*-ového prídavia (**VL**) + *by* (**Y**) + formálne sloveso *byť* v prítomnej forme (**VK**), ak je prítomné
- tvary **minulého času** (*bol/-la/-lo by som písal/-la/-lo*; *bol/-la/-lo by si písal/-la/-lo*; *bol/-la/-lo by písal/-la/-lo*; *boli by sme písali*; *boli by ste písali*; *boli by písali*)
formálne sloveso *byť* vo forme *l*-ového prídavia (**VL**) + *by* (**Y**) + formálne sloveso *byť* v prítomnej indikatívnej forme (**VK**), ak je prítomné + *písať* vo forme *l*-ového prídavia (**VL**)

3.3.8. Slovesný vid s hodnotami dokonavý (**d**), nedokonavý (**e**) alebo obojvidový (**j**) sa nachádza pri slovesách na treťom mieste. Pri určovaní vidu sa riadime kvalifikátormi v KSSJ (2003).

Napr.:

<i>píšem</i>	l = <i>písať</i>	t = VKesa+
<i>napíšem</i>	l = <i>napísať</i>	t = VKdsa+
<i>počujem</i>	l = <i>počuť</i>	t = VKjsa+

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

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

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

Napr.:

<i>písal som</i>	l = <i>písať</i>	t = VLesam+
	l = <i>byť</i>	t = VKesa+
<i>písal</i>	l = <i>písať</i>	t = VLescm+
<i>písal by</i>	l = <i>písať</i>	t = VLescm+
	l = <i>by</i>	t = Y
<i>budem písať</i>	l = <i>byť</i>	t = VBesa+
	l = <i>písať</i>	t = VIe+

3.3.10. Kongruencia v rode s hodnotami: kongruencia v mužskom rode živ. (**m**) a neživ. (**i**); v ženskom rode (**f**); v strednom rode (**n**); neurčená (**o**), všeobecný rod (**h**). Kategória kongruencie v rode je relevantná pri verbách len so slovesnou formou *l*-ového prídavia a uvádza sa na šiestom mieste. Pri ostatných slovesných formách sa kongruencia neurčuje. Hodnotu neurčená kongruencia (**o**) zapíšeme, ak sa pri tvare nedá z kontextu jednoznačne určiť, s ktorým rodom je sloveso v kongruencii.

Napr.:

<i>Študent nepríde.</i>	l = <i>neprísť</i>	t = VKdsc-
<i>Študent neprišiel.</i>	l = <i>neprísť</i>	t = VLdscm-
<i>Vy ste neprišli.</i>	l = <i>neprísť</i>	t = VLdpbh-
<i>Ženy a deti neprišli.</i>	l = <i>neprísť</i>	t = VLdpc-

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

Špecifické je riešenie pri slovese *byť*. Zatiaľ čo pri tvaroch minulého i budúceho času je negácia vyjadrená synteticky (*nebol*, *nebude*), v prítomnom čase sa negácia vyjadruje pomocou častice *nie*. Prítomnosť tejto častice ovplyvňuje lematizáciu slovesa (*nebyť*) a zároveň naznačuje

negáciu (-).

Napr.:	<i>píšem</i>	l = <i>písať</i>	t = VKesa+
	<i>nepíšem</i>	l = <i>nepísať</i>	t = VKesa-
	<i>nie je</i>	l = <i>nie</i>	t = T
		l = <i>nebyť</i>	t = VKesc-

3.3.12. Pri číslovkách sa na druhom mieste uvádza znak X, ak ide o samostatné použitie číslovky, vyjadrenie samostatného počtu: číslovky v matematických úkonoch, v pomenúvacom nominatíve a pod. V takomto prípade značíme len slovnú triedu a typ použitia.

Napr.:	<i>v roku tisíc deväťsto päťdesiatštyri</i>	l = <i>tisíc</i>	t = NX
		l = <i>deväťsto</i>	t = NX
		l = <i>päťdesiatštyri</i>	t = NX
	<i>vchod č. deväť</i>	l = <i>deväť</i>	t = NX

3.3.13. Znak pre atribút **aglutinovanosť (g)** pridávame pri pronomínach na koniec tagu, ak je to relevantné. Teda pri tvaroch ako *oňho, preňho, naňho, oň, preň, zaň...* považujeme za lemu tvary *o_on, pre_on, na_on, o_ono, pre_ono, za_ono* a prítomnosť aglutinovanej predložky zdôrazníme aj znakom pre aglutinovanosť (g).

