

ERRKORP – anotačný manuál

upravená pilotná verzia (10/2020)

1. ÚVOD

Pri tvorbe tagsetu vnútornej anotácie sme sa inšpirovali prácami S. Pita Cordera (1981), C. Jamesa (1998), odbornými štúdiami českých autorov K. Šebestu a S. Škodovej (2012), B. Štindlovej (2013) a anotačným tagsetom korpusu CzeSL-SGT-cs.

Pri spracúvaní jazykového materiálu sme zvyčajne uplatnili kombináciu dvoch anotačných prístupov: povrchovú typológiu chýb a funkčnú typológiu chýb.

a) V rámci **povrchovej typológie chýb** sa ako relevantné ukazujú tri možnosti:

- **vynechanie** alebo **absencia** (*omission*),
- **pridanie** (*addition*),
- **nahradenie** alebo **substitúcia** (*substitution*).

Na základe tohto delenia používame tam, kde je to relevantné, tri druhy značiek: 0, 1 a subst. Napríklad ak je v texte slovo, ktoré je redundantné (**súhlasím sa – súhlasím*), použijeme na jeho označenie značku word1. Ak isté slovo chýba (**opýtal ma – opýtal sa ma*), uplatníme značku word0, a ak študent v texte použil nevhodnú lexému s iným významom (**idem do lekára – idem k lekárovi*), máme k dispozícii značku substword. Niektoré potenciálne možné prípady chýb však do tagsetu nie sú zaradené vzhľadom na nízku textovú frekvenciu – napr. máme síce značky comma0 a comma1, no už nie značku substcomma, ktorá sa pri testovaní anotovaní vôbec nevyskytla.

b) **Funkčná typológia** spočíva v začlenení chyby spravidla podľa funkčnolingvistickej klasifikácie:

- či je chyba na úrovni jedného segmentu = týka sa jednej grafémy, diakritického znamienka alebo aj interpunkčného znamienka,
- na úrovni časti slova – takto identifikovateľné chyby sa vyskytujú najčastejšie v prípade gramatickej relačnej morfémy alebo derivačnej morfémy (prefixálnej alebo sufixálnej) alebo v prípade kmeňa,
- na úrovni slova a jeho začlenenia do vety – ide najmä o chyby týkajúce sa syntaktickej, lexikálno-sémantickej a štylistickej roviny jazyka (zlá kongruencia, zlá valencia, zlý slovosled enklitik, ale aj nesprávne použitie ustálených fráz a frazeologizmov),
- na úrovni textu – takáto chyba presahuje rámec vety – ide o chyby textotvornej povahy.

Rozlišujeme dva spôsoby značkovania chýb: 1) jednoduché (týkajúce sa jedného slova/tokenu) a 2) sekvenčné (týkajúce sa spojenia aspoň dvoch slov/tokenov) - tie zapisujeme pomocou zavináčov (viac nižšie).

1. Chyby na úrovni jedného segmentu v rámci jedného slova

Chyby na úrovni jedného znaku (grafémy, diakritických a interpunkčných znamienok) nepresahujú rozsah jedného slova (jedného gramatického tvaru) – ide o vypadnutie, nadbytočné použitie alebo o zámenu znaku alebo diakritického či interpunkčného znamienka. Nepatria sem morfosyntaktické ani štylistické chyby. Chyby opravujeme: 1) s ohľadom na zmysel textu a na vystihnutie pravdepodobného zámeru autora a 2) a s ohľadom na vykonanie minimálneho zásahu (v texte pri oprave robíme čo najmenej formálnych úprav).

Detailnejšie vymedzenie jednotlivých značiek je opísané v kapitole 3.

Značka	Vysvetlenie	Príklady
char0	znak chýba	prtože, s študentom
char1	znak navyše	súčasne, pekingský
charmata	prehodenie znakov v rámci jedného slova	Sbrsko, horovím
substvow	zámena vokálu (netýka sa zámen vokálov v gramatických relačných morfénoch)	meso, piatok, piesnička
substcons	zámena konsonantu (okrem konsonantov v gramatických relačných morfénoch)	lyba (ryba), yazyk (jazyk), chudba (hudba), Atlantic
substdiph	zámena diftongu (netýka sa zámen diftongov v gramatických relačných morfénoch)	pátek („piatok“), kvoň („kôň“), deťa („dieťa“), nevim („neviem“), o osmej
cap0	veľké písmeno chýba	ministerstvo financií SR
cap1	veľké písmeno navyše	Minister financií, Európska Únia

quantbase0	kvantita chýba v tvarotvornom základe slova	krasny, stretavali, vstavali, prinos, vystavba,
quantbase1	kvantita navyše v tvarotvornom základe slova	výsoký, hrád, prénos, výhodiť, predstavujem
quantsuf0	kvantita chýba v gramatickej relačnej morféme (substantíva, adjektíva...) a v tematickej morféme (slovesá)	novy, <u>tuto</u> osobu, skonči, teším, trva, robim, čakam, sadam, musim, konči, stojim, urobim, konči, učim, vola
quantsuf1	kvantita navyše v gramatickej relačnej morféme (substantíva, adjektíva...) a v tematickej morféme (slovesá)	krásny, budová, dám <u>svojím</u> kamarátkam, spáli sme, predstavujem
y1	zámena i – y	medzynárodný, boly tu
y0	zámena y – i	sir, midlo
substchar	zámena znakov	jedlo. ktoré sa volá
caron0	mäkčeň chýba	student, vecera,
caron1	mäkčeň navyše	več, ďesať, dieťa, ťžká
comma0	čiarka chýba	Ahoj Eva; Nebol tu tak neviem
comma1	čiarka navyše	V prípade zlého počasia, pôjdeme do múzea
dot0	bodka chýba	13 januára 2016
dot1	bodka navyše	v roku 2014. tu postavili
hyph0	spojovník chýba	krížomkrážom

hyph1	spojovník navyše	rímsko-katolícky
defdiacr	tri a viac chýb v diakritike ("chatové písanie")	paňnat, dolezity
defword	tri a viac zámen znakov v slove s výnimkou chybné diakritiky + slová z jazyka L1	samuslina (zmrzlina), bybarený (vybavený), nemám novac (peniaze), mám house (dom)
space	medzera medzi slovami je redundantná alebo chýba	na príklad (napríklad), Akosa voláš (Ako sa voláš)

2. Chyby na úrovni jednej morfémy alebo jedného slova

Chyby presahujú rozsah jedného znaku a nepresahujú úroveň slova. Najčastejšie ide o chyby týkajúce sa gramatickej, lexikálno-sémantickej a štylistickej roviny jazyka.

