

## Určovanie zmyslupnosti českej vety

### 1. Úvod

Na nasledujúcich stránkach je uvedená podrobná užívateľská dokumentácia k ročníkovému projektu Určovanie zmyslupnosti českej vety. Obsahuje popis ovládania programu, jeho vstupy a výstupy a popis riešenia procedúry na analýzu zmyslupnosti.

### 2. Zadávateľ a riešiteľ

**Zadávateľ:**

Mgr. Barbora Vidová-Hladká, Ph.D.  
ÚFAL MFF UK  
E-mail: hladka@ufal.mff.cuni.cz

**Riešiteľ:**

Vincent Kríž  
E-mail: vincent.kriz@gmail.com

### 3. Anotácia

*Cieľom ročníkového projektu je implementácia automatickej procedúry, ktorá v reálnom čase rozozná, či je zhluk textových reťazcov českou vetou. Automatická procedúra bude používať výstupy externých automatických modulov, ktoré sa týkajú tvaroslovia (slovné druhy a ich kategórie – rod, číslo, pád, osoba a iné). Vyriešenie tejto úlohy je využiteľné vo fulltextovom vyhľadávaní, kde je (okrem iného) dôležité rozoznať zhluk kľúčových slov určených pre zmätenie vyhľadávača od vety určenej užívateľom. Procedúra bude implementovaná pre UNIXovo orientované operačné systémy ako aj pre platformu OS Windows. Bude realizovaná riadkovo-orientovaným programom napísaným v jazyku C++.*

### 4. Definovanie „morfologickej zmyslupnosti“

Prvým krokom k úspešnému riešeniu zadaného problému je presné definovanie *zmyslupnosti*, teda stanovenie kritérií, podľa ktorých program prehlási vetu za zmyslupnú alebo nezmyslupnú. Program rozhoduje o zmyslupnosti vety len na základe morfologickej informácie – teda na základe skúmania slov, o ktorých vie ich slovný druh a ďalšie gramatické kategórie (napríklad slovný poddruh, rod, číslo, pád, osobu, čas, spôsob a podobne).

Podstatné je pozorovanie, že program pracuje len s morfologickou informáciou. To znamená, že sa môže úplne odpútať od sémantiky (toho, čo slovo pomenúva) a od syntaxe vety. Vďaka tomu neanalyzuje slová ako také ale len morfologickú informáciu a dve (sémanticky) rôzne slová s rovnakými gramatickými kategóriami považuje za totožné.

Príkladom zmyslupných viet sa môžu stať všetky „sémanticky zmyslupné“

vety, napríklad

*Nová komedie Marie Poledňákové Líbáš jako Bůh  
je jasně lepší než její předchozí film Jak se krotí krokodýli.*

Za zmysluplné vety však procedúra označí aj „sémanticky nezmysluplné“ vety, ktoré vyhovujú stanoveným morfológickým pravidlám, napríklad

*Krátký kabel během letadla zlepší vaši mrkev i výkonnost.*

Príkladom „morfológicky nezmysluplnej“ vety je napríklad

*Vhodná sněhu nastala pro kolo, kdo chce  
od nemovitost zhodnotit někdo úspory či anděl změnit bydlení.*

## 5. Idea riešenia

Hlavnou myšlienkou určovania zmysluplnosti je vyhľadávanie jazykových konštrukcií jazyka, ktoré sú morfológicky nesprávne. Program teda nekontroluje *správnosť* vety ale namiesto toho vyhľadáva možné *chyby*. Takýto prístup je výhodnejší z hľadiska kvantity pravidiel, ktoré bude musieť program viesť. Predpokladá totiž, že chybných konštrukcií je menej ako všetkých možných správnych.

Pôvodným zámerom programu bolo vystavať procedúru v štýle dynamického programovania. Jej podstatou malo byť postupné overovanie najskôr dvojíc (za sebou idúcich) slov, následné spojenie korektných dvojíc do väčších celkov a podobne ďalej – spájanie čoraz väčších zoskupení slov až do jediného objektu, ktorý by reprezentoval celú (zmysluplnú) vetu. Ukázalo sa však, že už na druhej úrovni analýzy, podľa zamýšľaného postupu, sú objekty príliš všeobecné a v podstate všetky možné kombinácie môžu byť zmysluplné.

Program preto určuje zmysluplnosť vety len podľa vyhľadávania nepovolených kolokácií za sebou stojacich slov. Pre prehľadnosť sú pravidlá kolokácií rozdelené a testované podľa prvého slovného poddruhu v danej kolokácii. Následne sú ešte slovné poddruhy združené do tzv. *modulov*, ktoré združujú poddruhy do rovnakých slovných druhov.

Ak modul zistí chybu vo vete, analýza končí s výsledkom, že veta je nezmysluplná. Ak každý z použitých modulov vráti *true*, veta je prehlásená za zmysluplnú.

Aplikovaním všetkých modulov by sa mala docieľať vyššia pravdepodobnosť objavenia chybných konštrukcií. Výberom len istých modulov zasa analýza viet len na určitý druh chýb.

## 6. Ovládanie programu

Program, ktorý implementuje procedúru, ktorá určuje zmysluplnosť viet, pracuje štýlom, ktorý je veľmi častý pre programy s textovým výstupom typickým pre terminál. Na začiatku očakáva definovanie vstupov a určenie (nepovinných) možností, podľa ktorých prispôsobí svoj beh.

Program sa spúšťa zavolaním preloženého binárneho súboru v jeho aktuálnom

adresári, nasledovaným povinnými (a nepovinnými) parametrami oddelenými medzerami.

Napríklad spustenie programu s dátami v súbore *input.csts*, ktoré sú umiestnené v rovnakom adresári ako program, s použitím všetkých modulov a v móde *verbose* sa vykoná nasledujúcim príkazom:

```
./moran -i input.csts -v
```

Príklad spustenia v *non-verbose* móde s použitím len dvoch modulov (X a R) je uvedený na nasledujúcom riadku:

```
./moran -i input.csts -m XR
```

V nasledujúcej tabuľke sú uvedené všetky parametre, ktorými môže užívateľ ovplyvniť beh programu.

Parameter	Typ parametra	Popis parametra
-i <i>súbor</i>	povinný	Definuje cestu k vstupnému súboru. Príklad: -i input.csts
-m <i>moduly</i>	nepovinný	Definuje ktoré moduly má program na analýzu viet aplikovať. Moduly sa zadávajú svojimi názvami ako jedno slovo oddelené od parametra -m medzerou. Bez zadania parametra sa implicitne použijú všetky moduly, ktorými program disponuje. Príklad: -m ARX
-v	nepovinný	Určuje formát výstupu programu. S uvedeným parametrom program pracuje v režime <i>verbose</i> , bez uvedenia v <i>non-verbose</i> režime.

Za parametrom -m môžu nasledovať tieto moduly:

Názov modulu	Popis
A	kontrola prídavných mien
C	kontrola čísloviek
J	kontrola spojok
N	kontrola podstatných mien
P	kontrola zámen
R	kontrola predložiek
V	kontrola slovík
X	kontrola slov, ktorým tool_chain neurčil gramatické kategórie

Viac o vstupných súboroch je uvedené v časti *Vstup programu*. Podrobné informácie o moduloch sú v časti *Moduly* a *verbose* a *non-verbose* výstup je podrobne popísaný v časti *Výstup programu*.