Napr.:	<i>Išiel poňho.</i>	l = <i>po_on</i>	t = PFms4g
--------	---------------------	------------------	------------

3.3.14. Na druhom mieste v tagu sa pri formálnych participiách (G) nachádza znak pre **druh príchastia** s hodnotami aktívne (k) a pasívne (t). Ďalej pokračujeme určením rodu, čísla, pádu a formálneho stupňa príchastia.

Napr.:	<i>od písaného denníka</i>	l = písaný	t = Gtis2x
	<i>píšucemu peru</i>	l = píšuci	t = Gkns3x

3.3.15. Pri prepozíciách (E) uvádzame na druhom mieste **formu predložky**: vokalizovaná (v) alebo nevokalizovaná (u). Na treťom mieste je uvedené spojenie s pádom. Hodnoty sú zhodné s kategóriou pádu, bližšie 3.3.5.

Napr.:	<i>s tebou</i>	l = s	t = Eu7
	<i>so mnou</i>	l = s	t = Ev7

3.3.16. Pri spojkách (O) alebo časticiach (T) *keby, aby, žeby, akoby, s'aby* je možné označiť, že ide o aglutinované tvary, teda že obsahujú i kondicionálnu morfému *by* tak, že pridáme znak Y (pozri aj 3.2.).

Napr.:	<i>Povedal by, keby vedel.</i>	l = <i>keby</i>	t = OY
	<i>Keby si vedel!</i>	l = <i>keby</i>	t = TY

3.4. Všeobecné zásady značkovania

3.4.1. Znak uvádzame vtedy, keď je kategória pre daný tvar relevantná. Napr. pre zámeno *ako* je nerelevantná kategória rodu, čísla, pádu, osoby, určíme teda iba **slovnú triedu** a **paradigm**: l = ako, t = PD. Pri slovesách v neurčitku (*vniknúť*) nemožno určiť kongruenciu v rode, číslo ani osobu, určíme teda iba **slovnú triedu** (sloveso), **slovesnú formu** (infinitív), **slovesný vid** (dokonavý), **afirmáciu** (afirmatív): l = *vniknúť*, t = VId+.

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

<i>Nechcem cestovať v tom kupé.</i>	(neutr., sg., L)	l = <i>kupé</i>	t = SUns6
<i>Dozvedel sa to od päť/piatich chlapov.</i>	(mask. živ., pl., G)	l = <i>päť</i>	t = NUmp2
<i>Pozdravil sa jeho sestru.</i>	(femin., sg., D)	l = <i>jeho</i>	t = PUfs3
<i>Pre koľko (koľkých) ľudí to má pripraviť?</i>	(neutr., sg., A)	l = <i>koľko</i>	t = PUns4

Potešil sa pätoro/piatim nohaviciam. (femin., pl., D) l = pätoro t = NUfp3
Nenašiel som ani kúsok. (dok., 1. os., sg., mask.) l = nenájsť t = VLdsam-

3.4.3. V prípadoch ako *kupé, jeho* 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 zámene *koľko* zachytíme A pl. v podobách: *pre koľko ľudí – pre koľkých ľudí*. Frekvenčná analýza potom ukáže pomer v ich výskyte. Alebo pri substantíve *Philips* (l = philips) sme v SNK zachytili takéto tvary G sg.:

Cieľom transakcie je transformácia Philips...
...hovori O. Š. z Philips Slovakia...
...s kapitálovou pomocou Philipsu...
...veľkých spotrebičov od Philipsa...

3.4.4. Podobne ako pri zjednotnení absolútnych morfológických homonym postupujeme aj pri homonymii len niektorých tvarov. Aj v týchto prípadoch zjednotňujeme tvarovú homonymiu z kontextu:

Napr.: tvar *pekné* – možnosti: N, A pl. mask. neživ. t = AAmp1x / AAmp4x
 N, A pl. fem. t = AAfp1x / AAfp4x
 N, A sg. neutra t = AAns1x / AAns4x
 N, A pl. neutra t = AAnp1x / AAnp4x

Kontext: *Dievčatá sú pekné*. Jedinou možnosťou je riešenie: t = AAnp1x.

3.4.5. Pri slovnodruhovej homonymii sa v zásade riadime kodifikačnými príručkami (KSSJ) a sémantikou. Napr.:

Tuším vo vzduchu búрку. l = tušiť t = VKesa+
Tuším budú problémy. l = tuším t = T
Prosím si vodu. l = prosiť t = VKesa+
Podťe, prosím. l = prosím t = T
Lepšie to nebude. l = dobrý t = AAns1y
Vieš to aj lepšie. l = dobre t = Dy
Bolo zima. l = zima t = Dx
Prišla/Bola zima. l = zima t = SSfs1

3.4.6. Príslušné gramatické kategórie možno priradiť aj tým skratkám a značkám (**W**), ktoré sa ustálili ako skratky a značky od ohybných slovných druhov (*nám., ul., č., l, cm*), alebo tým, ktoré majú v texte syntakticko-sémantickú funkciu porovnateľnú s funkciou ohybných slovných druhov (*do SND, v SR*). Bližšie pozri 3.5.15.