Značka	Vysvetlenie	Príklady
word0	slovo chýba	opýtal (sa) ma, ja (som) začala študovať
word1	slovo navyše	píšem s perom, ja sa s tebou súhlasím
substword	zámena celého slova	tajná vôňa (tajomná), prezentovať parfum (darovať)
morph	chyba v tematickej morféme verb + nešpecifická chyba v gram. morféme	sadnul si, pozriť (pozrieť), z galérii (z galérie), staršieho (staršieho), všetkých (všetkých), z izboj (z izby), študujem slovenčin (slovenčinu), v roke (v roku)
substderiv	zámena derivačnej alebo modifikačnej morfémy	historické (historické), Británska (Britská), zvariť obed (uvariť)

num	zámena čísla	koláč s ovociami, <u>ryže</u> sú na tanieri
gend	zámena rodu	v esejoch, z teritórie
asp	chyba v aspekte	budem prísť (prídem), budem dokončiť (dokončím), <u>vydanie</u> literárnych diel (vhodnejšie vydávanie), idem jesť, čo mama <u>pripravovala</u> (vhodnejšie pripravila), zaujíma ma, ktoré šaty si <u>oblieka</u> (vhodnejšie oblečie) zajtra
temp	chyba v čase	zavolať, že o dva dni musel (musí) odísť
styl	chyba v štylistickom použití slova	na konferencii papkáme (jeme) obed o jednej

3. Chyby na úrovni slovného spojenia v rámci jednej vety

Ide o chyby v spojení slov pri výstavbe syntagiem - predovšetkým chyby morfosyntaktické alebo štylistické. Pri takýchto chybách, ktoré presahujú úroveň jedného slova, zapisujeme začiatok a koniec úseku, v ktorom sa chyba nachádza – robíme to napríklad pri značkách **space, order, phrase**: pred chybovú sekvenciu dávame značku @@@ + typ chyby (napríklad značka pre chybu v slovoslede je @@@order), potom nasleduje značka @@@*, za ktorú vpíšeme opravu sekvencie, a napokon zapíšeme značku *@@@, ktorá opravenú sekvenciu ukončuje.

Značka	Vysvetlenie	Príklady
congr	chyby v syntaktickej zhode	ten centrum, zaujímavý esej, veľa ľudí boli, päť študentov majú
dep	chyby v syntaktickej závislosti (dependencii)	hľadám sestra, káva s mlieko, deväť študenti, človek, ktorý tu čakám
neg	nesprávne vyjadrenie negácie	nie budem, nikto prišiel, nie som bol
order	porušenie rytmického a gramatického princípu slovosledu	tu v minulosti nachádzalo sa veľa parkov
phrase	chyba v spojeniach, ustálených spojeniach a frazémach	nemá rozprávania (niet o čom)

4. Chyby na úrovni textu

Ide o chyby, ktoré vznikajú pri budovaní textu a presahujú úroveň jednej vety. Týkajú sa nesprávneho použitia textových konektorov (slová odkazujúce na predchádzajúcu vetu), a sémantického princípu slovosledu (aktuálneho vetného členenia). Tieto chyby sú nadslovné, preto treba zapisovať začiatok a koniec úseku, v ktorom sa chyba nachádza. Pokiaľ je tento úsek príliš dlhý, sekvenciu nevyznačujeme.

Značka	Vysvetlenie	Príklady
theme	chyba v aktuálnom členení výpovede	Mám malú izbu. Len posteľ, dve skrine, kreslo a písací stolík sú v izbe. (Mám malú izbu. V izbe je len posteľ, dve skrine, kreslo a písací stolík.)
connect	chyba v textovom konektore (nadvetná syntax)	V texte sa zaoberáme slovenčinou. Jeho (jej používanie v rôznych situáciách je ovplyvnené...

2. ZÁSADY ANOTOVANIA

1) Anotovanie textov robíme "vertikálne" = **v exceli v 3 stĺpcoch** – v 1. stĺpci je originálny text študenta, v 2. stĺpci je oprava (korektné tvary/slová, doplnené slová), v 3. stĺpci sú príslušné chybové značky, napr.

Videlo	Videla	congr
som		
octov	otcov	charmeta
pohľad		
.		

2) Pri opravovaní textov a značkování chýb sa riadime týmito princípmi:

- opravujeme s ohľadom na zámer autorky/autora a na predpokladaný kontext vzniku textu = **princíp rešpektovania kontextu**
nemám rad halušky – nemám rád halušky (neopravíme na *rada*, ak vieme, že autorom textu je muž)
- opravujeme tak, aby vznikla gramaticky a sémanticky korektná veta, ktorá však nemusí byť štylisticky perfektná = **princíp gramaticko-sémantickej korektnosti**
každý človek môže nosiť každú farbu; počas minulých mesiacov; teraz študujem na UK a to je môj prvý pobyt v Bratislave.
= necizelujeme, neopravujeme štylistiku až takto: *každý človek môže nosiť akúkoľvek/hocijakú farbu; počas uplynulých mesiacov; teraz študujem na UK a toto je môj prvý pobyt v Bratislave.*
- riadime sa morfosyntakticky nadradeným slovom, t. zn. opravujeme podradené slovo = **princíp dominancie**
hovorili o študentoch a o študentkami – hovorili o študentoch a o študentkách (z kontextu je zrejmé, že študent chce povedať spojenie “hovorí o niekom/niečom”, preto neopravíme na *hovorili o študentoch a so študentkami*)
 - riadime sa kontextom, ak je jasné, čo chcel autor povedať, a nie morfosyntakticky nadradeným slovom, t. zn. opravujeme nadradené slovo:
idem na autobusovú zastávku (idem na autobusovú zastávku)
- opravujeme tak, aby vzniklo podľa možnosti čo najmenej zásahov do textu = **princíp minimálnej opravy**
aby som sa trochu oddychovala – aby som trochu oddychovala (neopravíme na *aby som si trochu oddýchla*, lebo v prvom prípade urobíme jednu opravu – word1, v druhom prípade viac opráv)