## 7. Vstup programu

Morfologickú informáciu program nevytvára sám. Používa údaje z externých modulov, ktoré sú súčasťou Českého akademického korpusu 2.0. Konkrétne sa jedná o nástroj *tool\_chain*, ktorý združuje viacero nástrojov do jedného programu. Pre morfologický výstup, ktorý bude zároveň vstupom programu, budeme používať *tokenizátor*, *morfologickú analýzu* a *tagger*. Viac o programe je uvedené v [1].

Program na vstupe (definovanom povinným parametrom *-i*) očakáva vety určené na analýzu vo formáte CSTS. Formát CSTS je vedľajším údajovým formátom ČAK 2.0. Jedná sa o formát SGML používaný v PZK 1.0 a tiež v ČNK. Medzi jeho výhody patrí dobrá čitateľnosť pre človeka, jeho jednoduché spracovanie nástrojmi a tiež fakt, že niektoré nástroje z ČAK 2.0 pracujú výhradne s týmto formátom. Stručný úvod do CSTS je napríklad v [2], podrobná špecifikácia potom v [3].

Formát CSTS je jedným z výstupných formátov nástroja *tool\_chain*.

Zo vstupného súboru program čerpá morfologickú informáciu o jednotlivých slovách vo vetách. Tá je obsiahnutá v tagu `<mat>`. Údaje, ktoré program získava zo súboru CSTS, sú reprezentované v 15-znakových tagoch CPMT (The Czech Positional Morphological Tags). Podrobná špecifikácia je uvedená v [4].

Program vstupný súbor vyparsuje do pomocného súboru (ukladá ho do adresára */tmp*). Počas väčšiny času behu programu program číta vytvorený pomocný súbor. Zo súboru prečíta informáciu o jednej vete, aplikuje na ňu procedúru a ak je v súbore ďalšia veta, pokračuje jej načítaním. Inak pomocný súbor vymaže a program končí.

## 8. Výstup programu

Program dokáže výstup formátovať v dvoch módoch. V móde *verbose* program vypisuje množstvo informácií o aktuálnych testoch viet, ktoré uskutočňuje a v prípade objavenia nezmyslupnej vety vypíše ktorý test chybu objavil (a tým sa identifikuje aj druh chyby). V móde *non-verbose* sa program obmedzuje len na vypísanie identifikácie vety a text OK ak je veta analyzovaná ako zmyslupná a KO v opačnom prípade.

Na nasledujúcich riadkoch sú uvedené príklady a opisy oboch druhov výstupu.

### 1. Verbose výstup

*Verbose* výstup podrobne opisuje priebeh analyzovania vety. Najskôr celú vetu vypíše slovo za slovom vrátane morfologických informácií o jednotlivých slovách. Následne pokračuje aplikáciou jednotlivých modulov.

Každý modul obsahuje minimálne jeden test, ktorým vetu analyzuje. Na výstupe sa postupne objavia hlásenia každého z týchto testov. Ak daný jazykový jav testom prejde, test vypíše text OK, v opačnom prípade slovo WRONG. Následne program vypíše ktorý test objavil chybu. Analýza vety potom končí textom, že veta bola analyzovaná ako nezmyslupná.

Tu je príklad zmyslupnej vety analyzovanej modulmi X, R a A.

```
-----  
Veta: input.txt:001-pls1  
      NNMP3-----A---- Policistům
```

```

P7-X4----- se
VpNS---XR-NA--- nepodařilo
Vf-----A---- prokázat
Z:----- ,
J,----- že
Vc----- by
VpNS---XR-AA--- došlo
RR--3----- k
AAIS3----1A---- trestnému
NNIS3----A---- činu
Z:----- .
Startuje testovanie vety na tag X*-----
Kontrolujem tag NNMP3----A---- ... OK
Kontrolujem tag P7-X4----- ... OK
Kontrolujem tag VpNS---XR-NA--- ... OK
Kontrolujem tag Vf-----A---- ... OK
Kontrolujem tag Z:----- ... OK
Kontrolujem tag J,----- ... OK
Kontrolujem tag Vc----- ... OK
Kontrolujem tag VpNS---XR-AA--- ... OK
Kontrolujem tag RR--3----- ... OK
Kontrolujem tag AAIS3----1A---- ... OK
Kontrolujem tag NNIS3----A---- ... OK
Kontrolujem tag Z:----- ... OK
Startuje test korektnosti predloziek...
... test RF
... test RZ
... test nasledovnikov
      Kontrolujem tag RR--3----- ... OK
Startuje test korektnosti adjektiv...
... test A2
... test AA
      Kontrolujem tag AAIS3----1A---- ... OK
... test AC
Veta je zmysluplna. Vsetky testy prebehli uspesne.

```

Nasleduje príklad vety, ktorú program analyzoval ako nezmysluplnú.

```

-----
Veta: input.txt:001-pls1
      AAIS1----1A---- Názorný
      NNIS1----A---- postup
      AAIP2----1A---- popsaných
      AANS2----1A---- umělého
      NNNS2----A---- dýchání
      VB-S---3P-AA--- je
      RR--6----- na
      NNIS6----A---- obrázku
      Z:----- .
Startuje testovanie vety na tag X*-----
Kontrolujem tag AAIS1----1A---- ... OK
Kontrolujem tag NNIS1----A---- ... OK
Kontrolujem tag AAIP2----1A---- ... OK
Kontrolujem tag AANS2----1A---- ... OK
Kontrolujem tag NNNS2----A---- ... OK
Kontrolujem tag VB-S---3P-AA--- ... OK
Kontrolujem tag RR--6----- ... OK
Kontrolujem tag NNIS6----A---- ... OK
Kontrolujem tag Z:----- ... OK

```

```

Startuje test korektnosti predloziek...
... test RF
... test RZ
... test nasledovnikov
      Kontrolujem tag RR--6----- ... OK
Startuje test korektnosti adjektiv...
... test A2
... test AA
      Kontrolujem tag AAIS1----1A---- ... OK
      Kontrolujem tag AAIP2----1A---- ... WRONG
Tag je v konflikte s AANS2----1A----
Veta nie je zmysluplna. Testovací modul: A

```

## 2. Non-verbose výstup

*Non-verbose* výstup je výstup programu bez podrobných informácií. Obmedzuje sa len na kľúčovú informáciu – či veta je alebo nie je zmysluplná. Príklad výstupu v *non-verbose* móde je na nasledujúcich riadkoch.

```

Veta: input.txt:001-pls1 ... OK
Veta: input.txt:001-pls2 ... OK
Veta: input.txt:001-pls3 ... OK
Veta: input.txt:001-pls4 ... KO

```

## 9. Moduly

Procedúra, ktorá analyzuje zmyslupnosť viet je implementovaná v podobe viacerých nezávislých modulov. Každý z nich sa zameriava na iný slovný druh a obsahuje minimálne jeden test, ktorý skúma konkrétny gramatický jav alebo aké slovné poddruhy môžu za testovaným slovom nasledovať. (Napríklad modul R kontroluje všetky zistené javy spojené s predložkami, modul A s adjektívami a podobne.)

V nasledujúcom prehľade sú uvedené všetky moduly a všetky ich testy, ktoré program v aktuálnej verzii obsahuje. Pri popise väčšiny pravidiel je uvedená prehľadná tabuľka, ktorá obsahuje pravidlá, ktoré testujeme medzi testovaným slovným poddruhom a za nim stojacim slovom. Na nasledujúcom príklade takejto tabuľky sú vysvetlené všetky značky, ktoré sa v tabuľke vyskytujú.