3.5. Príklady značkovania a riešenie značkovania niektorých zložených tvarov v SNK

3.5.1. Substantíva:

Winston l = winston t = SSms1:r
poznám vedúceho oddelenia l = vedúci t = SAms4
od nejakej gazdinej l = gazdiná t = SFfs2
sediť v útulnom kupé l = kupé t = SUn6

3.5.2. Adjektíva:

mám neprijemný pocit l = neprijemný t = AAis4x
senzi dievča l = senzi t = AUms1x
pre otcovho priateľa l = otcov t = AFms4x

Analytické stupňovanie niektorých adjektív (*menej lukratívny*) je v tagsete SNK spojením (formálneho) adverbia a adjektíva. Napr.:

*Je to **menej lukratívny** obchod.* l = málo t = Dy
l = lukratívny t = AAis1x

3.5.3. Pronominá:

*odo **mňa*** l = ja t = PPhs2
*Bojím sa o **ňu**.* l = ona t = PFfs4
***Ktorých** ľudí poznáte?* l = ktorý t = PAmp4
***Žiadne** utierky tam neboli.* l = žiaden t = PFfp1
*od **jeho** nástupu* l = jeho t = PUis2
***Koľko** ľudí prišlo?* l = koľko t = PUns1
***Tak** to bolo.* l = tak t = PD

Analytické zámená (*taký istý, ten istý, tak isto*) anotujeme ako samostatné tokeny. Napr.:

*Dostal sa do **takého istého** problému.* l = taký t = PAis2
l = istý t = PAis2
***koľký** raz* l = koľký t = PAis4
l = raz t = NSis4
***koľko** ráz* l = koľko t = PUns4
l = raz t = NSip2

3.5.4. Numeráliá:

*vstup č. **9*** l = 9 t = 0
*vstup č. **deväť*** l = deväť t = NX
*Z **milióna** neostalo nič.* l = milión t = NSis2
*Pozreli sa na **1.** pretekárku.* l = 1 t = 0
l = . t = Z
*Rozprávali sa o **dvoch** stromoch.* l = dva t = NNip6
*od **jedného** dievčaťa* l = jeden t = NFns2
*pre **piatich** chlapov* l = päť t = NUmP4
*videl **päť** chlapov* l = päť t = NUns4
***mnoho** strachu* l = mnoho t = NUns1
*o **tisíc** stromoch* l = tisíc t = NUip6
*neprišiel **trikrát*** l = trikrát t = ND
*neprišiel ***trirazy*** l = trirazy t = ND

Zložené číslovky sú rozdelené na samostatné tokeny oddelené medzerou, preto ich značíme zvlášť. Napr.:

*s **tisíc*** l = tisíc t = NUmP7
deväťsto l = deväťsto t = NUmP7
osemdesiatimi l = osemdesiat t = NUmP7
***štyrmi** obyvateľmi* l = štyri t = NNmp7
*s **1 984** obyvateľmi* l = 1 t = 0
l = 984 t = 0
*neprišiel **3-krát*** l = 3 t = 0
l = - t = Z
l = krát t = Q
*neprišiel ***tri krát*** l = tri t = NNip4
l = krát t = Q
*neprišiel **tri razy*** l = tri t = NNip4
l = raz t = NSip4

3.5.5. Verbá:

Analytické a syntetické tvary slovíes značíme takto:

<i>písať</i>	l = <i>písať</i>	t = Vle+
<i>píšu</i>	l = <i>písať</i>	t = VKepc+
<i>písal som</i>	l = <i>písať</i> l = <i>byť</i>	t = VLesam+ t = VKesa+
<i>budeš písať</i>	l = <i>byť</i> l = <i>písať</i>	t = VBesb+ t = Vle+
<i>napíšem</i>	l = <i>napísať</i>	t = VKdsa+
<i>nenapíšme</i>	l = <i>nenapísať</i>	t = VMdpa-
<i>napísali by sme</i>	l = <i>napísať</i> l = <i>by</i> l = <i>byť</i>	t = VLdpam+ t = Y t = VKepa+
<i>boli by ste napísali</i>	l = <i>byť</i> l = <i>by</i> l = <i>byť</i> l = <i>napísať</i>	t = VLepbm+ t = Y t = VKepb+ t = VLdpbm+
<i>píšuc</i>	l = <i>písať</i>	t = VHe+