- 3) Ak sú v jednom slove **dve až tri chyby**, zapíšeme značky **vedľa seba do jednej bunky v 3. stĺpci a oddelíme ich čiarkou BEZ medzery**, napr.

ranó	ráno	quantbase0,quantsuf1
------	------	----------------------

- 4) Ak sú v jednom slove tri a viac chýb, označujeme slovo značkou defword, napr.

bybarena	vybavená	defword
----------	----------	---------

- 5) **Ak isté slovo alebo istý znak** (patrí sem aj interpunkcia, úvodzovky atď.) v texte študenta **chýbajú, doplníme na príslušné miesto v 1. stĺpci NONE** (vygenerujeme nový riadok v exceli a vpíšeme tam slovo NONE), do 2. stĺpca doplníme chýbajúce slovo a do 3. stĺpca príslušnú značku (napr. word0, comma0), napr.

Opýtal		
NONE	sa	word0
ma		
na		
cestu		
.		

- 6) **Ak chýba viac slov, ktoré vieme identifikovať, každé musí mať svoj doplnený riadok so slovom NONE**, napr.

Opýtal		
sa		
,		
prečo		
NONE	by	word0
NONE	som	word0
neprišla		
,		
keď		
by		
NONE	som	word0
mohla		

.		
---	--	--

- 7) Ak nejaké slovo v texte študenta chýba a nevieme ho jednoznačne doplniť, uvádzame do 2. stĺpca značku **unk** (od angl. unknown) – v tomto prípade môže byť totiž viacero možností doplnenia – *pani/kolegyňa/pani XY/pani docentka...*, napr.

Vážená		
NONE	unk	word0
,		
obraciam		
sa		
na		
Vás		

- 8) V prípade chyby nad rámec jedného slova označíme začiatok a koniec sekvencie, ktorá je chybná. Začiatok sekvencie označíme **@@@**, za ňou za značku **@@@*** umiestnime korektnú podobu sekvencie a uzavrieme to značkou ***@@@** – za ňu umiestnime značku chyby, napr.

@@@order		
Som		
včera		
všimla		
si		
@@@*		
Všimla		
som		
si		
včera		
*@@@		
pohľadnicu		
.		

- 9) Ak sú v istej sekvencii **zároveň** chyby v jednotlivých slovách, značia sa do príslušných stĺpcov zvyčajným spôsobom, napr.:

@@@order		
Som		
vcera	včera	caron0
všimla	všimla	quantbase1
si		
@@@*		
Všimla		
som		
si		
včera		
*@@@		
pohľadnicu		

- 10) Ak sú slová zapísané nesprávne ako 1 slovo (t. zn. správne majú byť v texte 2 slová a viac), použijeme zápis pomocou zavináčov ako pri nadslovnej chybe, napr.

@@@space		
Akosa		
@@@*		
Ako		
sa		
*@@@		
hovorí		
na		
Slovensku		
.		

- 11) Ak sú slová zapísané nesprávne ako 1 slovo a zároveň je jedno z nich použité nesprávne, použijeme zápis pomocou zavináčov ako pri nadslovnej chybe a zároveň príslušnú značku, napr.

@@@space		
umivamsa	umývam	quantbase0,y0
@@@*		
umývam		
sa		
*@@@		

@@@space		
----------	--	--

Zoberamsa	Zbalím	defword
@@@*		
Zbalím		
sa		
*@@@		

- 12) V prípade, že je **slovo zapísané oddelene, ale správne má byť zapísané ako jedno slovo, použijeme nadslovný zápis, napr.**

Hovoril		
@@@space		
do		
nekonečna		
@@@*		
donekonečna		
*@@@		
nezmysly		
.		

- 13) Ako samostatné vymedzujeme pri anotovaní **chyby, ktoré opravujeme v dôsledku inej vykonanej opravy – sekundárne chyby**. Sú to sekundárne chyby v kongruencii a v dependencii, teda podľa pôvodnej kongruencie alebo dependencie sú tieto gramatické tvary správne, napr.

ideme		
do	na	substword_dep
fakulty	fakultu	secdep

ideme		
na		
filozofickej	filozofickú	seccongr
fakulte	fakultu	dep

- 14) **Chyby, ktoré sú prienikom dvoch javov/rovín, značujeme dvoma značkami:**

Peter		
,		
kde		
NONE	si	word0,congr
bol		
?		

- 15) **Chyby neopravujeme v prípadoch, ak nie je z predchádzajúceho kontextu zrejmé, ako majú byť opravené:**

v		
tom		
zmysle		
treba		
NEČITATEĽNÉ	unk	defword
produkcia		
a		
priemysel		

Ďalšie príklady anotovania:

kvôli		
citam	citom	gend

dlho		
očavané	očakávané	char0,char0

mám		
rada		
cestovať	cestovanie	substword

dám		
tebe	ti	substword
pokoj		

ide		
s	so	char0
študentmi		

plaká	plače	substcons,morph
-------	-------	-----------------

veľa		
kavyaren	kaviarní	substdiph,dep

veľa		
krajín	krajín	quantbase0

v		
mojej		
obycej	obývacej	char0,char0,quantbase0
izbe		

3. VYMEDZENIE CHYBOVÝCH ZNAČIEK

char0

Vypadnuté je akékoľvek písmeno (konsonant, vokál, diftong):

prtože (pretože), vyší (vyšší),

Zaraďujeme sem aj prípady, ak je použitá nevokalizovaná prepozícia namiesto vokalizovanej:

s študentom (so študentom), k mne (ku mne)

NIE:

- V prípade chýbajúcej i-ovej časti diftongu (patok namiesto *piatok*) ide o **substdiph**.
- Prípady ako *Čakám kamarát* (namiesto *kamaráta*), *Chcem dva čaj* (namiesto *čaje*) značkujeme **dep**, keďže ide o zlú väzbu.
- Prípady ako *Ona bol* (*Ona bola*), *Mám rád* (*Mám rada*), *Volám sa Alexandra a má 36 rokov* (*Volám sa Alexandra a mám 36 rokov*) označujeme ako **congr**.