C=													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	C1	○	Db	○	P5	○	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	○	PW	○	Vq	○		

Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	o	Cw	o	NN	o	PE	o	PZ	o	Vt	o		

Vzorová tabuľka zobrazuje pravidlá pre dvojicu slov, kde prvá z dvojice je číslovka písaná číslom (C=). V šedých stĺpcoch potom nasleduje zoznam možných morfológických značiek a v susednom stĺpci značka vyjadrujúca pravidlo.

Okrem morfológických značiek je v tabuľke uvedená aj značka <s>, ktorá vyjadruje pravidlo, či sa daný slovný poddruh môže vyskytovať na konci vety.

Pravidlá sú vyjadrené niekoľkými symbolmi, najčastejšími sú:

Symbol	Význam
o	Slovný poddruh môže nasledovať, žiadne ďalšie pravidlo sa netestuje.
x	Poddruh nemôže nasledovať.
345	Slovný poddruh, ktorý nasleduje, sa musí zhodovať v rode, čísle a páde
P5	Slovný poddruh, ktorý nasleduje, sa musí zhodovať v páde, ale len za podmienky, že testované slovo je v množnom čísle.
S5	Nasledovník sa musí s testovaným slovom zhodovať v páde, ale len za podmienky, že je v jednotnom čísle.
*5	Slová sa musia zhodovať v páde, ale len za špeciálnej podmienky (uvedenej mimo tabuľku v ďalšom komentári).

## 1. Modul X (Neanalyzovateľné vety)

Modul X obsahuje jediný test, ktorý vetu prehlási za chybnú ak sa v nej nachádza slovo, ktorému *tool\_chain* nepriradil žiadnu gramatickú informáciu.

Vstup vo formáte CSTS je generovaný nástrojom *tool\_chain*, ktorý využíva tokenizátor, morfológickú analýzu a tagger. Je možné, že tento nástroj nerozpozna niektoré zo slov na vstupe a označí ich tagom *xe-----*. Vety, ktoré obsahujú slová, ktoré *tool\_chain* nedokázal analyzovať, program automaticky vyhodnocuje ako nezmyselné. O danom slove totiž nemá potrebnú morfológickú informáciu a nemôže aplikovať ďalšie moduly.

## 2. Modul R (Kontrola predložiek)

Modul R obsahuje tri testy, ktoré kontrolujú korektnosť predložiek vyskytujúcich sa v danej vete. Nižšie je uvedený ich podrobný popis.

### Test kontrola\_RF

Tagom RF----- sú označované tzv. *nepravé predložky* – slová, ktoré nikdy nestoja osamote a vždy po nich nasleduje predložka, napríklad *vzhľadom (k)*, *nehľadě (na)*. Test skontroluje, či po každom z týchto slov predložka skutočne nasleduje.

Tabuľka uvádza uvedené pravidlá v prehľadnej forme.

RF													
A2	x	Ch	x	Cy	x	P0	x	PH	x	VB	x	RF	x
AA	x	Cj	x	Cz	x	P1	x	PJ	x	Vc	x	RR	o
AC	x	Ck	x	C}	x	P4	x	PK	x	Ve	x	RV	o
AG	x	C1	x	Db	x	P5	x	PL	x	Vf	x	TT	x
AM	x	Cn	x	Dg	x	P6	x	PP	x	Vi	x	ZZ	x
AO	x	Co	x	II	x	P7	x	PQ	x	Vm	x	<s>	x
AU	x	Cr	x	J*	x	P8	x	PS	x	Vp	x		
C=	x	Cu	x	J,	x	P9	x	PW	x	Vq	x		
Ca	x	Cv	x	J^	x	PD	x	PY	x	Vs	x		
Cd	x	Cw	x	NN	x	PE	x	PZ	x	Vt	x		

### Test kontrola\_RZ

Predložky vo väčšine prípadov rozvíjajú slovný druh, ktorý sa skloňuje a nasleduje za predložkou. Je málo pravdepodobné, ak sa predložka nachádza na konci vety. Test tento kontroluje práve tento jav.

Program chybné označí vety s predložkou na konci vety alebo časti súvetia. Takýchto viet je však málo a je väčšia pravdepodobnosť chyby ako takejto vetnej konštrukcie.

### Test kontrola\_nasledovnikov

Predložky rozvíjajú za nimi stojaci slovný druh, ktorý sa skloňuje. Navyše sa viažu z pádom. Tento test sleduje aký slovný druh nasleduje za predložkou. V nasledujúcej tabuľke je uvedený prehľad akým spôsobom procedúra ďalej spracúva pár predložka a za ňou stojaci slovný poddruh.

RR, RV													
A2	o	Ch	5	Cy	5	P0	x	PH	5	VB	x	RF	x
AA	5	Cj	5	Cz	5	P1	5	PJ	o	Vc	x	RR	o
AC	o	Ck	5	C}	o	P4	5	PK	5	Ve	x	RV	o
AG	5	C1	5	Db	o	P5	5	PL	5	Vf	x	TT	x
AM	5	Cn	5	Dg	o	P6	5	PP	5	Vi	x	ZZ	x
AO	x	Co	o	II	o	P7	o	PQ	5	Vm	x	<s>	x
AU	5	Cr	5	J*	o	P8	5	PS	5	Vp	x		
C=	o	Cu	x	J,	o	P9	5	PW	5	Vq	x		
Ca	5	Cv	o	J^	o	PD	5	PY	o	Vs	x		
Cd	5	Cw	5	NN	5	PE	5	PZ	5	Vt	x		

V málo pravdepodobných prípadoch môžu nasledovať dve predložky za sebou. Takáto vetná konštrukcia je však málo pravdepodobná a program ju označí za chybnú.

## 3. Modul A (Kontrola prídavných mien)

Modul obsahuje testy, ktoré kontrolujú prídavné mená. Obsahuje samostatný



test pre každý slovný poddruh.

### Test kontrola\_A2

Test kontroluje špecifické prídavné mená, ktoré *tool\_chain* taguje značkou A2------. Ide o prídavné mená spojené s ďalším prídavným menom pomocou spojovníka a vyparované ako samostatný token, napríklad *technicko-*. Test kontroluje, či za každým z takýchto prídavných mien nasleduje ďalšie prídavné meno.

### Test kontrola\_AA

Test si všíma „štandardné“ prídavné mená. Tie sú otagované značkou začínajúcou AA. Test kontroluje zhodu v páde, rode a čísle pre vybrané poddruhy podstatných mien, čísloviek či sloviess.

Tabuľka uvádza prehľad kontrolovaných pravidiel.

AA													
A2	o	Ch	o	Cy	345	P0	o	PH	o	VB	4	RF	o
AA	5	Cj	5	Cz	x	P1	o	PJ	o	Vc	4	RR	o
AC	o	Ck	5	C}	o	P4	o	PK	o	Ve	o	RV	o
AG	5	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	5	Cn	o	Dg	o	P6	o	PP	o	Vi	4	ZZ	o
AO	4	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	5	Cr	345	J*	o	P8	5*	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	5	Cw	o	NN	345	PE	o	PZ	o	Vt	x		

\* Privlastňovacie zámená P8 sa musia zhodovať v páde len ak pred AA je predložka.

### Test kontrola\_AC

Test sleduje menné prídavné mená (napríklad *rád*, *schopen*) a kontroluje pravidlá popísané v tabuľke kolokácií.

