

**Univerzita Karlova v Praze**  
**Filozofická fakulta**  
Ústav teoretické a počítačnické lingvistiky

Tomáš Jelínek

**Forma a funkce u substantiv v češtině:  
vztah pádu a syntaktické funkce**

*Na materiálu korpusu současné psané češtiny (SYN2005)*

**Form and function of nouns in Czech:  
relation between nominal case and syntactic function**

*Based on a synchronic written corpus of Czech (SYN2005)*

Disertační práce

Studijní program: Filologie

Studijní obor: Matematická lingvistika

Vedoucí práce – doc. RNDr. Vladimír Petkevič, CSc.

Praha 2012



Prohlašuji, že jsem disertační práci vykonal samostatně s využitím uvedených pramenů a literatury.

## **Poděkování**

*Děkuji svému školiteli, docentu Vladimíru Petkevičovi, za trpělivé vedení práce a cenné připomínky.*

*Děkuji kolegům v Ústavu teoretické a počítačové lingvistiky za podporu.*

*Děkuji své ženě Marii za podporu, inspiraci a motivaci k dokončení práce.*

## Shrnutí

Pád je v češtině základním prostředkem morfologické roviny, jímž substantiva vyjadřují svou funkci ve větě. Cílem této práce je popsat z frekvenčního hlediska vztah mezi formou a funkcí substantiv, přesněji řečeno, jak často se prosté a předložkové pády substantiv používají k realizaci syntaktických funkcí ve větě. Práce je založena na rozsáhlém korpusu synchronní psané češtiny SYN2005.

Abychom získali údaje o frekvencích syntaktických funkcí substantiv ve vztahu k jejich pádům, opatřili jsme korpus SYN2005 závislostním syntaktickým značkováním, jehož formát jsme převzali z analytické roviny Pražského závislostního korpusu. Syntaktickou anotaci jsme uskutečnili pomocí stochastického MST parseru. Spolehlivost syntaktické anotace však nebyla dostatečně vysoká, vytvořili jsme proto automatický opravný modul, který vyhledává chyby syntaktické anotace ve výstupu stochastického parseru a na základě lingvistických pravidel tyto chyby opravuje. Implementovali jsme 26 různých pravidel, počet chyb anotace se však podařilo snížit jen o 6–8 %. Opravný modul je však možné dále rozvíjet. Lze jím korigovat výstup kteréhokoli závislostního parseru natrénovaného na datech Pražského závislostního korpusu.

Syntakticky anotovaný korpus SYN2005 jsme využili jako základ výzkumu frekvence syntaktických funkcí substantiv se zřetelem k jejich pádu. Manuálními sondami do korpusu jsme ověřili a korigovali získané statistické údaje. Podrobně jsme analyzovali relevantní kombinace syntaktické funkce a prostého nebo předložkového pádu. Frekvence syntaktických funkcí a jejich realizací pády substantiv jsme představili v přehledných tabulkách, a to pro celý korpus SYN2005 i pro jeho tři subkorpusy: beletrie, odborné literatury a publicistiky. Tyto subkorpusy jsme porovnali jak podle celkové frekvence syntaktických funkcí substantiv, tak podle použití některých kombinací syntaktické funkce a pádu.

## Abstract

The case in Czech is the basic morphological means by which nouns express their function in a sentence. The objective of this thesis is to describe, from a frequency point of view, the relation between form and function of nouns, or, more precisely, how frequently cases (both simple and prepositional) are used to realise syntactic functions in sentences. The thesis is based on one of the largest corpora of written synchronic Czech: 100-million-token corpus SYN2005.

In order to obtain data on frequencies of syntactic functions of nouns in relation to their cases, we annotated the corpus SYN2005 with a dependency syntactic annotation. For this annotation, we adopted the format of the analytical layer of the Prague Dependency Treebank. The syntactic annotation has been performed by a stochastic parser: the MST parser. Since the reliability of this annotation was not high enough, we have built an automatic correction module, which identifies errors of syntactic annotation in the output of the stochastic parser and corrects these errors by means of linguistic rules. We have implemented 26 different rules, but annotation errors have been reduced by merely 6–8%. However, this correction module can be further developed. It can be used to correct the output of any dependency parser trained on the data from the Prague Dependency Treebank.

We used the syntactically annotated corpus SYN2005 as the basis for our research of frequencies of syntactic functions of nouns in relation to their cases. Using manual probes into the corpus, we have verified and corrected the statistical data obtained. We analysed and described in detail all the relevant combinations of syntactic functions and simple or prepositional cases. The frequencies of the syntactic functions of nouns and their realisation have been presented in synoptic tables, presenting data from the corpus SYN2005 and its three subcorpora of fiction, scientific/technical literature and journalism. We compared these three subcorpora according to the frequencies of syntactic functions of nouns, as well as according to the details of usage of several combinations of syntactic function and case.

# Obsah

Úvod.....	11
<b>Část první. Syntakticky anotovaný korpus.....</b>	<b>13</b>
Úvod.....	13
I. Korpus SYN2005 a systém závislostní anotace PDT.....	14
1. Korpus SYN2005.....	14
1.1 Morfologická anotace a lematizace korpusu.....	14
1.2 Systém automatické morfologické anotace.....	15
2. Syntaktická anotace korpusu.....	15
2.1 Syntaktická anotace v Pražském závislostním korpusu.....	16
2.1.1 Pražský závislostní korpus.....	16
2.1.2 Analytická rovina PDT.....	16
2.1.2.1 Závislostní struktura na analytické rovině PDT.....	16
2.1.2.2 Syntaktické funkce na analytické rovině PDT.....	17
2.1.3 Nástroje pro práci s anotovanými texty.....	18
2.1.3.1 TectoMT.....	18
2.1.3.2 TrEd.....	19
II. Automatická syntaktická anotace.....	20
1. Automatická syntaktická anotace založená na pravidlech.....	20
2. Stochastická syntaktická anotace.....	20
2.1 Algoritmus pro určení závislostí v MST parseru.....	21
3. Analýza chyb automatické anotace korpusu SYN2005.....	22
III. Opravný modul.....	23
1. Základní charakteristiky opravného modulu.....	23
1.1 Prostředí a programovací jazyk.....	23
1.2 Koncepce programu a pravidel.....	24
1.3 Běh programu.....	24
1.4 Použité seznamy slov a jejich vlastností.....	24
1.5 Určení hranic klauzí.....	26
1.6 Určení sloves v určitém tvaru a reprezentantů složených slovesných tvarů.....	26
1.7 Určení hlavních a vedlejších vět.....	26
1.8 Oprava základní struktury věty.....	27
1.9 Oprava jednotlivých větných členů.....	27
1.10 Uložení věty a zápis vertikály.....	27
2. Jednotlivá opravná pravidla.....	28
2.0 Způsob prezentace jednotlivých pravidel.....	28
2.0.1 Příklady chyb.....	28
2.0.1.1 Textové zobrazení.....	29
2.0.2 Počet zásahů v korpusu SYN2005 a hodnocení úspěšnosti.....	31
2.0.3 Rozdělení pravidel podle zaměření.....	32
2.1 Pravidla korigující chybně určenou základní závislostní strukturu složených vět či souvětí.....	32
2.1.1 Pravidlo pro opravu koordinace nekompatibilních členů, která se považuje za řídicí člen závislostní struktury.....	33
2.1.2 Pravidlo pro slovesa ve vedlejších větách chybně považované za řídicí člen závislostní struktury.....	34
2.1.3 Pravidlo pro opravu závislostí souřadných hlavních vět.....	36
2.1.4 Pravidlo pro ověření závislostí a funkcí vztažných vět.....	38
2.1.5 Pravidlo pro opravu závislostí a funkcí u sloves se spojku <i>-li</i> .....	40
2.2 Obecná pravidla ověřující závislosti a syntaktické funkce větných členů.....	41

2.2.1 Pravidlo pro větné členy, jejichž závislost chybně překračuje hranici klauzí	42
2.2.2 Pravidlo pro spojení se spojky jako	43
2.2.3 Pravidlo pro určení syntaktické funkce reflexiva <i>se</i>	44
2.2.4 Pravidlo pro promítnutí změny morfologické značky substantiva na jeho shodné přívlastky	46
2.3 Pravidla pro opravu závislostí a syntaktických funkcí syntaktických substantiv	47
2.3.1 Pravidlo pro opravu shodných substantivních přívlastků	47
2.3.1.1 Pravděpodobně chybně určené shodné substantivní přívlastky	48
2.3.1.2 Pravděpodobné shodné substantivní přívlastky neurčené jako takové	49
2.3.2 Pravidlo pro opravu dvou subjektů závislých na jednom slovese	50
2.3.2.1 Dva subjekty závislé na tranzitivním slovese, oba uvnitř klauze	51
2.3.2.2 Dva subjekty závislé na slovese <i>být</i> , oba uvnitř klauze	52
2.3.2.3 Dva subjekty těsně vedle sebe, pojmenování osoby	52
2.3.2.4 Dva subjekty závislé na jednom slovese, jeden přes jednu hranici klauze	52
2.3.2.6 Osamostatněné přívlastky	54
2.3.2.7 Dva subjekty závislé na jednom slovese, jeden subjekt přes hranici klauzí	55
2.3.3 Pravidlo pro opravu podmětu v bezpředložkovém akuzativu	55
2.3.4 Pravidlo pro opravu koordinace nekompatibilních syntaktických funkcí	57
2.3.5 Pravidlo pro opravu přímého předmětu závislého na netranzitivním slovese	58
2.3.6 Pravidlo pro opravu předmětu závislého na modálním či fázovém slovese	60
2.4 Pravidla pro opravu závislostí a syntaktických funkcí předložkových frází	61
2.4.1 Pravidlo pro opravu syntaktických funkcí větných členů závislých na předložce	62
2.4.1.1 Chybné syntaktické funkce závislé přes předložku na (syntaktickém) substantivu	62
2.4.1.2 Chybné syntaktické funkce závislé přes předložku na slovese nebo adjektivu	63
2.4.2 Oprava chybného určení syntaktické funkce substantiva v předložkové frázi závislé na slovese	64
2.4.3 Pravidlo pro určení závislosti předložkových frází na rekčních substantivech	66
2.4.4 Předložkové fráze jako přívlastky substantiv bez odpovídající reke	68
2.4.5 Pravidlo pro opravu koordinací závislých na předložkách	68
2.4.6 Pravidlo pro opravy chyb u víceslovných předložkových výrazů	70
2.4.7 Pravidlo pro doplnění neuspokojených rekčních požadavků předložek	72
2.4.8 Pravidlo pro opravu závislosti v předložkové frázi se zájmenem	74
2.4.9 Pravidlo pro opravu subjektu v předložkové frázi	76
2.4.10 Pravidlo pro předložkovou frázi s předložkou <i>o</i> před komparativem	77
2.4.11 Pravidlo pro neprojektivní spojení předložky, adjektiva, číslovky a slovesa	78
3. Celková úspěšnost opravného modulu	80
3.1 Úspěšnost opravného modulu v korpusu SYN2005	80
3.1.1 Tabulka úspěšnosti jednotlivých pravidel a celého modulu	80
3.1.2 Tabulka typu změn pro jednotlivá pravidla i celý opravný modul	82
3.2 Úspěšnost pravidel měřená na testovacích datech PDT	84
3.2.1 „Accuracy“ použitého modelu MST parseru a opravného modulu	84
3.2.2 Tabulka úspěšnosti jednotlivých pravidel a celého modulu	84
3.2.3 Tabulka typu změn pro jednotlivá pravidla i celý opravný modul	86
3.3 Závěr	88
4. Další možný rozvoj opravného modulu	88

4.1 Oprava chybujících pravidel a doplnění algoritmů.....	89
4.2 Doplnění dalších pravidel.....	89
4.3 Rozšíření používaných seznamů.....	89
4.4 Zapojení frazémového modulu.....	90
IV. Jiné možnosti zlepšení automatické syntaktické anotace.....	91
1. Lepší využití existujících parserů, kombinace parserů.....	91
2. Doplnění a zlepšení pravidlového parseru.....	92
3. Pravidlové předzpracování dat a úprava tagsetu.....	92
3.1 Hlavní překážky kvalitního parsingu.....	92
3.1.1 Morfologické značky.....	92
3.1.2 Trénovací data.....	93
3.1.3 Složité struktury (koordinace).....	93
3.2 Návrh předzpracování trénovacích i nových dat za účelem zlepšení parsingu....	94
3.2.1 Přizpůsobení morfologického tagsetu potřebám syntaktického parseru.....	94
3.2.1.1 Zúžení a zjednodušení morfologické značky.....	94
3.2.1.2 Syntaktická substantiva a syntaktická adjektiva.....	94
3.3 Doplnění lingvistických informací.....	95
3.3.1 Valence a další vlastnosti slova.....	95
3.3.2 Hranice klauzí a charakteristika koordinační spojky.....	96
4. Závěr.....	96
<b>Část druhá. Syntaktické funkce substantiv a jejich realizace pády.....</b>	<b>97</b>
Úvod.....	97
I. Syntaktické funkce, pády a funkční styly.....	98
1. Syntaktické funkce substantiv.....	98
1.1 Podmět ( <i>Sb</i> ).....	99
1.2 Jmenná část verbonominálního přísudku ( <i>Pnom</i> ).....	99
1.3 Přívlastek ( <i>Atr</i> ).....	100
1.4 Předmět ( <i>Obj</i> ).....	100
1.5 Přísluvečné určení ( <i>Adv</i> ).....	101
1.6 Elipsa ( <i>ExD</i> ).....	101
2. Pády substantiv.....	102
2.1 Pojetí pádu.....	102
2.2 Prosté a předložkové pády substantiv.....	103
3. Funkční styly.....	104
4. Zpracování dat pro statistické účely.....	105
4.1 Ověření a korekce údajů.....	105
4.2 Zjednodušení dat.....	106
II. Syntaktické funkce substantiv realizované prostými a předložkovými pády.....	107
1. Tabulka frekvence syntaktických funkcí substantiv a jejich realizace pády v korpusu SYN2005.....	107
2. Popis jednotlivých realizací syntaktických funkcí pády.....	108
2.1 Podmět ( <i>Sb</i> ).....	109
2.1.1 Podmět v prostém nominativu ( <i>Sb_1</i> ).....	109
2.1.2 Podmět v prostém genitivu ( <i>Sb_2</i> ).....	109
2.1.3 Podmět v předložkových pádech ( <i>Sb_p2</i> , <i>Sb_p3</i> , <i>Sb_p6</i> ).....	110
2.2 Jmenná část verbonominálního přísudku ( <i>Pnom</i> ).....	111
2.2.1 Jmenná část verbonominálního přísudku v prostém nominativu ( <i>Pnom_1</i> ).....	111
2.2.2 Jmenná část verbonominálního přísudku v prostém genitivu ( <i>Pnom_2</i> ).....	112
2.2.3 Jmenná část verbonominálního přísudku v prostém instrumentálu ( <i>Pnom_7</i> ).....	112
2.3 Přívlastek ( <i>Atr</i> ).....	112
2.3.1 Shodný přívlastek ( <i>Atr</i> ).....	113

2.3.2	Přívlastek v prostém nominativu ( <i>Atr_1</i> ).....	115
2.3.3	Přívlastek v prostém genitivu ( <i>Atr_2</i> ).....	116
2.3.4	Přívlastek v předložkovém genitivu ( <i>Atr_p2</i> ).....	118
2.3.5	Přívlastek v prostém dativu ( <i>Atr_3</i> ).....	119
2.3.6	Přívlastek v předložkovém dativu ( <i>Atr_p3</i> ).....	120
2.3.7	Přívlastek v prostém akuzativu ( <i>Atr_4</i> ).....	122
2.3.8	Přívlastek v předložkovém akuzativu ( <i>Atr_p4</i> ).....	122
2.5.9	Přívlastek v prostém vokativu ( <i>Atr_5</i> ).....	124
2.5.10	Přívlastek v předložkovém lokálu ( <i>Atr_p6</i> ).....	124
2.5.11	Přívlastek v prostém instrumentálu ( <i>Atr_7</i> ).....	126
2.5.12	Přívlastek v předložkovém instrumentálu ( <i>Atr_p7</i> ).....	128
2.6	Předmět ( <i>Obj</i> ).....	130
2.6.1	Předmět v prostém nominativu ( <i>Obj_1</i> ).....	130
2.6.2	Předmět v prostém genitivu ( <i>Obj_2</i> ).....	130
2.6.3	Předmět v předložkovém genitivu ( <i>Obj_p2</i> ).....	133
2.6.4	Předmět v prostém dativu ( <i>Obj_3</i> ).....	135
2.6.5	Předmět v předložkovém dativu ( <i>Obj_p3</i> ).....	137
2.6.6	Předmět v prostém akuzativu ( <i>Obj_4</i> ).....	139
2.6.7	Předmět v předložkovém akuzativu ( <i>Obj_p4</i> ).....	141
2.6.8	Předmět v předložkovém lokálu ( <i>Obj_p6</i> ).....	143
2.6.9	Předmět v prostém instrumentálu ( <i>Obj_7</i> ).....	146
2.6.10	Předmět v předložkovém instrumentálu ( <i>Obj_p7</i> ).....	148
2.7	Příslovečné určení ( <i>Adv</i> ).....	150
2.7.1	Příslovečné určení v prostém nominativu ( <i>Adv_1</i> ).....	151
2.7.2	Příslovečné určení v prostém genitivu ( <i>Adv_2</i> ).....	152
2.7.3	Příslovečné určení v předložkovém genitivu ( <i>Adv_p2</i> ).....	152
2.7.4	Příslovečné určení v prostém dativu ( <i>Adv_3</i> ).....	154
2.7.5	Příslovečné určení v předložkovém dativu ( <i>Adv_p3</i> ).....	154
2.7.6	Příslovečné určení v prostém akuzativu ( <i>Adv_4</i> ).....	155
2.7.7	Příslovečné určení v předložkovém akuzativu ( <i>Adv_p4</i> ).....	157
2.7.8	Příslovečné určení v prostém vokativu ( <i>Adv_5</i> ).....	158
2.7.9	Příslovečné určení v předložkovém lokálu ( <i>Adv_p6</i> ).....	158
2.7.10	Příslovečné určení v prostém instrumentálu ( <i>Adv_7</i> ).....	159
2.7.11	Příslovečné určení v předložkovém instrumentálu ( <i>Adv_p7</i> ).....	160
2.8	Aktuální elipsy a řídicí větné členy výpovědi bez slovesa ( <i>ExD</i> ).....	161
2.8.1	Elipsa v prostém nominativu ( <i>ExD_1</i> ).....	162
2.8.2	Elipsa v prostém vokativu ( <i>Exd_5</i> ).....	163
2.8.3	Substantiva se syntaktickou funkcí <i>ExD</i> v jiných pádech než v nominativu a vokativu .....	164
2.9	Závěr.....	165
III.	Frekvence syntaktických funkcí a jejich realizací pády substantiv v korpusu SYN2005 .....	166
1.	Tabulky frekvence syntaktických funkcí a pádů substantiv.....	166
1.1	Podíl realizací syntaktických funkcí pády na všech substantivech.....	168
1.2	Podíl pádů na syntaktických funkcích substantiv.....	170
1.3	Podíl syntaktických funkcí na pádech substantiv.....	172
1.4	Podíl realizací syntaktických funkcí pády na všech substantivech (ppm).....	174
2.	Srovnání subkorpusů podle syntaktických funkcí a jejich realizace pády.....	176
2.1	Srovnání žánrů podle frekvence syntaktických funkcí.....	176
2.2	Nominální skupiny v subkorpusech BEL, ODB a PUB.....	177
2.3	Srovnání žánrů v korpusu podle realizace verbonominálního predikátu ( <i>Pnom</i> ) .....	178
2.4	Srovnání žánrů v korpusu podle užití přívlastku v prostém genitivu ( <i>Atr_2</i> )....	180
2.5	Předmět v prostém instrumentálu ( <i>Obj_7</i> ).....	181

2.6 Další rozdíly mezi žánrovými subkorpusy.....	182
2.7 Závěr.....	183
3. Srovnání frekvencí syntaktických substantiv v korpusu SYN2005 a v Kvantitativních charakteristikách současné češtiny.....	183
3.1 Kvantitativní charakteristiky současné češtiny.....	183
3.2 Srovnání frekvencí syntaktických funkcí substantiv v SYN2005 a v K.Ch.....	185
3.2.1 Přizpůsobení údajů o syntaktických funkcích substantiv z této práce Kvantitativním charakteristikám současné češtiny.....	185
3.2.2 Společné tabulky syntaktických funkcí substantiv ze SYN2005 a z K.Ch	186
4. Závěr.....	187
<b>Závěr.....</b>	<b>189</b>
<b>Literatura.....</b>	<b>191</b>

# Úvod

Cílem této práce je získat a představit statistické údaje o formě a funkcích substantiv v češtině, o vztahu pádu, popř. předložky a pádu substantiv a jejich syntaktické funkce ve větě na materiálu rozsáhlého textového korpusu současné psané češtiny, konkrétně na korpusu SYN2005. Statistika tohoto typu pro češtinu dosud chybí, srovnatelná starší práce (Těšitelová et al. 1985) vychází z korpusu řádově menšího, který nezahrnuje beletrii, a zanedbává rozdíl mezi prostými a předložkovými pády. Statistiky vhodně doplňují publikaci *Statistiky češtiny* (Bartoň et al. 2009), která vychází ze stejného korpusu a uvádí údaje z nižších rovin jazykového popisu. Práce je rozdělena do dvou částí: v první části je popsána metoda automatické syntaktické anotace použitého korpusu, druhá část popisuje vztah syntaktických funkcí substantiv a jejich pádů a předložek.

Statistické údaje o pádech a funkcích substantiv vycházejí z korpusu SYN2005, který jsme opatřili závislostní syntaktickou anotací. Anotace vychází z formalismu a anotovaných dat Pražského závislostního korpusu. K automatickému značkování byl použit stochastický parser, jehož výsledky byly korigovány opravným modulem založeným na lingvistických pravidlech. Tento modul je vlastním tvůrčím přínosem autora k automatické syntaktické anotaci, jeho popis a hodnocení jeho úspěšnosti tvoří většinu první části práce.

Vztah syntaktických funkcí substantiv, jejich pádů a použití předložek je popsán ve druhé části práce. Funkce a forma substantiv jsou představeny v obecných přehledných tabulkách, kde lze najít údaje o tom, jak často je určitá funkce reprezentována kterými formami, i o tom, jak často je která forma použita k reprezentaci určité funkce. Tabulky umožňují mimo jiné i srovnání mezi funkčními styly (styl publicistický, odborný a beletrie). Jednotlivými kombinacemi funkce a formy se pak práce zabývá podrobněji, uvádí přesnější rozdělení kategorií včetně dílčích frekvencí, nejčastější představitele určité kategorie nebo jejich nejčastější řídicí větné členy aj. Ve statistikách jsou zahrnuty jen základní, nejčastější funkce, a to jednak pro přehlednost, jednak proto, že spolehlivost automatické anotace málo častých, specifických syntaktických funkcí je výrazně menší než u funkcí základních.

Tato práce má primárně praktický charakter, zabývá se anotací korpusu a analýzou excerpovaných dat. Přestože je pád substantiv ústředním pojmem této práce, neuvádíme zde podrobný přehled různých teoretických pojetí této kategorie. V české jazykovědné literatuře se teorii pádu již věnovalo nemalé úsilí, podrobný přehled poskytuje například (Uličný 2000), další podobný souhrn není zapotřebí. Odkazy na teoretické přístupy, které byly pro tuto práci inspirativní, uvádíme na příslušném místě.



# Část první

## Syntakticky anotovaný korpus

### Úvod

Za základ pro statistický výzkum vztahu funkce a formy substantiv v češtině jsme v této práci zvolili syntakticky anotovaný korpus psané češtiny. Srovnatelné údaje by bylo možné získat i z korpusu opatřeného pouze morfologickými značkami a lemmaty, ale taková práce by vyžadovala buď velké množství manuální práce, nebo vlastní automatické zpracování, které by se v mnohém podobalo automatickému syntaktickému značkování. Automatická syntaktická anotace, která se může opřít o pečlivě zpracovaný formalismus existujícího závislostního korpusu a o manuálně syntakticky označovaná data, tedy byla logickou volbou. Snaha o vytvoření co nejkvalitnější automatické syntaktické anotace nebude navíc zúročena jen ve výsledcích této práce, může být využita i v různých dalších projektech. Sama práce na zlepšování syntaktické anotace přispívá k lepšímu porozumění jazykovému systému a k jeho přesnějšímu popisu.

V první kapitole stručně představíme korpus SYN2005, na němž je tato práce založena, a systém povrchové syntaktické anotace, který jsme převzali z Pražského závislostního korpusu. Ve druhé kapitole popíšeme systém automatické syntaktické anotace. Třetí kapitola představuje vlastní přínos autora k automatickému syntaktickému značkování: systém pro automatickou opravu výstupů stochastického parseru založený na lingvistických pravidlech. Ve čtvrté kapitole navrhneme další možné postupy, jimiž lze zvýšit spolehlivost automatické syntaktické anotace.

# Kapitola I

## Korpus SYN2005 a systém závislostní anotace PDT

V této kapitole představíme korpus SYN2005, z něhož práce vychází, včetně použitého morfologického značkování a lemmatizace. Zdůvodníme využití syntaktické anotace a popíšeme zvolený systém závislostní anotace, který přebíráme z Pražského závislostního korpusu.

### 1. Korpus SYN2005

Korpus SYN2005 je synchronní reprezentativní korpus psané češtiny o rozsahu 100 milionů textových slov. Patří mezi rozsáhlé textové korpusy řady SYN, které vybudoval Ústav Českého národního korpusu.

Korpus SYN2005 jsme jako datový zdroj pro tuto práci zvolili proto, že z něj vychází i publikace Statistiky češtiny (Bartoň et al. 2009), která poskytuje přehled o fonologické, grafémické, morfemické a morfologické rovině jazyka. Korpus SYN2005 se skládá ze 40 % z beletristických textů, z 27 % z textů odborné literatury a z 33 % z publicistických textů, je žánrově vyváženější než starší korpus SYN2000 (15 % beletrie, 25 % odborné literatury a 60 % publicistiky).

Korpus obsahuje publicistické texty z let 2000–2004, odbornou literaturu z let 1990–2004 a beletrii publikovanou v letech 1990–2004 (s několika staršími výjimkami publikovanými v letech 1959–1989), přičemž podmínkou zařazení do korpusu je narození autora po roce 1880. Korpus zahrnuje jak původní české texty, tak texty překladové.

#### 1.1 Morfologická anotace a lemmatizace korpusu

Všechny slovní tvary v korpusu jsou opatřeny morfologickými značkami a základními tvary slov (lemmaty). Morfologické značky (tagy) jsou poziční, zahrnují údaje o slovním druhu a relevantních morfologických kategoriích, jako je jmenný rod, číslo, pád, osoba aj. (Hajič 2004b). Podrobný popis morfologických značek se nachází mimo jiné na <http://ucnk.ff.cuni.cz/bonito/znacky.php>. V první části této práce budeme v příkladech syntaktické anotace příležitostně využívat i části morfologických značek, obvykle prvních pět pozic ze značky, které zde stručně popíšeme.

První pozice tagu označuje slovní druh. Možné hodnoty první pozice jsou: **N** (podstatné

jméno), **A** (přídavné jméno), **P** (zájmeno), **C** (číslovka), **V** (sloveso), **D** (příslowce), **R** (předložka), **J** (spojka), **T** (částice), **I** (citoslovce), **Z** (interpunkce) a **X** (neznámé slovo, popř. slovo chybné či cizí).

Druhá pozice tagu označuje podtyp slovního druhu (např. rozlišuje zájmena osobní, neurčitá, negativní apod.) nebo podrobněji určuje morfologické kategorie (např. tvary slovesa). Možných hodnot je více než 70 (volba je omezena první pozicí), pro porozumění příkladům v této práci není tato pozice nutná.

Třetí pozice tagu se používá pro rozlišení jmenného rodu. Možné hodnoty (v použité verzi morfologické anotace) jsou: **M** (mužský životný rod), **I** (mužský neživotný rod), **F** (ženský rod), **N** (střední rod) a znak „-“ (tato kategorie není pro daný slovní tvar relevantní).

Čtvrtá pozice označuje číslo. Hodnoty mohou být: **S** (singulár), **P** (plurál), **D** (duál) a znak „-“ (kategorie není relevantní).

Pátá pozice tagu představuje pád. Hodnotou této kategorie může být číslo od **1** do **7** (čísla odpovídají konvenčnímu řazení pádů) nebo znak „-“ (kategorie není relevantní).

## 1.2 Systém automatické morfologické anotace

Rozsah korpusu SYN2005 neumožňuje manuální morfologickou anotaci a lemmatizaci, proto byl korpus označován automaticky. Pro podrobný popis postupu automatické morfologické anotace zde není prostor. Korpus SYN2005 v publikované verzi byl označován s využitím složité kombinace stochastického značkování a disambiguace založené na lingvistických pravidlech (Jelínek 2008). Pro tuto práci (i pro výše zmíněnou publikaci Statistiky češtiny) jsme využili jednodušší a mírně spolehlivější kombinaci, v níž je většina disambiguace provedena pomocí pravidlové disambiguace, po ní následuje stochastický tagger, který disambiguaci dokončí (Jelínek et al. 2011).

## 2. Syntaktická anotace korpusu

Cílem této práce je popsat a vyčíslit využití forem substantiv ve větných strukturách: jak často jsou jejich jednotlivé pády (prosté či předložkové) obligatorními rozvitými sloves, jak často se používají jako fakultativní určení okolností a podobně. K tomuto účelu je nezbytná syntaktická anotace.

Teoreticky by bylo možné provést stejný statistický výzkum i bez syntaktické anotace. Bylo by nutné vytvořit složité postupy pro určování předložky, s níž je určité substantivum spojené, postupy pro zjišťování řídicího větného členu a podobně, které by se zakládaly

pouze na morfoložické anotaci. Ve výsledku by však tento postup představoval větší množství práce s méně spolehlivým výsledkem než automatická syntaktická anotace.

## **2.1 Syntaktická anotace v Pražském závislostním korpusu**

Systém syntaktické anotace, který jsme zvolili pro tuto práci, je závislostní syntax z analytické roviny Pražského závislostního korpusu (PDT). PDT je korpus psané češtiny opatřený několika rovinami lingvistického popisu. Je to nejrozsáhlejší syntakticky anotovaný korpus češtiny, systém syntaktické anotace na rovině povrchové syntaxe (analytická rovina) je velmi podrobně propracovaný a výborně vyhovuje účelům této práce. Kromě vlastního syntaktického systému poskytuje pracoviště PDT i širokou škálu nástrojů pro práci s texty a zvlášť pro automatickou syntaktickou anotaci.

### **2.1.1 Pražský závislostní korpus**

Pražský závislostní korpus (Prague Dependency Treebank, PDT) obsahuje texty novin a odborných časopisů z let 1990–1995. Texty jsou doplněny třemi rovinami anotace, a to v různém rozsahu: 2 milióny slovních jednotek jsou anotovány na morfoložické rovině, 1,5 miliónu na rovině syntaktické a 0,8 miliónu na rovině sémantické. Sémantická rovina je doplněna rovněž o anotaci aktuálního členění a koreference. Podrobnější popis korpusu poskytuje mnoho článků a publikací, např. (Hajič et al. 2003), (Hajič 2004a) a webové stránky PDT: <http://ufal.mff.cuni.cz/pdt2.0>.