3.5.6. Formálne participiá:

<i>mám skľučujúci pocit</i>	l = <i>skľučujúci</i>	t = Gkis4x
<i>kvety boli vyšité</i>	l = <i>vyšitý</i>	t = Gtip1x

3.5.7. Adverbiá:

<i>Herec hral výborne.</i>	l = <i>výborne</i>	t = Dx
<i>Dole sa nepozerať.</i>	l = <i>dole</i>	t = Dx
<i>Lepšie to neviem.</i>	l = <i>dobre</i>	t = Dy

Príslovky typu *na bielo*, *za čerstva*, *do ružova*, *po slovensky*:

<i>na bielo</i>	l = <i>na</i>	t = Q
	l = <i>bielo</i>	t = Dx

3.5.8. Prepozície:

<i>Stáli na streche.</i>	l = <i>na</i>	t = Eu6
<i>Rozprávaj sa so mnou!</i>	l = <i>s</i>	t = Ev7

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:

<i>na rozdiel od</i>	l = <i>na</i>	t = Eu4
	l = <i>rozdiel</i>	t = SSis4
	l = <i>od</i>	t = Eu2

3.5.9. Konjunkcie:

Aglutináciu spojky s kondicionálovou morféomou vyjadríme:

<i>Povedal, aby si sadli.</i>	l = <i>aby</i>	t = OY
-------------------------------	----------------	--------

Zložené spojky anotujeme ako samostatné tokeny:

<i>a teda</i>	l = <i>a</i>	t = O
	l = <i>teda</i>	t = O

3.5.10. Partikuly:

Aglutináciu častice s kondicionálovou morféomou vyjadríme:

<i>Keby si netáral!</i>	l = <i>keby</i>	t = TY
<i>koniec koncov</i>	l = <i>koniec</i>	t = T
	l = <i>koncov</i>	t = T

3.5.11. Interjekcie:
Hop, a už som tu! l = *hop* t = J

3.5.12. Reflexívne sa/si
rozprávali sme sa l = *sa* t = R
povedali sme si l = *si* t = R

3.5.13. Morféma by
písali by l = *by* t = Y

3.5.14. Interpunkcia
hovoril, že l = , t = Z
Dobre. l = . t = Z
Čo? l = ? t = Z
"Čo ty na to?" l = " t = Z
Mladosť – pochybnosť. l = – t = Z
2-krát l = - t = Z

3.5.15. Abreviácie

Gramatické kategórie pri skratke určujeme, ak je to relevantné:

SND	l = <i>snd</i>	t = Wns1
atď.	l = <i>atď</i>	t = W
	l = .	t = Z
25 km	l = <i>km</i>	t = Wip2
XML	l = <i>xml</i>	t = W
H₂O	l = <i>h2o</i>	t = W
na Panskej ul.	l = <i>ul</i>	t = Wfs6
	l = .	t = Z
do výšky 60 cm	l = <i>cm</i>	t = Wip2
do SND	l = <i>snd</i>	t = Wns2
v SR	l = <i>sr</i>	t = Wfs6

3.5.16. Neurčiteľný slovný druh

bielo-červený	l = <i>bielo</i>	t = Q
2-krát	l = <i>krát</i>	t = Q
česko-slovenský	l = <i>česko</i>	t = Q
po slovensky	l = <i>po</i>	t = Q

3.5.17. Neslovný element

* (uvádza poznámku pod čiarou) l = * t = #

3.5.18. Citátový výraz

Take it easy!	l = <i>take</i>	t = %
	l = <i>it</i>	t = %
	l = <i>easy</i>	t = %
	l = !	t = Z

3.5.19. Číslice

25 km	l = 25	t = 0
do výšky 60 cm	l = 60	t = 0
2-krát	l = 2	t = 0

4. Tabuľky

4.1. Substantíva

Pozícia 1: Substantívum – S

Pozícia 2: Paradigma

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

Pozícia 3: Rod

Znak	Hodnota	Príklad
m	mužský životný	<i>chlap, hrdina, kuli, vedúci, hlavný, Borský</i>
i	mužský 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>
o	neurčený	<i>djembe, sitar, ska</i>

Pozícia 4: Číslo

Znak	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

Znak	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, Hlavatú</i>
5	vokatív	<i>pane, mamo, mami, tati, oci, 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! (N alebo A sg.)</i>