AC													
A2	o	Ch	o	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	o	Cz	o	P1	o	PJ	o	Vc	o	RR	o
AC	o	Ck	o	C}	o	P4	o	PK	o	Ve	o	RV	o
AG	o	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	x	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	x	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	o	Cr	o	J*	x	P8	o	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	o	Cw	o	NN	o	PE	o	PZ	o	Vt	x		

### Test kontrola\_AG

Test sleduje prídavné mená odvodené od prítomného prechodníku a kontroluje pravidlá popísané v tabuľke kolokácií.

AG													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	x	P8	5*	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	345	PE	○	PZ	○	Vt	x		

\* Privlastňovacie zámená P8 sa musia zhodovať v páde len ak pred AA je predložka.

### Test kontrola\_AM

Test sleduje prídavné mená odvodené od minulého prechodníku a kontroluje pravidlá popísané v tabuľke kolokácií.

AM													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	x	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	x	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	x	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	x	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	x	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	345	PE	○	PZ	○	Vt	x		

### Test kontrola\_AO

Test sleduje samostatne stojace zámená (napríklad *svůj, nesvůj, tentam*) a kontroluje pravidlá popísané v tabuľke kolokácií.

AO													
A2	x	Ch	x	Cy	x	P0	x	PH	x	VB	○	RF	○
AA	x	Cj	x	Cz	x	P1	x	PJ	x	Vc	○	RR	○
AC	x	Ck	x	C}	x	P4	x	PK	x	Ve	○	RV	○
AG	x	Cl	x	Db	○	P5	x	PL	x	Vf	○	TT	○
AM	x	Cn	x	Dg	○	P6	○	PP	x	Vi	○	ZZ	○
AO	x	Co	x	II	○	P7	○	PQ	x	Vm	○	<s>	○
AU	x	Cr	x	J*	○	P8	x	PS	x	Vp	○		
C=	x	Cu	x	J,	○	P9	x	PW	x	Vq	x		
Ca	x	Cv	x	J^	○	PD	x	PY	x	Vs	○		
Cd	x	Cw	x	NN	○	PE	x	PZ	x	Vt	x		

### Test kontrola\_AU

Test si všíma privlastňovacie prídavné mená. Tie sú otagované značkou začínajúcou AU. Test kontroluje zhodu v páde, rode a čísle pre vybrané poddruhy podstatných mien a čísloviek.

Tabuľka uvádza prehľad kontrolovaných pravidiel.

AU													
A2	o	Ch	o	Cy	5	P0	x	PH	o	VB	o	RF	o
AA	5	Cj	o	Cz	o	P1	x	PJ	o	Vc	o	RR	o
AC	o	Ck	o	C}	o	P4	x	PK	o	Ve	o	RV	o
AG	5	C1	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	5	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	5	Cr	5	J*	x	P8	o	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	5	Cw	o	NN	345	PE	o	PZ	o	Vt	x		

### 4. Modul C (Kontrola čísloviek)

Modul na kontrolu čísloviek je relatívne obsiahly a vykonáva samostatný test pre každý slovný poddruh. V nasledujúcom popise uvádzame všetky testy, ktoré obsahuje.

#### Test kontrola\_cisel

Tento test sa zameriava na číslovky napísané číslom. Tag popisujúci čísla neobsahuje žiadnu ďalšiu morfológickú informáciu. Test však využíva to, že pri číslach môžeme z ich hodnoty určiť či ide o jednotné alebo množné číslo. Bohužiaľ, túto gramatickú informáciu nemožno využiť hromadne a je známy len jeden prípad kedy ju využijeme.

Ide o prípad, ak za číslom nasleduje číslovka vyjadrujúca zlomok (tag Cy-----). V tomto prípade ak je číslo väčšie ako 1, musí byť zlomok v množnom čísle, inak v jednotnom.

Test následne kontroluje, aký slovný poddruh môže za číslom nasledovať a aký nie. Tabuľka uvádza všetky pravidlá v prehľadnej forme.

C=													
A2	o	Ch	o	Cy	4	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	x	Cz	x	P1	o	PJ	o	Vc	o	RR	o
AC	o	Ck	x	C}	o	P4	o	PK	x	Ve	o	RV	o
AG	o	C1	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	o	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o

AO	o	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	o	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	x	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	x	Vs	o		
Cd	o	Cw	o	NN	o	PE	o	PZ	o	Vt	x		

### Test kontrola\_Ca

Kontroluje neurčité číslovky, napríklad „mnoho“, „málo“, „tolik“ atď. Z morfológického hľadiska tag pre neurčité číslovky obsahuje len informáciu o páde. Ten sa však zhoduje s nasledovníkom jedine ak nasledovník je ďalšia neurčitá číslovka.

Existuje však niekoľko slovných poddruhov, ktoré za neurčitou číslovkou nemôžu nasledovať.

Ca													
A2	o	Ch	x	Cy	o	P0	x	PH	o	VB	o	RF	o
AA	o	Cj	x	Cz	x	P1	x	PJ	x	Vc	o	RR	o
AC	o	Ck	x	C}	o	P4	o	PK	o	Ve	x	RV	o
AG	o	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	o	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	x	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	o	Cr	o	J*	x	P8	o	PS	o	Vp	o		
C=	o	Cu	x	J,	o	P9	x	PW	o	Vq	x		
Ca	5	Cv	o	J^	o	PD	o	PY	x	Vs	o		
Cd	o	Cw	x	NN	o	PE	o	PZ	o	Vt	x		

### Test kontrola\_Cd

Druhové číslovky s adjektívnym skloňovaním (napríklad „dvojí“, „desaterý“) sa zhodujú v páde s niektorými adjektívami, číslovkami, substantívami a zámenami. Existuje tiež niekoľko slovných poddruhov, ktoré za nimi nemôžu nasledovať.

Cd													
A2	o	Ch	x	Cy	5	P0	x	PH	o	VB	o	RF	x
AA	5	Cj	x	Cz	o	P1	x	PJ	x	Vc	o	RR	o
AC	x	Ck	x	C}	o	P4	x	PK	x	Ve	o	RV	o
AG	5	Cl	x	Db	o	P5	x	PL	x	Vf	o	TT	o
AM	x	Cn	x	Dg	o	P6	x	PP	o	Vi	o	ZZ	o
AO	x	Co	o	II	o	P7	o	PQ	x	Vm	o	<s>	o
AU	5	Cr	5	J*	x	P8	o	PS	o	Vp	o		
C=	o	Cu	x	J,	o	P9	x	PW	x	Vq	x		
Ca	x	Cv	o	J^	o	PD	5	PY	x	Vs	o		
Cd	5	Cw	x	NN	5	PE	x	PZ	o	Vt	x		

### Test kontrola\_Ch

Druhové číslovky (napríklad „jedny“, „nejedny“) nesú morfológickú informáciu o rode, čísle a páde. Nie je však známe žiadne pravidlo, kde by sa táto informácia mohla využiť.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Ch													
A2	○	Ch	×	Cy	○	P0	○	PH	○	VB	○	RF	×
AA	○	Cj	×	Cz	×	P1	×	PJ	×	Vc	○	RR	○
AC	○	Ck	×	C}	○	P4	×	PK	×	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	×	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	×	Co	○	II	×	P7	○	PQ	×	Vm	○	<s>	○
AU	○	Cr	○	J*	×	P8	○	PS	○	Vp	○		
C=	○	Cu	×	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	○	PY	×	Vs	○		
Cd	○	Cw	×	NN	○	PE	×	PZ	○	Vt	×		