### **2.1.2 Analytická rovina PDT**

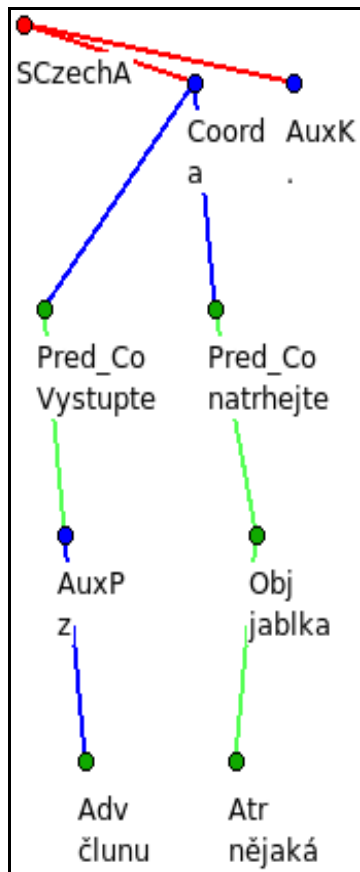
Syntaktická anotace použitá v této práci odpovídá analytické rovině PDT. Na této rovině je každá věta reprezentována závislostní strukturou, což je orientovaný strom (každé slovo odpovídá jednomu uzlu a má právě jeden uzel řídící) s jedním kořenem (kořen je formální, technický, stojí mimo větu). Každému slovu je přiřazena syntaktická funkce. Jako hrany ve stromě se zobrazují i vztahy, které nejsou závislostní (koordinace a apozice; vztah předložky a syntaktického substantiva; grafické znaky ve větě aj.).

Syntaktická anotace na analytické rovině vychází ze syntaktické koncepce Vladimíra Šmilauera (Šmilauer 1966), s drobnými úpravami teoretického i technického rázu.

#### **2.1.2.1 Závislostní struktura na analytické rovině PDT**

Závislostní struktura na analytické rovině reprezentuje vztahy ve větě formou orientovaného stromu. Hrany (větve) ve stromu zobrazují dva různé typy vztahů:

gramatickou závislost (např. vztah slovesa a jeho doplnění nebo vztah substantiva a jeho přívlasků) a vztahy technického rázu (např. součásti koordinace nebo vztah předložky a syntaktického substantiva). Na obrázku je zobrazen strom, který představuje v PDT závislostní strukturu příkladové věty: *Vystupte z člunu a natrhejte nějaká jablka*. Červeně je vyznačen formální kořen věty (SCzechA) a vztahy k němu, modře jsou vyznačeny „technické“ vztahy a pomocné uzly (gramatická slova), zeleně jsou vyznačeny uzly plnovýznamových slov a vztahy mezi nimi. Toto barevné rozlišení jsme použili pouze pro lepší orientaci v tomto obrázku, v dalších příkladech, kde budeme využívat obrázky závislostních struktur, ho používat nebudeme.



### 2.1.2.2 Syntaktické funkce na analytické rovině PDT

Každému textovému slovu včetně interpunkce je na analytické rovině PDT přiřazena syntaktická funkce. Tyto syntaktické funkce lze rozčlenit na základní syntaktické funkce odpovídající tradičním větným členům (subjekt, predikát, objekt, příslovečné určení aj.) a syntaktické funkce pomocné (označení předložky, pomocného slovesa, reflexiva tantum typu se aj.). Kombinované funkce řeší případy, kdy nelze ani podle kontextu a významu věty rozhodnout mezi dvěma potenciálními řídicími členy. V tabulce, kterou přebíráme z anotačního manuálu pro analytickou rovinu PDT (Hajič et al. 1999), uvádíme všechny syntaktické funkce, které mohou být přiřazeny na analytické rovině. Červeně zvýrazňujeme nejčastější syntaktické funkce substantiv, jimž se budeme nejvíce věnovat ve druhé části této práce. Modře zvýrazňujeme další důležité syntaktické funkce, které budeme opakovaně používat v příkladech v první části práce.

funkce	popis
<b>Pred</b>	Predikát, který nezávisí na jiném uzlu (kromě koordinace); věší se na kořen věty
<b>Sb</b>	Subjekt (podmět)
<b>Obj</b>	Objekt (předmět)
<b>Adv</b>	Adverbiale (příslovečné určení, bez dalšího rozlišení)
Atv	Doplněk (jen tzv. určující), technicky zavěšen na neslovesném členu
AtvV	Doplněk (jen tzv. určující), visící na slovese (chybí druhý řídicí člen)

funkce	popis
<b>Atr</b>	Atribut (přívlastek)
<b>Pnom</b>	Predikát nominální, resp. jmenná část přísudku se sponou <i>být</i>
<b>AuxV</b>	Pomocné sloveso být (Auxiliary Verb)
<b>Coord</b>	Koordinací uzel (souřadné spojení)
Apos	Aposice (hlavní uzel)
<b>AuxT</b>	Zvratné se, neoddělitelné se – u slovesa reflexiva tantum
<b>AuxR</b>	Zvratné se, které není Obj ani AuxT (tvoří pasivum reflexivní)
<b>AuxP</b>	Předložka primární, části předložky sekundární
<b>AuxC</b>	Spojka (podřadicí)
AuxO	Nadbytečný (odkazovací, emotivní) element
AuxZ	Zdůrazňovací slovo
AuxX	Čárka (ne však nositel koordinace)
AuxG	Jiné grafické symboly, které neukončují větu
AuxY	Příslovce a částice, které nelze zařadit jinam
AuxS	Kořen stromu (#)
AuxK	Koncová interpunkce věty
<b>ExD</b>	Náhradní funkce pro technické hrany vedoucí místo od elidovaného členu k „pseudořídícímu“ slovu nebo pro hlavní člen věty bez predikátu (Ex-Dependent)
AtrAtr	Řídícím slovem atributu může být díky strukturní víceznačnosti kterékoli z bezprostředně předcházejících (syntaktických) substantiv
AtrAdv	Strukturní víceznačnost mezi závislostí adverbální (přísllovečnou) a adnominální (zavěšení na jméno) bez sémantických důsledků
AdvAtr	Dtto, s opačnou preferencí
AtrObj	Strukturní víceznačnost mezi závislostí objektovou a adnominální (zavěšení na substantivum) bez sémantických důsledků
ObjAtr	Dtto, s opačnou preferencí

### 2.1.3 Nástroje pro práci s anotovanými texty

Na tomto místě musíme ještě zmínit nástroje, které pracovníci Ústavu formální a aplikované lingvistiky MFF UK vyvinuli pro práci s anotovanými texty a jež jsme využívali v této práci. Mezi tyto nástroje se řadí především softwarové prostředí pro počítačové zpracování textů přirozených jazyků (NLP) **TectoMT** a nástroj pro zobrazování a anotaci závislostních struktur **TrEd**.

#### 2.1.3.1 TectoMT

TectoMT (Žabokrtský et al. 2008) je modulární softwarové prostředí pro automatické zpracování přirozeného jazyka implementované v programovacím jazyce PERL. Primárně je určeno k automatickému překladu, ale díky tomu, že umožňuje provádět samostatně

jednotlivé kroky směřující k automatickému překladu, jako je morfologická a syntaktická analýza nebo různé konverze formátů, jsme ho využili jako vhodné prostředí pro automatickou syntaktickou analýzu korpusu a její opravu pomocí lingvistických pravidel. V prostředí TectoMT je také implementováno nastavení syntaktického analyzátoru, o němž pojednáme v příští kapitole: MST parseru Ryana McDonalda.

### 2.1.3.2 TrEd

**TrEd** je program pro vytváření a úpravu závislostních stromů. Pomocí TrEdu lze prohledávat anotované texty. TrEd nabízí grafické prostředí, v němž je možné editovat závislostní struktury a zobrazovat je. Všechna zobrazení závislostních struktur v této práci byla vytvořena s využitím editoru TrEd. Podrobnější popis tohoto editoru závislostních stromů se nachází na webových stránkách <http://ufal.mff.cuni.cz/tred/> .

# Kapitola II

## Automatická syntaktická anotace

Rozsáhlé textové korpusy, jako je SYN2005, není možné anotovat manuálně, morfologicky ani syntakticky. Texty je nutné anotovat automaticky. Pro automatickou anotaci obecně existují dvě základní varianty: anotace stochastická, která využívá manuálně označovaná data k „trénování“ anotačního programu (parseru), a anotace založená na (lingvistických) pravidlech. Obě varianty se dají i různými způsoby kombinovat.

V této kapitole představíme možnosti automatické syntaktické anotace, algoritmus syntaktického parseru, který jsme k anotaci vybrali, a celý postup provedené automatické syntaktické anotace korpusu SYN2005.

### 1. Automatická syntaktická anotace založená na pravidlech

Program pro automatickou anotaci, který morfologicky označovaný text doplňuje o závislostní syntaktické značkování ve formalismu PDT a který se zakládá na „manuálně“ vytvořených pravidlech, vyvinul Z. Žabokrtský (Holan et al. 2008). Program zahrnuje cca 40 pravidel. Úspěšnost tohoto programu v určení závislostí (bez ohledu na syntaktické funkce) je přibližně 76 %, což je cca o 8 procentních bodů méně než úspěšnost nejlepších stochastických parserů. Pokoušet se duplikovat tuto práci a překonat úspěšnost stochastických parserů pomocí jiného pravidlového parseru se nejevilo jako efektivní.

### 2. Stochastická syntaktická anotace

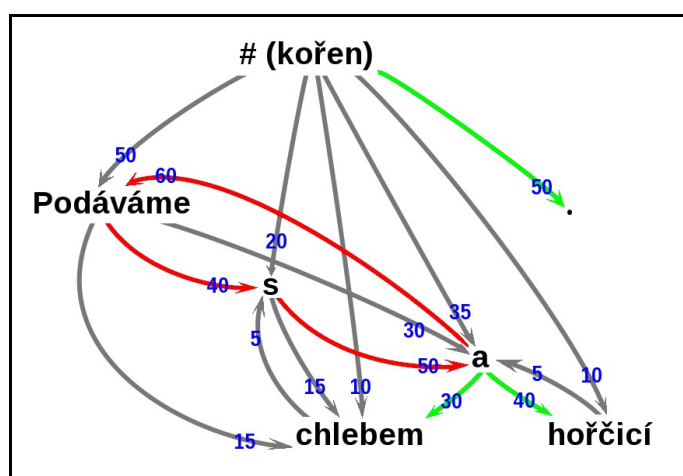
Stochastická syntaktická anotace je založena na principu „trénování“ parseru na vzorku správně označovaných dat a následné anotace nového textu. Parser na trénovacích datech zjišťuje frekvence (pravděpodobnosti) různých vztahů mezi slovy (větnými členy) podle předem daného nastavení (features). Získané informace využívá při anotaci neznámého textu.

Na PDT bylo natrénováno několik stochastických závislostních parserů s různými algoritmy a také s různou úspěšností anotace (mezi cca 50 % a 80 % úspěšnosti v přiřazení správné závislosti i správné syntaktické funkce). Jeden z nejlepších výsledků byl dosažen s MST parserem Ryana McDonalda (McDonald et al. 2005) v nastavení V. Nováka (Novák et. al 2007). Tento parser je implementován v prostředí TectoMT, je dostatečně rychlý i pro anotaci rozsáhlého, stamiliónového korpusu, a proto jsme ho pro syntaktickou anotaci korpusu SYN2005 využili.

## 2.1 Algoritmus pro určení závislostí v MST parseru

Algoritmus využitý v Maximum Spanning Tree (MST) parseru je podrobně vysvětlen ve výše citované publikaci (McDonald et al. 2005). Protože tento parser používáme, je na místě alespoň zjednodušeně popsat tento algoritmus i zde. Parser pracuje s větou jako s orientovaným grafem, v němž je každé (orientované) hraně, tj. možné závislosti jednoho slova na druhém, přiřazena určitá (pravděpodobnostní) hodnota. Cílem je vyhledat strom s maximální celkovou hodnotou, tedy acyklický orientovaný graf zahrnující všechny uzly původního grafu, přičemž součet hodnot hran je nejvyšší možný: **maximum spanning tree** (MST).

Algoritmus MST přiblížíme s pomocí příkladové věty *Podáváme s chlebem a hořčicí*. Hodnoty hran jsme dosadili víceméně náhodně tak, aby nakonec vyšla správná závislostní

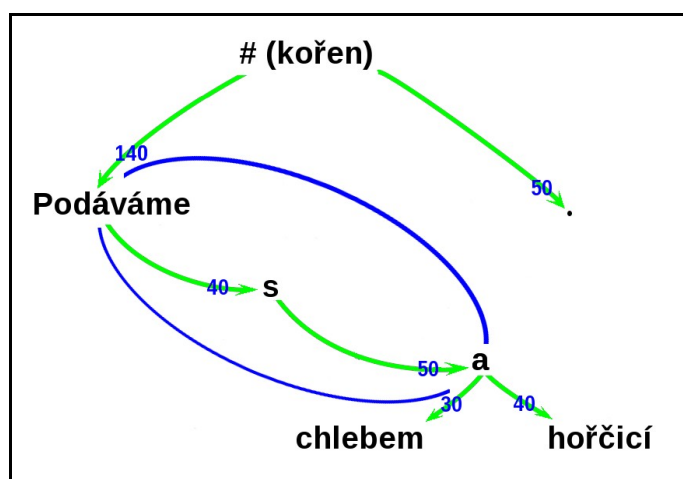


struktura. Na prvním obrázku je věta, kde jsou slova propojena orientovanými hranami s hodnotami (propojeny by měly být všechny uzly, ale pro přehlednost příkladu jsme zachovali jen některé hrany). Šipka směřuje od řídicího k závislému slovu.

Algoritmus nejprve vybere pro každé slovo vstupující hranu

s nejvyšší hodnotou. Pokud tyto hrany tvoří strom (nebo součást stromu), je vše v pořádku (označeno zeleně). Pokud hrany s nejvyššími hodnotami vytvoří cyklus (označeno červeně),

zachází se s každým cyklem samostatně jako s jedním uzlem. Pro tento uzel se najde vstupující hrana tak, že se sečte vstupující hrana do jednotlivého uzlu a zbytek cyklu (například vstup do červeně označeného cyklu přes slovo „s“ by měl hodnotu  $130 = 20 + 50 + 60$ : „kořen“ – 20 – „s“ – 50 – „a“ – 60 –



„Podáváme“; vstup do cyklu přes slovo „Podáváme“ hodnotu  $140 = 50 + 40 + 50$  apod.). Zvolíme nejvyšší vstup do cyklu, ostatní části cyklu zachováme. Takto se vyřeší všechny cykly v grafu, dokud není výsledkem strom. Výsledný strom má nejvyšší možnou celkovou

hodnotu hran.

Tento zjednodušený popis algoritmu nebere v potaz ani syntaktické funkce, ani další možná složitější nastavení parseru, ale umožňuje základní porozumění principu, na němž je parser založen.

MST parser se lépe než jiné stochastické parsery vyrovnává s neprojektivitou a se závislostmi na větší vzdálenosti; má naopak nižší úspěšnost v těsných, projektivních vztazích mezi větnými členy.

### 3. Analýza chyb automatické anotace korpusu SYN2005

Korpus SYN2005 jsme opatřili syntaktickým značkováním s využitím MST parseru. Při práci s takto anotovaným korpusem se však ukázalo, že přes použití poměrně spolehlivého parseru obsahuje syntaktická anotace příliš mnoho chyb. Hledali jsme tedy metodu, jak spolehlivost anotace zvýšit. V rámci dlouhodobého projektu morfologické disambiguace založené na lingvistických pravidlech vytvořilo naše pracoviště rozsáhlou databázi poznatků o českém jazyce, zdálo se tedy vhodné využít tyto znalosti ke zlepšení anotace syntaktické. Vlastní syntaktickou anotaci založenou na lingvistických pravidlech jsme vyloučili (viz oddíl 1. v této kapitole), místo toho jsme zvolili kombinaci stochastické a pravidlové syntaktické anotace: automatickou opravu výsledků stochastického parseru využívající lingvistická pravidla. Pro takovou opravu bylo nejprve nutné zjistit, jakých chyb se MST parser nejčastěji dopouští a zda je možné je automaticky opravit.

Podrobili jsme rozsáhlé vzorky syntakticky anotovaného korpusu (cca 10 000 vět z různých textů) manuální analýze, přičemž jsme vyhledávali opakující se typy chyb syntaktické anotace. Na základě tohoto předběžného výzkumu jsme vytvořili metody pro automatické vyhledávání chyb v syntaktické analýze, programy byly schopny vyhledat cca 40 typů pravděpodobných chyb.

Chyby, které se v korpusu nejčastěji opakovaly, jsme dále analyzovali: manuálně, cílenými počítačovými programy i v korpusovém manažeru. Mezi nejčastější patřil například výskyt sloves s funkcí **Pnom** (tj. řídicí slovesa hlavní věty) závislých na substantivech nebo výskyt dvou nekoordinovaných slov s funkcí **Sb** závislých na jednom slovesu. Často se podařilo najít spolehlivý algoritmus pro opravu takové chyby, na základě takového algoritmus jsme pak vytvořili opravná „pravidla“ pro automatickou opravu syntaktické anotace. Popis těchto pravidel a počítačového programu, v němž byly implementovány, je předmětem následující kapitoly.

# Kapitola III

## Opravný modul

Podrobný rozbor chyb parseru ukázal, že značnou část chyb lze automaticky identifikovat a často také opravit s využitím znalostí o jazyce, které parseru nejsou dostupné. Je také možné částečně kompenzovat nedostatečný rozsah trénovacích dat (tento nedostatek je samozřejmě relativní, trénovací data v PDT jsou velmi rozsáhlá na to, že byla anotována manuálně; pro časté jevy poskytují dostatek údajů) a do anotace tak zapojit seznamy slov a jejich vlastností, které lze získat z korpusů řádově větších. Podle původních odhadů možné úspěšnosti vyplývající z analýzy výsledků parseru bylo možné zvýšit úspěšnost anotace až o 3 % (přesnost přiřazení správného řídicího uzlu i syntaktické funkce, tzv. „labeled accuracy“, z cca 77 % na 80 %), po implementaci 26 opravných pravidel je však nárůst úspěšnosti znatelně menší (cca 1,3–1,9 %, viz oddíl 3. této kapitoly).

### 1. Základní charakteristiky opravného modulu

Opravný modul je počítačový program umožňující opravy chyb v syntaktické anotaci. Opravy jsou založeny na lingvistických pravidlech a seznamech vlastností slov, přičemž v současné době je implementováno 26 různých opravných pravidel. Opravná pravidla jsou koncipována tak, aby spolehlivě opravovala určitou chybnou strukturu. Program je implementován jako perlůvský modul v prostředí TectoMT. Opravná pravidla jsou sice primárně zaměřena na výstup McDonaldova MST parseru, opravují jeho typické chyby, ale principiálně mohou korigovat výsledky jakéhokoli parseru trénovaného na PDT a implementovaného ve stejném prostředí, možná s menší účinností (parser založený na jiném principu bude dělat jiné chyby), ale výstup parserulepší.

#### 1.1 Prostředí a programovací jazyk

Opravný modul byl naprogramován v jazyce PERL, který je zvláště vhodný pro práci s texty. Modul je implementován v prostředí TectoMT (Žabokrtský et al. 2008), což je modulární programové prostředí pro počítačové zpracování přirozených jazyků, jež jsme představili v předchozí kapitole. Prostředí TectoMT je v poslední době nahrazováno novějším prostředím Treex, převod opravného modulu do tohoto prostředí bude však jednoduchou technickou operací.

## 1.2 Koncepce programu a pravidel

Opravný modul se skládá ze základní části, která analyzuje jednotlivé věty textu, a z jednotlivých opravných pravidel, která jsou samostatnými podprogramy, jež v případě potřeby spouští hlavní program. Všechna pravidla jsou koncipována tak, aby bylo minimalizováno riziko nevhodného zásahu do správné struktury, pravidla jsou při opravách velmi opatrná, a pokud není (téměř) jisté, že nepoškodí správnou strukturu, raději neprovedou žádnou změnu.

## 1.3 Běh programu

Program ve svém průběhu nejprve načítá lingvistická data, jež využije pro ověření všech analyzovaných větných struktur. Potom postupně načítá jednotlivé věty textu s jejich závislostní strukturou, morfologickými značkami, lemmaty atd. U některých větných členů zjišťuje hodnoty dalších lingvistických charakteristik, které nejsou v původní struktuře uvedeny, aby je mohl využít při analýze věty. Tyto parametry vyplývají z kontextu a přímo z vlastností jednotlivých slov, program například rozlišuje mezi koordinačními spojkami, které tvoří hranice klauzí, a těmi, které stojí uvnitř věty.

Potom program opakovaně prochází strukturu věty, přičemž nejprve vyhledává možné chyby v základní závislostní struktuře (hlavní a vedlejší věty, závislosti sloves a spojek) a potom vyhledává ostatní pravděpodobně chybné struktury slovo po slově. Narazí-li na „podezřelou“, pravděpodobně chybnou strukturu, spustí opravné pravidlo (což je samostatný podprogram), které strukturu ověří a případně se ji pokusí opravit. Zjištění chyby je většinou snazší než její automatická oprava. Pravidlo má obvykle více možností, jak chybu opravit, pokud ale žádná z nich není v dané situaci použitelná, oprava se neprovede.

## 1.4 Použité seznamy slov a jejich vlastností

Opravný modul využívá několik seznamů slov, v nichž jsou upřesněny některé jejich syntakticky relevantní vlastnosti. Seznamy pocházejí z rozsáhlých korpusů (SYN2005 a SYN2010, v některých případech i další korpusy Českého národního korpusu z řady SYN), a mohou tak být výrazně obsáhlejší než údaje, které parser může získat z trénovacích dat. Seznamy zahrnují valenci sloves a adjektiv, slovesa reflexiva tantum, rekcii substantiv, typické adverbialní předložkové fráze, substantiva označující osoby, složené předložky, adjektiva, která mohou být syntaktickými substantivy aj. Krátké seznamy slov (jako modální a fázová slovesa) jsou zabudovány přímo v opravném modulu, nejsou načítány z externího seznamu.

Valence sloves zohledňuje reflexivitu, valence může být uvedena jak s prostým, tak s předložkovým pádem (pád je vyjádřen číslem): *zabránit 3, ptát se 2, přemýšlet o 6, pustit se do 2*, v některých případech je valence uvedena i negativně (netranzitivní slovesa). Seznam obsahuje cca 5000 údajů o valenci.

Valence adjektiv rozlišuje pouze mezi prostými a předložkovými pády: *dbalý 2, zodpovědný za 4*, v seznamu je cca 400 adjektiv.

U substantiv jsou v jednom seznamu uvedena jak „valenční“ (deverbativní) substantiva s valencí jako *nakládání s 7*, tak další substantiva často rozvíjená neshodným přívlastkem v určitém prostém nebo předložkovém pádu: *atentát na 4, spor o 4, hráč na 4, věrnost 3*. Seznam zahrnuje cca 400 substantiv.

Slovesa, která jsou označena jako reflexiva tantum, se rozdělují do dvou kategorií: patří sem jednak slovesa, která jako nereflexivní vůbec neexistují (*smát se*), jednak slovesa, která mají jako zvrtná jiný význam (*hodit se*; zde označení „reflexivum tantum“ není zcela namístě), celkem je v seznamu cca 1000 sloves.

Seznam typických adverbialních předložkových frází se využívá především pro určení syntaktické funkce v rámci předložkové fráze (**Adv** oproti **Obj**), někdy i pro určení závislosti. Seznam obsahuje spojení předložka – pád – lemma/forma substantiva (*o 6 přestávce, na 4 oplátku, za 4 rok*), celkem to je cca 1400 spojení s adverbialním významem.

Substantiva označující osoby jsou zařazena do zvláštního seznamu především kvůli spojení typu *pan Novák, sopranistka Eva Urbanová*. V seznamu je cca 1200 apelativ (*pan, sopranistka*); 900 křestních jmen (*Jan, Eva*) a 2000 příjmení (*Novák, Urbanová*).

Víceslovné předložkové výrazy jako *v souvislosti s, na základě* nebo *směrem k* se v PDT označují jako složené předložky, opravný modul proto potřebuje jejich seznam pro rozlišení, se kterými předložkovými obraty je třeba takto zacházet, a se kterými ne. V seznamu je cca 100 složených předložek.

Některá adjektiva mohou ve větě hrát roli syntaktického substantiva, aniž by bylo ve větě jasně elidované syntaktické substantivum. Část z nich by měla být disambiguována jako substantiva, např. *pracující, hovězí*, část je bližší zájmenům, např. *ostatní, další*, část si zachovává adjektivní charakter, ale často se používá bez substantiva, elidované je obecné apelativum *člověk, muž, žena: přítomný, umírající*. Pokud nejde o elipsu, nemají mít taková adjektiva funkci **ExD** (funkce slova, jehož původní řídicí člen byl elidován), ale syntaktickou funkci podle vlastní funkce ve větě. Seznam zahrnuje cca 100 adjektiv s touto charakteristikou.

## 1.5 Určení hranic klauzí

Jedním z důležitých lingvistických parametrů, které opravný modul na začátku rozboru každé věty přidává ke spojkám a interpunkčním znaménkům, je rozlišení mezi slovy, jež od sebe oddělují dvě klauze, a těmi, které stojí uvnitř věty. Mnoho chyb, kterých se parser dopouští, vyplývá právě z jeho neschopnosti pracovat s hranicemi klauzí. Systém dokáže určit, že daná spojka či interpunkce je hranicí klauzí, například když vlevo i vpravo od slova stojí sloveso v určitém tvaru. Dokáže také vyloučit hranici klauzí, např. když na koordinační spojce závisejí dvě syntaktická substantiva ve stejném pádu a spojka není jedinou potenciální hranicí klauzí mezi dvěma slovesy.

## 1.6 Určení sloves v určitém tvaru a reprezentantů složených slovesných tvarů

V PDT jsou věty (klauze) reprezentovány slovesy v určitém tvaru nebo plnovýznamovým slovesem ve složeném slovesném tvaru (složený minulý čas, složený budoucí čas, kondicionál, pasivum). Pro efektivní práci se základní strukturou věty je třeba tato slovesa rozpoznat a odlišit je od pomocných sloves (s funkcí **AuxV**) či infinitivních předmětů modálních či fázových sloves (u obou typů má infinitiv funkci **Obj**). Rozlišení nevyžaduje žádný sofistikovaný algoritmus.

## 1.7 Určení hlavních a vedlejších vět

Slovesa v určitém tvaru (tak budeme nadále nazývat i slovesné tvary, které jen reprezentují složený slovesný tvar, např. infinitiv u budoucího času), program označí jako řídicí slovesa klauzí (reprezentují celou klauzi). Pro práci s větnou strukturou je třeba je podrobněji rozlišit na slovesa reprezentující hlavní a vedlejší věty.

Je-li na slovese v rámci klauze přímo či nepřímo závislé předcházející vztažné zájmeno, před nímž stojí přijatelný antecedent (syntaktické substantivum se stejným rodem a číslem), bude sloveso považováno za reprezentanta vztažné věty.

Pokud je sloveso v jedné klauzi s předcházející podřadicí spojkou, která stojí těsně po čárce či souřadicí spojce, bude označeno jako sloveso reprezentující vedlejší větu. Sloveso by správně mělo být závislé na podřadicí spojce, ale ne vždy parser strukturu takto skutečně označí). Algoritmus je složitější, např. spojky *jako* a *než* těsně po čárce nemusí vždy uvozovat vedlejší větu (1).

(1) *Poklidné doby akciových investorů jsou pryč , **jako** mávnutím kouzelného proutku sletěly kapitálové trhy v letních měsících o desítky procent .*

Slovesa, která nejsou označena jako slovesa ve vztažné či spojkové vedlejší větě, jsou považována za potenciální slovesa v hlavní větě. Algoritmus nedokáže spolehlivě identifikovat slovesa, která patří do vedlejší věty přerušené jinou vloženou klauzí, jako třeba sloveso *rozmazlován* v příkladu (2).

(2) *Byla pevně přesvědčená o tom , že Michael , kterého tiše zbožňovala , byl rozmazlován v tom nejhorším slova smyslu ,*

### 1.8 Oprava základní struktury věty

Po získání všech potřebných informací o větě začne opravný modul ověřovat základní strukturu věty složenou z formálního syntaktického kořene, spojek a sloves reprezentujících věty. Základní strukturu je nutné opravit nejdříve, protože chyba zde by mohla zablokovat opravu jednotlivých větných členů. Opravný modul rozpozná několik typicky chybných struktur. Pokud na ně u dané věty narazí, spustí odpovídající opravné pravidlo.

### 1.9 Oprava jednotlivých větných členů

Po opravách základní struktury prochází opravný modul postupně všechna slova věty. Pokud svým kontextem, závislostí, syntaktickou funkcí, morfologickou značkou nebo jejich kombinací odpovídá jedné z typických chyb větných členů, opět se spustí odpovídající opravné pravidlo, ověří, zda je struktura skutečně chybná, a případně se pokusí chybu napravit. Pokud je struktura změněna, načtou se parametry věty znovu, aby další pravidla již mohla pracovat s novou, opravenou strukturou.

### 1.10 Uložení věty a zápis vertikály

Po ověření celé věty je věta uložena do opraveného souboru, zároveň se zapíše ve formátu tzv. vertikály (jednotlivá slova následují na řádcích pod sebou, ke každému slovu jsou na řádce připojeny jeho atributy oddělené tabulátory). Hranice vět jsou vyznačené prázdným řádkem. Oproti „klasické“ vertikále obsahující pro každé slovo tvar, lemma a morfologickou značku má tato vertikála navíc několik dalších údajů: syntaktickou funkci, identifikaci řídicího členu, lemma a slovní druh řídicího členu (je-li řídicím členem předložka, pak údaje o řídicím členu předložky), lemma předložky nebo prázdný údaj, není-li přímým řídicím členem předložka, v případě aplikace nějakého opravného pravidla také údaje o změně a původních vlastnostech slova atd. Tato komplikovaná vertikála je pak zdrojem dat pro korpus

ve formátu pro korpusový manažer Bonito a je jedním z možných způsobů přístupu k syntakticky anotovanému korpusu.

## **2. Jednotlivá opravná pravidla**

V následující části představíme jednotlivá implementovaná opravná pravidla, rozdělená do skupin podle svého zaměření. V souladu s cílem této práce (získat statistické údaje o syntaktických funkcích substantiv ve vztahu k jejich předložce a pádu) se největší důraz klade na opravu závislostí a syntaktických funkcí substantiv. Některá pravidla se ale zaměřují na obecnější jevy nebo na základní závislostní strukturu složených vět (souvětí), protože možnost provést dílčí opravu závislosti či funkce je často podmíněna správnou celkovou strukturou. Každá jednotlivá oprava tak může přispět k větší účinnosti pravidel specifitějších.

Pravidla jsou implementována jako samostatné podprogramy opravného modulu. Když hlavní program modulu narazí na podezřelou strukturu, vyvolá odpovídající pravidlo (podprogram).

### **2.0 Způsob prezentace jednotlivých pravidel**

U popisovaných pravidel vždy uvádíme nejprve stručnou motivaci, potom příklady chyb, které by pravidlo mělo opravit, podmínky, za kterých se z hlavního programu spouští, a základní postup, kterým chyby opravuje (pokud je algoritmus velmi složitý, může být výklad rozčleněn do několika částí), dále příklady řešení, počet zásahů v korpusu SYN2005 a celkovou úspěšnost pravidla.

#### **2.0.1 Příklady chyb**

Uvedené příklady chyb parseru vždy pocházejí z korpusu SYN2005 označovaného MST parserem. Příklady oprav představují opravy provedené pomocí pravidlového opravného modulu, a to popisovaným pravidlem. Zvýrazňujeme pouze chyby týkající se popisovaného pravidla stejně jako opravy (pokud jiné pravidlo opraví jinou část struktury, oprava může být v příkladu zobrazena, ale nebude nijak zvýrazněna). Příklady chyb a jejich řešení jsou uváděny ve dvou formátech podle složitosti. Složitě struktury jsou představeny graficky, jednodušší textově. Opravný modul často musí zpracovávat a úspěšně opravuje mnohem rozsáhlejší struktury, než zobrazujeme v příkladech, nemůžeme je však na omezeném prostoru adekvátně znázornit.

### 2.0.1.1 Textové zobrazení

V textovém formátu je uvedena věta nebo část věty z korpusu (bez dalších úprav, tj. zachováváme oddělenou interpunkci, malá a velká písmena, případně i textové chyby aj.). U slov, kterých se oprava týká, jsou za lomítkem uvedeny relevantní údaje: závislost, syntaktická funkce nebo část morfologické značky (obvykle prvních pět znaků pro identifikaci pádu, popř. čísla a rodu substantiva). Závislost je reprezentována číselným údajem nebo znakovým pro formální kořen: závislost na formálním syntaktickém kořeni je označena dvojitým křížkem (#), závislost na jiném větném členu je vyjádřena číslem vyjadřujícím vzdálenost řídicího větného členu (záporná čísla doleva, kladná doprava).