char1

Isté písmeno je redundantné:

súčasne (súčasne), rozrástili sa (rozrástli sa), zamrzlina (zmrzlina), exkurziju (exkurziu), mikrovlnku (mikrovlnku), na oddelení (na oddelení)

Zaraďujeme sem aj prípady, ak je použitá vokalizovaná prepozícia namiesto nevokalizovanej:

ku tomu (k tomu), torta so jahodom (torta s jahodami)

NIE:

- V prípade ako *Vidím domu* (namiesto *Vidím dom*), *Nemám počítača* (namiesto *Nemám počítač*) použijeme značku **dep**, keďže ide o zlú väzbu.
- Prípady ako *On bola* (namiesto *On bol*), *Mám rada* (namiesto *Mám rád*, ak je autorom muž) označujeme ako **congr**.

charmata

Ide o vzájomnú výmenu dvoch znakov v rámci toho istého slova:

Sbrsko (Srbsko), fotografei (fotografie), ked (kde), horovím (hovorím)

NIE:

Nasledujúci príklad značkujeme ako **substword**:

do roku 2016 do roku 2018 (od roku 2016 do roku 2018)

substvow

Ide o prípady, keď je namiesto vokálu nesprávne použitý diftong alebo iný vokál:

meso (mäso), piatok (piatok), piesnička (pesnička), diesať (desať)

NIE:

V prípade zámeny vokálov v gramatickej relačnej morféme, ide o chybu **morph**:

v roka (v roku), pozvam (pozvem)

substcons

Ide o prípady, keď je namiesto istého konsonantu nesprávne použitá iná graféma (zvyčajne iný konsonant):

lyba (ryba), yazyk (jazyk), chudba (hudba), Atlantic (Atlantik)

NIE:

V prípade zámeny konsonantov v gramatickej relačnej morféme, ide o chybu **morph**:

práva občanou (občanov), v mesiacoch (mesiacoch)

substdiph

Ide o prípady, keď je namiesto diftongu nesprávne použitý vokál alebo skupina segmentov.

pátek (piatok), kvoň (kôň), djeťa (dieťa), d'jeťa (dieťa)

y0

Ide o chybu, keď je namiesto ypsilonu použité i (mal byť použitý ypsilon, no je i):

sir (syr), vidatá (vydatá), umivať (umývať)

y1

Ide o chybu, keď je ypsilon použitý nesprávne = použitý je ypsilon namiesto i:

knyžnica (knižnica), vytamín

quantbase0

Tvarotvorným základom v anotačnom manuáli chápeme tú zložku slova/gramatického tvaru, ktorá ostane po odčlenení gramatickej relačnej morfémy a aj tematickej morfémy. Quantbase0 označuje chýbajúcu dĺžku v tvarotvornom základe slova:

krasny (krásny), pisal som (písal som), kupim (kúpim)

NIE:

Zle použitá kvantita v tematickej morféme slovesa patrí k chybe v **quantsuf0**:

priznam (priznám), chodim (chodím)

quantbase1

Quantbase1 označuje nadbytočnú dĺžku v tvarotvornom základe slova:

predstavujem, výsoký (vysoký), mesto (mesto), vólam (volám)

V prípade neohybných slovných druhov je celé slovo báza:

veľmí (veľmi), hovorím po nemecký (nemecky)

NIE:

Nadbytočná kvantita v tematickej morféme slovesa patrí k chybe v **quantsuf1**:

kúpím (kúpim), čítám (čítam)

quantsuf0

Quantsuf0 označuje chýbajúcu dĺžku v gramatickej relačnej morféme gramatického tvaru.

Rovnako sem patrí aj chýbajúca kvantita v tematickej morféme slovesa:

novy (nový), tuto osobu (túto osobu), skonči (skončí), učim sa (učím sa), vola (volá)

quantsuf1

Quantsuf1 označuje nadbytočnú dĺžku v gramatickej relačnej morféme gramatického tvaru. Ide

aj o zle použitú (nadbytočnú) kvantitu v tematickej morféme slovesa:

krásny (krásny), rýchlého (rýchleho), kúpím (kúpim), spál som (spal som)

caron0

Ide o prípady chýbajúceho mäkčeňa:

student (šstudent), veceral som (večeral som)

caron1

Ide o prípady redundantného výskytu mäkčeňa:

učiteľia (učitelia), bola som u Viední (bola som vo Viedni),

comma0

Prípady chýbajúcej čiarky:

Ahoj Jana (Ahoj, Jana); Nebol som tu preto neviem (Nebol som tu, preto neviem)

comma1

Prípady nadbytočnej čiarky:

V prípade zlého počasia, pôjdeme do múzea (V prípade zlého počasia pôjdeme do múzea)

dot0

Ide o prípady absentujúcej bodky:

13 januára 2016 (13. januára 2016)

dot1

Ide o prípady redundantnej bodky:

v roku 2014. prišiel (v roku 204 prišiel)

hyph0

Prípady chýbajúcej pomlčky alebo chýbajúceho spojovníka:

krížomkrážom (krížom-krážom)

hyph1

Prípady redundantnej pomlčky alebo redundantného spojovníka:

rímsko-katolícky (rímskokatolícky)

cap0

Táto značka sa používa v prípadoch:

- ak je pri propriálnom názve použité malé písmeno namiesto veľkého.
ministerstvo financií SR (Ministerstvo financií SR)
- ak sú použité malé písmená v prípade abreviátov: *v Sľuk-u (v SĽUK-u)*

cap1

Táto značka sa používa v prípadoch:

- ak je pri apelatívnom názve nevhodne použité veľké písmeno:
bol to Britský maklér (bol to britský maklér), na Užhorodskom Národnom Universitita (na Užhorodskej národnej univerzite), PortoRika (Portorika)

defdiacr

V prípade, ak sú tri alebo viaceré chyby v diakritike (ak je zápis viacerých grafém v jednom slove nesprávny), no slovo je zrozumiteľné (zvyčajne "chatové/eseemeskové písanie" bez diakritiky):

paťnast (pätnásť), dolezity (dôležitý)

defword

V prípade, ak sú aspoň tri chyby v slove, takže slovo samo osebe je neidentifikovateľné. Význam slova vieme zvyčajne dešifrovať z kontextu vety alebo na základe poznania jazyka L1 (iné chyby už netagujeme):

samuslina (zmrzlina), dni sa dlhšajú (dni sa predlžujú), posdno (neskoro), makši (mäkši)

word0

Touto značkou tagujeme prípady, ak:

- chýba celé slovo (plnovýznamové slovo),
- chýba pronomen pri reflexívach tantum,
- chýba pomocné verbum pri tvaroch préterita a kondicionálu.