### Test kontrola\_Cj

Druhové číslovky väčšie alebo rovné 4 v substantívnom postavení (napríklad „čtvero“, „desatero“) sa vyskytujú jedine v jednotnom čísle a nesú aj informáciu o páde.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Cj													
A2	○	Ch	×	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	×	Cz	×	P1	○	PJ	○	Vc	○	RR	○
AC	×	Ck	×	C}	○	P4	○	PK	×	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	×	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	×	Co	○	II	×	P7	○	PQ	×	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	×	J,	○	P9	×	PW	○	Vq	×		
Ca	×	Cv	○	J^	○	PD	○	PY	×	Vs	○		
Cd	×	Cw	×	NN	○	PE	×	PZ	○	Vt	×		

### Test kontrola\_Ck

Druhové číslovky väčšie alebo rovné 4 v adjektívnom postavení (napríklad „čtvery“) sa zhodujú v páde s niektorými adjektívami, číslovkami a podstatnými menami.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií a kontroly zhody v páde.

Ck													
----	--	--	--	--	--	--	--	--	--	--	--	--	--

A2	o	Ch	x	Cy	5	P0	o	PH	o	VB	o	RF	x
AA	5	Cj	x	Cz	x	P1	x	PJ	o	Vc	o	RR	o
AC	x	Ck	x	C}	o	P4	x	PK	x	Ve	o	RV	o
AG	5	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	x	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	x	Co	o	II	x	P7	o	PQ	o	Vm	o	<s>	o
AU	5	Cr	x	J*	x	P8	o	PS	o	Vp	o		
C=	o	Cu	x	J,	x	P9	x	PW	o	Vq	x		
Ca	x	Cv	o	J^	o	PD	o	PY	x	Vs	o		
Cd	x	Cw	x	NN	5	PE	o	PZ	o	Vt	x		

### Test kontrola\_Cl, kontrola\_Cn

Základné číslovky s nesubstantívnym skloňovaním (napríklad „jeden“, „dva“) sa zhodujú v páde s niektorými slovnými poddruhmi adjektív a čísloviek.

Rovnaké pravidlá platia aj pre základné číslovky väčšie ako 5.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Cl, Cn													
A2	o	Ch	x	Cy	5	P0	x	PH	o	VB	o	RF	x
AA	5	Cj	x	Cz	x	P1	x	PJ	o	Vc	o	RR	o
AC	o	Ck	x	C}	o	P4	o	PK	o	Ve	o	RV	o
AG	5	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	5	Cn	5	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	x	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	5	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	x	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	S5	PY	o	Vs	o		
Cd	o	Cw	o	NN	o	PE	x	PZ	o	Vt	x		

### Test kontrola\_Co

Násobné neurčité číslovky neposkytujú žiadne morfológické informácie, známe sú len možné kolokácie s ďalšími poddruhmi uvedené v tabuľke.

Co													
A2	o	Ch	x	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	o	Cz	x	P1	x	PJ	o	Vc	o	RR	o
AC	o	Ck	o	C}	o	P4	x	PK	x	Ve	o	RV	o
AG	o	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	o	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	x	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	o	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	x	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		

Cd	o	Cw	o	NN	o	PE	x	PZ	o	Vt	x		
----	---	----	---	----	---	----	---	----	---	----	---	--	--

### Test kontrola\_Cr

Radové číslovky sa zhodujú v rode čísla a páde s adjektívami, číslovkami, podstatnými menami a zámenami.

Tabuľka uvádza všetky pravidlá, ktoré program kontroluje.

Cr													
A2	o	Ch	x	Cy	5	P0	o	PH	o	VB	o	RF	o
AA	345	Cj	x	Cz	x	P1	o	PJ	x	Vc	o	RR	o
AC	o	Ck	x	C}	o	P4	345	PK	5	Ve	o	RV	o
AG	345	Cl	P5	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	345	Cn	P5	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	x	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	345	Cr	345	J*	x	P8	o	PS	o	Vp	o		
C=	o	Cu	x	J,	o	P9	x	PW	o	Vq	x		
Ca	5	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	x	Cw	o	NN	345	PE	o	PZ	o	Vt	x		

### Test kontrola\_Cu

Opytovacia číslovka „kolikrát“ nenesie morfológickú informáciu. Tabuľka uvádza iba prehľad povolených a zakázaných kolokácií.

Cu													
A2	o	Ch	o	Cy	o	P0	o	PH	o	VB	o	RF	x
AA	o	Cj	o	Cz	x	P1	x	PJ	x	Vc	o	RR	o
AC	o	Ck	o	C}	o	P4	o	PK	x	Ve	o	RV	o
AG	o	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	x	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	x	II	o	P7	o	PQ	x	Vm	o	<s>	o
AU	o	Cr	o	J*	x	P8	o	PS	o	Vp	o		
C=	o	Cu	x	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	x	J^	o	PD	o	PY	x	Vs	o		
Cd	o	Cw	x	NN	o	PE	x	PZ	o	Vt	x		

### Test kontrola\_Cv

Násobné číslovky nenesú žiadnu s použiteľných morfológických informácií, známe sú len niektoré zo zakázaných kolokácií.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Cv													
A2	o	Ch	x	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	x	Cz	x	P1	o	PJ	o	Vc	o	RR	o
AC	o	Ck	x	C}	o	P4	o	PK	x	Ve	o	RV	o

AG	○	C1	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	×	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	×	Co	×	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	×	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	○	PY	×	Vs	○		
Cd	○	Cw	×	NN	○	PE	○	PZ	○	Vt	×		

### Test kontrola\_Cw

Neurčité číslovky s adjektívnym skloňovaním sa zhodujú v rode čísla a páde s podstatnými menami, niektorými zámenami a adjektívami .

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Cw													
A2	○	Ch	×	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	345	Cj	○	Cz	×	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	C1	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	×	Cn	○	Dg	○	P6	○	PP	345	Vi	○	ZZ	○
AO	×	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	345	PS	345	Vp	○		
C=	○	Cu	×	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	345	PY	○	Vs	○		
Cd	×	Cw	×	NN	345	PE	×	PZ	○	Vt	×		

### Test kontrola\_Cy

Zlomky zakončené na „-ina“ sú vždy v ženskom rode, nesú informáciu o čísle a páde. Nie je však známe žiadne pravidlo, kde by sa poskytnuté informácie vedeli využiť.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Cy													
A2	○	Ch	×	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	×	Cz	×	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	×	C}	○	P4	○	PK	×	Ve	○	RV	○
AG	○	C1	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	×	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	×	Cw	×	NN	○	PE	○	PZ	○	Vt	×		



## Test kontrola\_Cz

Opytovacia číslovka „kolikátý“ neobsahuje žiadnu morfológickú informáciu, ale jej špecifické postavenie umožnilo vypracovať viacero pravidiel zakázaných kolokácií.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Cz													
A2	o	Ch	x	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	x	Cz	x	P1	o	PJ	o	Vc	o	RR	o
AC	o	Ck	x	C}	o	P4	o	PK	x	Ve	o	RV	o
AG	o	Cl	x	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	o	Cn	x	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	X	II	x	P7	o	PQ	o	Vm	o	<s>	o
AU	o	Cr	x	J*	x	P8	o	PS	o	Vp	o		
C=	o	Cu	x	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	x	J^	o	PD	o	PY	o	Vs	o		
Cd	x	Cw	x	NN	o	PE	x	PZ	o	Vt	x		

## 5. Modul N (Kontrola podstatných mien)

Podstatné mená nie sú rozdelené do poddruhov. Procedúra preto uvažuje o všetkých podstatných menách v jednom teste. Nasledujúca tabuľka uvádza prehľad povolených a zakázaných kolokácií a niektoré pravidlá, ktoré sa na podstatné mená dali aplikovať.