U příkladů chyb s opravou jsou za lomítkem dva údaje, nejprve chybný, potom opravený. Závislosti, funkce a tagy jsou barevně zvýrazněny. Původně **správné údaje** jsou zobrazeny **modrou barvou**, **chyby** jsou zvýrazněny **červeně**. U příkladů oprav je **zeleně** zvýrazněna **správná oprava**, **oranžově** je zvýrazněna **oprava chybná** (tj. oprava, jejíž výsledek není zcela správný bez ohledu na správný nebo chybný vstup).

Příklad chyby:

- (1) Vyšplhá/**Pred**/2 -/1 li/# na konec druhého dílu , získává/**Adv**/-6 2 body .

Sloveso *vyšplhá* má chybnou syntaktickou funkci **Pred**, je správně závislé na spojce *li* (o dvě slova doprava). Spojovník je správně závislý na následujícím slově: spojce *li*. Spojka *li* je chybně určena jako řídicí uzel celé závislostní struktury (závisí přímo na formálním kořenu věty). Sloveso *získává* má chybnou syntaktickou funkci **Adv** a je chybně závislé na spojce *li* (šest slov doleva).

Příklady chyby s opravou:

- (1) Vyšplhá/**Pred/Adv**/2 -/1 li/#/6 na konec druhého dílu , získává/**Adv/Pred**/-6/# 2 body .  
(2) když jsme ji o/6/2 to/**Obj/Adv**/-1 souvisleji/1 a trochu stranou požádali .

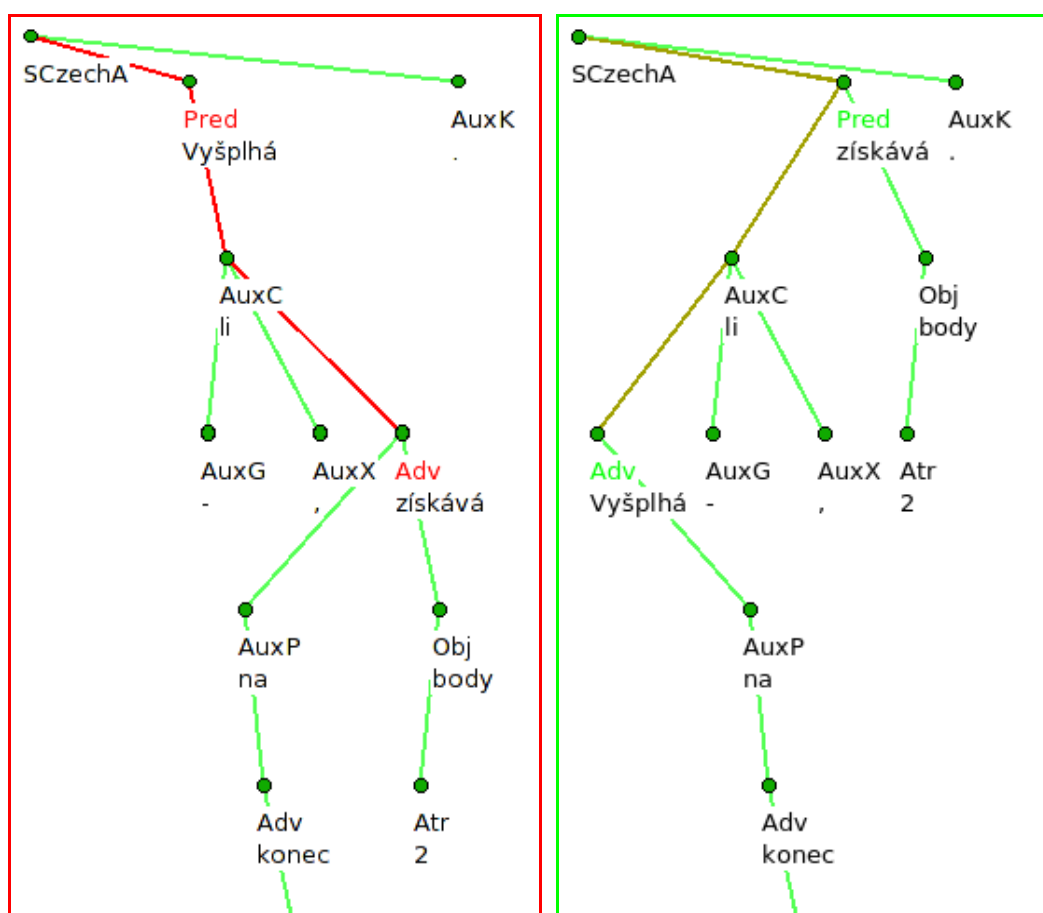
Kromě předešlého popisu chyb je **zeleně** doplněna oprava syntaktických funkcí a závislostí. V příkladu (2) došlo k chybné opravě slov *o* a *to* (oprava ze správné funkce či závislosti na špatnou), chybné opravy jsou zvýrazněny **oranžovou** barvou.

### 2.0.1.2 Grafické zobrazení

U složitějších případů je do textu zařazen obrázek graficky znázorněné závislostní struktury, a to celé věty nebo její části, vytvořený pomocí programu TrEd. Vždy se zobrazují dva obrázky: obrázek výchozí struktury označované MST parserem a obrázek závislostní struktury po opravě. Výchozí struktura je červeně orámována, relevantní chyby jsou zvýrazněny červenou barvou. Chyby, na něž není dané pravidlo zaměřené, se nezvýrazňují. Opravená struktura je orámována zeleně, správně opravené závislosti a funkce jsou zvýrazněny zeleně (funkce světlezelenou barvou, závislosti žlutozelenou). Opět se nezvýrazňují chyby, které se netýkají komentovaného pravidla, ani jejich opravy. Formální syntaktický kořen věty je označen **SCzechA**. U příkladu vždy uvádíme nejprve větu v textové podobě.

Příklad grafického znázornění chybné struktury s opravou:

(1) *Vyšplhá – li na konec druhého dílu , získává 2 body .*



V uvedeném příkladu prezentované pravidlo opravilo závislosti slov *vyšplhá*, *li* a *získává*. Jiné pravidlo zároveň opravilo závislost předložkové fráze reprezentované předložkou *na*,

tato oprava však v příkladu není zvýrazněna. Zatímco textových příkladů uvádíme pro každé pravidlo několik, a to nejprve typické chyby, potom chyby s opravami, u grafických příkladů většinou uvádíme jen jeden typický příklad s opravou.

## 2.0.2 Počet zásahů v korpusu SYN2005 a hodnocení úspěšnosti

Po příkladech chyb a jejich řešení uvádíme u každého pravidla počet zásahů v korpusu SYN2005, přepočítaný na 1 000 000 tokenů. Jako jeden zásah se počítá změna u jednoho tokenu. Změny rozdělujeme podle typu na 1) změny závislosti, 2) kombinované změny závislosti a syntaktické funkce (včetně případné změny morfologické značky), 3) změny syntaktické funkce (beze změny závislosti), 4) kombinované změny syntaktické funkce a morfologické značky a 5) změny morfologické značky (bez dalších změn). Pokud v průběhu zpracování věty zasáhne týž token více pravidel, počítá se jen první změna (první aplikované pravidlo).

Nakonec představujeme odhad úspěšnosti pravidla spočítaný na vzorku 100 náhodných aplikací pravidla (každý zásah pochází z jiného textu, texty pocházejí z různých funkčních stylů). Změny hodnotíme třemi stupni: úspěšná změna, neutrální zásah, chybná změna. Za **úspěšný zásah** považujeme změny, které relevantní části chybné struktury opraví na zcela správné (závislost i funkce).

Jako **neutrální zásah** označujeme změny, které chybnou strukturu změní, ale neopraví (nebo ne úplně) a přitom nezpůsobí závažné zhoršení struktury (např. změnu závislosti ze správné na chybnou). Mnoho **neutrálních zásahů** přinejmenším odstraní závažné rozpory s obecnými syntaktickými pravidly (např. odstraní situaci, kdy jsou na jednom slovese *být* závislé dva nekoordinované podmínky, tak, že jeden označí za jmennou část verbonominálního predikátu, ovšem změní syntaktickou funkci u nesprávného podmínky, takže ve větě je sice už jen jeden podmět, ale ne ten správný).

Za **negativní, chybný zásah** považujeme jednak zhoršení chybné struktury (např. změna závislosti ze správné na chybnou), jednak jakoukoli změnu původně správně interpretované struktury. Pravidla byla koncipována tak, aby se chybným zásahům do správných struktur vyhnula i za cenu menšího počtu zásahů, ale struktury, s nimiž opravný modul pracuje, jsou natolik variabilní, že nebylo vždy možno se závažné chybě vyhnout. Z principu je možné chybná pravidla vždy opravit a chyb se vyvarovat, uvedená úspěšnost odráží současný stav opravného modulu.

Přehled zásahů vždy představujeme v tabulce, jejíž příklad zde uvádíme. Na prvním řádku je název pravidla, na druhém a třetím řádku jsou počty zásahů v korpusu SYN2005 přepočtené na 1 000 000 tokenů, na řádku čtvrtém a pátém jsou uvedena procenta úspěšnosti oprav (celkem, bez ohledu na typ změny).

Pravidlo pro opravu závislostí a funkcí u sloves se spojkou <i>-li</i>						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	295	226	0	0	0	521
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	88 %		12 %		0 %	

### 2.0.3 Rozdělení pravidel podle zaměření

Pravidla představíme ve čtyřech skupinách podle zaměření. První skupinu tvoří pravidla, jejichž cílem je opravit základní závislostní strukturu složené věty (jedna hlavní věta, věty vedlejší) či souvětí (více hlavních vět, případně věty vedlejší). Ve druhé skupině jsou obecná pravidla, která opravují chybné struktury uvnitř klauzí, ale nejsou přímo zaměřena na substantiva a jejich funkce. Třetí skupinu tvoří pravidla, jež opravují konkrétní chybné syntaktické funkce nebo závislosti především syntaktických substantiv. Čtvrtá, nejpočetnější skupina se skládá z pravidel pro předložkové fráze, kde se použitý parser nejčastěji dopouští chyb ovlivňujících funkce či závislosti substantiv.

## 2.1 Pravidla korigující chybně určenou základní závislostní strukturu složených vět či souvětí

České psané texty se z velké části skládají ze souvětí a složených vět (v korpus SYN2005 připadá na větu v průměru cca 13 slov a 3 interpunkční znaménka). Struktury jsou často složité, obsahují větší množství interpunkce a spojek s různými funkcemi. Často se obtížně rozlišuje mezi větou a členskou koordinací. Ve větách jsou navíc často elidované větné členy, slovesa, syntaktická substantiva aj., což zpracování vět dále komplikuje. V trénovacích datech je řídicím členem věty či souvětí jednou sloveso, jindy souřadící spojka, interpunkce, nebo dokonce spojka podřadící (samostatně stojící vedlejší věty). Pro parser je pak velmi obtížné správně zvolit řídicí člen závislostní struktury a správně určit závislosti všech vět (hlavních i vedlejších), jež jsou reprezentovány slovesy.

Pět implementovaných pravidel se pokouší řešit některé typické chyby parseru. Kromě

posledního pravidla, které je úzce zaměřené na jednu konkrétní strukturu, jsou pravidla poměrně obecná, odpovídají velkému množství různých potenciálně chybných struktur. V implementované verzi jsou pravidla velmi opatrná: nenajdou-li spolehlivé řešení chyby, raději chybnou strukturu ponechají nezměněnou. Pravidla jsou možná příliš opatrná, a jsou proto poměrně málo účinná.

### 2.1.1 Pravidlo pro opravu koordinace nekompatibilních členů, která se považuje za řídicí člen závislostní struktury

Ve složitých větných strukturách obsahujících koordinaci a další interpunkci je pro parser někdy obtížné rozlišit, zda je koordinace větná nebo členská a zda slovesa ve větě patří do hlavních, nebo vedlejších vět. Parser v použitém nastavení neověřuje sousední uzly při přiřazování závislosti, dochází tedy k tomu, že na jedné koordináční spojce závisí ve výsledku syntaktická substantiva i slovesa.

Když se taková smíšená koordinace považuje za řídicí člen závislostní struktury jako v příkladech (1), (2) a (3), mělo by ji toto pravidlo opravit (znak # označuje formální kořen věty, čísla vzdálenost řídicího uzlu v počtu slov, záporné číslo doleva).

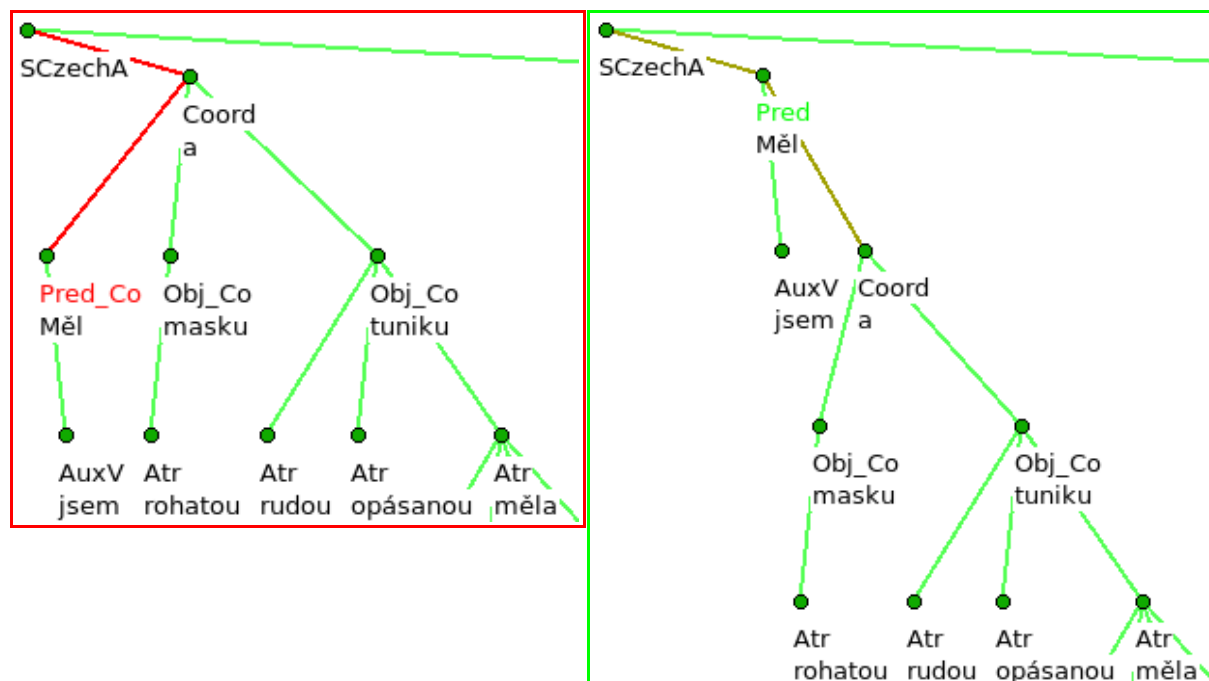
- (1) *Přikryl/4 jsem/-1 ji prostěradlem/1 a/Coord/# bavlněnou dekou/-2 .*
- (2) *Sbírali/13 rákosí a listy orobince , vrbové pruty , dlouhé tenké kořeny/2 smrků a/# všechno/-1 , co mohla/-3 Ajla použít na pletení košů a provazů .*
- (3) *Nezval byl/3 mimo prostor/1 a/Coord/# čas/-1 .*

Pravidlo se aktivuje ke zpracování koordinace závislé přímo na formálním kořenu věty, na nichž jsou zároveň závislá alespoň dvě syntaktická substantiva a alespoň jedno sloveso v určitém tvaru. Pravidlo je opatrné, dokáže opravit jen jeden snadno řešitelný typ struktury: všechna substantiva závislá na koordinaci se shodují v pádě a na koordinaci je kromě toho závislé jedno sloveso neoznačené jako sloveso, které patří do vedlejší věty.

Za těchto podmínek opravné pravidlo označí pravděpodobné sloveso z hlavní věty (tj. sloveso, které je závislé na koordinaci a zřejmě nepatří do vedlejší věty) za řídicí člen závislostní struktury (bude záviset přímo na formálním kořenu věty), pro koordinaci pak najde vhodný řídicí uzel podle kontextu a pádu substantiv závislých na koordinaci: vhodným uzlem může být těsně předcházející předložka, blízké sloveso či předcházející substantivum, popř. sloveso, jež bylo dosud závislé na koordinaci a nyní je řídicím uzlem celé struktury.

(3b) *Nezval byl/3/# mimo/-1 prostor/Atr/Adv a/#-2 čas/Obj/Adv .*

(4) *Měl/4/# jsem rohatou masku a/#-4 rudou opásanou tuniku , která měla vzadu přišitý kus provazu jako ocas .*



Pravidlo pro opravu koordinace nekompatibilních členů považované za řídicí člen závislostní struktury

počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	247	38	70	0	0	355
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	86 %		6 %		8 %	

### 2.1.2 Pravidlo pro slovesa ve vedlejších větách chybně považované za řídicí člen závislostní struktury

Správné určení hlavní věty (a slovesa, které ji reprezentuje) je pro parser často velmi obtížné, znamená to zvolit z několika sloves jedno, které nepatří do vedlejší věty, přičemž větné struktury jsou někdy velmi složité, včetně vložených vedlejších vět aj. Vedlejší věta se ve formátu PDT, tedy v trénovacích datech, může projevit různým způsobem (sloveso závisí na pořadici spojce nebo na slovesu závisí vztažné zájmeno či adverbium). Parser tedy v tomto určení často chybí a označuje vedlejší věty za věty hlavní (1) nebo koordinuje vedlejší věty s hlavními (2).

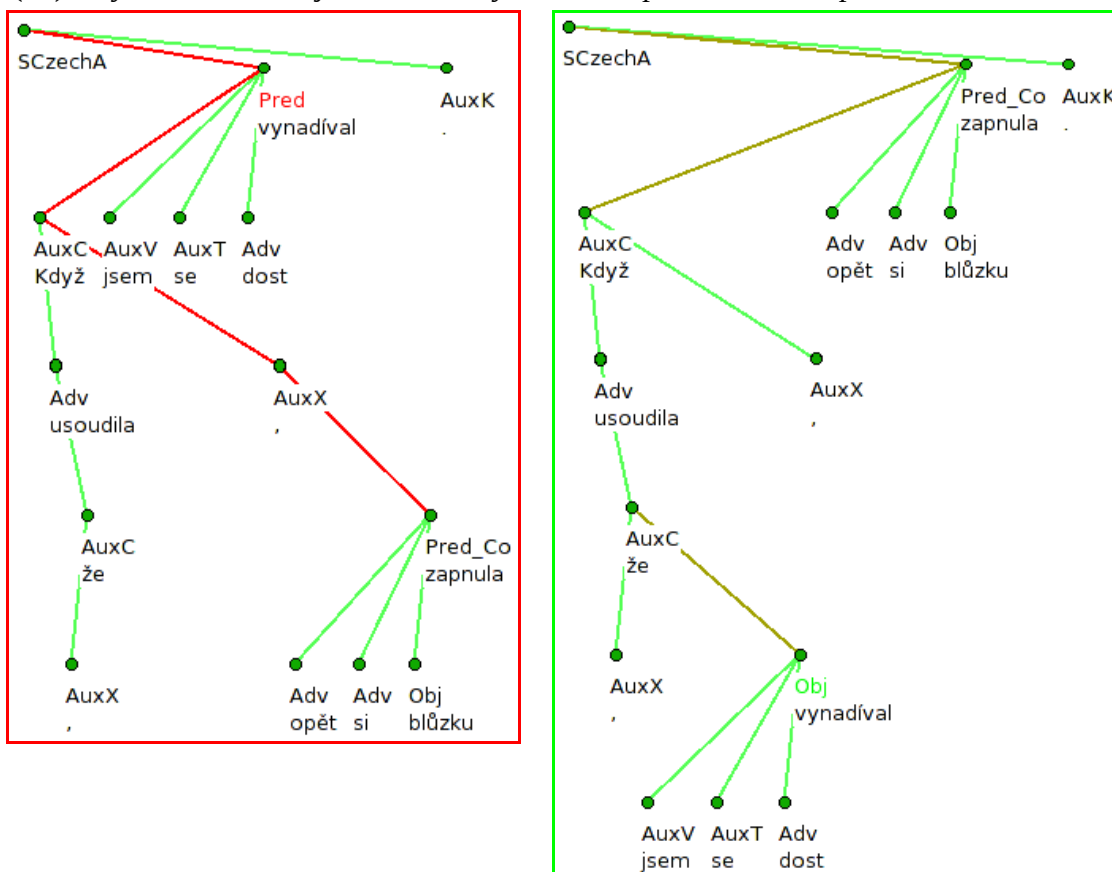
(1) *Když/AuxC/7 usoudila/Adv/-1 , že/AuxC/-2 jsem se dost vynadával/Pred/# , opět si blůzku zapnula/Pred/-4 .*

(2) *Snažil/Pred/10 jsem se ti naznačit , abys/4 ho zadržela/Pred/2 , ale/Coord/# nechtěl/Pred/-1 jsem křičet a vyplašit je .*

V základní části opravného modulu je každému slovesu v určitém tvaru (nebo reprezentantu složeného slovesného tvaru, jímž může být např. infinitiv u složeného futura) přiřazen parametr, který rozlišuje, zda sloveso patří do hlavní nebo vedlejší věty (vztažné nebo s podřadící spojkou). Když se pak v průběhu ověřování věty setká opravný modul se slovesem, které podle přiřazeného parametru patří do vedlejší věty, ale parser ho interpretuje jako řídicí člen celé závislostní struktury (sloveso závisí na formálním kořenu) nebo jako člen koordinace, která je řídicím členem závislostní struktury, spustí se toto opravné pravidlo.

Pravidlo rozděluje opravu na dvě větve: v jedné zpracovává struktury s vedlejší větou závislou přímo na formálním kořeni věty, ve druhé větvi zachází s koordinací označenou za řídicí člen závislostní struktury, která obsahuje vedlejší větu. V první větvi pravidlo zjistí, zda se ve větě nachází jiné sloveso, které není označeno jako sloveso ve vedlejší větě, a je tedy pravděpodobně správným řídicím členem věty. Najde-li takové sloveso, změní ho na řídicí člen věty. Pro vedlejší větu potom hledá vhodný řídicí uzel: nejčastěji jím bude předcházející sloveso v určitém tvaru, tj. slovo, které reprezentuje předcházející větu (1b).

(1b) *Když usoudila , že jsem se dost vynadával , opět si blůzku zapnula .*



Ve druhé větvi pravidlo ověří parametry všech sloves, která jsou závislá na koordinaci. Pokud alespoň jedno z nich (lépe dvě) program nemá označené jako slovesa reprezentující vedlejší věty, stačí vyřadit vedlejší větu z koordinace a nalézt pro podřadicí spojku ve větě vhodný řídicí uzel, přičemž sloveso má být závislé na podřadicí spojce. Vhodným řídicím uzlem je obvykle předcházející sloveso v určitém tvaru, i když v některých případech musí být pravidlo ve volbě řídicího uzlu sofistikovanější, bere v potaz například infinitivy závislé na modálních, fázových a dalších podobných slovesech (2b) a odkazovací obraty typu *na to*, *aby* (3).

(2b) *Snažil jsem se ti naznačit , abys/4/-2 ho zadržela/Pred/Obj/2/-2 , ale nechtěl jsem křičet a vyplašit je .*

(3) *Jsem/13 už stará na to , abych/-4/-2 svištěla/6/-1 po sjezdovkách jako kdysi , a/# nehodlám/-1 se otravovat s nějakou novou technikou .*

Pravidlo pro slovesa ve vedlejších větách chybně považované za řídicí člen závislostní struktury						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	134	96	29	0	0	259
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	76 %		15 %		9 %	

### 2.1.3 Pravidlo pro opravu závislostí souřadných hlavních vět

Musí-li parser pro řídicí uzel celého souvětí volit mezi slovesem a souřadicí spojkou, často se rozhoduje chybně, někdy vytváří syntakticky nesmyslné kompromisní struktury. Příkladem takové struktury může být věta, kde je jedno sloveso závislé přímo na formálním kořeni, na něm je závislá koordinační spojka, na této spojce další sloveso. Ve skutečnosti měla být obě tato slovesa, popř. více sloves, koordinována, tj. řídicím uzlem struktury měla být spojka (1). Do koordinace jsou často chybně zapojeny další větné členy (2) a (3).

(1) *Byla/# hrozně pomalá a/-3 se vším se děsně babrala/-5 ;*

(2) *Anna obešla/# stůl/1 a/-2 chvíli/1 trvalo/-2 , než si ti dva všimli ,*

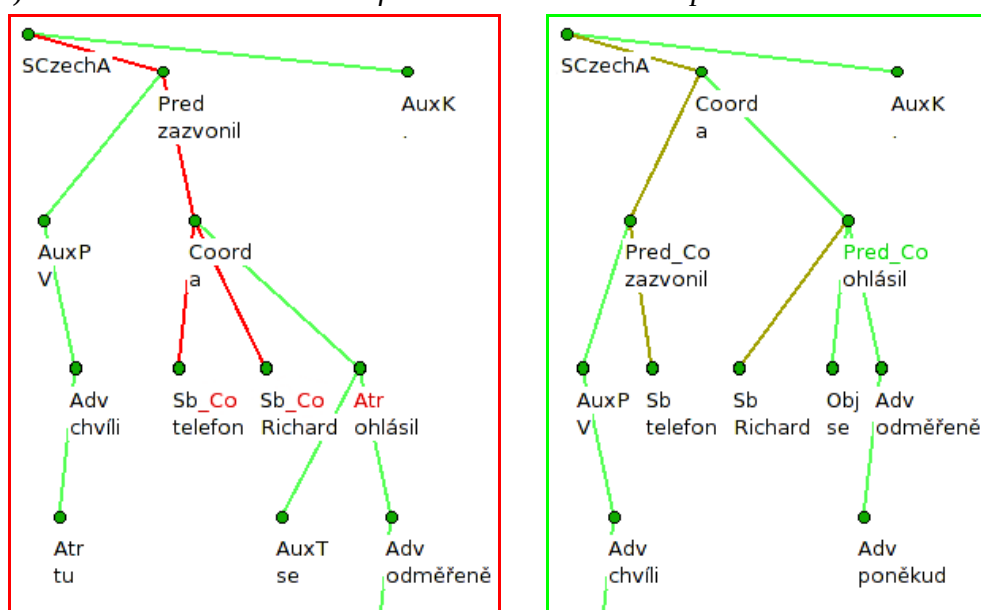
(3) *Každý Čech rozumí/# Slovákov/1 a/-2 Slováka/1 rozumí/-2 Čechovi .*

Pravidlo se spustí, když hlavní program při ověřování struktury věty narazí na sloveso označené jako řídicí uzel celé závislostní struktury (visí na formálním kořeni věty), na tomto slovesu je závislá koordinační spojka, na této spojce je závislé sloveso, které nepatří do vedlejší věty (nejsou na něm závislá vztažná zájmena, v jeho klauzi nestojí podřadicí spojka).

Ani určení hranic klauzí (čárky, spojky aj., které oddělují klauze) na základě lingvistických pravidel není vždy zcela spolehlivé, protože se musí opírat o disambiguaci (s několika procenty chyb) a částečně i o závislostní strukturu vytvořenou parserem, kterou se snaží opravit. Aby nevnášelo do struktury další chyby, je tedy pravidlo opatrné a počítá pouze s jediným možným separátorem klauzí mezi dvěma hlavními slovesy: koordinační spojkou.

Pravidlo ověří, že mezi slovesem označeným jako řídicí člen struktury a slovesem závislým na koordinační spojce nestojí kromě této spojky žádná jiná potenciální hranice klauzí. V takové konfiguraci pak označí spojku za řídicí uzel celé struktury (#), původní hlavní sloveso bude závislé na této spojce. Mezi případnými dalšími členy koordinace pravidlo ponechá beze změny pouze další slovesa, která reprezentují hlavní věty. Slovesa, která reprezentují věty vedlejší (vztažné nebo uvozené podřadicí spojkou), budou převěšena na jiné vhodné uzly (vztažné na předcházející antecedent, spojkové na předcházející sloveso, popř. na jedno z hlavních sloves ve větě). Ostatní větné členy (vyjma grafické znaky) pravidlo převěsí na sloveso, které je jim nejbliž: ty, které stojí vlevo od koordinace, na sloveso vlevo, ty, které stojí vpravo od koordinace, na sloveso na pravé straně. Pravidlo nemění syntaktické funkce jednotlivých větných členů, to ponechává na následujících pravidlech (4).

(4) *V tu chvíli zazvonil telefon a Richard se ohlásil poněkud odměřeně .*



Pravidlo pro opravu závislostí souřadných hlavních vět						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	1105	15	58	0	0	1078
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	96 %		4 %		0 %	

#### 2.1.4 Pravidlo pro ověření závislostí a funkcí vztažných vět

Problémy s určením správné závislosti a funkce vztažných vět mají v zásadě stejné příčiny jako ostatní chyby týkající se celé větné struktury: složité struktury obsahující interpunkční znaménka, koordinace členské i větné aj. K tomu se přidávají obtíže s hledáním náležitého antecedentu vztažného zájmena dané mj. i tím, že mezi relativem a antecedentem stojí v závislostní struktuře sloveso (relativum je závislé na slovesu; sloveso, jež reprezentuje vztažnou větu, je závislé na antecedentu). Chyba může být méně závažného charakteru, kdy sloveso ve vztažné větě závisí na větném členu z předcházející věty, i když ne na správném antecedentu (1). Chyba v závislosti vztažných vět může také odrážet závažnější chybu v konstrukci závislostní struktury: vedlejší vztažná věta může být například chybně považována za hlavní větu (2).

(1) *Ihned zpozoroval , že kromě klíče od vozu tam bylo několik dalších , které/1 vypadaly/Pred/-5 jako klíče od domu .*

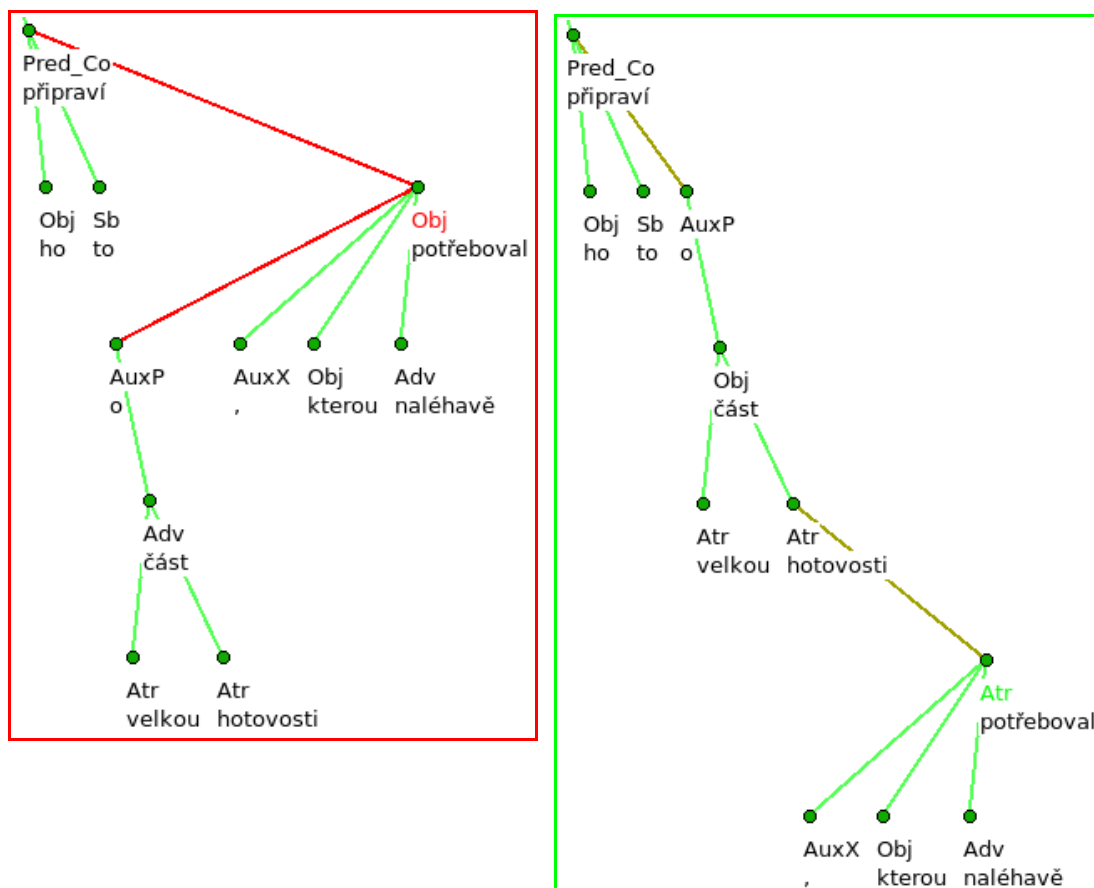
(2) *Tolik Shawovo oznámení/5 o svatbě , které/1 zaslal/Pred/# k otištění do londýnských novin 2 . června 1898 .*

Pravidlo se zavolá pro každé sloveso, na němž je (přímo či nepřímo) závislé vztažné zájmeno. Před relativem musí stát čárka (popř. čárka a předložka), na vhodném místě před čárkou musí stát větný člen, který může sloužit jako antecedent vztažného zájmena (substantivum, některá zájmena, adjektiva, číslovky).