Ak v prípade chýbajúceho plnovýznamového slova nevieme v 2. stĺpci uviesť jediné vhodné doplnenie, vpisujeme **unk** ("unknown").

NIE:

Nie je potrebné opravovať napr. prípady, ktoré sú aj v SJ variantné:

myslím, že som inteligentná - neopravujeme na *myslím si, že som inteligentná*

word1

Značku používame, ak je isté slovo navyše. Ide o prípady, ak je nevhodne (nadbytočne) použité:

- plnovýznamové slovo,
- reflexívne pronomen pri verbách,
- pomocné verbum pri tvaroch préterita a kondicionálu.

Nadužívanie osobných pronomín v pozícii subjektu neopravujeme.

Ja		
som		
študent		
vo		
Varšave		

substword

Značku používame v prípadoch, ak je zamenené slovo za iné slovo. Zvyčajne ide o zámenu významu slova:

som Čína (som Čiňanka), izba je neveľká alebo útulná (izba je neveľká, ale útulná)

Pri opravách minimalizujeme štylistické korekcie.

Patria sem aj prípady, keď ide o zámenu v slovnodruhovej príslušnosti slov:

starý		
otec		
je		
zo		
slovensky	Slovenska	substword

Značku používame aj pri zámene prepozícií v dôsledku zlej dependencie, vtedy ju pri anotovaní spájame so značkou **dep** – **substword_dep**:

idem do fakulty (idem na fakultu)

ideme		
do	na	substword_dep
fakulty	fakultu	secdep

Musím		
sa		
NONE	o	word0_dep
ju	ňu	substword, <u>secdep</u>
deliť		

Dvoma spôsobmi – v závislosti od kontextu – môžeme opraviť nesprávne spojenie typu **mám rada cestovať**:

@@@phrase		
mám		
rada		
cestovať		
@@@*		
rada		
cestujem		
*@@@		

mám		
rada		
cestovať	cestovanie	substword

Treba zvážiť kontext - či ide o vyjadrenie aktívneho vykonávania obľúbenej aktivity alebo o vyjadrenie pozitívneho vzťahu k nejakej aktivite. Sémantický rozdiel je medzi opravami spojenia napr. *mám rád športovať*:

1) rád športujem, 2) mám rád športovanie/šport.

NIE

Neopravujeme prípady používania osobných prívlastňovacích pronomín namiesto zvratných prívlastňovacích pronomín!

dám		
to		
mojej		
kamarátke		

substderiv

Značku používame v prípade zámeny derivačnej morfémy (zlý slovotvorný prostriedok) alebo v prípade modifikačnej morfémy pri aspekte:

nástolná lampa (stolová lampa), sezónové zľavy (sezónne zľavy), odrelaxovať sa (zrelaxovať sa)

gend

Značku používame, ak je nevhodne použitý rod. Používame ju **len pri substantívach**, ktoré sú použité v „korektnom“ páde, ale v nesprávnom rode (t. zn., že ak by toto substantívum malo iný rod, uvedený tvar by bol správny, napr. ak by sme mali *tá teritória*, tak tvar genitívu singuláru z *teritórie* by bol korektný), napr.:

z teritórie (z teritória), v esejoch (v esejach)

NIE:

Nepoužívame pri adjektívach (resp. adjektíváliách), pretože tie nemajú kategóriu rodu – pri nich použijeme značku **seccongr**:

z tej zvláštnej teritórie značujeme takto:

z		
tej	toho	seccongr
zvláštnej	zvláštneho	seccongr
teritórie	teritória	gend

Nepatria sem ani prípady nevhodnej kongruenčnej morfémy pri I-participiách (tvaroch préterita alebo kondicionálu) – v tomto prípade ide o **congr**, napr.:

Eva		
bol	bola	congr
už		
doma		

num

VLOŽENÉ: Označujeme len pri tých slovných druhoch, ktoré kategóriu čísla majú = pri substantívach. Ide o prípady, ak je použité substantívum v nesprávnom čísle, ide o:

- príznakové použitie singuláru alebo plurálu: *po toľkých časoch (po toľkom čase)*
- tvaru singuláru v prípade plurálií tantum: *nemám peniaz (nemám peniaze)*
- tvaru plurálu v prípade singulárií tantum (pri materiáliách, abstraktách, kolektívach) v primárnom význame: *nemám rád ovocia ani zeleniny (nemám rád ovocie ani zeleninu)*

V prípade adjektív, ktoré dané substantíva rozvíjajú, nejde o **num**, ale o chybu v sekundárnej kongruencii:

po		
toľkých	toľkom	seccongr
časoch	čase	num,morph

V prípade, ak je tvar singuláru v prípade vyjadrenia kvantity 2, 3, 4 - napr. *chcem dva čaj (chcem dva čaje)*.