NN													
A2	o	Ch	o	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	o	Cz	o	P1	4	PJ	o	Vc	o	RR	o
AC	o	Ck	o	C}	o	P4	34	PK	o	Ve	o	RV	o
AG	5*	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	o	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	o	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	o	Cw	o	NN	5**	PE	o	PZ	o	Vt	x		

\* Zhoda v páde sa pri adjektíve odvodenom od prítomného prechodníku kontroluje len ak pred podstatným menom nie je predložka.

\*\* Podstatné meno, ktoré nasleduje za podstatným menom musí byť v druhom páde. (Okrem situácie, ak za dvojicou nasleduje sloveso.)

## 6. Modul J (Kontrola spojok)

V tomto module je implementovaný jediný test, ktorý kontroluje, či spojka (podradňovacia alebo priraďovacia) nebola na konci vety.

## 7. Modul P (Kontrola zámen)

Zámená procedúra kontroluje podobne ako v prípade čísloviek v niekoľkých testoch, každý pre špecifický slovný poddruh. Okrem nich modul obsahuje ešte test, ktorý sleduje, či sa zámeno nenachádza na konci vety. Niektoré špecifické zámená sa tam vyskytovať nemôžu.

### Test kontrola\_P0

Predložka s pripojeným „-ň“ (napríklad „proň“, „naň“) neobsahuje žiadne morfologické informácie, známe sú len niektoré zakázané kolokácie, ktoré test sleduje.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

P0													
A2	○	Ch	○	Cy	○	P0	×	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	○	PZ	○	Vt	×		

### Test kontrola\_P1

Vzťažné privlastňovacie zámeno, napríklad „jehož“, „jejíž“ obsahuje bohatú morfologickú informáciu, nie je však známe žiadne pravidlo, v ktorom by mohla byť využitá. Súčasná verzia programu preto obsahuje len pozorovania o kolokáciách s ostatnými slovnými druhmi.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

P1													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	×
AA	○	Cj	○	Cz	○	P1	×	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	×	PK	×	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	×	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	×	P7	○	PQ	○	Vm	○	<s>	×
AU	○	Cr	○	J*	×	P8	×	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	○	PY	×	Vs	○		
Cd	○	Cw	○	NN	○	PE	×	PZ	○	Vt	×		

### Test kontrola\_P4

Vzťažné/opytovacie zámeno s adjektívnym skloňovaním (napríklad „jaký“,

„který“) obsahuje informácie o rode, čísle a páde, ale nepodarilo sa zistiť žiadne pravidlo na ich aplikovanie.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

P4													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	x	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	x	PZ	○	Vt	x		

### Test kontrola\_P5

Zámeno „on“ po predložke, podobne ako iné zámená, obsahuje morfológické informácie, vo vzťahu k za nimi stojacimi slovami je ale nepodstatná a nevieme ju využiť.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

P5													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	x	PZ	○	Vt	x		

### Test kontrola\_P6

Pre reflexívne zámeno „se“ v dlhých tvaroch procedúra opäť nedisponuje žiadnym pravidlom, ktoré by využívalo jeho morfológickú informáciu.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

P6													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○

AM	○	Cn	○	Dg	○	P6	x	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	x	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	○	PZ	○	Vt	x		

### Test kontrola\_P7

Pre reflexívne zámená „se“, „si“, „ses“ a „sis“ nie sú známe žiadne pravidlá využívajúce morfológickú informáciu. Test sa preto zameriava len na kontrolu povolených a nepovolených kolokácií.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

P7													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	x	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	○	PZ	○	Vt	x		

### Test kontrola\_P8

Privlastňovacie zámeno „svůj“ obsahuje morfológické informácie o rode, čísle a páde a tie test využíva pri testovaní správnosti kolokácií s podstatnými menami, niektorými adjektívami a zámenami.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

P8													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	345	Cj	○	Cz	○	P1	x	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	x	PK	○	Ve	○	RV	○
AG	345	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	345	Cn	○	Dg	○	P6	x	PP	○	Vi	○	ZZ	○
AO	x	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	345	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	345	PY	○	Vs	○		
Cd	○	Cw	○	NN	345	PE	x	PZ	○	Vt	x		

### Test kontrola\_P9

Vzáložné zámená „jenž“, „již“ po predložke sú testovacie na nasledujúce možné kolokácie uvedené v tabuľke.

P9													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	C1	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	x	PZ	○	Vt	x		

### Test kontrola\_PD

Ukazovacie zámená sa zhodujú v rode, čísle a páde s adjektívami, podstatnými menami a niektorými zámenami...

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PD													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	345	Cj	○	Cz	○	P1	345	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	345	PK	○	Ve	○	RV	○
AG	345	C1	○	Db	○	P5	x	PL	345	Vf	○	TT	○
AM	345	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	345	Cr	○	J*	○	P8	○	PS	345	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	345	PE	○	PZ	○	Vt	x		

### Test kontrola\_PE

Za vzložným zámenom „což“ nemôžu nasledovať len niektoré špecifické druhy zámen a archaické druhy slovíes.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PE													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	x	PJ	x	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	x	PK	○	Ve	○	RV	○
AG	○	C1	○	Db	○	P5	x	PL	○	Vf	○	TT	○

AM	o	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	o	II	o	P7	o	PQ	x	Vm	o	<s>	o
AU	o	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	x	Vs	o		
Cd	o	Cw	o	NN	o	PE	x	PZ	o	Vt	x		

### Test kontrola\_PH

Krátke tvary osobných zámen majú podobné vlastnosti ako zámeno „což“ - nasledovať po nich nemôžu len niektoré zo zámen a archaické slovesá.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PH													
A2	o	Ch	o	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	o	Cz	o	P1	x	PJ	o	Vc	o	RR	o
AC	o	Ck	o	C}	o	P4	o	PK	o	Ve	o	RV	o
AG	o	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	o	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	o	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	o	Cw	o	NN	o	PE	o	PZ	o	Vt	x		

### Test kontrola\_PJ

Vzťažné zámeno „jenž“, „jíž“ a pod. bez predložky má okrem niektorých zakázaných kolokácií aj špecifickú vlastnosť a síce, že nemôže stáť na konci vety...