Pravidlo nejprve ověřuje možnou shodu relativu s antecedentem v rodě a čísle. Protože však disambiguace není zcela spolehlivá, nevyžaduje se shoda morfologických značek. Program ověřuje, že tvary relativu a antecedentu umožňují shodu v rodě a čísle, a to bez ohledu na přiřazené morfologické značky. Jestliže se relativum s antecedentem shoduje a sloveso není ve vztažné větě chybně označeno jako řídicí člen závislostní struktury, opraví se závislost slovesa a případně i jeho syntaktická funkce.

Je-li však sloveso chybně určeno jako řídicí uzel (3), je nejprve nutno vyhledat nový, správný řídicí uzel pro celou závislostní strukturu a opravit podle toho závislosti ve větě. Tímto novým řídicím uzlem by mělo být v nejlepším případě sloveso (takové, které nestojí ve vedlejší větě, tj. v klauzi s daným slovesem není ani podřadicí spojka, ani vztahné zájmeno). Nenajde-li se takové sloveso, může jím být i antecedent relativna nebo jemu nadřazený větný člen.

(3) *Ale připraví ho to o velkou část hotovosti, kterou naléhavě potřeboval.*



Pravidlo pro ověření závislostí a funkcí vztahných vět						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	207	346	6	0	0	559
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	83 %		7 %		10 %	

### 2.1.5 Pravidlo pro opravu závislostí a funkcí u sloves se spojkou *-li*

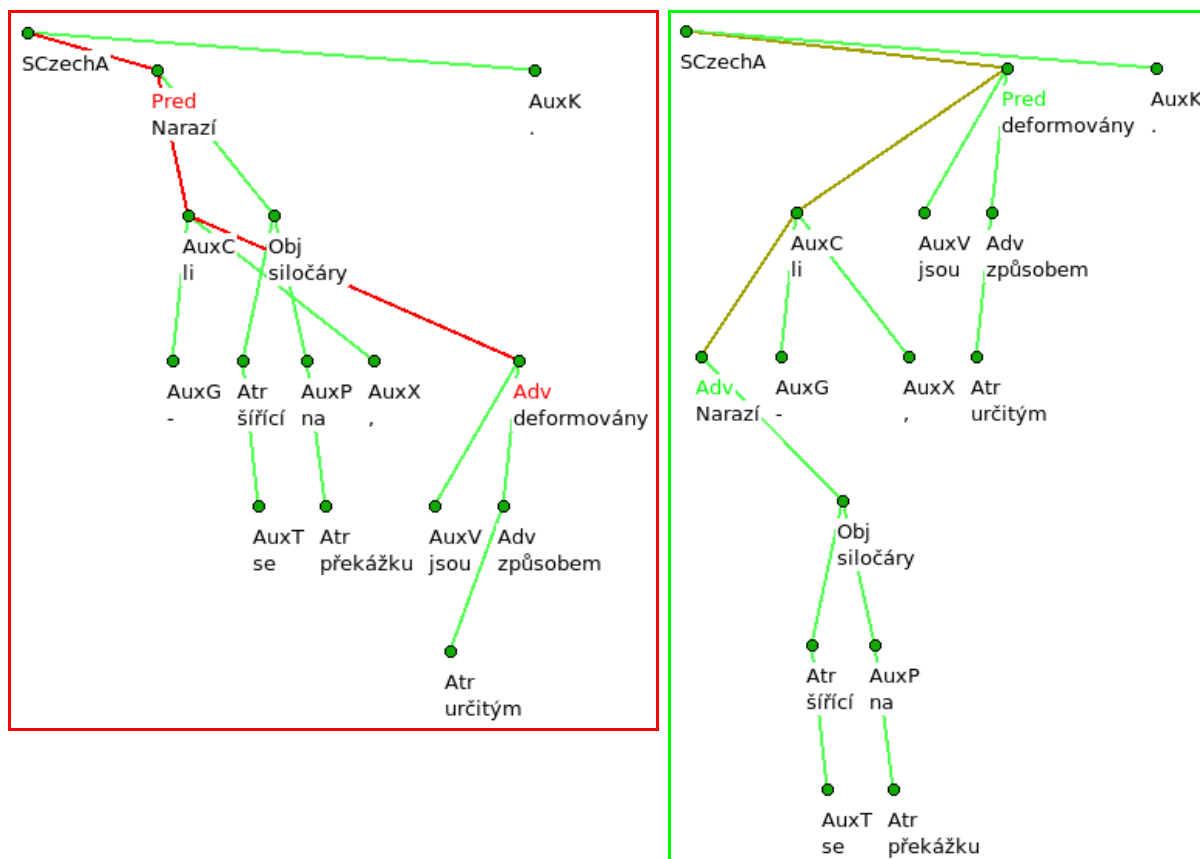
Sloveso následované podřadicí spojkou *-li* je specifická struktura, která vyžaduje samostatné pravidlo. V PDT je (stejně jako u ostatních podřadicích spojek) spojka *li* řídicím uzlem celé věty, sloveso i spojovník závisí na spojce, spojka na slovese nadřazené věty. Struktura je poměrně častá, objevuje se cca třístokrát v trénovacích datech, přesto ji parser přibližně v polovině případů interpretuje chybně (závislosti nebo syntaktické funkce, často obojí). Někdy je jen prohozena závislost mezi slovesem a spojkou *li*, případně jsou chybně označené syntaktické funkce (1). Často je chybná širší závislostní struktura (2) a opravený příklad (3).

- (1) *Je/Obj/13 -/1 li/-2 v čele odboru skutečně člověk pouze se základním vzděláním , porušila radnice zákon !*
- (2) *Vyšplhá/Pred/2 -/1 li/# na konec druhého dílu , získává/Adv/-6 2 body .* (Oprava tohoto příkladu byla textově i graficky znázorněna v odstavci 2.0.2).

Pravidlo se volá vždy, když po slovese následuje spojovník a slovo *li*. Zároveň musí platit jedna z následujících podmínek: sloveso nebo spojovník nejsou závislé na spojce *li*, sloveso má chybnou syntaktickou funkci nebo je spojka *li* závislá přímo na kořeni věty a ve větě (souvětí) je více sloves.

Není-li sloveso závislé na spojce, změní pravidlo závislost slovesa na spojkou s výjimkou pomocných sloves, která mají být závislá na hlavní části složeného slovesného tvaru ( v tom případě se také ověří závislost druhé části složeného slovesného tvaru). Přitom se může také opravit závislost spojovníku, ale většinou to není třeba. Byla-li spojka závislá na předcházejícím slovese nebo přímo na kořeni věty, závislost se změní, pokud se ve větě najde jiné vhodnější sloveso (sloveso, v jehož klauzi nejsou podřadicí spojky ani vztažná zájmena).

(3) *Narazí-li šířící se siločáry na překážku, jsou určitým způsobem deformovány.*



Pravidlo pro opravu závislostí a funkcí u sloves se spojkou <i>-li</i>						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	295	226	0	0	0	521
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	88 %		12 %		0 %	

## 2.2 Obecná pravidla ověřující závislosti a syntaktické funkce větých členů

Skupina pěti obecných pravidel, která představíme v následujícím oddíle, je nesourodá. Spojuje ji jen to, že není primárně zaměřená ani na základní závislostní strukturu (složenou ze sloves a spojek) jako skupina předchozí, ani přímo na syntaktická substantiva nebo předložky (jako dvě následující skupiny).

První pravidlo opravuje závislosti větých členů, které chybně překračují hranici klauzí. Druhé pravidlo ověřuje závislosti a syntaktické funkce spojení se slovy *jako*. Třetí pravidlo se soustředí na zvrtné zájmeno *se* a jeho syntaktické funkce. Poslední pravidlo, které řadíme do této skupiny, není vlastně samostatným pravidlem, ale pouze ověřuje, zda změny

morfologických značek prováděných jinými pravidly nemají ovlivnit shodné přívlastky slov, u nichž změny probíhají.

### 2.2.1 Pravidlo pro větné členy, jejichž závislost chybně překračuje hranici klauzí

Jak bylo řečeno výše, MST parser nedokáže dobře pracovat s hranicemi klauzí. Někdy tak vytváří chybné struktury, v nichž je slovo závislé na řídicím větném členu za hranicí klauzí, přestože se přímo v klauzi nachází jiný, vhodný řídicí uzel, neoddělený žádnou hranicí klauzí. Pravidlo se zaměřuje především na syntaktická substantiva, syntaktická adjektiva a předložky, které jsou závislé na slovese v sousední klauzi místo na slovese ve své klauzi.

V mnoha případech je zaváhání parseru pochopitelné, protože zkoumané slovo je od „svého“ slovesa odděleno jiným slovem, které se někdy vyskytuje na hranici klauzí: v příkladu (1) je to slovo *až*; popř. mezi slovesem a větným členem stojí neznámé slovo: v příkladu (2) slovo *Knightonem*. Často je ale taková chyba spíše překvapivá, protože hranice klauzí jasná je a správné řídicí sloveso je blíže než to, na němž je zkoumané slovo chybně závislé: příklad (3) a opravené příklady (4) a (6).

- (1) *tvrzení , že to ještě nedokazuje Baileyho verzi , podle které on narazil/-5 do oběti až jako/3 druhý , byla neudržitelná .*
- (2) *Když byl na chvílku s Knightonem/X@--- sám/2 , poznamenal :*
- (3) *Prosím tě/2 , nemluv v samých hádankách !*

Opravné pravidlo zaměřené na tento typ chybných konstrukcí patří k obecným, nespecializovaným opravným pravidlům. Vyhledává a opravuje struktury, v nichž je větný člen závislý na slovese, od kterého je oddělen nejméně jednou hranicí klauzí, přestože se v jeho klauzi nachází jiné sloveso v určitém tvaru, neoddělené od dotyčného žádnou hranicí klauzí. Změní závislost větného členu ze slovesa v jiné klauzi na závislost na nejbližším slovese v téže klauzi, případně také změni syntaktickou funkci.

Pravidlo se spouští na předložky, syntaktická substantiva a adjektiva, jejichž závislost překračuje hranici klauzí. Nalezne-li pravidlo jiné sloveso v určitém tvaru, které od zkoumaného slova není odděleno hranicí klauzí, je závislost zkoumaného slova určena chybně, pravidlo se pokusí vyhledat pro zkoumané slovo vhodný řídicí větný člen (může jím být nalezené sloveso nebo jiný vhodný větný člen z klauze podle kontextu, slovního druhu a pádu slova, jehož závislost pravidlo mění, popř. podle valence okolních slov).

- (4) Tento postřeh si zapamatujme , vrátíme se k/4/-2 němu totiž a začleníme jej do širšího kontextu později
- (5) Tři motory se zážehovým paprskem ( celkem mají být čtyři/2/-2 ) dávají elektrický výkon 320 kW .
- (6) Pórky očistíme a světlou část/-3/1 nakrájíme na dílky .

Pravidlo pro větné členy, jejichž závislost chybně překračuje hranici klauzí						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	550	15	0	0	0	565
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	91 %		7 %		2 %	

### 2.2.2 Pravidlo pro spojení se spojkami jako

V PDT jsou obraty se slovy *jako* značkovány různě podle významu. Příklady pocházejí z manuálu PDT<sup>1</sup>. Ve srovnávacích výrazech má *jako* funkci spojky **AuxC**, srovnávací výraz po spojce *jako* je závislý na spojce a má obvykle funkci **ExD** (většinou lze výraz interpretovat jako elipsu predikátu): *spí jako/AuxC/-1 zabítý/Exd/-1*. Jinak má slovo *jako* syntaktickou funkci **AuxY** (což je kategorie zahrnující několik různých funkcí jako součást víceslovné spojky či částice, „pokleslá vsuvka“ aj.) a je závislé na následujícím substantivu či adjektivu (předmětu či doplňku): *odmítl nabídku jako/AuxY/1 málo atraktivní/Atv/-3 ; připadá mi to jako/AuxY/1 neskutečné/Obj/-4*.

Toto pojetí je z lingvistického hlediska možná opodstatněné, ale pro stochastický parser je nevhodné, protože často závisí na významových nuancích, parserem nerozeznatelných. Rozlišení mezi doplňkem a předmětem možné je, předměty může parser rozeznávat podle sloves, např. *označit, chápat, vnímat*, i když v trénovacích datech je takových předmětů málo (cca 30). V důsledku složité situace chybí parser u spojek *jako* a *než* a u slov po nich následujících velmi často. Někdy jen chybně určí funkci a závislost substantiva (adjektiva) po spojce (1), (4) a (5), jindy chybně interpretuje celou strukturu: jak spojku, tak následující slovo (2) a (3).

- (1) Snažil se vypadat jako/AuxY/1 mírotvůrce/Atv/-3 , zatímco v podstatě vytvářel příznivé podmínky

1 <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s07s09.html>

(2) možná i proto si ho jako/**AuxY/2** horký brambor/**Obj/1** přehazuje z jednoho okrsku na druhý .

(3) boj lidí proti strojům , z něhož nevyhnutelně vycházeli lidé jako/**AuxC/-2** poražení/**ExD/-1** .

Pravidlo se nesnaží rozlišit automaticky mezi srovnávacími obraty na jedné straně a doplňky a předměty na druhé, tj. nerozhoduje mezi **AuxC** a **AuxY**, jen napravuje závažné chyby v rámci těchto kategorií. Pravidlo se zavolá pro všechny případy nesouladu mezi funkcí a závislostí spojky a substantiva / adjektiva, které k ní patří: spojka je označena jako **AuxY** a je závislá na následujícím slově, které nemá funkci **Obj**, **Atv** nebo **AtvV**; spojka je označena jako **AuxC** a je na ní závislé následující slovo, které nemá funkci **ExD** nebo **Adv**.

Pravidlo uvádí do souladu syntaktickou funkci substantiva či adjektiva s funkcí spojky, kterou nemění. Je-li spojka označena **AuxY**, zvolí mezi **Obj** a **Atv/AtvV** podle řídicího uzlu, funkci **Obj** může dostat pouze slovo závislé na slovesech či adjektivech z odpovídajícího seznamu (*chápat, jevit se, označit; označovaný, sloužící*), ostatním bude přiřazena funkce **Atv** (pokud je slovo závislé na substantivu) nebo **AtvV** (pro slova závislá na slovesech). Je-li spojka označena jako **AuxC**, přiřadí se substantivu, které je na ní závislé, funkce **ExD**, adjektivu funkce **Adv**. Spolehlivý algoritmus pro automatickou volbu mezi označením spojky **AuxY** a **AuxC** nebyl bohužel nalezen, pravidlo tak často pouze mění syntaktickou funkci substantiva ze zcela nepřipustné na jinou, ale bohužel také nesprávnou.

(4) čímž chtěl naznačit , že jako/**AuxY** obyčejný voják/**Sb/AtvV/1** není odpovědný ,

(5) Miluj svého bližního jako/**AuxC/-1** sám/**AtvV/ExD/-1** sebe .

Pravidlo pro spojení se spojkami jako						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	0	1	403	0	0	404
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	51 %		49 %		0 %	

### 2.2.3 Pravidlo pro určení syntaktické funkce reflexiva se

Reflexivu se jsou ve formalismu PDT přiřazeny čtyři různé syntaktické funkce podle jejich

použití ve větě: **Atr** pro *se* závislé na deverbativním substantivu; **AuxT** pro *se* závislé na slovese reflexivu tantum (*smát se*), popř. na slovese, jež má znatelně odlišný význam jako zvrtné (*hodit se* vs. *hodit*); **AuxR** pro reflexivní pasivum (*Tyto změny se/AuxR* v roce 1915 uskutečnily postupně všude) a **Obj** tam, kde plní zájmeno funkci předmětu (*Ministr Škromach se/Obj už teď chytá za hlavu*). Rozpoznání těchto funkcí je však pro parser velmi obtížné, při určování funkcí je často nutné opřít se o vlastnosti sloves (z trénovacích dat například nelze získat dostatečná data o reflexivech tantum či o netranzitivních slovesech), o strukturu věty či o význam slov. Syntaktické funkce slova *se* parser tedy často určuje chybně.

Dosud implementované pravidlo opravuje jen chybně přiřazenou funkci **AuxT**, nepůsobí v opačném směru, nevyhledává aktivně reflexiva *se* pro zvrtná slovesa ani neopravuje funkci slova *se* v sousedství deverbativních reflexivních substantiv, i když by to bylo vhodné (např. ve spojení *ztotožnění se* je tvar *se* správně interpretován jako přívlastek substantiva pouze asi v jedné třetině výskytů v korpusu). V budoucnu s doplněním těchto oprav do pravidla počítáme.

Pravidlo se spustí, když hlavní program narazí na slovo *se* označené **AuxT**, které je závislé na slovese, jež není reflexivum tantum (ani sloveso typu *hodit se*, *domluvit se*...). Je-li slovo *se* závislé na slovese *být* nebo na modálním či fázovém slovese, pokusí se pravidlo najít v klauzi jiné zvrtné sloveso, jemuž by mohlo slovo *se* přiřadit. Pokud se to nepodaří nebo slovo *se* nezávisí na takovém slovese, změní pravidlo jeho syntaktickou funkci na **Obj** nebo **AuxR** podle kontextu a vlastnosti slovesa: je-li sloveso netranzitivní, je v reflexivním pasivu a slovo *se* má mít funkci **AuxR** (1). Je-li sloveso tranzitivní a je v první či druhé osobě, má *se* funkci předmětu. Tranzitivní sloveso ve třetí osobě s neurčeným rodem nebo mužského životného rodu také pravděpodobně není reflexivní pasivum, slovo *se* pak bude mít funkci **Obj** (2). U ostatních rodů je rozhodování obtížné, častěji je *se* součástí reflexivního pasiva než předmětem, podle toho pravidlo upravuje funkce (3) a (4). Podíl změn, které funkci v konečném důsledku neopraví, je poměrně vysoký (5): oranžově označená funkce byla opravena nevhodně.

- (1) *Pěšími vojáky se/AuxT/AuxR jako pomocnými sbory pohrdalo .*
- (2) *Všichni jsme se/AuxT/Obj namačkali do svatyně předků proti těm dvěma portrétům*
- (3) *na neutrálních lodích budou považovány za kontraband , pokud se/AuxT/AuxR neprokáže opak .*
- (4) *Vyjednávání se/AuxT/AuxR protahovala a díky publicitě , které se jim dostalo , ještě*

zvyšovala

(5) část britské královské rodiny se/**AuxT/AuxR** přejmenovala z Battenbergů na Mountbatteny .

Pravidlo pro určení syntaktické funkce reflexiva se						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	60	12	4351	0	0	4423
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	86 %		12 %		2 %	

#### 2.2.4 Pravidlo pro promítnutí změny morfologické značky substantiva na jeho shodné přívlastky

Uvádíme zde i dílčí pomocné pravidlo, které samo neopravuje žádnou chybnou strukturu, ale jen promítá změnu morfologické značky z řídicího substantiva na jeho shodné přívlastky. Protože však změny provedené tímto pravidlem evidujeme samostatně, hodnotíme samostatně i úspěšnost změn, které se provedou jeho prostřednictvím.

Pravidlo se zavolá vždy, když se jiné pravidlo chystá změnit morfologickou značku syntaktického substantiva. Ověří, zda tvary všech shodných přívlastků syntaktického substantiva jsou homonymní tak, aby se i jejich morfologická značka mohla změnit (pád, rod, číslo). Pokud homonymní nejsou, změna je zablokována a neprovede se ani v pravidle, které původně změnu iniciovalo. Pokud homonymní jsou, změna se provede a zároveň je umožněna i změna morfologické značky řídicího syntaktického substantiva. Úspěchy i chyby tohoto pravidla plně odpovídají úspěchům a chybám nadřazeného pravidla, které je využívá.

V příkladech (1), (2) a (3) nadřazené pravidlo opravuje struktury se dvěma nekoordinovanými subjekty závislými na jednom slovese. Příklady (1) a (2) ukazují správnou změnu, příklad (3) změnu chybnou (modře původně správná morfologická značka, oranžově chybná oprava).

- (1) *Historici předkládající tato/**PDNP1/PDNP4** tvrzení připouštějí , že vzrůstající síla*
- (2) *Včely čalounice vykusují polokruhovitě/**AAIP1/AAIP4** kousíčky listů růží a vystylají jimi buňky pro larvy .*
- (3) *Velké kusy horniny drtí mamutí/**AAIP1/AAIP4** drtiče do velikosti tenisových míčků .*

Pravidlo pro promítnutí změny morfologické značky substantiva na jeho shodné přívlasky						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	0	0	0	0	121	121
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	68 %		0 %		32 %	

## 2.3 Pravidla pro opravu závislostí a syntaktických funkcí syntaktických substantiv

Cílem této práce je získat co nejspolehlivější data o syntaktických funkcích substantiv, většina implementovaných pravidel je proto zaměřena na jejich ověření a opravu. Parser nejčastěji chybí v předložkových frázích, pravidla, která je opravují, budou představena až v následujícím oddíle. Mnoho chyb spočívá ale také v určení závislostí a syntaktických funkcí bezpředložkových pádů. Často je jejich příčinou i chybná disambiguace, kterou se pravidla s větším či menším úspěchem také pokoušejí opravit. Opakovaně se také objevují struktury, které se v trénovacích datech vůbec nemohly objevit a které vyplývají z omezení použitého parseru (nemožnost ověření sesterských uzlů aj.), například na jednom slovese často závisejí dva nekoordinované subjekty (**Sb**) nebo jmenné části verbonominálního přísudku (**Pnom**), dvě různé syntaktické funkce jsou koordinované (formálně závislé na jedné koordinaci) atd. Do této skupiny patří šest opravných pravidel, uvádíme je v pořadí podle počtu zásahů v korpusu.

### 2.3.1 Pravidlo pro opravu shodných substantivních přívlasků

V PDT se souřadné struktury typu *pan Novák* považují za spojení shodného substantivního přívlasku s řídicím jménem: *pan* je přívlask, *Novák* je řídicí substantivum. V případě složitější struktury (více apelativ, jméno a příjmení: *pan ministr zemědělství Jan Fencel*) jsou všechna shodná substantiva ve skupině závislá na posledním: *pan/Atr/4 ministr/Atr/3 zemědělství/Atr/-1 Jan/Atr/1 Fencel*. Je to arbitrární rozhodnutí, jak v jednoduché závislostní struktuře zacházet s těmito složitými skupinami. Protože se však v trénovacích datech objevuje jen omezený počet takovýchto struktur, který neumožňuje jejich bezpečné rozpoznání, parser zde poměrně často chybí, a to v obou směrech: neznačkuje substantiva ve vhodných strukturách jako shodné přívlasky (1) a značkuje substantiva jako shodné přívlasky ve strukturách, které takto interpretovat nelze (2).

- (1) na nástěnce vedle nejšpinavějšího portrétu předsedy/**Atr/-1** Maa/**Atr/-1** , jaký jsem kdy viděl .
- (2) během těch tří dnů/**Atr/1** mise/**Adv/-4** postoupili Američané více než o dvacet kilometrů

Pravidlo se spouští ve dvou případech, jimž také odpovídají větve řešení. V prvním případě je substantivum závislé na následujícím substantivu (ne nutně těsně následujícím), a to přímo, ne přes spojku či předložku, ale není to proprium mužského životného nebo ženského rodu ani apelativum ze seznamu cca 1200 substantiv, které se v takovýchto strukturách vyskytují (*pan, docentka, ředitel, sopranistka...*), popř. se neshoduje v rodě či pádě.

Ve druhém případě vedle sebe stojí dvě substantiva v singuláru, shodují se v rodě i pádě, první patří do výše zmíněné kategorie apelativ (typických „shodných substantivních přívlastků“) nebo je to křestní jméno (ze seznamu cca 900 jmen), druhé je také apelativum z téhož seznamu, křestní jméno nebo příjmení (ze seznamu cca 2000 jmen). První substantivum není závislé na druhém ani na dále následujícím substantivu stejného typu. Druhý případ je tedy ve svých požadavcích restriktivnější, požaduje splnění více podmínek, aby pravidlo do struktury zasáhlo a označilo sousedící substantiva za strukturu se shodným substantivním přívlastkem.

Pravidlo pak ve dvou větvích ověřuje a opravuje pravděpodobně chybné shodné substantivní přívlastky a struktury, které by pravděpodobně měly být označeny jako shodné substantivní přívlastky.

### 2.3.1.1 Pravděpodobně chybně určené shodné substantivní přívlastky

Vstupem této větve pravidla jsou pravděpodobně chybně určené substantivní přívlastky. Možné řešení záleží na pádu substantiv a jejich okolí, tedy zda pravidlo dokáže nalézt vhodné řídicí uzly pro substantiva. Pro první substantivum (dosud závislé na následujícím) je nutné najít nový řídicí uzel, může to být původní řídicí uzel druhého substantiva, sloveso v klauzi s odpovídající valencí nebo chybějícím subjektem, předcházející substantivum či předložka aj.

Pokud je druhé substantivum v genitivu nebo pokud odpovídá jmenná skupina shodnému typu *město Praha* (kde se v PDT řeší závislost opačně) nebo je druhé substantivum v nominativu a první slovo umožňuje následující nominativ jmenovací (*řekou Indus*), označí pravidlo druhé substantivum za přívlastek **Atr** prvního substantiva (3).

Když se první a druhé substantivum neshodují v pádě a druhé může být podmětem či

předmětem slovesa v klauzi, změní se závislost druhého substantiva na blízké sloveso spolu se změnou funkce (4).

Není-li nic z toho možné (pro první ani pro druhé substantivum), žádná oprava se neprovede. Pravidlo se nepokouší o opravu případně chybné morfologické značky, na to je struktura příliš složitá a variabilní.

(3) , blízko byla řeka/Atr/Sb/1/-1 Nežárka/Sb/Atr/-2/-1 s mlýnem a do města vedly staré , bohaté aleje .

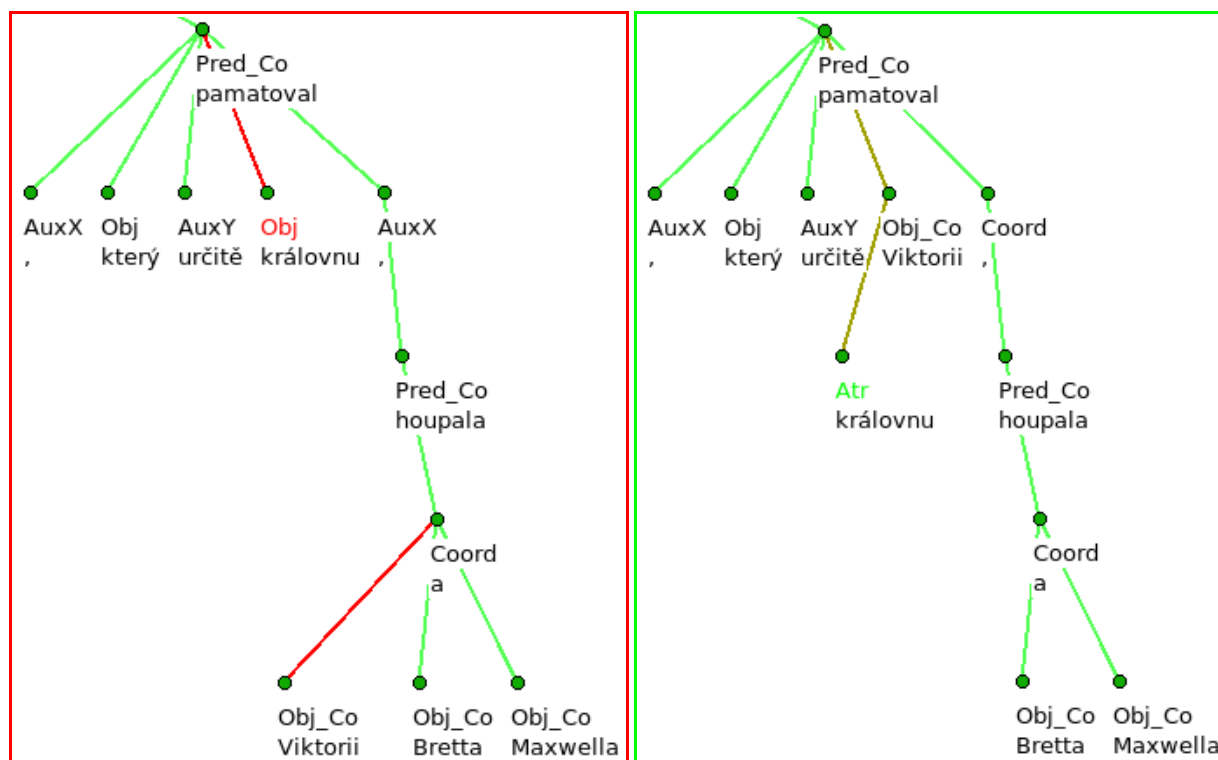
(4) Když minuli vlnolam , navedl několika jistými posunky/Atr/Obj/1/-3 člun/Obj/-4 ke královskému molu

### 2.3.1.2 Pravděpodobné shodné substantivní přívlastky neurčené jako takové

Vstupem druhé větve pravidla jsou vedle sebe stojící substantiva (nebo oddělené přísně vymezenými neshodnými přívlastky jako v následujících skupinách: *mistr světa Tomáš Dvořák*, *ministr zahraničí David Levy*), jež se shodují v pádu, rodu i čísle a patří do výše zmíněných skupin apelativ či proprií. První ze substantiv (a postupně všechna další) by mělo být závislé na posledním a mít syntaktickou funkci **Atr**. Pro druhé (poslední) substantivum je třeba zvolit vhodný řídicí uzel. Tím může být řídicí uzel prvního (5) a (6) nebo druhého substantiva (pokud druhé nebylo závislé na prvním), jestliže splňují podmínky (stejná klauze, vhodná kombinace slovesa a substantiva z hlediska pádu, valence atd., pád a pozice předložky atd.). Není-li ani jeden z původních řídicích uzlů vhodný, pokusí se pravidlo najít jiný řídicí uzel v nejbližším okolí. Jestliže vhodný řídicí uzel nenajde, opraví se jen závislost prvního substantiva na druhé a druhé převezme řídicí uzel prvního (přestože tak konstrukce zůstane pravděpodobně chybná, opraví se jen její vnitřní struktura).

(5) Zjevně zaskočen byl z včerejších myšlenkových obrátů ministra/Atr/-1/2 obrany/Atr/-1 Tvrdíka/Atr/-1/-3 premiér Vladimír Špidla .