Chcem		
dva		
čaj	čaje	num

Kedže v praxi v niektorých kontextoch preferujeme v SJ skôr formu singuláru alebo skôr formu plurálu (treba overiť v Slovenskom národnom korpuse), hoci existuje úplná paradigma - treba veľmi citlivo zvážiť, či ide o chybu:

*Bluetooth ako prepojovacia technológia je spôsobilá na prenos dôverných informácií
ale: prudký rozvoj vedy a technologie (technológií)*

asp

Ide o nesprávne použitie kategórie aspektu, napr.:

budem prísť (prídem), budeme napísať (napíšeme)

budem	NONE	word1,asp
skončiť	skončím	morph

NIE

V prípade opozície determinatívne vs. indeterminatívne verbum ide o **substword**:

každý človek môže niest' hocijakú farbu – každý človek môže nosiť hocijakú farbu

temp

Ide o prípady, ak je zrejmé, že mala byť použitá iná forma času, napr.:

zavolať, že o dva dni chcel prísť (chce prísť)

V prípade analytických tvarov použijeme sekvenčnú značku, napr.:

dúfam		
,		
že		
vždy		
@@@temp		
žijeme		
@@@*		
budeme		
žiť		
*@@@		
spolu		

morph

Touto značkou označujeme:

- chyby v tematických morfédoch pri verbách:
spám (spím), rozumiel som (rozumel som), sadnul by si (sadol by si)
- iné, nešpecifické chyby v použití gramatických morfédoch, t. j. ak ich nevieme zaradiť napr. k chybám v dependencii, kongruencii, čísle a pod.:
na Slovenske (na Slovensku), v roke (v roku), v mestu (v meste), študujem slovenčin (slovenčinu), o nezodpovednom správaní (správaní), veľa slovov (slov), s rodičami (s rodičmi)

V prípade neistoty pri určovaní toho, čo je tematická morféma, je vhodné overiť si morfematické členenie v morfematickom slovníku.

styl

Značku používame, ak je použité slovo, ktoré síce existuje, no je nevhodne štylisticky použité, napr.:

na konferencii papkáme obed o jednej (na konferencii jeme obed o jednej)

congr

Značka sa týka chýb v realizácii morfosyntaktickej kategórie zhody. Ide o nasledujúce prípady:

- chyby v zhodnom atribúte: *ten centrum (to centrum), tento vec (táto vec), nový esej (nová esej), dlhý báseň (dlhá báseň), dve náhrdelníky (dva náhrdelníky), jedne tričko (jedno tričko), hlavné mesto Bratislava, ktorá (hlavné mesto Bratislava, ktoré), v tejto oblečení (v tomto oblečení), o spisovateľovi Dante Alighieri (o spisovateľovi Dantom Alighierim)*
- chyby v rode a čísle v tvaroch slovesného predikátu: *žena bol (žena bola), dievča bola (dievča bolo), čas je peniaze (čas sú peniaze), prišli niekoľko detí a ich matka (prišlo niekoľko detí a ich matka), som Alexandra a má 36 rokov (som Alexandra a mám 36 rokov), ja nerozumieť (ja nerozumiem), rozprávajú, čo radi robiť (rozprávajú, čo radi robia), na konci bol aj dlhá diskusia (na konci bola aj dlhá diskusia), budem študovám (budem študovať), musím robím (musím robiť)*
- chyby v rode a čísle v mennej časti predikátu: *báseň je pekný (báseň je pekná), domy sú veľkí (domy sú veľké)*
- chyby v zhode predikátu so subjektom vety v prípade, že obsahujú kvantitatívne výrazy *päť* a *viac*, *veľa*, *málo*, *dost* a podobne: *päť študentov majú (päť študentov má), veľa ľudí boli (veľa ľudí bolo), niekoľko budov boli postavené (niekoľko budov bolo postavených)*
- chyby v zhode doplnku so subjektom alebo s objektom vety: *rodičia prišli unavený (rodičia prišli unavení)*
- chyby vo vzťažnom zámene vo vedľajšej vete atributívnej: *muž, ktorí tu už bol (muž, ktorý tu už bol), mali by tam byť produkty, aký potrebujem (mali by tam byť produkty, aké potrebujem)*

NIE:

K chybám v kongruencii nepatria prípady, ak je menná časť slovesno-menného predikátu v nesprávnom páde, v tomto prípade ide o **dep**, napr.:

Ona je učiteľky (Ona je učiteľka)

seccongr

Značku používame v prípadoch, ak ide o sekundárnu chybu v zhode (kongruencii). Používame ju vtedy, ak je isté substantívum v zlom tvare a potrebná je sekundárne tiež oprava atribútu, ktorý dané substantívum rozvíja: *túžiť po vysoké* → **seccongr** *ciele* → **dep** (*túžiť po vysokých cieľoch*), *mám jedna* → **seccongr** *sestra* → **dep** (*mám jednu sestru*),

Sekundárnu chybu v kongruencii značujeme takto:

študujem		
na		
Užhorodskom	Užhorodskej	caron0,seccongr
Narodnom	národnej	cap1,quantbase0,seccongr
Universitita	univerzite	defword,dep

NIE:

Značku seccongr nepoužívame v prípade, ak je nadradené substantívum v zlom páde (chyba v dependencii), no chyba bola aj v samotnom atribúte v rámci atributívneho spojenia – vtedy ide o chybu **congr**, napr.:

boli sme tam v tým istým rok značujeme takto:

boli sme tam v tým → **congr** *istým* → **congr** *rok* → **dep** (*boli sme tam v tom istom roku*)

dep

Ide o chyby v syntaktickej závislosti inej ako kongruencia. Ide o chyby vo vyjadrení pádovej determinácie podradeného vetného člena vo vzťahu k nadradenému vetnému členu:

- chyby vo valencii slovesa: *čakám kamarát* (*čakám kamaráta*), *nemám brat* (*nemám brata*), *zaujímam sa o politike* (*zaujímam sa o politiku*), *idem na fakulte* (*idem na fakultu*), *bola som v kaviarne* (*v kaviarni*), *išiel som do doktora* (*išiel som k doktorovi*), *je to použité pre reklamy* (*je to použité v reklame*), *idey týkajúce sa animované filmy* (*idey týkajúce sa animovaných filmov*)
- chyby v predložkových väzbách, napr. *chlieb s maslo* (*chlieb s maslom*), *cestovanie v Slovensku* (*cestovanie po/na Slovensku*)
- prípady zlého použitia substantíva po číslovkách 5 a viac a po iných výrazoch kvantity, napr. *osem študenti má* (*osem študentov má*), *veľa ľudia robí* (*veľa ľudí robí*), *je tam menej auto* (*je tam menej áut*)
- chyby v použití slovných druhov v príslovkových určeníach a vo fundamente: *potom krátke povedal* (*potom krátke povedal*), *kde je tiché a pokoj* (*kde je ticho a pokoj*)
- chyba v pádovej forme subjektu vety (ak je subjekt v inom páde ako nominatív, resp. genitív): *spolužiačka býva na internáte dlhšie* (*spolužiačka býva na internáte dlhšie*)
- chyba v použití podradovacích spojok a v tvare verba po podradovacej spojke: *povedala som jej vrátiť sa* (*povedala som, aby sa vrátili*), *povedala som, aby sa vráti* (*povedala som, aby sa vrátila*)