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PJ													
A2	o	Ch	o	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	o	Cz	o	P1	x	PJ	x	Vc	o	RR	o
AC	o	Ck	o	C}	o	P4	x	PK	o	Ve	o	RV	o
AG	o	Cl	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	o	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	x
AU	o	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	o	Cw	o	NN	o	PE	x	PZ	o	Vt	x		

### Test kontrola\_PK

Zámeno vzťažné a opytovacie „kdo“ (vrátane tvarov s „-ž“ a „-s“) má podobné vlastnosti ako krátke tvary osobných zámen a zámeno „což“.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PK													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	×	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	×	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	×	PZ	○	Vt	×		

### Test kontrola\_PL

Neurčité zámená sa s niektorými adjektívami, s podstatnými menami a s vybranými zámenami zhodujú v rode, čísle a páde.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PL													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	345	Cj	○	Cz	○	P1	×	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	345	Cl	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	345	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	345	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	345	PY	○	Vs	○		
Cd	○	Cw	○	NN	345	PE	×	PZ	○	Vt	×		

### Test kontrola\_PP

Pre osobné zámená procedúra neobsahuje žiadne testy, ktoré by využívali gramatické informácie.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PP													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		

C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	o	Cw	o	NN	o	PE	o	PZ	o	Vt	x		

### Test kontrola\_PQ

Vzážné opytovacie zámená typu „co“, „copak“ a „čožpak“ sú typom pravidiel podobné väčšine zámen – majú široké možnosti na kolokáciu s ostatnými slovnými druhmi.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PQ													
A2	o	Ch	o	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	o	Cz	o	P1	o	PJ	o	Vc	o	RR	o
AC	o	Ck	o	C}	o	P4	o	PK	o	Ve	o	RV	o
AG	o	C1	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	o	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	o	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	o	Cw	o	NN	o	PE	x	PZ	o	Vt	x		

### Test kontrola\_PS

Test kontrolujúci privlastňovacie zámená obsahuje najviac pravidiel pre kolokácie s adjektívami, podstatnými menami, číslovkami a predložkami.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PS													
A2	o	Ch	345	Cy	345	P0	x	PH	o	VB	o	RF	o
AA	345	Cj	45	Cz	o	P1	x	PJ	o	Vc	o	RR	o
AC	x	Ck	45	C}	o	P4	o	PK	x	Ve	o	RV	o
AG	345	C1	345	Db	o	P5	x	PL	x	Vf	o	TT	o
AM	345	Cn	345	Dg	o	P6	o	PP	x	Vi	o	ZZ	o
AO	x	Co	o	II	o	P7	o	PQ	x	Vm	o	<s>	o
AU	345	Cr	345	J*	o	P8	x	PS	x	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	x	Vq	x		
Ca	5	Cv	o	J^	o	PD	o	PY	x	Vs	o		
Cd	345	Cw	45	NN	345	PE	x	PZ	x	Vt	x		

### Test kontrola\_PW

Záporné zámená (napríklad „nic“, „nikdo“, „žádný“) obsahujú na rozdiel od privlastňovacích zámen málo pravidiel, aj napriek tomu, že nesú pomerne mnoho morfológických informácií.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.



PW													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	C1	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	○	PZ	○	Vt	x		

### Test kontrola\_PY

Vzťažné zámeno „co“ spojené s predložkou (napríklad „oč“, „nač“, „zač“) neobsahuje žiadnu podstatnú morfológickú informáciu.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PY													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	○	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	C1	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	x	Vs	○		
Cd	○	Cw	○	NN	○	PE	○	PZ	○	Vt	x		

### Test kontrola\_PZ

Neurčité zámená typu „nějaký“, „některý“, alebo „číkoli“ a „cosi“ sa zhodujú v rode, čísle a páde s niektorými druhmi adjektív a zámen a s podstatnými menami.

Morfologická informácia o páde a čísle však môže pri niektorých zámenách tohto poddruhu chýbať. Napriek tomu však je možné aplikovať viaceré morfológické pravidlá.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

PZ													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	345	Cj	○	Cz	○	P1	x	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	x	PK	○	Ve	○	RV	○
AG	345	C1	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	345	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○

AO	34	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	345	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	345	PY	o	Vs	o		
Cd	o	Cw	o	NN	345	PE	o	PZ	o	Vt	x		

## 8. Modul V (Kontrola slovies)

Modul obsahuje testy kolokácií pre každý zo slovesných poddruhov. Pri väčšine z nich nie sú známe žiadne pravidlá, ktoré by využívali bohatú morfológickú informáciu, ktorú väčšina slovies obsahuje. Výnimku tvoria poddruhy VB (všeobecné slovesá v prítomnom a budúcom čase) a Vp a Vs (trpné a činné prídavné). Každé zo slovies sa môže nachádzať na konci vety, preto tento test modul neobsahuje.

### Test kontrola\_VB

Test kontroluje všeobecné slovesá v prítomnom alebo budúcom čase. Pri kolokáciách s určitými druhmi adjektív sa testuje zhoda v čísle. Známych je tiež niekoľko zámen, ktoré za slovesom nemôžu nasledovať.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

VB													
A2	o	Ch	o	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	o	Cz	o	P1	x	PJ	o	Vc	o	RR	o
AC	4	Ck	o	C}	o	P4	o	PK	o	Ve	o	RV	o
AG	o	C1	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	4	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	4	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o
AU	o	Cr	o	J*	o	P8	o	PS	o	Vp	o		
C=	o	Cu	o	J,	o	P9	x	PW	o	Vq	x		
Ca	o	Cv	o	J^	o	PD	o	PY	o	Vs	o		
Cd	o	Cw	o	NN	o	PE	o	PZ	o	Vt	x		

### Test kontrola\_Vc

Kondicionál slovesa „být“ (napríklad „by“, „bych“, „bys“, „bychom“) poskytuje morfológickú informáciu o osobe a čísle. Bohužiaľ, ani jedna sa pri testovaní správnosti kolokácií nedá použiť.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Vc													
A2	o	Ch	o	Cy	o	P0	o	PH	o	VB	o	RF	o
AA	o	Cj	o	Cz	o	P1	x	PJ	o	Vc	x	RR	o
AC	o	Ck	o	C}	o	P4	o	PK	o	Ve	o	RV	o
AG	o	C1	o	Db	o	P5	x	PL	o	Vf	o	TT	o
AM	o	Cn	o	Dg	o	P6	o	PP	o	Vi	o	ZZ	o
AO	o	Co	o	II	o	P7	o	PQ	o	Vm	o	<s>	o

AU	○	Cr	○	J*	x	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	x	PZ	○	Vt	x		

### Test kontrola\_Ve

Prítomný prechodník napriek morfolologickej informácie, ktorú nesie neurčuje žiadne pravidlá, ktorými sa musí riadiť slovo stojace vo vete za ním. Známe sú však niektoré špecifické poddruhy slov, ktoré za ním nemôžu nasledovať.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Ve													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	x
AA	○	Cj	○	Cz	○	P1	x	PJ	x	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	x	Co	○	II	○	P7	○	PQ	○	Vm	x	<s>	○
AU	○	Cr	○	J*	x	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	x	Vs	○		
Cd	○	Cw	○	NN	○	PE	x	PZ	○	Vt	x		

### Test kontrola\_Vf

Infinitív disponuje jedinou morfolologicou informáciou, ktorú, podobne ako pri iných typoch slovies nevieme použiť pri určení správnosti kolokácie.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Vf													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	x	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	○	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	○	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	○	PZ	○	Vt	x		

### Test kontrola\_Vi

Rozkazovací spôsob slovies patrí do skupiny slovies, ktoré neurčujú žiadne pravidlo pre nasledujúci slovný druh. Jedinou výnimkou je jeho kolokácia so záporným slovesom, ktorá je korektná iba za podmienky, že sloveso je v negovanom tvare.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Vi													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	x
AA	○	Cj	○	Cz	○	P1	x	PJ	x	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	x	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	*	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	x	PZ	○	Vt	x		

\* Zámeno záporné (PW, napríklad „nic“, „nikdo“, „žiadny“) môže za rozkazovacím spôsobom nasledovať jedine ak sloveso je v negovanom tvare (má predponu „-ne“)

### Test kontrola\_Vm

Minulý prechodník určuje len zopár slovných poddruhov, ktoré nemôžu nasledovať. Uvádzame ich v tabuľke.