4.2. Adjektíva

Pozícia 1: Adjektívum – A

Pozícia 2: Paradigma

Znak	Hodnota	Príklad
A	adjektívna	<i>pekný, pekná, pekné, cudzí, cudzia, cudzie, pávi</i>
F	zmiešaná	<i>otcov, matkin</i>
U	neúplná	<i>super, fajn, nanič; hoden, vinen, rád (rada, rado, radi, rady)</i>

Pozícia 3: Kongruencia v rode

Znak	Hodnota	Príklad
m	mužský životný	<i>pracovitý (chlap), besný (pes), besní (psi), otcov (priateľ)</i>
i	mužský neživ.	<i>staré (duby), besné (psy), nový (stroj), super (nápad)</i>
f	ženský	<i>štithe (ženy), široké (ulice), prázdna (dlaň), biela (kost)</i>
n	stredný	<i>veľké (mestá), dobré (vysvedčenie), fajn (dievča)</i>
o	neurčený	nie je možné zjednotiť na základe kontextu: <i>smutné (ženy a deti)</i>

Pozícia 4: Kongruencia v čísle

Znak	Hodnota	Príklad
s	singulár	<i>pracovitý (chlap), besný (pes), korpusový (manažér), super (nápad)</i>
p	plurál	<i>pracoviti (chlapí), korpusovi (lingvisti), fajn (ľudia), matkini (známi)</i>
o	neurčené	nie je možné zjednotiť na základe kontextu: <i>(Kúpili) super (kanoe.)</i>

Pozícia 5: Kongruencia v páde

Znak	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)</i>
6	lokál	<i>o pracovitom (chlapovi), besnom (psovi), super (nápade)</i>
7	inštrumentál	<i>s pracovitým (chlapom), besným (psom), super (nápadom)</i>
o	neurčený	<i>Pekné/Super (mesto)! (N alebo A sg.)</i>

Pozícia 6: Stupeň

Znak	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: Pronominum – P

Pozícia 2: Paradigma

Znak	Hodnota	Inventár (podľa KSSJ a doplnené)
S	substantívna	<i>kol'kátka, všetučko</i>
A	adjektívna	<i>oný, (oná, oné), onaký, onakvý, inší (inšia, inšie), iný (iná, iné), inakší aký, ký, kýže, bárs-, bár-, bohvie-, čertvie-, daj-, (dáky), hocij-, hoc-, kadej-, kdej-, ktovie-, neviem-, vol'aj-, (vol'áky), všakov-, (vsakový), všelij-, nej- -aký- -koľvek, -si nijaký, nijakovský ktorý, bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, neviem-, nie-, poda-, ponie-, sotva-, málo-, voľa-, zriedka- -ktorý- -koľvek, -si čí (čia, čie), ničí, da-, nie-, voľa- -čí- -koľvek, -si, taký, takýto, hentaký koľký (koľkí, koľké), koľkoraký, koľkonásobný, koľkáty toľký (toľkí, toľké), toľkoraký, toľkonásobný, toľkýto, toľčizný, toľkánsky každý, každučký, každučičký, samý (samá, samé, sami, samé) bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, leda-, neviem-, nie-, voľa-, všeli- -koľký-, -koľkoraký-, -koľkonásobný- -koľvek</i>
P	zámenná	<i>ja, ty, my, vy, seba (sebe, sebou)</i>
F	zmiešaná	<i>on, ona, ono, onô kto, nikto (nik), bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, kade-, kde-, ktovie-, málo-, neviem-, nie-, sotva-, voľa-, všeli-, zriedka- -kto- -koľvek, -si čo, nič, pranič, bárs-, bár-, bohvie-, čertvie-, da-, hoci-, hoc-, kade-, kde-, ktovie-, leda-, málo-, neviem-, nie-, sotva-, voľa-, všeli- -čo- -koľvek, -si ten, tá, to, tento, táto, toto, tamten, tamtá, tamto, henten, hentá, hento, tenže, táže, tože môj (moja, moje), tvoj, náš, váš, svoj onen (oná, ono), všetok (všetka, všetko, všetci, všetky), sám (sama, samo), žiaden (žiadny, žiadna, žiadne, žiadni, žiadne)</i>
U	neúplná	<i>koľko (koľkí, koľké), koľkože, toľko (toľkí, toľké), toľkoto, bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, leda-, neviem-, nie-, voľa-, všeli- -koľko- (-koľvek) jeho, jej, ich</i>
D	príslovková	<i>koľkonásobne, koľkokrát, koľkýkrát, toľkonásobne, toľkokrát, toľkýkrát bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, leda-, neviem-, nie-, voľa-, všeli- -koľkonásobne-, -koľkokrát-, -koľkýkrát- -koľvek ako, nijako, nijakovsky, tak, takto, onak, ináč, inak, inakšie, takisto, všakovak, bárs-, bár-, bohvie-, čertvie-, daj- (dáko), hocij-, hoc-, ktovie-, neviem-, nej-, vol'áj-, (vol'áko), všelij- -ako- -koľvek, -si 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 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 kam, nikam, inam, tam, tamto, tamže, sem, ta, tu, tuto, tuná, hen, hentam, bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, neviem-, nie- -kam- -koľvek, -si kadiaľ, odkadiaľ, skadiaľ, skadiaľže, znikadiaľ, odnikadiaľ, bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, neviem-, nie-, odnie-, voľa-, všeli-, zda-, znie-, zvol'a- -kadiaľ- -koľvek, -si, tadiaľ, tadiaľto, odtadiaľ, stadiaľ, stadiaľto, všadiaľ, odvšadiaľ, zovšadiaľ dokiaľ, dokiaľže, dokiaľkoľvek, odkiaľ, odniekiaľ, pokiaľ, odnikiaľ, odiniaľ, bohvie-, čertvie-, ktovie-, neviem- -odkiaľ- -koľvek, -si dotiaľ, zatiaľ, odtiaľ, odtiaľto, potiaľ, potiaľto kedy, dokedy, nakedy, odkedy, inokedy, bohvie-, čertvie-, da-, hoci-, hoc-, ktovie-, málo-, neviem-, nie-, sotva-, voľa-, zriedka-, bárs-, bár- -kedy- -si, -koľvek vtedy, dovtedy, odvtedy, onehdy, nikdy, vždy, vždycky, dosiaľ, posiaľ, nakoľko, natoľko, načo, prečo, prečože, začo (zač), nato, preto, zato, bohvie-, ktovie-, čertvie-, neviem- -prečo</i>