(6) *Postávala před školou a v kočárku , který určitě pamatoval královnu Viktorii , houpala Bretta a Maxwella .*



Pravidlo pro opravu shodných substantivních přívlastků						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	682	373	8	0	0	1063
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	81 %		16 %		3 %	

### 2.3.2 Pravidlo pro opravu dvou subjektů závislých na jednom slovese

Parser při přiřazování uzlů ve struktuře, kterou vytváří, nemůže ověřovat sousední uzly, pouze uzly nadřazené a podřazené. Častým typem chyby parseru je tak struktura, v níž jsou na jednom slovese závislé dva nekoordinované subjekty, parser totiž nemůže ověřit, zda již danému slovesu jeden subjekt nepřihradil. Oprava této snadno identifikovatelné chyby je však relativně složitá, podrobnou analýzou věty je totiž třeba určit, které z mnoha možných řešení je nejlepší. Poměrně často je chyba způsobena chybným morfologickým značkováním: i tu může pravidlo odstranit. V současné verzi opravného modulu je sedm možností, jak nalezenou chybu opravit (ne vždy úspěšně a ne každou takovou chybu je modul v současnosti schopen opravit, přestože ji identifikuje).

Pravidlo se spustí, když hlavní program narazí na sloveso, na němž jsou závislé dva

nekoordinované subjekty. Pravidlo potom vyhodnocuje charakteristiku slovesa a pozici subjektů ve větě a podle toho volí větev, v níž se pokusí o nápravu chybné struktury. Postup a příklady řešení uvádíme u jednotlivých větví řešení.

### 2.3.2.1 Dva subjekty závislé na tranzitivním slovese, oba uvnitř klauze

Nejčastější podtyp chyby má příčinu v chybném morfologickém značkování, popř. v textové chybě (2): jedno substantivum má chybně určený pád, nominativ místo akuzativu. Parser na základě chybných dat určil chybně funkci (**Sb** místo **Obj**):

- (1) *Neštěstí/NNNS1/Sb také prověřuje Inspektorát/Sb bezpečnosti práce ,*
- (2) *Jakmile byla 28 . červenec/NNNS1/Sb vyhlášena válka/Sb mezi Rakouskem a Srbskem*

Tato větev opravy je zvolena, pokud jsou oba subjekty ve stejné klauzi jako sloveso a alespoň jeden ze subjektů je pádově homonymní (akuzativ – nominativ). Dále musí být sloveso v činném rodu a na slovese nesmí být závislý jiný předmět ve čtvrtém pádě ani reflexivum *se* (3), nebo musí být substantivum vyjádřením míry nebo času (4). Za těchto podmínek změní modul morfologickou značku (z nominativu na akuzativ) i funkci homonymního substantiva (ze **Sb** na **Obj**, popř. **Adv**).

- (3) *Důležitost/NNFS1/NNFS4/Sb/Obj těchto pokusů vědci/Sb zdůrazňují i tím , že dospívající mládež*
- (4) *Celé následující století/NNNS1/NNNS4/Sb/Adv byl stroj/Sb v Marly považován za další div světa .*

Jsou-li obě substantiva homonymní a mohou se přitom shodovat se slovesem v rodě i čísle, opraví pravidlo značku a funkci **druhého** substantiva, kde je o něco vyšší pravděpodobnost úspěchu, nicméně cca ve 30 % případů tak strukturu neopraví správně, ale zhorší celkovou disambiguaci věty (5).

- (5) *Význam/NNIS1/Sb má však i prostředí/NNNS1/NNNS4/Sb/Obj jeslí , školy a zaměstnání , prostředí v době volna .*

### 2.3.2.2 Dva subjekty závislé na slovese *být*, oba uvnitř klauze

Jsou-li na jednom slovese *být/bývat* závislé dva subjekty (v nominativu) a na slovese přitom není závislý jmenný přísudek, lze předpokládat, že jeden ze subjektů by správně měl být označován jako jmenný přísudek. Je-li jedním ze subjektů zájmeno *to*, bude jako jmenný přísudek označen druhý subjekt bez ohledu na shodu se slovesem (6), jinak bude opravena funkce u toho syntaktického substantiva, které se neshoduje se slovesem v rodě a čísle (7). Shodují-li se obě syntaktická substantiva se slovesem, bude opravena funkce u druhého substantiva v pořadí, opět nejde o opravu jistě správnou, ale pravděpodobněji správnou (8).

- (6) *Bude to/Sb asi něco/Sb/Pnom o kytíčkách , myslela jsem si*
- (7) *Byli ti tvorové vůbec ještě ženy/Sb/Pnom ?*
- (8) *On/Sb je totiž ten typ/Sb/Pnom člověka , který může přemýšlet , jenom když mluví .*

### 2.3.2.3 Dva subjekty těsně vedle sebe, pojmenování osoby

Tvoří-li dva nekoordinované subjekty stojící ve větě těsně vedle sebe dvojici, jež se shoduje v rodě, čísle i pádě, a oba subjekty jsou substantiva, a to buď vlastní jména, nebo substantiva ze seznamu generických označení osob (jako *pan*, *paní*, *soudruh*, *doktorka*, *prezident* aj.), musí být v rámci formalismu PDT první z nich označen jako shodný přívlástek závislý na druhém jménu (podrobně se těmto skupinám věnuje pravidlo 2.3.1). Jsou-li obě substantiva v nominativu, parser někdy nedokáže správně identifikovat jejich syntaktickou funkci a závislost, obě zavěsí na sloveso a oběma přiřadí funkci **Sb**. Správnou opravou je v tomto případě změna závislosti i funkce prvního substantiva (9) a (10).

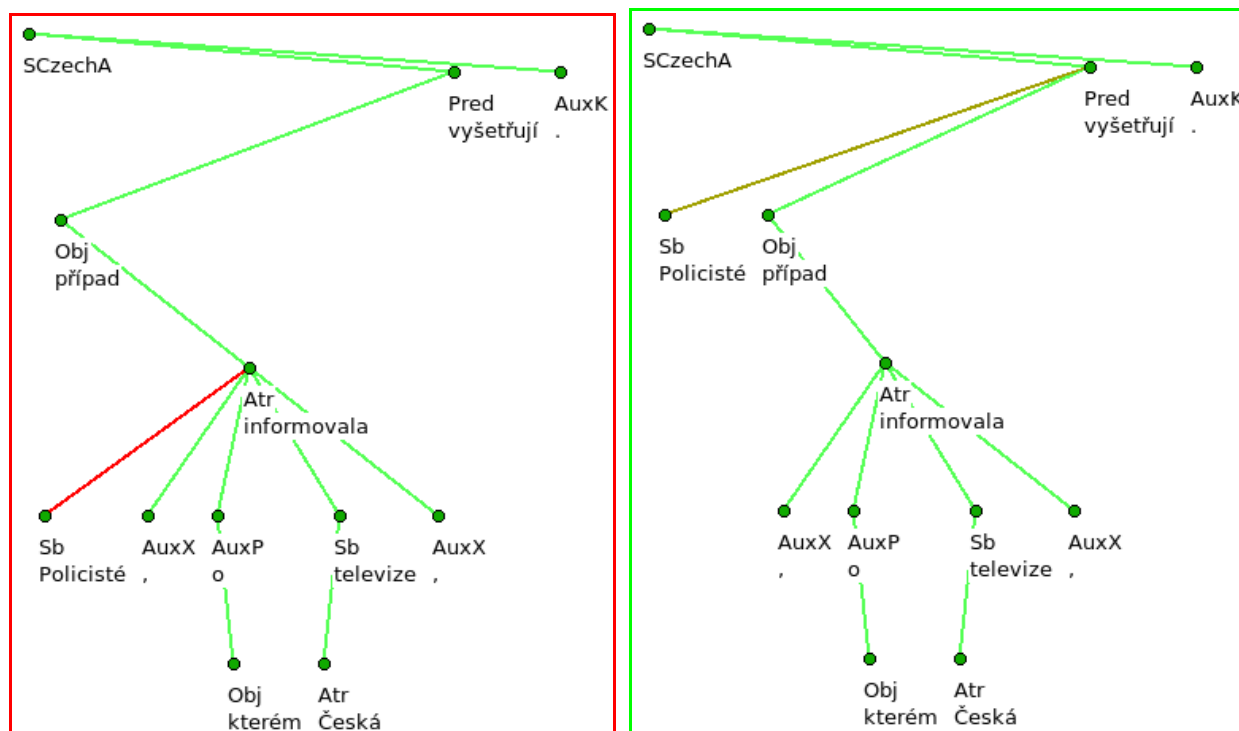
- (9) *Za Davidem Moravcem vyrazila jeho dívka/Sb/Atr/-2/1 Jana/Sb .*
- (10) *( slyšel ji říkat , jak moc se jí líbila Anna/Sb/Atr/-1/1 Kareninová/Sb )*

### 2.3.2.4 Dva subjekty závislé na jednom slovese, jeden přes jednu hranici klauze

Je-li větný člen závislý na slově, jež patří do sousední klauze, je závislostní struktura pravděpodobně chybná (ve formátu PDT kromě spojek a sloves, které v souvětích reprezentují celé vedlejší věty). Závislost může přecházet přes hranice klauzí jen tehdy, když mezi větnými členy stojí vnořená vedlejší věta, jež by sama měla být ukončena další hranicí

klauze. Pokud tedy vazba jednoho ze subjektů v identifikované chybě přechází přes právě jednu hranici klauzů a tento subjekt nemá ve „své“ části věty žádné sloveso, je třeba hledat jiné, vhodnější sloveso, na kterém by měl být subjekt závislý. Jinak je třeba hledat řídicí sloveso po vložené vedlejší větě. Takové sloveso nesmí být samo ve vnořené vedlejší větě (musí tedy následovat po čárce, po níž nestojí hypotaktická spojka ani vztažné zájmeno či příslovce), nesmí mít jiný subjekt a musí se se subjektem, pro nějž hledáme řešení, shodovat v čísle i v rodě (11 a 12). Nalezne-li opravný modul takové sloveso, může na něj subjekt převést, a chybu se dvěma subjekty tak opravit, jako v následujícím příkladu, kde sice nebyla opravena celá chybná struktura, ale chyba se dvěma subjekty byla odstraněna.

- (11) *Přesná doba/Sb/4/8 , kdy posily/Sb dorazí , zatím však není známa , neboť si budou*  
 (12) *Policisté případ , o kterém informovala Česká televize , vyšetřují .*



### 2.3.2.5 Dva subjekty závislé na slovese *být*, jeden subjekt přes dvě hranice klauzů

Je-li na jednom slovese *být/bývat* závislý jeden subjekt uvnitř klauze a druhý přes dvě hranice klauzů, přičemž vzdálený subjekt není ve vedlejší větě (uvozené hypotaktickou spojkou či vztažným zájmenem nebo příslovcem), pravděpodobně jsou správně určeny závislosti, ale chybně určena funkce. Tato větev pravidla změní syntaktickou funkci druhého subjektu na **Pnom**, bez ohledu na shodu. Shoda u dvou nominativů závislých na slovese *být* nerozhoduje, viz (13) a (14).

(13) *A že tým/Sb , který ženu operoval , nebyli žádní mladí cucáci/Sb/Pnom ,*

(14) *Jediná voda/Sb , kterou v bytě našli , byl nepatrný doušek/Sb/Pnom v odtokové trubce pod dřezem .*

### 2.3.2.6 Osamostatněné přívlastky

Část nadbytečných subjektů ve strukturách se dvěma subjekty jsou chybně oddělené přívlastky (shodné i neshodné), jež parser zavěsil přímo na sloveso a přiřadil jim funkci **Sb**. Parser takto zachází se slovy, která často stojí samostatně; takové rozhodnutí má tedy určitou oporu v trénovacích datech. Opravné pravidlo identifikuje tuto chybu u některých spojení s čísly, u vztažných zájmen a za přísně omezených podmínek také u ukazovacího zájmena *ten*.

Číslicí zapsaná čísla označená jako **Sb**, která jsou závislá na slovese, na němž je závislý i jiný subjekt, a jež těsně následují po několika typických substantivech často rozvíjených číslovkou psanou číslicí (*rok, číslo, Praha, strana* aj., dále zkratky jako *odst., čl., str.*), označí tato větev pravidla jako **Atr** závislý na předcházejícím substantivu (15).

Vztažná zájmena přívlastňovací stojící před substantivem, s nímž se mohou shodovat v rodě, čísle i pádě, se mohou vždy převést na substantivum se změnou funkce (16). U zájmen *který* a *jaký* lze spolehlivě provést změnu, jen když před zájmenem stojí čárka a před ní sloveso, takže vedlejší věta pravděpodobně není vztažná, nýbrž předmětná (17). Opravu lze také provést uvnitř věty u obrátů jako *ten který* (18). Zde někdy parser označuje hned dva chybné subjekty (jak slovo *ten*, tak *který*).

Ukazovací zájmeno *ten (to)* je často označeno jako **Sb** správně. Opravu, tj. změnu závislosti na následující substantivum a změnu funkce na **Atr**, lze provést, jen pokud se toto zájmeno shoduje s následujícím substantivem a jsou vyloučeny jiné možnosti řešení: sloveso není tranzitivní (19) nebo na něm závisí jiný předmět v akuzativu (20); na slovesu *být* závisí jiné slovo označené jako **Pnom** (21).

(15) *Soudní spor/Sb , který začal před deseti lety , vyšel zatím Prahu 8/Sb/Atr/-3/-1 na téměř 31 tisíc korun .*

(16) *epicentrem zemětřesení , jehož/Sb/Atr/4/2 tektonické vlny/Sb dodnes zmítají světovou ekonomikou .*

(17) *Především nevíme , které/Sb/Atr/5/1 kompetence/Sb na nás budou převedeny .*

(18) *na jejímž základě je ten/Sb/Atr/3/2 který/Sb/Atr/2/1 jedinec/Sb přiřazován do předem známé skupiny ,*

- (19) *Mosca s úlevou pozoroval , že to/Sb/Atr/2/1 dítě/Sb vypadá skoro přesně jako jeho .*  
 (20) *bude to naprostá pravda , jenom to/Sb/Atr/2/1 město/Sb je jiné/Pnom ,*  
 (21) *O to víc mě/Obj to/Sb/Atr/3/1 jeho sobotní rozhodnutí/Sb šokovalo .*

### 2.3.2.7 Dva subjekty závislé na jednom slovese, jeden subjekt přes hranici klauzí

Relativně málo se vyskytují struktury, kde jsou dva subjekty závislé na témž slovese, přičemž závislost jednoho subjektu jde přes hranici klauzí, přestože v rámci jeho vlastní klauze sloveso je. Pokud je to možné (vzhledem ke shodě, jinému subjektu blízkého slovesa aj.), převěsí se subjekt na nejbližší sloveso (22), (23) a (24). Tento obecnější typ chyby systematicky opravuje pravidlo 2.2.1, ale za určitých okolností nemusí být předřazeno pravidlu pro dva subjekty, takže i zde se musí takové struktury samostatně opravovat.

- (22) *Bývala to posluchárna/Sb/6/-2 , ale zbytek/Sb budovy se proměnil v ruiny .*  
 (23) *Tři miny/Sb/7/1 byly neoznačené a 40/Sb jich bylo označeno jako cvičné .*  
 (24) *Proud už zase nejde a voda neteče a zběsilci/Sb/-2/6 s ostřelovacími puškami se dnes trefují do lidí*

Pravidlo pro opravu dvou subjektů závislých na jednom slovese						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	47	45	628	273	0	993
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	79 %		19 %		2 %	

### 2.3.3 Pravidlo pro opravu podmětu v bezpředložkovém akuzativu

Protože i trénovací data obsahují jisté procento chyb, mohou se ve výsledcích parseru objevit kombinace morfologických značek a syntaktických funkcí, které jsou z lingvistického pohledu zcela vyloučené, například funkce podmět u syntaktického substantiva v bezpředložkovém akuzativu. Někdy je skutečně chybná syntaktická funkce (1), ale v poměrně velké části těchto případů je syntaktická struktura (závislosti) i syntaktická funkce určena správně, chybná je morfologická značka. Obvykle se chybně disambiguuje akuzativ místo nominativu (2), výjimečně by byl namísto jiný pád (3).

- (1) *Ale nevěřil , že jsou přesně to , co/PQ--4/Sb o sobě říkají .*

(2) *porostem zežloutlých listnatých stromů , se prudce a neočekávaně hnal silný vítr/NNIS4/Sb .*

(3) *Podobné prezentace/NNFP4/Sb se jí dostalo od konkurenční stanice Capital FM*

Pravidlo se zavolá na všechny větné členy přímo závislé na slovese, které mají syntaktickou funkci **Sb** a zároveň jsou (podle morfologické značky) v akuzativu. Velkou většinu případů lze vyřešit změnou pádu z akuzativu na nominativ (vykazuje-li slovo homonymii akuzativ–nominativ), samozřejmě za přísně stanovených podmínek, jiné špatně disambiguované pádové homonymie jsou mnohem méně časté. Ve zbytku případů pravidlo mění syntaktickou funkci: na předmět nebo příslovečné určení. Bez možnosti sledovat širší kontext a s chybnou disambiguací jako vstupem (spolehlivé řešení homonymie nominativ–akuzativ patří k nejsložitějším úkolům při disambiguaci češtiny) nemůže být pravidlo stoprocentně spolehlivé, ale v celkovém součtu přispívá k celkové kvalitě značkování (morfologického i syntaktického).

Pravidlo nejprve ověřuje, zda je na řídicím slovese závislý jiný subjekt (včetně nevyjádřeného subjektu první nebo druhé osoby) nebo objekt v akuzativu, dále zda je zkoumaný subjekt homonymní s nominativem a zda by se nominativ shodoval se slovesem v čísle (popř. v čísle a rodě).

Jestliže je akuzativní subjekt homonymní s nominativem (stejně jako případné shodné přívlastky závislé na akuzativním subjektu), sloveso se s nominativem může shodovat a nemá jiný subjekt, změní pravidlo morfologickou značku subjektu z akuzativ na nominativ (4) a (5). V opačném případě pravidlo nejprve ověří, zda akuzativní subjekt nepatří do skupiny typicky časových durativních adverbialíí (*hodina, chvíle*); pokud ano, změní syntaktickou funkci na **Adv** (6).

Jinak ověří, zda sloveso nemá jiný objekt v akuzativu, zda je sloveso tranzitivní a není v pasivním tvaru; jestliže jsou všechny podmínky splněny, změní syntaktickou funkci na **Obj** (7). Není-li možná ani tato změna, neprovede se oprava vůbec. To, že na slovese není závislý jiný subjekt a že se sloveso může shodovat s (nominativem) subjektu, není samozřejmě zárukou, že je namístě změna morfologické značky, a ne změna syntaktické funkce, opravný algoritmus tedy není zcela spolehlivý (8) a (9): chybná oprava je vyznačena oranžovou barvou.

(4) *Povstání/NNNS4/NNNS1/Sb skončilo z vojenského hlediska katastrofou , ale mělo*

nemalý politický význam .

- (5) *Pro Rusy všechno/PLNS4/PLNS1/Sb skončilo špatně .*
- (6) *Každý večer/NNIS4/Sb/Adv nám připravila něco k jídlu a čekala v posteli*
- (7) *Pokud by ale návrh/NNIS4/Sb/Obj nedal sám majitel , udělali bychom to my*
- (8) *A je nám prd/NNNS4/NNNS1/Sb platný .*
- (9) *Přivolali koně/NNMP4/NNMP1/Sb , rychle z nich sundali cestovní koše*

Pravidlo pro opravu podmětu v bezpředložkovém akuzativu						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	0	13	573	0	301	887
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	68 %		24 %		8 %	

### 2.3.4 Pravidlo pro opravu koordinace nekompatibilních syntaktických funkcí

Když je větný člen závislý na koordinaci a tato koordinace na dalším větném členu, je pro parser určení syntaktické funkce závislého stromu obtížnější, protože se mezi skutečným řídicím členem a podřízeným větným členem nachází jedna úroveň navíc. Zároveň parser nedokáže ověřit sousední uzly. Ve výstupu parseru se tak často objevují koordinované větné členy, které mají různé, nekompatibilní syntaktické funkce (1) a (2).

Oprava se zaměřuje pouze na koordinovaná syntaktická substantiva, jež se shodují v pádu (tj. koordinace je pravděpodobně v pořádku, stačí opravit syntaktickou funkci). Opravu koordinace syntaktických substantiv v různých pádech bude nutné do opravného modulu doplnit, ale vzhledem k ne zcela spolehlivé disambiguaci a k tomu, že by pravděpodobně bylo nutné zasahovat do celé závislostní struktury, je to úkol mnohem složitější a nebyl dosud vyřešen.

- (1) *V říjnu/Atr a/Coord listopadu/Adv Francouzi zlikvidovali výběžek fronty ve středu německé linie.*
- (2) *starým Brunem , který ho naučil lovit oštěpem/Pnom a/Coord prakem/Atr i žít způsobem lidí klanu .*

Pravidlo se spustí, když hlavní program narazí na dvě koordinovaná syntaktická substantiva s rozdílnou syntaktickou funkcí. Pravidlo v tom případě ověří ostatní členy

koordinace. Jestliže je mezi členy koordinace nějaký nesourodý prvek (sloveso, jiný pád aj.), mělo by tuto zřejmě chybnou koordinaci opravit jiné pravidlo: pravidlo se neuplatní.

Jsou-li členy koordinace pouze syntaktická substantiva ve stejném pádě, ověří se, zda řídicí uzel koordinace, přiřazený parserem, je nejlepší možný, případně bude vyhledán vhodnější na základě valence sloves, adjektiv, reky substantiv aj. (3). Podle řídicího větného členu (řídicí člen koordinace většinou zůstává nezměněn) pak pravidlo zvolí vhodnou syntaktickou funkci (4). Je-li řídicím uzlem předložka, pak zvolí funkci podle řídicího větného členu předložky (5) a (6).

- (3) *a sepsal spisek Průvodce po/-3/-1 Písku/Atr/1 a okolí/Adv/Atr/-1*
- (4) *Jeden má ženu/Obj a dům/Atr/Obj a všecko , nač si vzpomene , a druhý nemá nic .*
- (5) *Také toužím po teplé vodě/Adv/Obj a splachovacím záchodě/Atr/Obj .*
- (6) *které se objevily v novinách/Atr/Adv a časopisech/Adv .*

Pravidlo pro opravu koordinace nekompatibilních syntaktických funkcí						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	24	1	576	0	0	601
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	82 %		12 %		6 %	

### 2.3.5 Pravidlo pro opravu přímého předmětu závislého na netranzitivním slovese

Předmět v bezpředložkovém akuzativu může být správně závislý pouze na tranzitivním slovese v činném rodu. Parser však struktury, které tomuto pravidlu odporují, opakovaně vytváří, zčásti proto, že mu trénovací data neposkytují dostatečné údaje o valenci méně frekventovaných sloves. Někdy dává chybně přednost závislosti akuzativního předmětu na pomocném slovese (1). Obtížně rozlišuje mezi valenčními předměty a časovými či měrovými určeními v bezpředložkovém akuzativu, zvláště v případě kvantifikace (2). Situaci dále komplikuje chybná disambiguace pádové homonymie nominativ–akuzativ (3).

- (1) *Proč jsi mě/Obj/-1 nikdy nevyslechla , proč ses o to ani jednou nepokusila .*
- (2) *Podlitiny přetrvávají asi deset/Obj dní .*
- (3) *Tady už z něj netryskal proud/NNIS4/Obj pražských legend a pařížských historek ,*

Opravný modul volá toto pravidlo, když narazí na netranzitivní sloveso (tuto vlastnost ověřuje v seznamu sloves s jejich valencemi) rozvíte předmětem v bezpředložkovém akuzativu (předmět závisí přímo na slovesu). Pokud zkoumaný předmět závisí na slovese *být*, jež má funkci pomocného slovesa (nemusí být nutně označeno **AuxV**, stačí, když např. tvar přítomného času slovesa *být* rozvíjí minulé nebo trpné příčestí), bude převěšen na plnovýznamové sloveso. Je-li toto sloveso tranzitivní, změna je dokončena (4), jinak pokračují změny jako pro jiná, plnovýznamová slovesa.

Jestliže zkoumaný předmět patří mezi substantiva vyjadřující čas nebo jde o číslovku (základní číslovku *pět* a více v akuzativu, popř. číslovku neurčitou), která kvantifikuje taková substantiva, bude opravena syntaktická funkce z **Obj** na **Adv** (5) a (6). Toto řešení však není zcela spolehlivé, v některých spojeních opravuje struktury špatně (7); bylo by vhodné ještě rozlišit, u kterých sloves má časové vyjádření obvykle funkci subjektu, a ne příslovečného určení (např. *ubíhat*, *plynout*), popř. přísně vyžadovat kvantifikaci časových údajů u plurálu.

Je-li tvar předmětu pádově homonymní (nominativ–akuzativ), sloveso nemá jiný subjekt, je ve třetí osobě a může se shodovat s potenciálním nominativem u předmětu, změní pravidlo morfologickou značku (akuzativ na nominativ) a syntaktickou funkci slova z **Obj** na **Sb** (8). Ani tato změna není zcela spolehlivá, zvláště ve vztahu ke slovu *to*, jak je zřejmé z příkladu (9), kde byla chybně opravena syntaktická funkce slova *to* místo jeho závislosti (chybná oprava je vyznačena oranžovou barvou), čímž byla zablokována správná oprava následujícího slova. Pro správné určení závislostí, funkcí, pádu či dokonce slovního druhu tvaru *to* by se však mělo vytvořit samostatné pravidlo; tato problematika je velmi složitá.

- (4) *kdo psal ten vzkaz , který/Obj/1/2 jsi našel na stole ,*
- (5) *Děj , který trval ve skutečnosti tři vteřiny/Obj/Adv , proběhne na projekční stěně*
- (6) *Lord seděl několik/Obj/Adv vteřin/Atr/-1 beze slova a sledoval rybky*
- (7) *Vteřiny/Obj/Adv ubíhaly , pět , možná deset , možná dvacet .*
- (8) *jednoho dne u mě zazvonil telefon/NNIS4/NNIS1/Obj/Sb a naše dispečerka nasadila*
- (9) *ale kdepak , to/PDNS4/PDNS1/Obj/Sb/2 děťátko/NNNS4/Obj hovořilo ke mně !*

Pravidlo pro opravu přímého předmětu závislého na netranzitivním slovese						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	13	0	109	188	0	310
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	85 %		15 %		0 %	

### 2.3.6 Pravidlo pro opravu předmětu závislého na modálním či fázovém slovese

Na některých modálních a fázových slovesech (*chtít, začít*) mohou být závislá syntaktická substantiva s funkcí předmětu. Pokud je však na modálním či fázovém slovesu závislé plnovýznamové sloveso v infinitivu, obvykle je předmět závislý na něm. Pro parser je takové rozhodování obtížné, parser často raději zvolí bližší sloveso nebo sloveso v určitém tvaru. Vznikají tak případy, kdy je syntaktické substantivum v bezpředložkovém akuzativu závislé na modálním či fázovém slovese, které má infinitivní předmět (1), nebo na modálním či fázovém slovese místo na slovese plnovýznamovém závisí předmět v předložkové frázi, plnovýznamové sloveso je přitom valenční s odpovídající valencí (2).

(1) Slova/**Obj/3** ze sebe musel přímo páčít .

(2) To je úroveň , o/**3** které/**Obj** si může většina Evropy od dob ropných krizí v sedmdesátých letech nechat jen zdát .

Pravidlo se spustí z hlavního programu, když je na modálním či fázovém slovese zároveň závislý předmět v infinitivu i syntaktické substantivum s funkcí **Obj**, a to buď v bezpředložkovém akuzativu, nebo v předložkové frázi. Pravidlo se dělí do dvou větví, jedna opravuje závislost či značkování předmětu v akuzativu, druhá se zaměřuje na předložkové fráze.

V případech s předmětem v bezpředložkovém akuzativu je náprava chyby složitější. Stejně jako u výše popsané chyby se dvěma subjekty závislými na jednom slovese je častou příčinou této chyby nesprávná disambiguace pádově homonymních slov, zvláště homonymie nominativ–akuzativ.

Jestliže modální či fázové sloveso nemá jiný podmět, je ve třetí osobě a potenciální tvar nominativu u zkoumaného předmětu se shoduje se slovesem v rodě a čísle, je pravděpodobnější, že je třeba opravit morfologickou značku a funkci předmětu spíše než měnit závislosti ve větě (3). Má-li plnovýznamové sloveso v infinitivu již jiný předmět v akuzativu, je toto řešení víceméně jediné možné (4). Takováto oprava opět není zcela spolehlivá, identifikovaná chyba může být způsobena jinou chybou na jiném místě (5): chybná oprava vyznačena oranžovou barvou.

Nelze-li změnu z nějakého důvodu provést (např. proto, že předmět nevykazuje homonymii ak.–nom., na slovese již je závislý jiný subjekt aj.), převěsí pravidlo předmět na plnovýznamové sloveso, ovšem pouze pokud na druhém slovese není závislý jiný přímý předmět a sloveso je tranzitivní (6).

- (3) *Jedná se o zcela novou trať , po které by vlaky/NNIP4/NNIP1/Obj/Sb/1 mohly jezdit rychlostí kolem 300 kilometrů za hodinu ,*
- (4) *Kvůli nepravidelnému rozložení váhy nemohou nohy/NNFP4/NNFP1/Obj/Sb/-2 zcela absorbovat otřesy/NNIP4/Obj/-1 vyvolané chůzí .*
- (5) *zda vyhlášenému hlavičkáři nemohly potíže/NNFP4/NNFP1/Obj/Sb/-1 způsobit například právě údery/NNIP4/Obj/-3 míče .*
- (6) *Musela jsem si Prahu/NNFS4/Obj/-3/4 teprve zase postupně ohmatat .*

Ve druhé větvi pravidla je oprava chyby jednodušší: je-li na jednom modálním či fázovém slovese závislá předložková fráze s funkcí předmětu a zároveň sloveso v infinitivu s odpovídající valencí (předložka a pád), změni opravný modul závislost předložkové fráze z modálního slovesa na valenční sloveso v infinitivu (7). V některých specifických případech je nutné převést předmět na infinitiv slovesa, který je v závislostní struktuře ještě o stupeň níže (8).

(7) *Stát se o/4/3 mě/Obj bude muset postarat - koneckonců jsem přece nežádal o to , abych se narodil .*

(8) *že by si o/2/8 tom/Obj mohl pravidelný konzument hororů nechat jenom zdát .*

Pravidlo pro opravu předmětu závislého na modálním či fázovém slovese						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	95	0	3	16	0	114
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	87 %		4 %		9 %	

## 2.4 Pravidla pro opravu závislostí a syntaktických funkcí předložkových frází

Správně určit závislosti a syntaktické funkce v předložkových frázích (především přímo závislosti předložek a větných členů na nich přímo závislých) je pro parser obtížnější než určení závislostí syntaktických substantiv v bezpředložkových pádech, mimo jiné i proto, že kombinací předložek a pádů je mnohem více než prostých pádů a v trénovacích datech se parser s jednotlivými kombinacemi předložky, pádu a syntaktické funkce setká méně často.

Obtížné je i určení (formální) závislosti. Na předložce obvykle závisí syntaktické substantivum ve stejném pádu, jako vyžaduje předložka, nebo na ní závisí koordinační

spojka, jež koordinuje syntaktická substantiva, která jsou ve stejném pádu, jako vyžaduje předložka. V případě elipsy substantiva však může být na předložce závislé i syntaktické adjektivum. Parser navíc nedokáže spolehlivě rozlišit syntaktická substantiva a adjektiva. Chyby týkající se předložkových frází jsou velmi rozmanité, zaměřuje se na ně nejvíc pravidel v celém opravném modulu.

#### 2.4.1 Pravidlo pro opravu syntaktických funkcí větných členů závislých na předložce

Velmi častou chybou použitého parseru jsou chybně přiřazené syntaktické funkce větných členů závislých na předložce, popřípadě chybně určená závislost celé předložkové fráze. Syntaktickou funkci nese člen závislý na předložce, mezi řídicím uzlem a závislým členem jsou tak dvě úrovně, což parseru určení funkce ztěžuje. Částečným řešením je poměrně jednoduché pravidlo na opravu takto chybných struktur. Následující tři příklady ukazují typy chyb, na něž se pravidlo zaměřuje: předložkové fráze závislé na substantivu s označením **Adv** nebo **Obj** (1) a (2) a předložkové fráze závislé na slovese či adjektivu s označením **Atr** (3).