Vm													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	x
AA	○	Cj	○	Cz	○	P1	x	PJ	○	Vc	○	RR	○
AC	○	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	x	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	○	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	x	PW	*	Vq	x		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	x	PZ	○	Vt	x		

### Test kontrola\_Vp

Činné prídavné sa musí zhodovať v čísle s niekoľkými druhmi adjektív a nemôžu za ním nasledovať vybrané zámená a archaické tvary sloves.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Vp													
A2	○	Ch	○	Cy	○	P0	○	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	x	PJ	○	Vc	○	RR	○
AC	4	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	x	PL	○	Vf	○	TT	○
AM	4	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	4	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○

AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	○	PZ	○	Vt	×		

### Test kontrola\_Vs

Trpné prídavné má podobné vlastnosti ako činné – determinuje číslo adjektív a zakazuje niektoré zo zámen.

Tabuľka uvádza prehľad povolených a zakázaných kolokácií.

Vs													
A2	○	Ch	○	Cy	○	P0	×	PH	○	VB	○	RF	○
AA	○	Cj	○	Cz	○	P1	×	PJ	○	Vc	○	RR	○
AC	4	Ck	○	C}	○	P4	○	PK	○	Ve	○	RV	○
AG	○	Cl	○	Db	○	P5	×	PL	○	Vf	○	TT	○
AM	4	Cn	○	Dg	○	P6	○	PP	○	Vi	○	ZZ	○
AO	4	Co	○	II	○	P7	○	PQ	○	Vm	○	<s>	○
AU	○	Cr	○	J*	○	P8	○	PS	○	Vp	○		
C=	○	Cu	○	J,	○	P9	×	PW	○	Vq	×		
Ca	○	Cv	○	J^	○	PD	○	PY	○	Vs	○		
Cd	○	Cw	○	NN	○	PE	×	PZ	○	Vt	×		

## 10. Výsledky na testovacích dátach

Pri testovaní procedúry sme mali k dispozícii dva súbory viet. Jedným zo súborov bola vzorka viet z Českého akademického korpusu, druhým potom sada viet dodaných spoločnosťou Seznam.cz. V nasledujúcej časti dokumentácie sú uvedené výsledky procedúry pre tieto súbory testovacích dát.

### 1. Český akademický korpus

Celkom bolo k testovaniu pripravených 28 575 viet. Z nich 24 341 bolo zmysluplných a 4 234 nezmysluplných. Nasledujúca tabuľka uvádza prehľad viet tak, ako ich vyhodnotila automatická procedúra. Súbor obsahujúci zmysluplné vety je označený ako OK, súbor s nezmysluplnými vetami ako KO

	Súbor OK	Súbor KO
Zmysluplné vety	78,28 %	61,43 %
Nezmysluplné vety	21,72 %	38,57 %

Z výsledkov vyplýva, že procedúra má veľmi peknú úspešnosť pri posudzovaní zmysluplných viet – len približne každú piatu zmysluplnú vetu označí chybné. Podstatne horšie výsledky však procedúra ukázala pri analyzovaní nezmysluplných viet.

Nelichotivé výsledky však procedúra dosiahla hlavne vďaka nie veľmi šťastnému výberu nezmysluplných viet. Tieto vety boli vybrané z Českého akademického korpusu 1.0, ktorý si nekládol za cieľ tagovať číslovky písané číslom a tie z textov rovno vyhadzoval. Namiesto nich do textu doplnil značku „?“ . Súbor nezmysluplných viet vznikol vynechaním týchto otáznikov a teda vety na rôznych pozíciách majú vynechané číslo. Nie všetky vety takto vytvorené však sú nezmysluplné. Príkladom môže byť napríklad nadpis „*Odstavec 34.*“. Veta po vynechaní čísla je v tvare „*Odstavec.*“, čo je však dostatočne zmysluplná veta.

Po ručnom prejení prvých 200 viet z KO súboru bolo objavených len 11 skutočne nezmysluplných viet.

K výsledkom procedúry neprispieva ani tagger *tool\_chainu*, ktorý konečný tag vyberá tak, aby sa zhodoval v páde so susednými slovami. Takže procedúra nezhdou v páde môže objaviť jedine v prípade, ak tagger nenájde tag (hoci zlý), ktorý by sa v páde zhodoval.

Z pomedzi prvých 200 viet KO súboru bolo objavených 14 viet, ktoré tagger vyhodnotil nesprávne. Medzi nesprávne analyzované slová patrili napríklad:

- vo vete *...uvedenými v indikačnóm seznamu pod a*. Slovo „a“ bolo označené ako spojka. Šlo o chybu taggera, ktorý z ponúkaných morfológických značiek nevybral správnu možnosť (podstatné meno).
- vo vete *...před skončením lhůty stanovené indikací*. Tagger zle vybral z pomedzi ponúkaných morfológických značiek značku pre slovo „indikací“. Uprednostil značku s množným číslom, ktorá sa zhodovala v páde s predchádzajúcimi slovami, pred značkou s jednotným číslom.
- slovné spojenie *odborné učiliště*, v ktorom slovo „odborné“ otagoval mužským rodom.

## 2. Seznam.cz

Dáta od spoločnosti Seznam.cz sa skladali z dvoch súborov. OK súbor obsahoval 158 viet, KO súbor zasa 182 viet. V tabuľke sú uvedené výsledky procedúry:

	Súbor OK	Súbor KO
<b>Zmysluplné vety</b>	71,52 %	37,36 %
<b>Nezmysluplné vety</b>	28,48 %	62,64 %

Procedúra si zachovala priaznivý podiel pri vyhodnocovaní zmysluplných viet a pekne si polepšila pri posudzovaní viet nezmysluplných.

Aj pri tomto súbore však vidno, aký je problém zostaviť súbor nezmysluplných viet, čo sa morfológického posudzovania týka. Tento súbor vznikol náhodným preberaním rôznych textov z webových stránok. A tak sa medzi vetami ocitla napríklad informácia o čase a dátume, rôzne hypertextové odkazy alebo dotazy, ktoré boli zadané do vyhľadávacieho formulára. Niektoré dokonca v angličtine. Tento súbor by bol však vhodnejší skôr na formálny test zmysluplnosti, nie na morfológický – vety

nezačínali veľkými písmenami, mnohé z nich boli len textami, ktoré na pôvodnej stránke tvorili menu (šlo o podstatné mená oddelené rôznymi *delimitermi*) a pod.

## 11. Literatúra a zdroje

- [1] Sprievodca ČAK 2.0, sekcia 3.3.5. Automatické zpracování textů  
<http://ufal.mff.cuni.cz/morce/cac/?chapter=3#nastroje-zprac>
- [2] Sprievodca ČAK 2.0, sekcia 3.2.1. Formát dat  
<http://ufal.mff.cuni.cz/morce/cac/?chapter=3#data-format>
- [3] Pražský závislostný korpus 2.0, sekcia 3.4.4. CSTS  
<http://ufal.mff.cuni.cz/pdt2.0/doc/pdt-guide/cz/html/ch03.html#a-data-formats-csts>
- [4] Pražský závislostný korpus 2.0, špecifikácia CPMT  
<http://ufal.mff.cuni.cz/pdt2.0/doc/data-formats/csts/>