Pozícia 3: Rod, resp. kongruencia v rode

Znak	Hodnota	Príklad
m	mužský životný	<i>oni, on (brat), môj, ten, aký, taký, každý, žiadny, nejaký chlap</i>
i	mužský neživotný	<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 (dieťa)</i>
h	všeobecný	<i>ja, ty, my, vy, seba</i>
o	neurčený	nie je možné zjednotiť z kontextu: <i>naše (ženy a deti)</i>

Pozícia 4: Číslo, resp. kongruencia v čísle

Znak	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 kanoé?</i>

Pozícia 5: Pád, resp. kongruencia v páde

Znak	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)</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: Aglutinovanosť

Znak	Hodnota	Príklad
g	aglutinované	<i>preňho, zaňho, naňho, doňho, oňho, oň, preň, zaň, naň, doň</i>

4.4. Numerálie

Pozícia 1: Numerále – N

Pozícia 2: Paradigma

Znak	Hodnota	Inventár (podľa KSSJ – doplnené a opravené)
S	substantívna	<i>nula, milión, bilión, miliarda, stá milióny, státisíce raz</i>
A	adjektívna	<i>jediný, mnohý 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, dvojstý, trojstý, štvorstý, päťstý, sedemstý, tisíci, dvojtisíci, stotisíci milióny, miliardty, meruôsmi, nultý entý, ostatný, predostatný, posledný, neposledný, predposledný dvojité, trojité, štvoritý, dvojný, jednorazový, dvojnásobný, trojnásobný, štvornásobný, päťnásobný, šesťnásobný, sedemnásobný, osemnásobný, deväťnásobný, desaťnásobný, dvanásťnásobný, dvadsaťnásobný, sedemdesiatnásobný, stonásobný, tisíc násobný, viacerázový, mnohonásobný, viacnásobný jednaký, dvojaký, trojaký, štvoraký, päťoraký, šestoraký, sedmoraký, osmoraký, deväťoraký, desatoraký, dvanásťoraký, dvadsaťoraký, storaký, tisícoraký, obojaký mnohoraký</i>
N	čísloková	<i>dva (dvaja, dve), oba (obaja, obe), obidva (obidvaja, obidve), tri (traja), štyri (štyria)</i>
F	zmiešaná	<i>jeden (jedna, jedno)</i>
U	neúplná	<i>sto, tisíc, dvesto, tristo, štyristo, päťsto, šesťsto, sedemsto, osemsto, deväťsto dvetisíc, tritisíc, štyritisíc, päťtisíc, šesťtisíc, sedemtisíc, osemtisíc, desaťtisíc, štyridsaťtisíc, sedemdesiatitisíc, stotisíc, dvestotisíc, sedemstotisíc päť (piati), šesť/šiesti, sedem/siedmi, osem/ôsmi, deväť/deviati, desať/desiati, jedenásť/jedenásti, dvanásť/dvanásti, trinásť/trinásti, štrnásť/štrnásti, pätnásť/pätnásti, šestnásť/šestnásti, sedemnásť/sedemnásti, osemnásť/osemnásti, deväťnásť/deväťnásti, dvadsať/dvadsiaty, tridsať/tridsiaty, štyridsať/štyridsiaty, päťdesiat/päťdesiati, šesťdesiat/šesťdesiati, sedemdesiat/sedemdesiati, osemdesiat/osemdesiati, deväťdesiat/deväťdesiati, dvadsaťpäť/dvadsiatipiati, päťadvadsať dvoje, troje, štvoro, päťoro, šestero, sedmero, osmero, deväťoro, desatoro, dvadsatoro, tisícero, obidvoje, oboje, viacero (viacerí, viaceré) byťu, byťku, habadej, hodne, veľa, neveľa, priveľa, 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, dosť, pol, štvrť, trištvrte, (tri) štvrte,</i>