- (1) *Pro vstup do/-1 války/Adv bylo takové vysvětlení podle jeho názoru nezbytné*
- (2) *Ve starých domech není o/2 výklenky/Obj nouze .*
- (3) *Němci dosahují průlomů u/-2 Gorlice/Atr a vytlačují Rusy z Karpat .*

Pravidlo se volá v případě, že na substantivu, adjektivu, slovese nebo vybraných zájmenech a číslovkách je přes předložku závislý větný člen, který nemá odpovídající syntaktickou funkci. Na substantivu musí být závislý větný člen s funkcí **Atr**. Je-li větný člen v genitivu plurálu závislý přes předložku *z* na adjektivu, číslovce či zájmenu typu *jeden*, *první*, *další*, *některý*, *každý* (jež jsou v daném kontextu syntaktickými substantivy), musí mít také funkci **Atr**. Větný člen závislý na slovese či adjektivu musí mít funkci **Obj** nebo **Adv** (subjekty v předložkové frázi řešíme samostatným pravidlem; jmenné části verbonominálních predikátů v předložkových frázích se v PDT označují jako **Adv**).

Pravidlo opravuje funkci tak, aby odpovídala slovnímu druhu větného členu, na němž je předložka závislá, pokud nemá zvláštní důvod ke změně závislosti předložky na jiný větný člen ve větě. K posouzení takové změny se pravidlo rozděluje na dvě větve: předložka závislá na syntaktickém substantivu a předložka závislá na slovese či přídavném jménu.

##### 2.4.1.1 Chybné syntaktické funkce závislé přes předložku na (syntaktickém) substantivu

Je-li podstatné jméno předcházející předložce „valenční“ (předložková fráze s danou

předložkou a pádem obvykle závisí na něm: *řeč o 6, debata o 6, kniha o 6...*), pravidlo změny syntaktickou funkci na **Atr** (4). Stejně tak pro větné členy závislé na syntaktickém substantivu typu *první, každý* přes předložku *z*.

Pokud řídicí substantivum není valenční, zjistí pravidlo, zda se v klauzi nenachází sloveso s odpovídající předložkovou valencí (*platit za 4*); pak změny závislost předložky, funkce bude **Obj** (5).

V opačném případě pravidlo ověří, zda se v klauzi nachází sloveso, na němž často závisí předložková fráze s daným pádem a předložkou s adverbialní funkcí (*schovat za 4*); pak změny závislost předložka, funkce bude **Adv**.

Dále se změny závislost předložky, pokud se v klauzi nachází sloveso a předložka závisí na propriu – propria jsou jen zřídka rozvíjena předložkovými frázemi (6), nebo předložková fráze patří k typicky adverbialním určení – zvl. časovým nebo místním (7). Tato poslední možnost ovšem není zcela spolehlivá (8). Ve zbývajících případech zůstane závislost zachována a syntaktická funkce bude opravena na **Atr** (9).

- (4) *před sebou však viděl velkou naději na/-1 medaili/Obj/Atr*
- (5) *V sále na/-1/3 šťastný pár/Adv/Obj/-2 čekaly tři stovky pozvaných hostů ,*
- (6) *Český Telecom v/-1/2 současnosti/Adv rozmisťuje telefonní automaty s cílem zvýšit jejich využívání .*
- (7) *která se narodila císařským řezem ve/-3/-1 čtvrtek/Adv v Ústavu péče o matku a dítě v Podolí*
- (8) *celého tohoto pokoření , v němž vidí znovu pokoření z/-1/-3 roku/Adv 1844*
- (9) *Bože na/-1 nebesích/Adv/Atr , " bědovala Maja-Lisa .*

#### 2.4.1.2 Chybné syntaktické funkce závislé přes předložku na slovese nebo adjektivu

Je-li sloveso nebo adjektivum, na němž je předložka závislá, valenční (*nutit k 3, přimět k 3; šetrný k 3, nutný k 3*), pravidlo změny syntaktickou funkci na **Obj** (10) a (11). Pokud řídicí sloveso patří do seznamu sloves, na nichž jsou často závislé předložkové fráze s danou předložkou a pádem s adverbialní funkcí, nebo je-li sama předložková fráze typicky okolnostní (12), změny se funkce na **Adv**.

Jinak pravidlo ověří, zda těsně před předložkou nestojí „valenční“ (rekční) substantivum s odpovídající předložkovou vazbou. Jestliže ano, změny se závislost předložky na toto substantivum a funkce na **Atr** (pokud již původní funkce nebyla **Atr**). Ve zbývajících případech zůstane závislost zachována a syntaktická funkce bude opravena na **Adv**; ne vždy

je to ale správně, viz (13): chybná oprava je zvýrazněna oranžovou barvou.

- (10) *V praxi však – a kůň o/2 tom/Atr/Obj věděl své – takových čtyřadvacet hodin*
- (11) *Rakousko-Uhersko označilo vrahy za/-2 loutky/Atr/Obj srbské zpravodajské služby*
- (12) *Ve Francii v/5 prvních třech letech/Atr/Adv války kráčela diktatura a koalice většinou ruku v ruce .*
- (13) *co se dělo s lidmi z/-3 jejich okolí/Atr/Adv*

Pravidlo pro opravu syntaktických funkcí větných členů závislých na předložce						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	699	13	7161	0	0	7873
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	70 %		29 %		1 %	

#### 2.4.2 Oprava chybného určení syntaktické funkce substantiva v předložkové frázi závislé na slovese

I když parser správně určí závislost předložkové fráze, označení syntaktické funkce bývá problematické. Parseru scházejí rozsáhlá data o předložkových valencích sloves, nadto není hranice mezi příslovečným určením a předmětem v případě předložkových frází zcela ostrá a ani v trénovacích datech PDT se nehodnotí všechny případy zcela jednotně. Přesto lze s využitím dat z rozsáhlých korpusů poměrně spolehlivě mezi předložkovými frázemi závislými na slovesech rozlišovat a vybrat jim vhodné syntaktické funkce.

To, že problémy parseru vycházejí primárně z relativně malých trénovacích dat, dokládá skutečnost, že u často se vyskytujících valenčních sloves i u frekventovaných příslovečných výrazů jsou chyby výjimečné (substantiva v lokálu v předložkové frázi s předložkou *o* závislé na slovese *mluvit* jsou v 99 % označeny jako předmět; spojení *o půlnoci* závislé na slovese je v 97 % příslovečné určení), protože se dostatečně často vyskytovaly v trénovacích datech, zatímco u méně častých valenčních sloves a příslovečných určení je procento chyb mnohem vyšší, například substantiva v lokálu v předložkové frázi s předložkou *po* závislé na slovese *dychtit* jsou v 17 % označena chybně jako příslovečná určení (1); *o Letnicích* je ve 36 % označeno chybně jako předmět (2).

- (1) *Jsou lidé , kteří dychtí po tom/Adv , udělat ze světa jeden jediný kriminál*

- (2) *Mimoto jste mi říkal , že býváte o letnicích/Obj v Praze u Mistra Bílka .*

Jedna větev pravidla se spouští na předložkové fráze označené **Adv**, jež jsou závislé na slovese, které má valenci s danou předložkou a pádem. Druhou větev volá hlavní program na předložkové fráze označené **Obj**, které jsou závislé na slovese, jež nemá odpovídající valenci.

Valence se ověřují v seznamu cca 2000 sloves získaného z korpusů SYN2005 a SYN2010. Seznam zahrnuje jednak slovesa s předložkovou valencí (např. *toužit po* 6), jednak slovesa, jež jsou typicky rozvíjena předložkovými frázemi s okolnostním významem (nevalenční, např. *schovat se za* 4). Seznam odlišuje slovesa, která mají určitou valenci pouze jako zvrtná (*pustit se do* 2, *vyslovit se k* 3). Kromě toho jsme sestavili seznamy typických adverbiálních předložkových frází (cca 1300 kombinací lemmat substantiv s předložkou a pádem, např. *o víkendu, o Vánocích; za chvíli, za týden*). Tyto seznamy slouží jako základ pro opravné pravidlo (používají se však i v mnoha dalších pravidlech).

Je-li jako příslovečné určení (**Adv**) označeno substantivum v předložkové frázi závislé na slovese, jež má odpovídající valenci, a předložková fráze nepatří mezi typická příslovečná určení s daným pádem a předložkou, bude syntaktická funkce změněna z **Adv** na **Obj** (3) a (4). Typická příslovečná fráze změněna nebude (5), přestože někdy tato opatrnost není namístě (6).

- (3) *Fischer by na toto vysvětlení/Adv/Obj pohlížel určitě skepticky .*  
(4) *Orlin Grabbe zjistil , že se na ní/Adv/Obj podílejí dvě skupiny*  
(5) *Daisy a Tom na sebe na okamžik/Adv mlčky pohlédli .*  
(6) *Celý Šumperk čeká na chvíli/Adv , kdy se domů vrátí nový hrdina*

Ve druhé větvi směřuje oprava opačným směrem: patří-li řídicí sloveso předložkové fráze do seznamu sloves, která nemají danou valenci, ale naopak jsou obvykle rozvíjena okolnostními určeními s danou předložkou a pádem, bude funkce syntaktického substantiva v předložkové frázi změněna z **Obj** na **Adv** (7). Stejná oprava funkce se provede, jestliže řídicí sloveso nepatří do žádného seznamu (tj. není u něj evidována daná valence) a předložková fráze patří mezi typicky adverbiální (8).

- (7) *Jedu na chalupu/Obj/Adv , je , mmm , čtvrtek večer , 8.47 .*  
(8) *Muž se s očividnou úlevou/Obj/Adv protahuje dovnitř .*

Stojí-li před předložkovou frází, která je závislá na nevalenčním slovese, substantivum s rekcí odpovídající předložce a pádu, převěsí pravidlo předložku na substantivum a změní funkci substantiva v předložkové frázi (9). Je-li předložková fráze závislá na modálním či fázovém slovese, na němž je závislé také sloveso v infinitivu, které má odpovídající valenci, převěsí pravidlo předložkovou frázi na sloveso v infinitivu (10). Obě změny by za normálních okolností měla provést jiná pravidla, ale protože pořadí prováděných oprav může být ovlivněno mnoha faktory, je nutné, aby byly možnosti takových změn zahrnuty i do tohoto pravidla.

(9) *V půlce koncertu pak na scénu zavítal hráč na/-2/-1 bicí/Obj/Atr , což bylo skutečně*

(10) *Nechci na/-1/3 to/Obj ani pomyslet .*

Výsledky testování pravidla ukazují, že je ještě nutné doplnit do opravného modulu změnu morfologické značky místo změny funkce, protože některé chyby jsou způsobeny chybnou disambiguací pádu: u pádově homonymního substantiva s pádově homonymní předložkou (např. akuzativ/lokál) závislého na slovese, které má valenci s danou předložkou, ale s jiným pádem, je třeba změnit pád v morfologické značce (11), dosud však tato možnost implementována nebyla.

(11) *Na zajištění/NNNS6/Adv tahače lany čekali i hasiči*

Oprava chybného určení syntaktické funkce substantiva v předložkové frázi závislé na slovese						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	65	5	2868	0	0	2938
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	87 %		5 %		8 %	

### 2.4.3 Pravidlo pro určení závislosti předložkových frází na rekčních substantivech

Určit závislost předložkové fráze je často obtížné i pro trénovaného anotátora, natož pro parser, který se může opřít jen o omezená trénovací data, a ne o jazykové kompetence rodilého mluvčího. Volba řídicího uzlu předložkové fráze je často ovlivněna nejen řídicím uzlem (jeho slovním druhem, valencí, rekcí aj.) a lemmatem a pádem předložky, ale také významem celé předložkové fráze (zjednodušeně lemmatem syntaktického substantiva

v předložkové frázi). Parser tak v určení závislosti předložkových frází chybuje častěji než u nepředložkových pádů syntaktických substantiv (1). Pravidlo vyhledává rekční substantiva (substantiva, jež jsou často rozvíjena určitou typickou kombinací předložky a pádu), po nichž následuje předložka s odpovídajícím lemmatem a pádem. Není-li na substantivu taková předložková fráze závislá, ve vhodných případech závislost předložky mění. Mezi chybnými strukturami z výstupů parseru se často objevují struktury, kde syntaktickou funkci substantiva (**Atr**) není nutné měnit, přestože předložka není závislá na substantivu. Je to tedy chyba, kterou by případně mohlo řešit i pravidlo pro opravu syntaktických funkcí větných členů závislých na předložce (2).

- (1) *Pravděpodobně však špatná manipulace s/2 nákladem/Adv způsobila , že vozík*  
 (2) *Arabské povstání vyrostlo z touhy po/-3 nezávislosti/Atr na Turecku , která byla*

Pravidlo se zavolá, když těsně po substantivu, které patří do seznamu rekčních substantiv (např. *touha po 6*), následuje odpovídající předložka (např. *po* s lokálem), přičemž tato předložka není závislá na rekčním substantivu. Pravidlo nezasáhne do struktury, je-li předložka závislá na valenčním slovese či adjektivu (jež také vyžadují tuto předložku). Pravidlo také neprovede žádnou změnu, když předložka se substantivem patří do seznamu předložkových frází s časovým či lokálním významem, např. *o Vánocích* (3).

Jestliže se žádná taková překážka neobjeví, pravidlo převěsí předložku na rekční substantivum (4), případně také změní syntaktickou funkci slova závislého na předložce (5). Když je předložka závislá na následující koordinační spojce, ponechá pravidlo ve většině případů závislost nezměněnou, pokud není v koordinaci jednoznačně chyba, např. když koordinační spojka zároveň koordinuje slovesa i předložky (6), ani pak ale nedokáže opravit strukturu celou.

- (3) *Návrhy sledují recepty používané s úspěchem v/-3 zemích/Adv třetího světa*  
 (4) *Řekl jsem mu , že jsem víru/2/-1 v Boha ztratil .*  
 (5) *Tuhle cestu do/6/-1 Česka/Adv/Atr si Bulhaři za rámeček nedají .*  
 (6) *Když obři zmizeli , otevřel se vchod do/2/-1 skály/Atr a/-3 v něm stál/-3 Henke .*

Pravidlo pro určení závislosti předložkových frází na rekčních substantivech						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	218	0	17	0	0	235
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	85 %		6 %		9 %	

#### 2.4.4 Předložkové fráze jako přívlastky substantiv bez odpovídající rekcce

Volba správného řídicího členu předložkové fráze je často problematická. Ve větě, kde se nachází sloveso s odpovídající předložkovou valencí je však pravděpodobné, že řídicím členem předložkové fráze má být sloveso. Parser však nemá k dispozici dostatečná data o valenci méně frekventovaných sloves ani o rekčních substantivech, nedokáže proto spolehlivě rozlišit mezi přívlastkem substantiva a předložkovým předmětem slovesa a chybně určuje závislost předložkové fráze na substantivu (1) a (2).

- (1) *Je tedy třeba , aby rodiče nebo lékař s/-1 učitelem/Atr promluvili*
- (2) *skoro každéj den koukáme v kostelech na/-1 svatby/Atr .*

Opravný modul spustí toto pravidlo, když identifikuje předložkovou frázi závislou na substantivu, které nemá odpovídající rekcce (není v seznamu rekčních substantiv nebo v daném seznamu nemá přiřazenu danou kombinaci předložky a pádu). To samo o sobě neznamená chybu. Pravidlo se pokusí najít sloveso ve stejné klauzi. Najde-li sloveso, které má předložkovou valenci s danou předložkou a pádem (*dívat se po 6, pohrávat si s 7*), změní závislost předložkové fráze na toto sloveso (3). Výjimkou je případ, kdy je na valenčním slovesu závislá jiná předložková fráze se stejnou předložkou a substantivem s funkcí **Obj** , tj. valenční (4). Pravidlo může také využít infinitivní předměty modálních a fázových sloves jako řídicí uzly předložkových frází, mají-li správnou předložkovou valenci (5).

- (3) *Dívala jsem se z okna tramvaje po/-1/-6 domovních číslech/Atr/Obj ,*
- (4) *Dál si pohrával s/-1 kšiltovkou/Obj s/-1 koženým štítkem/Atr .*
- (5) *Matka o/-1/4 morčeti/Atr/Obj nechtěla ani slyšet .*

Předložkové fráze jako přívlastky substantiv bez odpovídající rekcce						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	160	0	159	0	0	319
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	76 %		9 %		15 %	

#### 2.4.5 Pravidlo pro opravu koordinací závislých na předložkách

Určení závislosti koordinace následující po předložce a jejích členů je pro parser často

obtížné, parser nedokáže spolehlivě rozlišit (například na základě slovního druhu a pádu členů koordinace), zda má být koordinace závislá na předložce nebo na jiném větném členu. Parser také často přiřazuje předložce více samostatných závislých uzlů. V důsledku toho je pak někdy na jedné předložce závislé jak následující syntaktické substantivum, tak koordinace. Někdy by mělo být součástí koordinace i první syntaktické substantivum po předložce (1), jindy koordinace na předložce závislá být nemá (2). Často je koordinace také závislá na chybné (dřívější) předložce (3). Na nespolehlivé určení závislostí v případě koordinační spojky po předložce naráželo i předchozí pravidlo pro předložky s neuspokojenou rekcí.

- (1) *v povinnostech a věcech cti , ve víře v Boha/-1 a/-2 jeho zákony/-2 !*
- (2) *Nejčastěji se horší na podzim/-1 a/-2 na/7 jaře , při účasti pylové alergie i v létě ,*
- (3) *Najdeme ji nejen v Řecku , ale i jinde na Balkáně , v Itálii/1 a/-5 jižní Francii/-4 .*

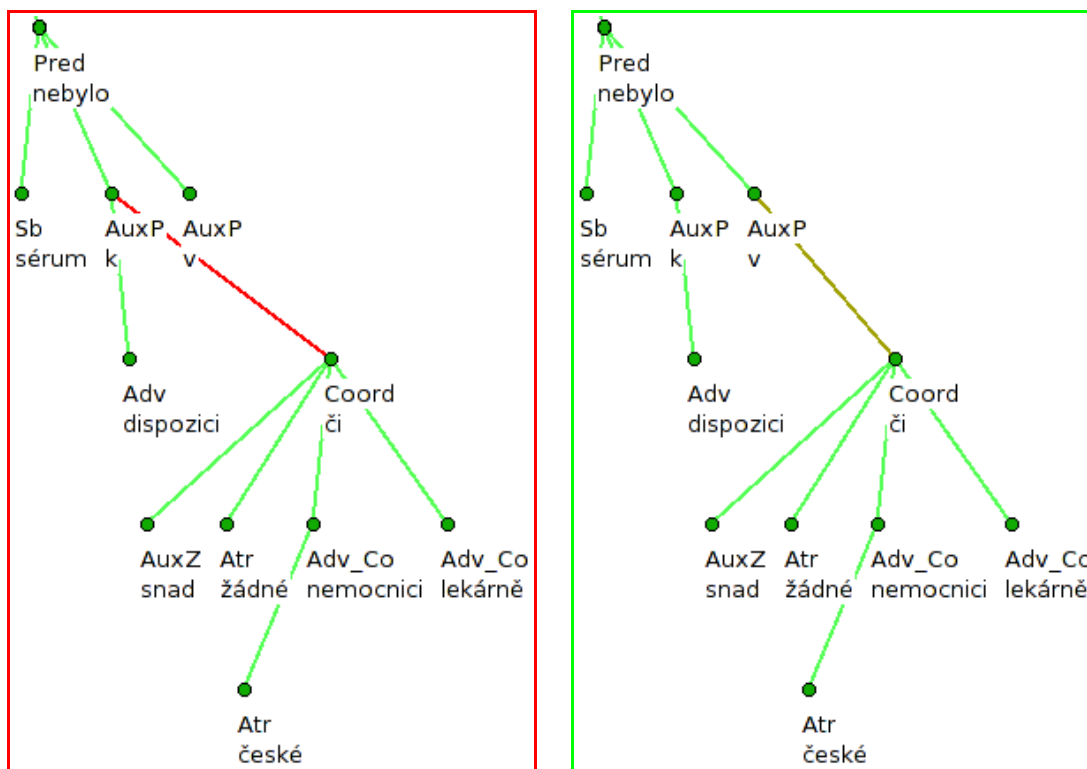
Pravidlo se spustí, kdykoli je na předložce závislá koordinace. Ověří, zda je struktura v pořádku: pokud je na předložce kromě koordinace závislý i jiný větný člen nebo pokud jsou na koordinaci závislé nesourodé prvky, pokusí se pravidlo zvolit vhodný postup opravy.

Je-li na jedné předložce závislé syntaktické substantivum ve správném pádu, které následuje po předložce a stojí před koordinací, ověří pravidlo ostatní členy koordinace a slova mezi tímto syntaktickým substantivem a koordinací. Pokud jsou členy koordinace pouze syntaktická substantiva se stejným pádem jako první syntaktické substantivum a mezi prvním syntaktickým substantivem a koordinací nestojí žádný nevhodný větný člen (sloveso, předložka, podřadící spojka aj.), převěsí syntaktické substantivum na koordinaci (4).

V ostatních případech koordinace nemá být závislá na koordinaci. Pravidlo vyhledá jiný vhodný řídicí člen, nejčastěji blízké sloveso, a koordinaci na něj převěsí (5). Jestliže se mezi předložkou a koordinací nachází jiná předložka, jejíž pád odpovídá pádu syntaktických substantiv pod koordinací, převěsí pravidlo koordinaci na tuto předložku (6).

- (4) *Znalost cizích právních řádů je pro legislativní záměr/-2/1 a/-3 jeho formulaci/-2 významná z řady důvodů .*
- (5) *Kromě toho/-1 vojáky/1 a/-3/3 děla teď potřeboval zcela jinde .*

(6) sérum nebylo k dispozici snad v žádné české nemocnici či lékárně !



Pravidlo pro opravu koordinací závislých na předložkách						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	83	4	40	0	0	127
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	70 %		19 %		11 %	

#### 2.4.6 Pravidlo pro opravy chyb u víceslovných předložkových výrazů

Pro víceslovné předložkové výrazy (*v souvislosti s, směrem k, na základě*) má formalismus PDT na analytické rovině zvláštní označení: části takového výrazu jsou označeny **AuxP** (stejně jako prosté předložky) a jsou závislé na poslední části výrazu: *v/AuxP souladu/AuxP s/AuxP*. V trénovacích datech je takto označeno 25 trojčlenných obrátů, celkem cca 300 výskytů (*v souvislosti s, s přihlédnutím k*) a několik desítek dvojčlenných obrátů, celkem cca 900 výskytů (*vzhledem k, směrem do, na základě, v oblasti*). Většina obrátů má velmi nízkou frekvenci, některé se v trénovacích datech objevují jen jednou. Ve verzi trénovacích dat, která byla použita pro trénování parseru, navíc není v označení víceslovných předložkových výrazů úplná shoda, některé z nich jsou jednou označeny jako víceslovný výraz, podruhé se s nimi zachází jako s běžným spojením předložky a substantiva. Nízká frekvence a

nejednotnost v trénovacích datech parseru neumožňují víceslovné předložkové výrazy značkovat správně. Parser chybje v obou směrech: často neoznačuje ani typické víceslovné předložkové výrazy (1) a (2) a naopak označuje jiná spojení předložky a substantiva (či předložky, substantiva a předložky) jako víceslovné předložkové výrazy (3).

- (1) *Nebo ho pod/AuxP/5 vlivem/Adv/-1 takové představy doopravdy zabil ?*
- (2) *Vzhledem/Obj/11 k/AuxP/-1 nejasněnosti pravidel publikace právních předpisů*
- (3) *letos vůbec poprvé počítá v/AuxP/2 rozpočtu/AuxP/1 s/AuxP/-3 částkou 600 tisíc korun na zajištění*

Hlavní program volá toto pravidlo, pokud narazí na víceslovný předložkový výraz (jednoznačný, sporné případy neopravuje), který není správně označen. Volá je i v opačném případě, když jsou po sobě následující slova označena jako víceslovný předložkový výraz, ale nepatří do seznamu složených předložek ani do seznamu výjimek (sporných případů). Názor autora na to, co by mělo nebo nemělo být značkováno jako víceslovný předložkový výraz, není důležitý, rozhodující je převládající značkování v PDT a seznam nepravých předložek uvedený v anotačním manuálu PDT. Sporné výrazy do tohoto seznamu sice nepatří, ale jsou analogické s jinými výrazy, které tam patří (*směrem k* je v seznamu nepravých předložek, *směrem na* nikoli). U sporných výrazů pravidlo nezasahuje ani v jednom směru.

Všem částem identifikovaných víceslovných předložkových výrazů přiřadí pravidlo funkci **AuxP**, syntaktickému substantivu, které je na výrazu závislé, bude přiřazena funkce **Adv** nebo **Atr** podle toho, na jakém slovním druhu je výraz závislý (4) a (5).

U víceslovných předložkových výrazů, které jako takové chybně označil parser, je oprava o něco složitější, je nutné najít vhodné řídicí větné členy pro jednotlivé části a vhodné syntaktické funkce pro substantiva ve výrazech. Pravidlo volí řídicí členy a funkce podle kontextu, pádu substantiv, valence okolních slov, původního řídicího členu výrazu aj. (6) a (7).

- (4) *V/AuxP/7/1 případě/Adv/AuxP/-1/6 nežádoucích účinků nebo pochybností se obraťte na svého lékaře nebo lékárníka .*
- (5) *Současně/Adv/AuxP/9/1 s/AuxP/8 uvažovanými legislativními změnami bude však nutné*
- (6) *O/AuxP/1/4 zájmu/AuxP/Obj/3/-1 Newcastleu nic nevím .*
- (7) *nastínil jako předehtu/AuxP/Adv/1/-2 ke/AuxP/-3/-1 svému putování napříč Texasem*

Pravidlo pro opravy chyb u víceslovných předložkových výrazů						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	28	48	3	0	0	79
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	76 %		13 %		11 %	

#### 2.4.7 Pravidlo pro doplnění neuspokojených rekčních požadavků předložek

Ve výstupu parseru se nezřídka objevují případy, kdy normální česká předložka stojící uvnitř věty nemá uspokojen svůj rekční požadavek (na předložce není závislý žádný vhodný větný člen), přestože po předložce následuje syntaktické substantivum s odpovídajícím pádem. Tato chyba málokdy nastane v jednoduchých, přehledných strukturách, spíše se týká vět s neznámými slovy, s větším počtem předložek, se složitější koordinací nebo s apozicí adjektiv.

Někdy je chyba jen důsledkem jiné, závažnější chyby v konstrukci celého závislostního stromu, kterou by mělo před aplikací tohoto pravidla odstranit jiné, obecnější pravidlo. Tak je tomu například v (1), kde je koordinační spojka (*a*) chybně považována za řídicí uzel celé struktury (#) místo správné závislosti na předložce (*na*), jíž chybí jakýkoli závislý člen.

Pravidlo se snaží správně naplnit rekční požadavky předložky i za cenu dalších změn struktury. Variabilita chybných struktur je však příliš velká, změna může být zablokována (například aby nedošlo k závislosti v kruhu). V tom případě se rekční požadavek naplní, pokud vůbec, nejbližším vhodným větným členem. Ostatní větné členy, například v koordinaci, která by měla být celá závislá na předložce, se pak neopraví (a struktura zůstane chybná).

(1) *Slabé světlo vrhalo odlesky/4 na/-2 skleněné pulty/1 a/# tmavé lesklé hlavně/-3 zbraní*

Pravidlo se spustí na všechny předložky, které nejsou součástí složené předložky, nejsou na konci věty a není na nich závislý žádný následující větný člen. Na začátku vlastního pravidla jsou ještě odfiltrovány předložky *vzdor*, *navzdory* a *vstříc*, u nichž není závislý člen vyžadován tak přísně, a kombinace jako *hlasovat pro*, *být proti* aj.

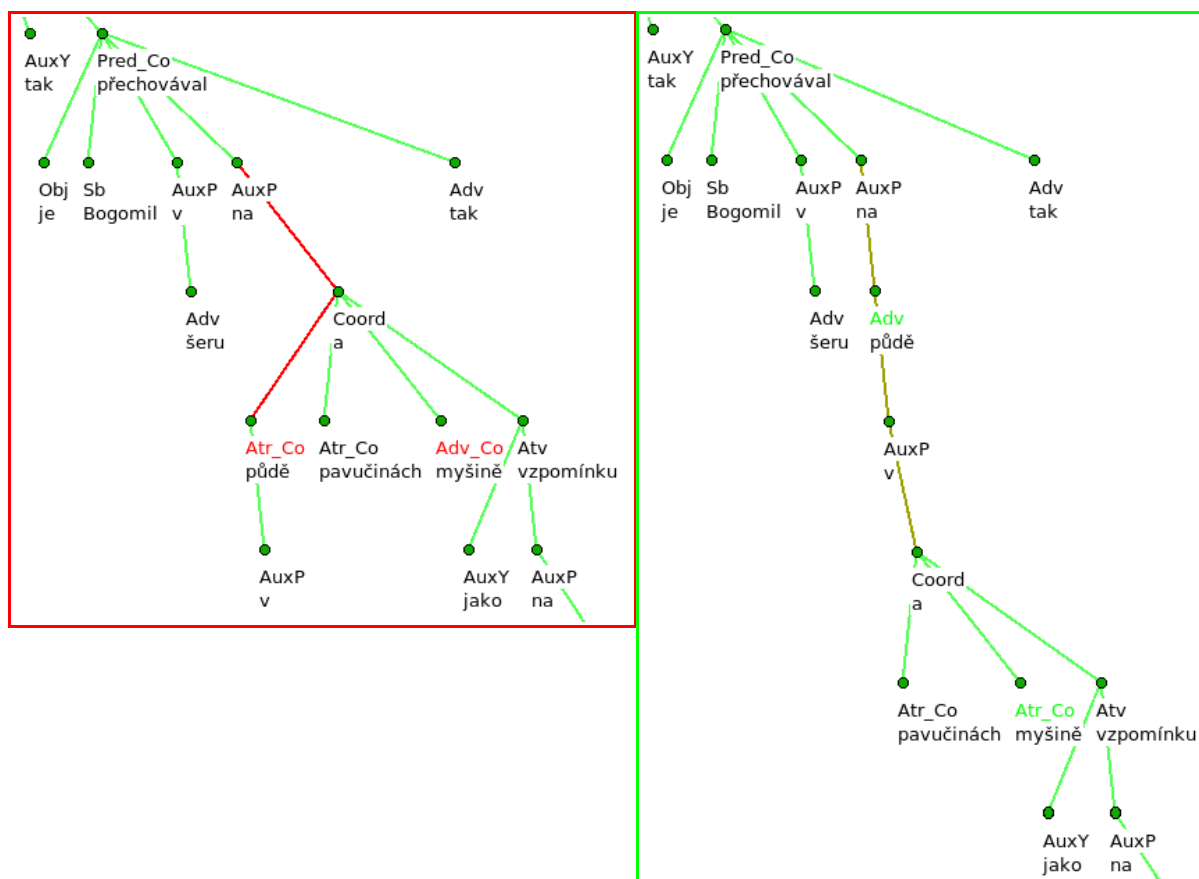
Pravidlo postupuje od předložky doprava, dokud nenarazí na pravděpodobnou pravou vnější hranici předložkové fráze (sloveso, některé typy interpunkce, další předložka,

nevhodný pád substantiva atd.). Pokud do té doby narazí na syntaktické substantivum ve vhodném pádě, ukládá si ho do zásobníku a stejně tak syntaktická adjektiva. Narazí-li před hranicí předložkové fráze na koordinaci, po níž následuje další syntaktické substantivum ve vhodném pádě, pokusí se převést substantiva na koordinací spojku a spojku na předložku (2) a (6). „Pokusí se“, protože nemůže vytvořit závislost v kruhu. Kdyby taková kruhová závislost hrozila, má pravidlo dvě možnosti složitější opravy struktury. Nevyjde-li ani jedna z těchto možností, není možné koordinaci na předložku zavěsit, rekční požadavky předložky uspokojí jen první vhodné substantivum.