Príklady anotovania:

idey		
týkajúce		
sa		
animované	animovaných	seccongr
filmy	filmov	dep

všetko		
ostatné		
považujem		
ako	za	substword,dep
vhodným	vhodný	seccongr
darčekom	darček	dep
k		
narodeninám		

Dependencia a špecifické chyby pri prepozíciách:

- a) prepozícia je správna vo vzťahu k nadradenému slovu (verbu/adjektívu/substantívu), ale nasledujúci výraz je v zlom páde vo vzťahu k prepozícii:

túžiť po vysoké ciele, bol to dlhý deň na fakultu (bol to dlhý deň na fakulte), ostal sám v kaviarne (ostal sám v kaviarni), vychádzam z domy, bol som v Michailovcach a Donovaly

anotujeme takto:

túžiť		
po		
vysoké	vysokých	seccongr
ciele	cieľoch	dep

bol		
to		
dlhý		
deň		
na		
fakultu	fakulte	dep

ostal		
sám		
v		
kaviarne	kaviarni	dep

- b) prepozícia je nesprávna vo vzťahu k nadradenému slovu (verbu/adjektívu/substantívu), ale nasledujúci výraz je v dobrom páde vo vzťahu k prepozícii

zaoberať sa o módu (zaoberať sa módou)

anotujeme takto:

VLOŽENÉ PRÍKLADY:

ideme		
do	na	substword_dep
fakulty	fakultu	secdep

cestovali		
sme		
do	na	substword_dep
Slovenska	Slovensko	secdep

- c) prepozícia je nesprávna vo vzťahu k nadradenému slovu (verbu/adjektívu/substantívu) a aj nasledujúci výraz je v zlom páde vo vzťahu k prepozícii:

chodím v Dastor, mám skúsenosť z cestovaníu

chodím		
v	do	substword_dep
Dastor	Dastora	dep

skúsenosť		
z	s	substword_dep
cestovaníu	cestovaním	dep

V prípade opravy prepozície za inú prepozíciu používame teda dvojznačku **substword_dep**.

- d) prepozícia chýba, ale nasledujúci výraz je v dobrom páde vo vzťahu k nadradenému slovu (verbu/adjektívu/substantívu) – hodnotíme to ako chýbajúce slovo a zároveň dependenciu = používame teda dvojznačku **word0_dep**:

zaujíam sa politiku (zaujíam sa o politiku)

anotujeme takto:

zaujíam		
sa		
NONE	o	word0_dep
politiku		

- e) prepozícia chýba a nasledujúci výraz je v zlom páde vo vzťahu k nadradenému slovu (verbu/adjektívu/substantívu) – hodnotíme to ako chýbajúce slovo a zároveň dependenciu = značka **word0_dep**:

pokračoval ďalej svoj príbeh, nahnevane ho pozrel

anotujeme takto:

pokračoval		
ďalej		
NONE	vo	word0_dep
svoj	svojom	seccongr
príbeh	príbehu	secdep

- f) prepozícia je nadbytočná a nasledujúci výraz je v dobrom páde vo vzťahu k nadradenému slovu (verbu/adjektívu/substantívu) = hodnotíme to ako redundantné slovo a zároveň dependenciu = značka **word1_dep**, napr.:

zelená evokuje na prírodu (zelená evokuje prírodu),

anotujeme takto:

zelená		
evokuje		
na	NONE	word1_dep
prírodu		

- g) prepozícia je nadbytočná a nasledujúci výraz je v zlom páde vo vzťahu k nadradenému slovu (verbu/adjektívu/substantívu) = hodnotíme to ako redundantné slovo a zároveň dependenciu = značka **word1_dep**, napr.:

zaoberať sa o módu (zaoberať sa módou)

zaoberať		
sa		
o	NONE	word1_dep
módu	módou	secdep

Dependencia a značkovanie pri výrazoch kvantity

Chyby pri číslovkách môžu byť dvojaké: buď **dep**, alebo **congr**.

Značku **dep** používame v nasledujúcich prípadoch:

- zlá forma substantíva po číslovkách 5 a viac a po kvantitatívnych výrazoch veľa, dosť, málo, niekoľko atď.: *prišlo päť ľudí (prišlo päť ľudí), je tu desať študenti (je tu desať študentov), ostalo tam ešte veľa ľudí (ostalo tam ešte veľa ľudí)*

NIE:

- ak ide o zlú formu číslovky pred substantívom: *dostala som dve náhrdelníky (dostala som dva náhrdelníky), hovoril s päť študentmi (hovoril s piatimi študentmi), kúpila som jedna knihu (kúpila som jednu knihu)* – značujeme ako **congr**
- ak ide o chybu verba vo vzťahu k subjektu vety: *vonku stáli niekoľko ľudí (vonku stálo niekoľko ľudí), veľa ľudí boli chorí (veľa ľudí bolo chorých)* – ide o **congr**

Dependencia a značkovanie pri konjunkciách a vedľajších vetách

Značku **dep** používame v prípadoch, ak si hlavná veta vyžaduje istú konjunkciu, prípadne aj gramatický tvar vo vedľajšej vete (najčastejšie ide o účelovú vetu vyjadrenú formou *aby + l-forma*) alebo ak predikát vo vedľajšej vete „diktuje“ pádovú formu vzťažného zámena

ktorý/aký:

povedala som jej sa vrátiť (povedala som jej, aby sa vrátila), povedala som, aby sa vrátila (povedala som, aby sa vrátila), to je ten muž, ktorý som už videla (to je ten muž, ktorého som už videla)

povedala	povedala	
som	som	
jej	jej	
NONE	,	comma0
NONE	aby	word0,dep
sa		
vrátiť	vrátila	secdep

povedala	povedala	
som	som	
jej	jej	
,		
aby		
sa		
vráti	vrátila	dep

to		
je		
ten		
muž		
,		
ktorý	ktorého	dep
som		
už		
videla		

NIE:

K chybe **dep** nezaraďujeme prípady, ak ide o chybu v zámene konjunkcie v súvetí, ktoré je inak lexikálno-gramaticky v poriadku – v takom prípade ide o chybu **substword**:

Nečakal Zuzanu, keď si sadla (Nečakal Zuzanu, kým si sadla)

secdep

Túto značku používame vtedy, ak ide o sekundárnu chybu v dependencii. Používame ju najmä v spojeniach s prepozíciami, napr.:

pokračoval ďalej svoj príbeh (pokračoval ďalej vo svojom príbehu), zaoberať sa o módu (zaoberať sa módou)

anotujeme takto:

pokračoval		
ďalej		
NONE	vo	word0_dep
svoj	svojom	seccongr
príbeh	príbehu	secdep

Pozri aj výklad o **dep** vyššie.

neg

Ide o chyby vo vyjadrení jednoduchej i viacnásobnej negácie, napr.:

nie budem (nebudem), nikto prišiel (nikto neprišiel), niebudem (nebudem)

Zvyčajne vyznačujeme celú sekvenciu, ktorej sa chyba týka:

@@@neg		
nie		
budem		
@@@*		
nebudem		
*@@@		

@@@neg		
nie		
som		
bol		
@@@*		
nebol		
som		
*@@@		

@@@neg		
nesom		
bol		
@@@*		
nebol		

som		
*@@@		

@@@neg		
nesom		
@@@*		
nie		
som		
*@@@		
doma		

Sekvenciu **ne**vyznačujeme v prípadoch, ako je tento - uvádzame iba opravu v riadku:

niebudem	nebudem	substvow,neg
----------	---------	--------------

nikto		
prišiel	neprišiel	neg
sem		

phrase

Ide o chyby, ktoré vznikajú vtedy, ak je použité isté slovné spojenie na vyjadrenie významu, pre ktorý už máme v SJ ustálené iné slovné vyjadrenie (frázu, frazému) – často ide o doslovné preklady (kalky) z L1, napr.:

nemá rozprávania (niet o čom)

V takomto prípade vyznačujeme a opravujeme celú chybnú sekvenciu:

@@@phrase		
nemá		
rozprávania		
@@@*		
niet		
o		
čom		
*@@@		

išiel		
som		
do		
Anglicka		
@@@phrase		
pre		
študovanie		
@@@*		
študovať		
*@@@		

NIE:

V podobných prípadoch, ako je nasledujúci, chybné spojenie *ísť pre moju prvú prácu* neopravujeme ako phrase, keďže ide o chybu v dep a následne v seccongr:

išiel		
som		
do		
Holandska		
pre	kvôli	substword_dep
moju	mojej	seccongr
prvú	prvej	seccongr
prácu	práci	secdep

order

Ide najmä o chyby v nesprávnom umiestnení enklitik vo vete = **rytmický princíp slovosledu** a o chyby v poradí členov postupne rozvitého kongruentného atribútu, pozíciu nekongruentného atribútu a podobne = **gramatický princíp slovosledu**.

Pri rytmickej oprave upravujeme pozície nasledujúcich slov:

SPRECIZOVANÉ:

- gramatických morféme slovesa *byť* – *som, si, sme, ste, sa, si* – v tvaroch präterita a kondicionálu,
- častice *by* v tvare kondicionálu,
- krátkych foriem zámen v genitíve, datíve a akuzatíve: *mi, ma, ti, ťa, ho, mu, ju, jej, nás, nám, vám, vás, ich, im*)
- demonstratíva *tu, tam, sem, to*
- reflexívnych zámen *sa* a *si*
- konjunkcia *však*

Pri gramatickej oprave slovosledu upravujeme pozície ako napr.:

Narodila som sa v 1986 roku (Narodila som sa v 1986 roku); v texte sa nachádza veľa mien podstatných (v texte sa nachádza veľa podstatných mien)

Slovosled opravujeme tak, aby sme čo najmenej zasahovali do vety a čo najmenej menili počet slov vo vete. Chybu vyznačujeme a opravujeme v príslušnej sekvencii:

veľmi mi tam sa páčilo (veľmi sa mi tam páčilo), potom rýchlo obliekam sa (potom sa rýchlo obliekam)

@@@order		
veľmi		
mi		
tam		
sa		
páčilo		
@@@*		
veľmi		
sa		
mi		
tam		
páčilo		
*@@@		

Príklad opravy gramatického princípu slovosledu (pozícia inkongruentného atribútu):

som		
študent		
@@@theme		
vo		
Varšave		
na		
univerzite		
@@@*		
na		
univerzite		
vo		
Varšave		
*@@@		

space

Ide o chyby vznikajúce buď nevhodným, redundantným delením slova, alebo nekorektným spojením dvoch slov do jedného slova, napr.:

Winton nie len vyviezol deti, ale (Winton nielen vyviezol deti, ale), má osemdesiat štyri rokov (má osemdesiatštyri rokov), akosa volá (ako sa volá)

@@@space		
osemdesiat		
štyri		
@@@*		
osemdesiatštyri		
*@@@		

@@@space		
akosa		
volá		
@@@*		
ako		
sa		
volá		
*@@@		

theme

Ide o chyby v aktuálnom vetnom členení vety = nesprávnu pozíciu novej a známej informácie vo vete v súvislosti s budovaním textu.

Kamaráti robia pre absolventa univerzity rôzne žarty. Všetko sa koná na verejnosti, aby to bolo čo najtrápnejšie pre absolventa. (... aby to bolo pre absolventa čo najtrápnejšie.)

som		
študent		
@@@theme		
vo		
Varšave		
na		
univerzite		
@@@*		
na		
univerzite		
vo		
Varšave		
*@@@		

connect

Ide o chyby v odkazovacích výrazoch v rámci budovania textu. Pri týchto chybách nevyznačujeme celé sekvencie, ktorých sa chyba týka.

V texte sa zaoberáme slovenčinou. Jeho používanie v rôznych situáciách je ovplyvnené... (V texte sa zaoberáme slovenčinou. Jej používanie v rôznych situáciách je ovplyvnené...)