D	príslovková	<i>prvýkrát, druhýkrát, tretíkrát *štvrtýraz, *piatyraz, *šiestyraz šesťkrát, sedemkrát, *osemráz, *deväťráz, dvoj-, troj-, štvor-, päť-, šesť-, sedem-, osem-, deväť-, desať-, dvanásť-, dvadsať-, sedemdesiat-, sto-, tisíc- -násobne dvojito, dvojmo, trojito, trojmo, štvormo viackrát, *mnohoráz, veľa-, nejeden-, pár-, zopár-, nesčíselne-, nasto- ráz mnoho-, viac- -násobne, neraz jedn-, dvoj-, troj-, štvor-, päťor-, šesor-, sedmor-, osmor-, deväťor-, desator-, dvanásťor-, dvadsaťor-, stor-, tisícor-, oboj- -ako, mnohorako</i>
X	samostatné použitie	<i>v roku dvetisíc, vchod č. šesť, a pod.</i>

Pozícia 3: Rod, resp. kongruencia v rode

Znak	Hodnota	Príklad
m	mužský životný	<i>jeden (pán), jedni, dvaja, traja, štyria, piati (páni), päť (chlapov)</i>
i	mužský neživotný	<i>jeden (stól), 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	<i>nie je možné zjednotiť na základe kontextu: tri ženy a deti</i>

Pozícia 4: Číslo, resp. kongruencia v čísle

Znak	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, resp. kongruencia v páde

Znak	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), 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</i>
7	inštrumentál	<i>s jedným (človekom), druhou (osobou), trojakým (vínom), štvornásobným</i>
o	neurčený	<i>dve ženy, dva stromy (N alebo A sg.)</i>

4.5. Verbá**Pozícia 1: Verbum – V****Pozícia 2: Slovesná forma**

Znak	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! buďte! anotujme!</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, poletí</i>

Pozícia 3: Vid

Znak	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, počuť</i>

Pozícia 4: Číslo

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

Pozícia 5: Osoba

Znak	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

Znak	Hodnota	Príklad
m	mužský životný	<i>(lingvista) anotoval, (lingvisti) anotovali</i>
i	mužský neživotný	<i>(korpus) sa rozrastal</i>
f	ženský	<i>(búrka) prišla, (búrky) prišli</i>
n	stredný	<i>(zrkadlo) viselo, (zrkadlá) viseli</i>
h	všeobecný	<i>(My sme) rástli.</i>
o	neurčený	<i>(Ženy a deti) rástli.</i>

Pozícia 7: Negácia

Znak	Hodnota	Príklad
+	afirmácia	<i>písať, napísať, pracovať, kresliť, behať, vidieť, chcieť, prosiť, byť, som</i>
-	negácia	<i>nepísať, nenapísať, nepracovať, nekresliť, nebehať, nevidieť, nebol, nie je</i>

4.6. Formálne particípiá

Pozícia 1: Particípium – G

Pozícia 2: Druh

Znak	Hodnota	Príklad
k	aktívne	<i>píšuci, pracujúci, kresliaci, bežiaci, vidiaci, prosiaci, vediaci, nosiaci</i>
t	pasívne	<i>písaný, napísaný, kreslený, videný, nosený, vyhodенý, založený</i>

Pozícia 3: Kongruencia v rode

Znak	Hodnota	Príklad
m	mužský životný	<i>kreslený, videný, vyhodенý, píšuci, pracujúci, kresliaci, bežiaci chlapec</i>
i	mužský neživ.	<i>kreslený, videný, vyhodенý, pracujúci, kresliaci, bežiaci stroj</i>
f	ženský	<i>kreslená, videná, vyhodенá, pracujúca, kresliaca, bežiaci žena</i>
n	stredný	<i>kreslené, videné, vyhodенé, pracujúce, kresliace, bežiaci dieťa</i>
o	neurčený	nie je možné zjednotiť na základe kontextu: <i>bežiaci (ženy a deti)</i>