Když od předložky po pravou hranici předložkové fráze nestojí žádná koordinací spojka nebo koordinací spojka nekoordinuje více slov se stejným pádem, jako vyžaduje předložka, ale najde se vhodné syntaktické substantivum, budou rekční požadavky předložky uspokojeny tímto syntaktickým substantivem (3) a (4). Pokud se najde pouze syntaktické adjektivum, bude na předložku převěšeno toto adjektivum a bude mu přiřazena syntaktická funkce **ExD** – elipsa (4). Jiné syntaktické funkce větných členů, které jsou nově závislé na předložce, se určí na základě větného členu, na němž je závislá předložka, jeho slovního druhu a dalších vlastností (valence).

- (2) *Cestu pro sebe/Obj/Atr/1 a/3/-3 čluny/Obj/Atr/-1 si museli prosekávat .*
- (3) *Nezajímala jsem se o/-3 tebe/Atr/Obj/3/-4 jako o člověka , pouze jako o matku .*
- (4) *Jsou banky při jednání s/-1 vámi/Adv/Atr/1/-1 vstřícné ?*
- (5) *Ty se dáš po/-1 hlavní/Atr/Exd/1/-1 a/# já to vezmu boční ulicí . . .*

(6) , a tak je Bogomil přechovával v šeru na půdě v pavučinách a myšíně tak jako vzpomínku na svou bývalou ženu



Pravidlo pro doplnění neuspokojených rekčních požadavků předložek						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	181	536	49	0	0	766
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	73 %		12 %		15 %	

#### 2.4.8 Pravidlo pro opravu závislosti v předložkové frázi se zájmenem

Ve spojeních, kde po sobě stojí předložka, zájmeno a substantivum, se parser dopouští chyb v určení závislosti, protože nedokáže rozlišit mezi zájmeny fungujícími jako syntaktická substantiva (*sebe, jemuž*), zájmeny fungujícími obvykle jako syntaktická adjektiva (*svému, nějaký*) a zájmeny, která mohou fungovat jako syntaktická substantiva i adjektiva podle kontextu (*všem, toho*), mimo jiné proto, že některé morfologické značky mezi těmito typy také nerozlišují (*nic* a *žádný* patří podle morfologických značek ke stejnému druhu zájmen).

Parser tak i zájmena, která jsou syntaktickými substantivy, označuje jako přívlastky následujících substantiv, která pak zavěšuje na předložku (1) a (2). K problematickému rozlišení syntaktických adjektiv a substantiv se často připojuje také chybná disambiguace pádu substantiva (3), jehož pád se chybně shoduje s pádem předložky.

- (1) *Předpokládám , že pro sebe/Atr/2 vhodnou postel/Adv/-2 má .*
- (2) *Udělal jsem pro všechny/Atr/1 kotel/Adv/-1 čaje s medem a rumem*
- (3) *Chodil se svým otcem na ryby a vyprávěl si s/RR--7 ním/P5MS7/Atr/1 vtipy/NNIP7/Adv/-2 .*

Pravidlo se zavolá, když hlavní program najde po předložce zájmeno, jež rozpozná jako syntaktické substantivum: buď je zájmeno vždy syntaktickým substantivem (*sebe, mě*), nebo je syntaktickým substantivem v daném kontextu v důsledku neshody s následujícím substantivem (*všechno, ten*). Toto zájmeno není závislé na těsně předcházející předložce, ale na následujícím substantivu.

Pravidlo změní závislost zájmena ze substantiva na předložku, substantivum pak převěsí na nejbližší sloveso. Syntaktickou funkci zájmena opraví podle pádu, předložky a případné valence slovesa. Funkci substantiva pravidlo opraví podle jeho pádu a podle valence slovesa (4), popř. podle významu substantiva, např. v (5) rozpozná substantivum jako časové určení a přiřadí funkci **Adv**.

Pravidlo se nepokouší o opravu disambiguace, nedokáže ani spolehlivě určit, že je oprava namísto (6), někdy ale kvůli chybné disambiguaci pádu substantiva chybně určí jeho syntaktickou funkci (7): chybná oprava funkce (**Obj**) je vyznačena oranžovou barvou.

- (4) *To by podle všeho/Atr/Adv/1/-1 krizi/Adv/Obj/-2/1 prohloubilo natolik ,*
- (5) *Richard na něj/Atr/Obj/1/-1 chvíli/Adv/-1/1 hleděl nechápavě .*
- (6) *vyznávají a s ním/Atr/Adv/1/-1 věky/NNIP7/Adv/1/-1 věků bydleti sobě vinšují*
- (7) *Proto vybíhaly ženy z baráků , a at' na ně dozorkyně/NNFP4/Adv/Obj/-1/1 křičely ,*

Pravidlo pro opravu závislosti v předložkové frázi se zájmenem						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	23	61	0	0	0	84
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	79 %		17 %		4 %	

## 2.4.9 Pravidlo pro opravu subjektu v předložkové frázi

Formalismus PDT umožňuje v souladu s pojetím Šmilauerovým (Šmilauer, 1966) považovat za vyjádření podmětu předložkové fráze ve výrazech přibližnosti a podílnosti: *Své lampy rozžehlo na/-1 sta/Sb mušek/Atr v trávě* (Šmilauerův příklad citovaný v manuálu PDT). V PDT je takto označeno cca 70 výrazů (nepočítaje chyby), všechny vyjadřují přibližný počet (*okolo 500/Sb dělostřeleckých granátů, na třicet/Sb domů, přes čtrnáct/Sb tisíc přízniců*) nebo distribuci (*po 28 žácích/Sb*). Pouhých 70 výskytů je ale na spolehlivé natrénování parseru málo. V trénovacích datech se navíc objevují složitější struktury jako *Na třicet/Sb ázerbájdžánských vojáků a dva Arménci/Sb byli zabiti v sobotu během bojů na severovýchodě Náhorního Karabachu*, kde je koordinována předložková fráze s nominativem a tato koordinace je závislá na slovese v plurálu.

Parser se tedy v trénovacích datech setká i s větou, která obsahující subjekt v předložkové frázi a její sloveso přitom není v singuláru. U nově anotovaných vět pak takové konstrukce považuje za přijatelné. Parser navíc není schopen ověřovat najednou větší množství podmínek, které subjekt v předložkové frázi vyžaduje.

Syntaktickou funkci **Sb** přiřazuje parser předložkovým frázím jen velmi zřídka (jen asi 0,2 % subjektů je podle parseru v předložkové frázi, zčásti přes chybně závislou koordinaci), ale toto přiřazení působí spíše náhodně a většinou je chybné (1) a (2). Někdy je značka přiřazena správně, chybně je ale určena závislost: substantivum vůbec nepatří do předložkové fráze (3).

- (1) *Například pro/3 francouzské banky/Sb/-2 byl nárůst podvodů hlavním důvodem pro přechod*
- (2) *Přes/3 900000/Sb/-1 dolarů vynaložil otcův sponzor za jediný rok na zorganizování jeho schůzek*
- (3) *Nikdy z něj jachtař/Sb/-2 nebude , řekl později Arne .*

Pravidlo, jež se zaměřuje na tyto chyby, je opatrné, opravuje jen přehledné konstrukce. Hlavní program je vyvolá, když se ve větě setká s předložkovou frází, v níž je syntaktickému substantivu přiřazena funkce **Sb**. Pravidlo ověří, zda předložková fráze jako celek splňuje podmínky pro takové označení: sloveso, na němž je předložková fráze závislá, musí být v singuláru, a jde-li o přičestí, musí být v singuláru neutra (struktura analogická výše uvedené výjimce, která se objevuje v PDT, nebyla v datech nalezena); na slovese nesmí být závislý jiný podmět, zvláště ne obvyklý podmět v nominativu; předložka řídící předložkovou frází

musí patřit mezi následující předložky: *okolo, kolem, k, přes, na* s akuzativem, *po* s lokálem; předložková fráze musí obsahovat kvantifikaci, ať už v řídicím členu fráze (*na padesát/Sb nejrůznějších map/Atr/-2*), nebo v přívlastku (*kolem 2000/Atr/1 vzorků/Sb/-2*), s výjimkou singuláru po předložce *po*, kde kvantifikace přímo vyjádřena být nemusí.

Nesplňuje-li předložková fráze všechny tyto podmínky zároveň, změní pravidlo její syntaktickou funkci na **Obj** nebo **Adv** v závislosti na pádu substantiva, lemmatu předložky a případné valenci slovesa (4) a (5). Pravidlo nedokáže správně rozpoznat nepřesnou kvantifikaci přímého předmětu (analogie s přibližností u **Sb**), a přiřazuje tedy funkci chybně (6). Opravu závislostí substantiv v nominativu, která jsou (správně) označena jako **Sb**, ale chybně závisejí na předložce, by měla zajistit jiná pravidla.

- (4) *Na ryby/Sb/Obj působí poškozením CNS a změnou krevního obrazu fenoly*  
 (5) *Likvidátor například komunikuje s centrálou výhradně přes přenosný počítač/Sb/Adv ,*  
 (6) *Slavia už nyní dluží firmě ENIC , která své injekce do klubové pokladny poskytuje formou půjček , kolem 400 miliónů/Sb/Adv korun .*

Pravidlo pro opravu subjektu v předložkové frázi						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	0	0	18	0	0	18
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	86 %		14 %		0 %	

#### 2.4.10 Pravidlo pro předložkovou frázi s předložkou *o* před komparativem

Poslední dvě pravidla ověřují méně časté, specifické struktury s předložkami, u nichž parser vykazuje vyšší chybovost. První strukturou jsou předložkové fráze s předložkou *o* s akuzativem stojící před komparativem adjektiva nebo adverbia, kdy předložková fráze upřesňuje míru rozdílu ve vlastnosti vyjádřené komparativem. Parser u takovýchto obrátů chybuje ve více než jedné třetině, a to jak v určení závislostí, tak v označení syntaktických funkcí (1), (2) a (3).

- (1) *Nikdy nepochopím , proč Anna měla o/-1 tolik/Adv/1 hodnější/Atr/1 děti než já .*  
 (2) *která zaznívá ve skutečnosti o/-1 oktávu/Atr/-1 níž/Atr/-1 ,*  
 (3) *V pátek přijel Rickard a vypadal o/-1 deset/Adv/-1 let mladší/Obj/-4 .*

Pravidlo se aktivuje, když v analyzované větě následuje po předložce *o* substantivum, zájmeno nebo číslovka v akuzativu a po něm komparativ. V omezené míře může mezi číslovkou v akuzativu a komparativem stát ještě genitiv plurálu substantiva. Pravidlo ověří, zda syntaktické substantivum či adverbium po předložce patří do seznamu cca 40 slov, které často upřesňují srovnání. Nejsou-li pak správně závislosti nebo syntaktické funkce (předložka závisí na komparativu, syntaktické substantivum závisí na předložce a má funkci **Adv**), pravidlo funkce a závislosti opraví. V případě číslovky následované substantivem v genitivu plurálu připouští pravidlo mezi substantivy a komparativy pouze omezený počet obrátů s vyjádřením času (*o několik hodin později, o pár let starší...*), vzdálenosti (*o deset kilometrů dál*) nebo ceny (*o tisíc korun dražší*). Parser zde nechybuje v určení závislosti substantiva, ale u předložky, výjimečně i u kvantifikátoru (6). Je-li komparativ (chybně) závislý na předložce nebo na slově po předložce, určí pravidlo jako řídicí uzel komparativu dosavadní řídicí uzel předložky (7).

Pravidlo zachází opatrně s předložkovými frázemi závislými na slovesech s valencí *o + ak.* obsahujícími adjektiva v komparativu, v (8) tedy nezasáhne, může však v podobném kontextu chybovat u adverbii v komparativu, jako v (9). Chybně opravená struktura v příkladu (9) je však syntakticky víceznačná.

- (4) *jsou hodně od sebe , s řasami o/-1/2 odstín/Atr/Adv/-1 tmavšími/-3 než vlasy .*  
 (5) *O/4/2 poznání/Obj/Adv/-1 dražším/1 pomocníkem jsou grily plynové ,*  
 (6) *Pak jsem se potkal v o/-1/3 pár/Adv/-2/-1 let/Atr/-1 starším/1 vydání .*  
 (7) *O/4/2 to/Obj/Adv/-1 výmluvnějšími/Obj/-2/2 se staly Neprašovy sochy .*  
 (8) *Nejde totiž o/-2 nic/Obj/-1 menšího než Ústavní soud , mocnou instituci ,*  
 (9) *když jsme ji o/6/2 to/Obj/Adv/-1 souvisleji/1 a trochu stranou požádali .*

Pravidlo pro předložkovou frázi s předložkou <i>o</i> před komparativem						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	38	9	50	0	0	97
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	94 %		4 %		2 %	

#### 2.4.11 Pravidlo pro neprojektivní spojení předložky, adjektiva, číslovky a slovesa

Neprojektivita není pro parser nijak závažnou překážkou pro vytvoření správné závislostní struktury, v případě dílčí struktury, na niž se toto pravidlo zaměřuje, však chybuje přibližně

v polovině jejích výskytů. Neprojektivní konstrukce typu *za posledních patnáct let* se v korpusu SYN2005 vyskytuje pouze asi 2600krát, pravidlo má tedy jen velmi malý efekt. V trénovacích datech je podobných struktur jen asi 20. Parser chybuje především v určení závislosti a funkce syntaktického adjektiva po předložce a před číslovkou, za jeho řídicí člen považuje někdy předložku (1), jindy číslovku (2). V menší části případů je konstrukce navíc chybně disambiguovaná (3).

- (1) *Pošlete na to svoje nejlepší lidi , na celých/ExD/-1 čtyřadvacet/Adv/-2 hodin/Atr/-1 a samozřejmě ozbrojené .*
- (2) *Na potřebných/Atr/1 šest/Adv/-2 miliónů/Atr/-1 korun se složilo ministerstvo kultury ,*
- (3) *Za/RR--2 prvních/Atr/3 patnáct/Cn--1/Sb/3 srpnových/Atr/1 dnů/Atr/-2 navštívilo centrum ve Falkultní nemocnici v Porubě 1714 dárců krve a plazmy .*

Hlavní program spustí toto pravidlo, když narazí na předložku s akuzativem (potenciálně s akuzativem, bez ohledu na aktuální disambiguaci), po níž následuje jedno syntaktické adjektivum v genitivu nebo lokálu plurálu (homonymní tvar), popř. více takových syntaktických adjektiv, dále číslovka *pět* a více nebo číslovka neurčitá typu *několik*, *pár* v akuzativu (popř. v nominativu, opět je to homonymní tvar), dále následují případně další syntaktická adjektiva v genitivu plurálu a konečně substantivum v genitivu plurálu.

Pravidlo ověří správnost závislostí, syntaktických funkcí a morfologických značek. Není-li struktura správně interpretována (na předložce závisí číslovka, na číslovce substantivum, na substantivu všechna syntaktická adjektiva), opraví závislosti, opraví také syntaktické funkce (4) a (5). Jsou-li ve spojení chybné morfologické značky, opraví také je (6).

- (4) *Za/6 prvních/Atr/1/2 pět/Adv/-2 měsíců/Atr/-1 roku 2002 získaly firmy obchodované na londýnské burze*
- (5) *v Německu se za/6 prvních/Atr/1/2 osm/Cn-S4/Sb/Adv/4/-2 měsíců/Atr/-1 roku 2000 zvedl knižní prodej o dvě procenta .*
- (6) *Po/RR--6/RR--4/4 celých/AAIP6/AAIP2/Atr/2 šest/Obj/Adv/5/-2 týdnů/NNIP2/Atr/-1 můžeš každou hodinu věnovat akcím .*

Pravidlo pro neprojektivní spojení předložky, adjektiva, číslovky a slovesa						
počet zásahů	závislosti	závislosti a funkce	synt. funkce	funkce a morf. zn.	morf. značky	celkem
	8	3	0	0	4	15
úspěšnost oprav	úspěšná oprava		neutrální změna		negativní zásah	
	96 %		4 %		0 %	

### **3. Celková úspěšnost opravného modulu**

V tomto oddíle shrneme úspěšnost jednotlivých pravidel a počet zásahů tak, aby bylo možné odhadnout celkové zlepšení syntaktického značkování, které opravný modul zajišťuje. Údaje získané z korpusu SYN2005 srovnáme také s podobnými údaji z testovacích dat PDT (e-test), kde bylo porovnání provedeno automaticky.

#### **3.1 Úspěšnost opravného modulu v korpusu SYN2005**

Úspěšnost opravného modulu v korpusu SYN2005 odvozujeme ze dvou doplňujících se údajů o každém aplikovaném pravidlu: počet změn v korpusu provedených každým pravidlem a odhad úspěšnosti pravidel odvozený ze vzorku aplikací pravidla.

##### **3.1.1 Tabulka úspěšnosti jednotlivých pravidel a celého modulu**

V první tabulce představujeme procento úspěšnosti jednotlivých pravidel spočítané ze vzorků aplikace pravidel (100 náhodných zásahů v různých textech) a celkový počet zásahů (přepočítáno na 1 000 000 tokenů). Uvedeno je všech 26 pravidel v pořadí, v němž byla představena, spolu s odkazy na číslo oddílu, kde je popsána jejich funkce. Údaje již byly uvedeny jednotlivě u každého pravidla, zde jsou jen shrnuty do jedné tabulky.

Úspěšnost pravidel v korpusu SYN2005		Procento úspěšnosti			
	Zaměření pravidla	úspěšných oprav (%)	neutrálních změn (%)	negativních zásahů (%)	celkem
2.1.1	Koordinace nekomp. vět. členů jako #	86 %	6 %	8 %	355
2.1.2	Vedlejší věty jako #	76 %	15 %	9 %	259
2.1.3	Závislost souřadných hlavních vět	96 %	4 %	0 %	1178
2.1.4	Vztažné věty	83 %	7 %	10 %	559
2.1.5	Slovesa se spojkou <i>-li</i>	88 %	12 %	0 %	521
2.2.1	Závislosti přes hranici klauzí	91 %	7 %	2 %	565
2.2.2	Spojka <i>jako</i>	51 %	49 %	0 %	404
2.2.3	Reflexivum <i>se</i>	86 %	12 %	2 %	4423
2.2.4	Změny tagu shodných přívlastků	68 %	0 %	32 %	121
2.3.1	Shodné substantivní přívlastky	81 %	16 %	3 %	1063
2.3.2	Dva subjekty na jednom slovese	79 %	19 %	2 %	993
2.3.3	Podmět v akuzativu	68 %	24 %	8 %	887
2.3.4	Koordinace nekomp. funkcí	82 %	12 %	6 %	601
2.3.5	Předmět netranzit. slovesa	85 %	15 %	0 %	310
2.3.6	Předmět mod./fáz. slovesa	87 %	4 %	9 %	114
2.4.1	Synt. funkce v předl. frázích	70 %	29 %	1 %	7871
2.4.2	Předl. fráze závislé na slovesech	87 %	5 %	8 %	2938
2.4.3	Předl. fr. závislé na rekcčních subst.	85 %	6 %	9 %	235
2.4.4	Předl. fr. závislé na nerekcčních subst.	73 %	12 %	15 %	766
2.4.5	Koordinace závislé na předložkách	70 %	19 %	11 %	127
2.4.6	Víceslovné předl. výrazy	76 %	13 %	11 %	79
2.4.7	Neuspokojené předložky	76 %	9 %	15 %	319
2.4.8	Předl. fráze se zájmeny	79 %	17 %	4 %	84
2.4.9	Subjekt v předložkové frázi	86 %	14 %	0 %	18
2.4.10	Předložka <i>o</i> před komparativem	94 %	4 %	2 %	97
2.4.11	Neprojektivní konstrukce s předl.	96 %	4 %	0 %	15
	Procento úspěšnosti	79 %	17 %	4 %	100 %
	Celkem zásahů (na 1 000 000 tokenů)	19667	4293	942	24902

Opravný modul tedy zasahuje přibližně 2,5 % tokenů v korpusu SYN2005, z toho úspěšných zásahů je cca 79 %, ale 4 % zásahů ovlivňují struktury negativně, a tyto negativní zásahy je nutné od úspěšných oprav odečíst. Výsledná úspěšnost je tedy cca 75 %.

### **3.1.2 Tabulka typu změn pro jednotlivá pravidla i celý opravný modul**

Ve druhé tabulce jsou představeny počty úspěšných zásahů opravného modulu podle jejich typu, tj. podle toho, co bylo při zásahu opraveno. Rozlišujeme pět typů opravy: opravu, která zasahuje pouze závislostní strukturu (určení řídicího větného členu slova); opravu závislostní struktury spolu se syntaktickou funkcí slova; opravu pouze syntaktické funkce; opravu syntaktické funkce a zároveň morfologické značky; opravu morfologické značky (bez dalších změn).

Uvedená čísla jsou součinem procenta úspěšnosti (počet úspěšných zásahů snižený o počet chybných změn) a evidovaného počtu zásahů jednotlivého druhu v celém korpusu, přepočteno na 1 000 000 tokenů. Z počtu zásahů je odvozeno také předpokládané zlepšení syntaktické anotace v procentech.

Rozdělení úspěšných zásahů podle typu		Typ opravy					
	Zaměření pravidla	záv.	z.+fce	funkce	fce+t.	tagy	celkem
2.1.1	Koordinace nekomp. vět. čl. jako #	193	30	55	0	0	278
2.1.2	Vedlejší věty jako #	90	64	19	0	0	173
2.1.3	Závislost souřadných hlavních vět	1061	14	56	0	0	1131
2.1.4	Vztažné věty	151	253	4	0	0	408
2.1.5	Slovesa se spojkou <i>-li</i>	260	199	0	0	0	459
2.2.1	Závislosti přes hranici klauzí	490	13	0	0	0	503
2.2.2	Spojka <i>jako</i>	0	1	206	0	0	207
2.2.3	Reflexivum <i>se</i>	50	10	3655	0	0	3715
2.2.4	Změny tagu shodných přívlasků	0	0	0	0	44	44
2.3.1	Shodné substantivní přívlasky	532	291	6	0	0	829
2.3.2	Dva subjekty na jednom slovese	36	35	484	210	0	765
2.3.3	Podmět v akuzativu	0	8	344	0	181	533
2.3.4	Koordinace nekompat. funkcí	18	1	438	0	0	457
2.3.5	Předmět netranzit. slovesa	11	0	93	160	0	264
2.3.6	Předmět mod./fáz. slovesa	74	0	2	12	0	88
2.4.1	Synt. funkce v předl. frázích	482	9	4940	0	0	5431
2.4.2	Předl. fráze závislé na slovesech	51	4	2266	0	0	2321
2.4.3	Předl. fr. závislé na rekčních subst.	166	0	13	0	0	179
2.4.4	Předl. fr. závislé na nerek. subst.	98	0	97	0	0	195
2.4.5	Koordinace závislé na předložkách	49	2	24	0	0	75
2.4.6	Víceslovné předl. výrazy	18	31	2	0	0	51
2.4.7	Neuspokojené předložky	105	311	28	0	0	444
2.4.8	Předl. fráze se zájmeny	17	46	0	0	0	63
2.4.9	Subjekt v předložkové frázi	0	0	15	0	0	15
2.4.10	Předložka o před komparativem	35	8	46	0	0	89
2.4.11	Neprojektivní konstrukce s předl.	8	3	0	0	4	15
	Celkem úspěšných zásahů	3994	1332	12792	382	228	18729
	Předpokládané zlepšení v %	0,40 %	0,13 %	1,28 %	0,04 %	0,02 %	1,87 %

Z tabulky vyplývá, že většina oprav ovlivňuje pouze syntaktické funkce, chyby v určení závislosti budou opraveny jen asi v polovině procenta tokenů v korpusu (předpokládaná redukce chyb v určení závislosti z 15,9 % na 15,4 %, tj. o 3 %). Ze sond vyplývá, že dohromady opravný modul redukuje chyby přibližně o 8 % (počet chyb v určení závislosti nebo syntaktické funkce se sníží z 23 % na 21,1 %).

## 3.2 Úspěšnost pravidel měřená na testovacích datech PDT

Úspěšnost pravidlového modulu byla měřena i samostatně na testovacích datech PDT. Zde uvádíme jednak tzv. accuracy (přesnost určení řídicího uzlu a přesnost určení syntaktické funkce), jednak tabulky úspěšnosti a typů zásahů pro jednotlivá pravidla, které jsou stejné jako tabulky pro korpus SYN2005, ale výpočet se provedl automaticky, mírně odlišnou metodou.

### 3.2.1 „Accuracy“ použitého modelu MST parseru a opravného modulu

Pro hodnocení kvality závislostního syntaktického značkování se používají dva parametry: „unlabeled accuracy“ a „labeled accuracy“. „Unlabeled accuracy“ označuje podíl správných určení řídicího uzlu u všech tokenů. „Labeled accuracy“ označuje podíl správných určení řídicího uzlu a zároveň syntaktické funkce u všech tokenů. V následující tabulce jsou uvedeny údaje pro samotný MST parser v nastavení, které jsme použili pro syntaktickou anotaci korpusu, a pro MST parser doplněný o opravný modul.

e-test PDT	MST parser	MST + opravný modul	rozdíl
unlabeled accuracy	84,12 %	84,48 %	0,36 %
labeled accuracy	76,95 %	78,28 %	1,33 %

Z tabulky vyplývá, že zlepšení naměřené na testovacích datech PDT je zřetelně menší než předpokládané zlepšení, které jsme vyvodili ze sond v korpusu SYN2005. Možné příčiny tohoto rozdílu uvádíme dále po rozboru výkonu jednotlivých pravidel, vcelku však lze konstatovat, že zlepšení opravným modulem není dostatečné, pro větší zlepšení kvality syntaktickou anotaci bude třeba hledat i jiné cesty. Podotýkáme, že samotný vstup, tj. výstup MST parseru bez opravného modulu, je zřetelně horší než nejlepší publikované výsledky téhož parseru. V následujících dvou tabulkách uvedeme výkon jednotlivých pravidel v e-testu PDT.

### 3.2.2 Tabulka úspěšnosti jednotlivých pravidel a celého modulu

V následující tabulce představujeme automaticky zjištěné hodnoty úspěšnosti jednotlivých opravných pravidel na e-testu PDT. Úspěšnost se vypočítává automaticky porovnáním s manuálně anotovanými daty, v některých ohledech je přísnější než v předchozí tabulce (nevhodná oprava syntakticky zcela nesmyslné struktury se může zařadit i mezi negativní zásahy aj.). Počet zásahů jsme přepočítali na 1 milión tokenů, aby byl srovnatelný s výše uvedenými tabulkami.

Úspěšnost pravidel v e-testu PDT		Procento úspěšnosti			
	Zaměření pravidla	úspěšných oprav (%)	neutrálních změn (%)	negativních zásahů (%)	celkem
2.1.1	Koordinace nekomp. vět. členů jako #	44 %	40 %	16 %	464
2.1.2	Vedlejší věty jako #	50 %	21 %	29 %	302
2.1.3	Závislost souřadných hlavních vět	88 %	8 %	4 %	799
2.1.4	Vztažné věty	75 %	15 %	10 %	994
2.1.5	Slovesa se spojkou <i>-li</i>	73 %	19 %	8 %	281
2.2.1	Závislosti přes hranici klauzí	77 %	13 %	10 %	518
2.2.2	Spojka <i>jako</i>	43 %	57 %	0 %	400
2.2.3	Reflexivum <i>se</i>	67 %	33 %	0 %	1760
2.2.4	Změny tagu shodných přívlastků	55 %	45 %	0 %	119
2.3.1	Shodné substantivní přívlastky	55 %	28 %	17 %	1631
2.3.2	Dva subjekty na jednom slovese	59 %	41 %	0 %	1199
2.3.3	Podmět v akuzativu	61 %	37 %	2 %	983
2.3.4	Koordinace nekomp. funkcí	46 %	50 %	4 %	583
2.3.5	Předmět netranzit. slovesa	80 %	20 %	0 %	108
2.3.6	Předmět mod./fáz. slovesa	60 %	40 %	0 %	54
2.4.1	Synt. funkce v předl. frázích	56 %	41 %	3 %	8867
2.4.2	Předl. fráze závislé na slovesech	79 %	21 %	0 %	2408
2.4.3	Předl. fr. závislé na rekčních subst.	83 %	10 %	7 %	324
2.4.4	Předl. fr. závislé na nerečních subst.	53 %	27 %	20 %	324
2.4.5	Koordinace závislé na předložkách	76 %	24 %	0 %	184
2.4.6	Víceslovné předl. výrazy	40 %	0 %	60 %	54
2.4.7	Neuspokojené předložky	65 %	16 %	19 %	810
2.4.8	Předl. fráze se zájmeny	100 %	0 %	0 %	22
2.4.9	Subjekt v předložkové frázi	100 %	0 %	0 %	11
2.4.10	Předložka <i>o</i> před komparativem	92 %	8 %	0 %	130
2.4.11	Neprojektivní konstrukce s předl.	100 %	0 %	0 %	22
	Procento úspěšnosti	62 %	32 %	6 %	100 %
	Celkem zásahů (na 1 000 000 tokenů)	14558	7527	1266	23351

Jak je z tabulky vidět, poměr úspěšných a chybných zásahů jednotlivých pravidel testovaných na e-testu PDT se značně liší od výsledků manuálního testování vzorků z korpusu SYN2005. Pravidlo pro víceslovné předložkové výrazy svými zásahy v konečném důsledku syntaktickou anotaci zhoršuje (na základě průzkumu trénovacích dat lze však soudit, že důvodem zhoršení může být nejednotné značkování víceslovných předložek v testovacích datech, které zvyšuje počet rozdílů automaticky anotovaného textu oproti

manuálnímu značkování testovacího vzorku). Jiná pravidla ve většině svých zásahů mění syntaktickou anotaci, aniž by se jim podařilo anotaci zlepšit.

### **3.2.3 Tabulka typu změn pro jednotlivá pravidla i celý opravný modul**

Ve druhé tabulce opět rozdělujeme úspěšné zásahy podle typu. Počet úspěšných zásahů se počítá jako rozdíl úspěšných zásahů a negativních, chybných zásahů. Je-li počet chybných zásahů v dané kategorii vyšší než počet zásahů správných, je číslo záporné (**červeně** zvýrazněno). Počet zásahů je přepočítán na 1 milión tokenů stejně jako v předchozí tabulce.

Rozdělení úspěšných zásahů podle typu		Typ opravy					
	Zaměření pravidla	záv.	z.+fce	funkce	fce+t.	tagy	celkem
2.1.1	Koordinace nekomp. vět. čl. jako #	129	0	0	0	0	129
2.1.2	Vedlejší věty jako #	-10	53	22	0	0	65
2.1.3	Závislost souřadných hlavních vět	648	0	22	0	0	670
2.1.4	Vztažné věty	159	476	12	0	0	647
2.1.5	Slovesa se spojkou <i>-li</i>	54	129	0	0	0	183
2.2.1	Závislosti přes hranici klauzí	346	0	0	0	0	346
2.2.2	Spojka <i>jako</i>	0	0	173	0	0	173
2.2.3	Reflexivum <i>se</i>	46	0	1142	0	0	1188
2.2.4	Změny tagu shodných přívlasků	0	0	0	0	65	65
2.3.1	Shodné substantivní přívlasky	430	172	13	0	0	615
2.3.2	Dva subjekty na jednom slovese	43	55	425	179	0	702
2.3.3	Podmět v akuzativu	0	22	366	0	184	572
2.3.4	Koordinace nekompat. funkcí	-16	11	253	0	0	248
2.3.5	Předmět netranzit. slovesa	0	0	22	64	0	86
2.3.6	Předmět mod./fáz. slovesa	32	0	0	0	0	32
2.4.1	Synt. funkce v předl. frázích	13	13	4628	0	0	4654
2.4.2	Předl. fráze závislé na slovesech	23	11	1845	0	0	1879
2.4.3	Předl. fr. závislé na rekčních subst.	216	0	32	0	0	248
2.4.4	Předl. fr. závislé na nerek. subst.	22	0	86	0	0	108
2.4.5	Koordinace závislé na předložkách	87	0	54	0	0	141
2.4.6	Víceslovné předl. výrazy	0	-10	0	0	0	-10
2.4.7	Neuspokojené předložky	107	236	35	0	0	378
2.4.8	Předl. fráze se zájmeny	0	22	0	0	0	22
2.4.9	Subjekt v předložkové frázi	0	0	11	0	0	11
2.4.10	Předložka o před komparativem	22	0	97	0	0	119
2.4.11	Neprojektivní konstrukce s předl.	22	0	0	0	0	22
	Celkem úspěšných zásahů	2373	1190	9238	243	249	13293
	Zaznamenané zlepšení v %	0,24 %	0,12 %	0,92 %	0,02 %	0,02 %	1,33 %

Z tabulky je zřejmé, že počet úspěšných zásahů v e-testu PDT je výrazně nižší než v korpusu SYN2005 (cca o 40 %). Na e-testu se ukazuje celkové zlepšení úspěšnosti o 1,33 %, na korpusu SYN2005 jsme vypočítali zlepšení úspěšnosti o 1,87 %. Rozdíly lze vysvětlovat mnoha způsoby, jež nelze ověřit, protože e-test musí zůstat pro manuální kontrolu nepřístupný. Některé z možných důvodů horších výsledků opravného modulu na e-testu PDT uvedeme v následujícím odstavci.