Pozícia 4: Kongruencia v čísle

Znak	Hodnota	Príklad
s	singulár	<i>kreslený, videný, vyhodенý, píšuci, pracujúci, kresliaci, bežiaci chlapec</i>
p	plurál	<i>kreslení, videní, vyhodení, píšuci, pracujúci, kresliaci, bežiaci chlapci</i>
o	neurčené	nie je možné zjednotiť na základe kontextu: <i>(Našli) vyhodенé (kanoe).</i>

Pozícia 5: Kongruencia v páde

Znak	Hodnota	Príklad
1	nominatív	<i>kreslený, videný, vyhodенý, píšuci, pracujúci, kresliaci, bežiaci chlapec</i>
2	genitív	<i>kresleného, videného, vyhodенého, píšuceho, pracujúceho chlapca</i>
3	datív	<i>kreslenému, videnému, vyhodенému, píšucemu, pracujúcemu chlapcovi</i>
4	akuzatív	<i>kresleného, videného, vyhodенého, píšuceho, pracujúceho chlapca</i>
5	vokatív	<i>Píšuci spisovatelia! Vyhodení zamestnanci!</i>
6	lokál	<i>o kreslených, videných, vyhodенých, píšucich, pracujúcich chlapcoch</i>
7	inštrumentál	<i>s kresleným, videným, vyhodенým, píšucim, pracujúcim chlapcom</i>
o	neurčený	<i>Videné (mesto)! (N alebo A sg.)</i>

Pozícia 6: Stupeň

Znak	Hodnota	Príklad
x	pozitív	<i>kresleného, videného, vyhodенého, píšuceho, pracujúceho chlapca</i>
y	komparatív	<i>presvedčujúcejší dôvod</i>
z	superlatív	<i>najpresvedčujúcejší dôvod</i>

4.7. Adverbiá

Pozícia 1: Adverbium – D

Pozícia 2: Stupeň

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

4.8. Prepozície

Pozícia 1: Prepozícia – E

Pozícia 2: Forma

Znak	Hodnota	Príklad
v	vokalizovaná	<i>so mnou, zo seba, odo dverí, podo mnou</i>
u	nevokalizovaná	<i>s nami, z neho, od okna, pod vrchom, medzi mnou, pomocou neho</i>

Pozícia 3: Spojenie s pádom

Znak	Hodnota	Príklad
2	genitív	<i>od teba, do teba</i>
3	datív	<i>ku kozubu</i>
4	akuzatív	<i>za koho, na neho</i>
6	lokál	<i>o kreslených, na večierku</i>
7	inštrumentál	<i>medzi nami, s kreslením</i>
o	neurčený	

4.9. Konjunkcie

Pozícia 1: Konjunkcia – O

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

Pozícia 2: Kondicionálnosť – Y

– pri spojkách *keby, aby, žeby, akoby, s'aby* vyjadruje, že ide o spojenie spojky a kondicionálovej morfémy *by*

4.10. Partikuly

Pozícia 1: Partikula – T

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

Pozícia 2: Kondicionálnosť – Y

– pri časticiach *keby, žeby...* vyjadruje, že ide o spojenie spojky a kondicionálovej morfémy *by*

4.11. Interjekcie

Pozícia 1: Interjekcia – J

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

4.12. Reflexívne sa/si

Pozícia 1: Reflexívum – R

sa, si

4.13. Kondicionálová morféma by

Pozícia 1: Y

by

4.14. Abreviácie a značky

Pozícia 1: W

l, km, H₂O, X569847, atď., tzv., t. j., pod., kt., i., XML, SND

Ďalšie pozície podľa potreby (rod, číslo, pád a pod.).

4.15. Interpunkcia

Pozícia 1: Z

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

4.16. Neurčitelný slovní druh

Pozícia 1: Q

Napr.: až60, po (slovensky), (3-)krát, česko(-slovenský), New (York), (Rio) de (Janeiro)

4.17. Neslovný element

Pozícia 1:

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

4.18. Citátový výraz

Pozícia 1: %

Napr.: *Take it easy!*; *Šaj pes dovakeras*; „správně vychlazená dvanáctka“; *Ta naše povaha česká; náměstí*

4.19. Číslice

Pozícia 1: 0

6, 1 224, 86, 2000, 3(-krát)