E-test PDT je řádově menšího rozsahu než korpus SYN2005, na němž jsme opravný modul

vyvíjeli a testovali (e-test je cca tisíckrát menší). Je možné, že některé jevy, které opravný modul úspěšně koriguje v korpusu SYN2005, se v e-testu neobjevují. Při testování opravného modulu na trénovacích datech PDT jsme narazili i na občasné chyby manuální syntaktické anotace, nelze tedy vyloučit ani chyby v e-testu. Zásahy opravného modulu, které nevhodným způsobem řešily zcela nesmyslnou syntaktickou strukturu, se na e-testu započítaly jako negativní, přestože výsledná struktura je „méně chybná“. Pokud by například pravidlo, které vyhledává a odstraňuje dva nekoordinované subjekty závislé na jednom slovese, špatně zvolilo a opravilo nevhodný subjekt, vznikla by sice lepší struktura s jedním **Sb** a jedním **Obj** či **Pnom**, ale při testování by se projevila jen změna ze správného **Sb** na chybný **Obj** či **Pnom**.

Skutečné výsledky by podle těchto hypotéz mohly být mírně lepší, než jak byly automaticky zjištěny, tyto úvahy však nelze ověřit a do výsledků započítat (testovací data musí zůstat „slepá“). Za prokázané považujeme tedy to, že opravný modul zlepšil výsledky „labeled accuracy“ MST parseru o 1,33 %.

### 3.3 Závěr

Zlepšení o 1,33 % v případě automaticky syntakticky anotovaného korpusu není zanedbatelné, ale zaostává za našimi očekáváními. Vstupní data navíc obsahovala o něco vyšší procento chyb než nejlepší publikované výsledky MST parseru (Novák et al. 2007), ty se nám ale dosud nepodařilo reprodukovat, celkovou úspěšností se tak naše výsledky řadí jen mezi lepší průměr syntaktické anotace češtiny, nejlepší dosažené výsledky jsme nepřekonali.

Přesto nelze vývoj tohoto opravného modulu považovat za slepou uličku vývoje: jak ukážeme dále, opravný modul lze dále vyvíjet a lze ho také aplikovat na výstupy parserů s lepšími výsledky. Přínos opravného modulu u lépe nastavených parserů bude sice nižší, ale celkový výsledek by měl přesáhnout i nejlepší dosažené výsledky automatické syntaktické anotace.

## 4. Další možný rozvoj opravného modulu

Výsledky opravného modulu zaostávají za původním očekáváním, přesto znatelně zlepšují výsledky stochastického parseru. Jednou z možných cest k dalšímu zlepšení kvality syntaktické anotace je tedy další rozvoj opravného modulu, který by měl zahrnout opravu současných pravidel, rozšíření seznamů slov, jichž pravidla využívají, doplnění nových pravidel a zapojení frazémového modulu.

## 4.1 Oprava chybu pravidel a doplnění algoritmů

Při testování jednotlivých pravidel na nových datech bylo zjištěno, že některá pravidla se poměrně často dopouštějí chyb: zasahují do správných struktur (zřídka) nebo chybně opravují špatné struktury a dále tak zhoršují jejich stav. Chybovost pravidel (procento negativních zásahů) se pohybuje mezi 0 % a 32 % (na e-testu PDT až 60 %). V pravidlech, která často chybují, je nutné upravit algoritmus tak, aby k chybám docházelo méně nebo vůbec. K dispozici je dostatečně rozsáhlý korpus, na kterém lze zásahy testovat, do pravidel lze snadno přidat omezení nebo složitější rozhodovací algoritmy. Podrobnější rozpracování pravidel se vyplatí pouze u těch, jež se aplikují dostatečně často.

Některá pravidla jsou naopak příliš opatrná a opravují jen jasně přehledné struktury, i když dokážou identifikovat mnohem více chybných struktur: nemají-li bezpečný algoritmus pro opravu, nezasahují do struktury vůbec. Pokud je ale takováto struktura (identifikovaná jako chybná, ale neopravovaná) dostatečně častá, vyplatí se upravit algoritmus tak, aby pravidlo opravovalo i další podtypy chyby, aniž by přitom dělalo více chyb.

## 4.2 Doplnění dalších pravidel

Rozbor výsledků parseru a také analýza výsledků opravného modulu ukázala, že oblastí, na něž by se pravidla mohla zaměřit (často se opakující typ chyby s jasným opravným algoritmem), je i po implementaci prvních 26 pravidel ještě dost, zvláště v oblasti závislostí okolo koordinací a interpunkce. Lze však předpokládat, že přínos rostoucího množství pravidel bude klesat, případné zdvojnásobení počtu pravidel by pravděpodobně vedlo k menšímu než dvojnásobnému zlepšení parsingu. Zvažujeme tedy zapojit jen několik dalších pravidel, u nichž předpokládáme dostatečně vysokou návratnost a malou chybovost a jejichž oblast působnosti již byla zmapována. U některých pravidel jsou již připravené opravné algoritmy, stačí pravidlo jen implementovat, jiná pravidla jsou nutná proto, že chybné struktury blokují opravy, které by mohla provést pravidla již implementovaná.

## 4.3 Rozšíření používaných seznamů

Seznamy, jež pravidla využívají pro práci s valencí sloves a adjektiv, pro určení reflexivity, pro zpracování skupin typu *pan Novák aj.*, pocházejí zčásti ze starších zdrojů, jako je projekt pravidlové morfologické disambiguace, zčásti byly nově vyvozeny z korpusových dat (korpusy SYN2005 a SYN2010). Při takovém poloautomatickém zpracování byly do seznamů zařazeny pouze dostatečně frekventované výskyty. Rozšíření seznamů o méně frekventovaná slova umožní spolehlivě opravovat větší procento chybných struktur.

V případě valence bude navíc nutné zjemnit valenční seznam. Tento seznam dosud obsahuje pouze informaci, že určité sloveso má valenci s jistým pádem nebo pádem a předložkou (*toužit po 6*), popř. u slovesa, kde je časté spojení s okolnostním určením, obsahuje seznam informaci, že spojení pádu či předložky a pádu s daným slovesem je adverbialního typu (*schovat za 4*). Pro spolehlivější určení závislostí předložkových frází by však bylo třeba rozlišovat valenci alespoň ve dvou stupních (fakultativní a obligatorní): jestliže se například určité sloveso s obligatorní předložkovou valencí objeví ve stejné klauzi s odpovídající předložkovou frází, téměř vždy bude předložková fráze záviset na slovesu (*toužit po 6, usilovat o 4*), zatímco vazba jiného slovesa s fakultativní předložkovou frází může být mnohem slabší (*začít s 7, bránit v 6*), i když obě předložkové fráze mohou být označeny jako **Obj**, tj. jako valenční předmět slovesa. Díky jemnějšímu rozlišení valenčních požadavků sloves by méně chybovala například pravidla zaměřená na změnu závislosti předložkových frází (v případě volby mezi závislostí na slovesu a na substantivu aj.).

#### 4.4 Zapojení frazémového modulu

Do projektu pravidlové morfologické disambiguace češtiny je zapojen také frazémový modul, který rozpoznává frekventované frazémy, frazeologismy, úsloví a podobná spojení, jež jsou často z hlediska morfologie i syntaxe specifické. V této chvíli se frazémy automaticky rozpoznají u cca 3 % tokenů. U části takto určených frazémů by bylo možné přímo určit správnou závislostní strukturu a syntaktické funkce, a to zvláště u těch, u nichž parser často chybuje a jejichž syntaktická struktura je v něčem výjimečná (pokud vůbec lze u frazémů o syntaktické struktuře uvažovat). Celkový přínos (zlepšení syntaktické anotace) by však pravděpodobně nepřesáhl desetiny procenta.

## Kapitola IV

### Jiné možnosti zlepšení automatické syntaktické anotace

V závěru práce na pravidlech pro automatické opravy výstupu parseru se ukázalo, že variabilita chyb stochastického parseru je tak velká, že omezený počet pravidel může výsledek zlepšit jen relativně málo, celkový přínos opravného pravidlového modulu tak značně zaostává za očekáváním. To vedlo k úvahám, jak výstup parseru zlepšit jiným způsobem. Tyto úvahy bohužel již nebylo možné uskutečnit před dokončením této práce. Přínos postupů, které budou dále popsány, tedy není vyzkoušený, postupy jsou však založené na rozsáhlé analýze chyb parseru a jejich pravděpodobných příčin.

Navrhované postupy zahrnují lepší využití existujících parserů, dosud jsme totiž s parserem pracovali jako s takzvanou „černou skříňkou“ bez vlastního trénování a vlastní optimalizace parametrů. Také je možné využít kombinace parserů. Dále by bylo možné na základě postupů použitých v automatickém opravném modulu doplnit a zlepšit výsledky staršího čistě pravidlového parseru (Holan et al. 2006). Posledním navrhovaným způsobem, jak zlepšit parsing, je pravidlové předzpracování dat pro parser, které zjednoduší morfologickou anotaci na vstupu a zároveň doplní další údaje nezbytné pro kvalitní syntaktickou analýzu.

#### 1. Lepší využití existujících parserů, kombinace parserů

Při přípravě dat syntakticky anotovaného korpusu jsme využili předem nastavený model používaný v rámci prostředí TectoMT. Až během zpracování a testování dat se ukázalo, že model má o cca 1,5 % nižší úspěšnost než nejlepší dokumentovaný model téhož parseru (Novák et al. 2007). Protože ale nebyl k dispozici jiný, lepší model a na vlastní trénování a optimalizaci parseru nezbýval v pokročilém stadiu projektu dostatek času, použili jsme pro syntakticky anotovaný korpus tento nedokonalý model. Pro získání spolehlivějších dat ale bude nutné optimalizovat nastavení nejméně dvou parserů: MaltParser Joakima Nivreho (Nivre et al. 2006) a MST parser Ryana McDonalda (McDonald et al. 2005), nově je natrénovat a porovnat jejich výsledky.

Poněvadž v automatické syntaktické analýze češtiny byly zaznamenány při využití kombinace parserů velmi dobré výsledky, bude nutné vyzkoušet i zmíněné kombinace. Pro syntaktickou analýzu rozsáhlých korpusů je nutné, aby celková rychlost procesů umožnila

označkovat celý korpus v rozmezí nejvýše několika týdnů.

Bude-li dosaženo znatelně lepších výsledků v parsingu samém, lze předpokládat, že účinnost opravného modulu poklesne. Přesto může po případné revizi dále zlepšit kvalitu syntaktické anotace.

## **2. Doplnění a zlepšení pravidlového parseru**

Jednoduchý závislostní parser pro češtinu založený na pravidlech, který vytváří závislostní strukturu odpovídající PDT, naprogramoval Zdeněk Žabokrtský (Holan et al. 2006). Tento parser je součástí nástrojů TectoMT. Jeho úspěšnost nedosahuje výsledků nejlepších statistických parserů, ale podrobná analýza jeho chyb a jejich oprava, rozšíření pravidlového parseru o další pravidla a zapojení některých postupů uplatněných v opravném modulu by mohly zlepšit jeho výsledky tak, aby byl „konkurenceschopný“ ve srovnání se stochastickými parsery a lépe využitelný například v kombinacích více statistických parserů.

## **3. Pravidlové předzpracování dat a úprava tagsetu**

Poslední způsobem, kterým navrhujeme zlepšit kvalitu syntaktického značkování, je pravidlové předzpracování dat pro parser, které by parseru zjednodušilo vstupní data pro trénování i pro vlastní anotaci nových textů, čímž je možné dosáhnout zrychlení parseru a zlepšení kvality anotace. Návrhy na předzpracování dat vycházejí z analýzy chyb parseru na rozsáhlých datech.

### **3.1 Hlavní překážky kvalitního parsingu**

Nejprve je třeba připomenout hlavní problémy, na které parsery (a snahy o jejich co nejlepší nastavení) narážejí. Jsou to především tři oblasti: složité morfologické značky nepřizpůsobené potřebám parserů, dále relativně malá trénovací data, z nichž parser nemůže získat spolehlivé údaje o méně častých slovech a strukturách, a konečně složité, nepřehledné struktury v psané češtině, zvláště kumulace koordinací a interpunkce uvnitř vět.

#### **3.1.1 Morfologické značky**

Morfologické značky používané v PDT nebyly vytvořeny pro potřeby syntaktického parseru. Informace v nich obsažené dobře popisují morfologické kategorie a další vlastnosti slova, v některých případech ale důležité syntaktické vlastnosti ze značky vyčíst nelze, nebo jen obtížně. Zvláště je to patrné v případě odlišení syntaktického substantiva a adjektiva. Například kategorie neurčitá zájmena (v morfologické značce je jejich slovní druh a podtyp

slovního druhu označen **PZ**) zahrnuje jak syntaktická substantiva (*něco, cosi, kdekdo*), tak syntaktická adjektiva (*nějaký, jakýkoli, všelijaký*). Parser se tak musí opřít i o lemma, ale u méně častých slov pro to nemá dostatečně velká trénovací data. Parsery navíc obvykle nemohou zpracovat celý rozsáhlý tag (byly by příliš pomalé), zaměřují se pouze na dílčí informace z tagu, např. první, druhou a pátou pozici (nejčastěji používané nastavení MST parseru pro češtinu používané v nástrojích MFF UK). Na těchto sledovaných pozicích se však u některých slovních druhů nenacházejí důležité údaje: např. slovesa pátou pozici tagu (téměř) nevyužívají, důležitější je zato pozice osmá (osoba).

Řešením tohoto problému by mohlo být zjednodušení morfologických značek pro účely syntaktické anotace, nové značky by na menším prostoru zobrazovaly právě ty informace, které jsou relevantní pro syntaktické značkování. Podrobnější popis navrhovaného zjednodušení následuje níže.

### 3.1.2 Trénovací data

Jak bylo řečeno výše, rozsah trénovacích dat neumožňuje parseru získat dostatečné množství údajů o vlastnostech slov kromě těch nejčastějších. Rozšíření trénovacích dat by si vyžádalo obrovské úsilí a velmi mnoho času, přesto by problémy s nedostatkem údajů o méně frekventovaných slovech a strukturách nevyřešilo. Řešením je dodat parseru v jednoduché automaticky zpracovatelné formě ty údaje (valenci aj.), které potřebuje ke správné interpretaci struktur. Tyto údaje lze poloautomaticky získat z rozsáhlých korpusů (např. z korpusu SYN Českého národního korpusu, který v roce 2011 zahrnuje cca 1,3 miliardy označkových textových slov).

### 3.1.3 Složité struktury (koordinace)

V českých psaných textech se často objevují relativně složité struktury zahrnující koordinace, apozice a různé výčty a seznamy. Objevují se jak v trénovacích datech, tak v datech zpracovávaných (korpus SYN2005). Protože parser nesleduje sesterské uzly slov, jimž přiřazuje závislosti, často vytváří chybné hybridní koordinace (např. koordinace sloves a syntaktických substantiv), a celá základní struktura větné závislosti je tak často chybná, což způsobuje další chyby ve vedlejších větvích. Překážky ve vytváření základní struktury mohou omezit jednoduchá pravidla, která v trénovacích datech i v nově zpracovávaných textech rozlišují mezi koordinací větnou a členskou, mezi hranicemi klauzí a oddělením větných členů.

## 3.2 Návrh předzpracování trénovacích i nových dat za účelem zlepšení parsingu

Výše uvedené překážky kvalitního parsingu lze zmírnit automatickým postupem, který zde popíšeme. Postup nebyl ověřen, ale na základě rozboru chyb parseru může chybovost znatelně snížit. Spočívá v automatickém přizpůsobení morfologického tagsetu a doplnění různých syntaktických informací do trénovacích dat. Parser bude přetrénován s nově označenými daty, a bude tak schopen využívat nové informace i v datech nově zpracovávaných, která budou stejným způsobem obohacena. Postup bude plně automatický a poměrně jednoduchý, rychlost parsingu by neměl výrazně ovlivnit.

### 3.2.1 Přizpůsobení morfologického tagsetu potřebám syntaktického parseru

Jak bylo řečeno výše, informací obsažených v morfologických značkách je pro syntaktický parser příliš mnoho na to, aby byl dostatečně rychlý a aby údaje, jež získává z trénovacích dat, nebyly příliš rozsáhlé. Na druhou stranu je výsek z morfologické značky běžně používaný k rychlému parsingu nedostatečný, jedna pozice u jednoho slovního druhu může nést relevantní informace, u druhého slovního druhu je zato zcela nezajímavá. Navíc je podrobné rozlišení slovních druhů a poddruhů ze syntaktického pohledu často nevýznamné.

#### 3.2.1.1 Zúžení a zjednodušení morfologické značky

Pro parsing je důležité jednak zjednodušit rozlišení mezi mnoha podtypy adjektiv, zájmen a číslovek na rozlišení, které je syntakticky relevantní, dále je třeba zúžit morfologickou značku na maximálně pět pozic tak, aby obsahovala všechny informace důležité pro daný slovní druh. U syntaktických substantiv a adjektiv je důležitý rod, číslo a pád. U sloves číslo, osoba a údaj rozlišující typ slovesného tvaru (tvar oznamovacího způsobu přítomného času; imperativ; přičestí činné nebo trpné). U dalších slovních druhů stačí menší množství údajů (pád a vokalizace u předložek, typ u spojek atd.).

#### 3.2.1.2 Syntaktická substantiva a syntaktická adjektiva

Pro parsing je zásadní rozlišení mezi syntaktickými substantivy a adjektivy, syntaktická substantiva mohou nést vlastní syntaktickou funkci (*Sb*, *Obj*, *Adv* atd.), být přímo závislá na předložkách a podobně, zatímco syntaktická adjektiva obvykle rozvíjejí substantiva jako shodné přívlastky, popř. mohou být jmennou částí verbonominálního predikátu; stojí-li samostatně bez substantiva, jde většinou o elipsu (*ExD*).

Je třeba přesněji rozlišovat mezi vlastními substantivy (jež často bývají rozvíjena shodnými i neshodnými přívlastky), jinými syntaktickými substantivy (kam řadíme zvl. zájmena osobní,

některá neurčitá, záporná či totalizující jako *někdo, nic, všechno* aj.), která bývají rozvíta jen zřídka (*my křesťané, někdo z dnešních čtenářů*), některými číslovkami v přímých pádech (základní číslovky *pět* a *více* a neurčité číslovky jako *několik, mnoho*) a syntaktickými adjektivy (sem patří většina adjektiv; některé číslovky, zvláště číslovky řadové a číslovky základní v nepřímých pádech; některá zájmena, např. přivlastňovací a některá neurčitá: *její, jakýsi, nijaký*).

U některých slov nevyplývá rozlišení přímo z kombinace morfologické značky a lemmatu, podle kontextu může být syntaktickým substantivem i adjektivem. Je tedy nutné využít kontextově zapojená pravidla (např. pro slova *každý, další, ostatní, jeden, všichni, to*): stojí-li po takovém ambivalentním slově substantivum ve stejném rodě, čísle a pádě, pravděpodobně jde o syntaktické adjektivum. V opačném případě lze slovo označit za syntaktické substantivum.

Takové rozlišení se parser může naučit využívat v trénovacích datech a při vlastním značkování mu umožní správně přiřazovat závislosti a syntaktické funkce.

### 3.3 Doplnění lingvistických informací

Jednou z možných cest ke zlepšení práce stochastického parseru je doplnění parametrů, které parser nemůže získat kvůli omezenému rozsahu trénovacích dat (např. valence méně frekventovaných sloves) nebo které nedokáže z textu sám získat (např. rozdíl mezi koordinačními spojkami tvořícími a netvořícími hranice klauzí). Tyto informace lze doplnit do trénovacích dat i do nově zpracovávaných textů tak, aby je parser mohl využít.

#### 3.3.1 Valence a další vlastnosti slova

Parser se ve svých obvyklých nastaveních opírá kromě výseku morfologické značky i o lemmata. U méně frekventovaných lemmat však informaci nemůže uchovat, protože by tak neúměrně vzrostl rozsah uchovávaných dat (modelu). O lemmatech, která se v trénovacích datech vůbec neobjeví, pak nemá vůbec žádné informace. Řešením je dát parseru k dispozici v trénovacích datech i v nově značkových datech údaje související s jednotlivými lemmaty, a to určitým parametrem, který bude moci využít. Parametry mohou zhruba odpovídat skupinám slov používaných v seznamech pro opravný modul, musí zahrnovat valenci sloves a adjektiv, frekventovanou rekcii substantiv, substantiva označující osoby (*pan, kolegyně*) atd. S množstvím a formou dodávaných informací by bylo nutné experimentovat.

### 3.3.2 Hranice klauzí a charakteristika koordinační spojky

MST parser v použitém nastavení nepracuje dobře s hranicemi klauzí, závislosti je často chybně překračují. Na základě jednoduchých kontextových pravidel lze často určit, zda určitá spojka nebo interpunkční znaménko představuje hranici klauzí, nebo ne. Pokud by se parser naučil informace o hranicích klauzí využívat, mohl by se tak vyhnout mnoha chybám. Závislost větného členu v případě vnořené vedlejší věty může překračovat hranice klauzí, v tom případě ale nejméně dvě hranice, ne jen jednu.

Parser také nerozlišuje mezi uzly závislými na jedné souřadné spojce, ve výsledku mohou být na jedné koordinační spojce (formálně) závislá jak substantiva, tak slovesa. Přitom u velké části spojek lze s pomocí jednoduchých pravidel určit, zda spojka koordinuje slovesa, syntaktická substantiva nebo syntaktická adjektiva, nemůže koordinovat zároveň slovesa i jiné kategorie (snad kromě infinitivu). Tuto informaci je možné parseru předat a naučit ho s ní pracovat.

## 4. Závěr

K tomu, abychom dosáhli spolehlivé syntaktické anotace s úspěšností alespoň 85–90 % (úspěšnost určení závislosti i syntaktické funkce), je nutné vyzkoušet všechny dostupné cesty ke zlepšení výkonu stochastických parserů a k opravě jejich výstupu, v případě úspěchu je využít a rozvinout. Kvalitně syntakticky anotované korpusy se pak mohou stát užitečným doplňkem současných, pouze lemmatizovaných a morfologicky anotovaných textových korpusů.

# Část druhá

## Syntaktické funkce substantiv a jejich realizace pády

### Úvod

Cílem celé práce je představit statistické údaje o vztahu funkce a formy u substantiv, jinak řečeno: jaké syntaktické funkce zaujímají substantiva v českých textech a jak často, a také jakými pády (prostými či předložkovými) se tyto funkce realizují.

V první části práce jsme popsali metodu, kterou byl syntakticky označován korpus SYN2005, jež jsme použili jako zdroj dat. V této druhé části práce představíme dosažené výsledky.

V první kapitole upřesníme, s jakými syntaktickými funkcemi budeme pracovat a jak jsou definovány. Vysvětlíme, jak budeme zacházet s pády, popíšeme tři žánrové subkorpora, které budeme srovnávat ve třetí kapitole. Ukážeme také postup, jímž jsme redukovali chybovost výsledků.

Ve druhé kapitole představíme získané údaje o syntaktických funkcích a jejich realizacích různými pády v celém korpusu SYN2005. Podrobně probereme každou relevantní realizaci syntaktické funkce prostým nebo předložkovým pádem substantiva (například předmět vyjádřený předložkovým akuzativem substantiv).

Třetí kapitola se pak zaměří na srovnání frekvencí syntaktických funkcí substantiv podle funkčních stylů v korpusu SYN2005. Porovnáme také zjištěné frekvence se statistikami publikovanými dříve (Těšitelová 1985).

# Kapitola I

## Syntaktické funkce, pády a funkční styly

Aby bylo možné porozumět údajům o frekvencích syntaktických funkcí substantiv realizovaných prostými či předložkovými pády, je nutné nejprve podrobněji představit jednotlivé syntaktické funkce i to, jak budeme dále v této práci zacházet s pády. Syntaktické funkce odpovídají pojetí používanému v PDT, vycházejí tedy z české lingvistické tradice, upřesnění a vyjasnění pojmů je však nezbytné. V práci budeme také pracovat s funkčními styly v korpusu, proto je zde velmi stručně představíme.

Údaje získané z automaticky syntakticky anotovaného korpusu nejsou zcela spolehlivé, museli jsme tedy vytvořit metodu, jejíž pomocí jsme vliv chybovosti značkování na statistické údaje omezili. Tuto metodu popisujeme v závěru kapitoly.

### 1. Syntaktické funkce substantiv

V této práci používáme tyto syntaktické funkce substantiv: podmět (**Sb**), jmenná část verbonominálního přísudku (**Pnom**), přívlastek (**Atr**), předmět (**Obj**), příslovečné určení (**Adv**) a aktuální elipsu (**ExD**). Ostatní syntaktické funkce, které ve formalismu PDT mohou být přiřazeny substantivům, jako doplněk (**Atv, AtvV**) nebo součásti víceslovných předložek či spojek (**AuxP, AuxY**), jsou příliš málo frekventované a většinou i nedostatečně spolehlivě automaticky značkovány (viz oddíl 4.2 **Zjednodušení dat** v této kapitole).

Vycházíme z pojetí Vladimíra Šmilauera (Šmilauer 1966) a z definice syntaktických funkcí v PDT na analytické rovině (Hajič et al. 1999), a to pouze v těch částech, které lze vztáhnout na substantiva. Vše ověřujeme na datech v manuálně anotovaném korpusu PDT a v automaticky syntakticky anotovaném korpusu SYN2005.

V práci vycházíme z řešení syntaktických struktur, jež byla zvolena při koncepci PDT. Tato řešení mohou být někdy diskutabilní, ale při popisu složité jazykové reality se osvědčila. Cílem této práce navíc není vytvářet vlastní syntaktické teorie, ale popsat s využitím dostupných prostředků jazykovou realitu v daném korpusu. Případy, na jejichž syntaktickou interpretaci se názory mohou různit, se snažíme důsledně kvantifikovat tak, aby bylo možné upravit si údaje podle té syntaktické teorie, jíž dává čtenář přednost.

Zde uvádíme pouze stručné charakteristiky a vymezení syntaktických funkcí s příklady, podrobněji se budeme realizacím syntaktických funkcí pády věnovat v příslušné kapitole.

V celé další práci uvažujeme o syntaktických funkcích **substantiv**: všechny syntaktické funkce mohou být realizovány i jinými slovními druhy, ale v dalším výkladu se budeme věnovat výhradně podstatným jménům.

### 1.1 Podmět (*Sb*)

Podmět se obvykle definuje jako člen predikační dvojice, o kterém se ve větě něco tvrdí přísudkem (druhým členem predikační dvojice). V aktivních větách má obvykle význam činitele, v pasivních předmětu zasaženého dějem. Význam však v definici podmětu nehraje důležitou roli, vymezení je spíše formální (tvar podmětu, shoda přísudku). Podmět je obvykle v nominativu, přísudkové sloveso se obvykle shoduje se subjektem ve jmenném rodě a čísle. Substantivum ve funkci podmětu je obvykle v bezpředložkovém nominativu (příklad 1), výjimečně v genitivu (genitiv záporový či partitivní, příklad 2) či v předložkovém pádu vyjadřujícím přibližnost (příklad 3).<sup>2</sup>

- (1) **Lidé/Sb jsou** neuvěřitelně vynalézaví .
- (2) Nikde **nebylo** ani živé **duše/Sb** , bylo posvátné ticho .
- (3) aby v radě **zasedlo po** čtyřech **zástupcích/Sb** z KDU-ČSL a z ODS .

### 1.2 Jmenná část verbonominálního přísudku (*Pnom*)

Verbonominální přísudek se skládá ze spony (*být, bývat*) a jmenné části, kterou je syntaktické substantivum či adjektivum (protože však představujeme syntaktické funkce substantiv, uvažujeme zde jen substantiva). Substantivum v této funkci může být v bezpředložkovém nominativu či instrumentálu: příklady (1) a (2); výjimečně v genitivu: příklad (3). Předložkové fráze závislé na slovese *být* se v PDT značkují jako příslovečná určení.

- (1) Asi **bude pravda/Pnom** , co jsi říkal prve ,
- (2) Významnou **součástí/Pnom** léčby **je** profylaxe příštích příhod .
- (3) Říkám , že je jasná - a soud **je** stejného **názoru/Pnom** .

---

2 Příklady ve druhé části práce představujeme odlišným způsobem než v první: modře je zbarveno **řídící slovo**, červeně **závislé substantivum**, popř. i předložka, na níž je substantivum formálně závislé, je-li v předložkovém pádě. Pro přehlednost je u závislého substantiva uvedena za lomítkem jeho **syntaktická funkce**, a to zeleně.

### 1.3 Přívlastek (*Atr*)

Přívlastek je větný člen závislý na syntaktickém substantivu. Mezi substantiva takto označená v PDT se řadí především typické neshodné přívlastky v genitivu (1), popř. v jiných prostých pádech (2), nebo v pádech předložkových (3). Dále sem patří substantiva závislá na kvantifikátorech, zvláště substantiva v genitivu plurálu (formálně) závislá na základních číslovkách *pět* a *více*, popř. na neurčitých číslovkách *několik* aj. (4). Řadí se sem také substantiva závislá (samostatně nebo s předložkou) na adjektivech, zájmenech a číslovkách, jež mají v dané větě funkci syntaktického substantiva (5). Kromě toho se jako *Atr* označují i shodné souřadné výrazy jako *pan* a *Arne* ve spojení *pan Arne Novák* či *Praha* ve spojení *město Praha*, což technicky umožňuje zachytit tyto souřadné konstrukce jako jeden celek v závislostní struktuře (6).

- (1) Od **konce roku/Atr** 1995 se jejich dluh zvýšil o 269 mld. Kč .
- (2) Jaký podíl v **zásobování** obyvatel **teplem/Atr** má dnes u nás teplárenství ?
- (3) **Vstup** České republiky **do** Evropské **unie/Atr** je ohrožen !
- (4) A Bedřich s náma žije **pět let/Atr** .
- (5) To byl **jeden z důvodů/Atr** , proč byl tak neoblíbený .
- (6) Nesejdeme se ani s **ministrem/Atr** vnitra **Grossem** ,

### 1.4 Předmět (*Obj*)

Předmět je větný člen, který je závislý na slovese nebo na adjektivu, výjimečně na adverbiu. Jeho tvar je určen řídicím členem. Z hlediska významu je předmět větší či menší měrou zasažen dějem slovesa (popř. vlastností vyjádřenou adjektivem). Předmět může být vyjádřen substantivem v přímém (1) a (2) i předložkovém pádě (3) a (4).

Volný dativ (prospěchový, přivlastňovací aj.) se v PDT řadí k předmětu jen zčásti, zčásti se řadí k příslovečným určením. Podrobněji představíme toto rozlišení ve výkladu o předmětu v bezpředložkovém dativu. Naopak zčásti se za předměty považují i některé typy předložkových frází obvykle chápané jako příslovečná určení: určení původu a výsledku děje (5). Jako předmět jsou také označována substantiva v instrumentálu závislá na slovese v opisném pasivu, která vyjadřují původce děje, tj. podmět před pasivizací (6), a substantiva závislá na deverbativním adjektivu se stejnou funkcí (7). V tradičním, školském výkladu se tyto případy interpretují jako příslovečné určení původce děje, *Adv* (Šmilauer 1966, s. 334).

- (1) *Ale kde tedy mám své děti/Obj ?*
- (2) *Podle zdrojů blízkých ministerstvu/Obj zahraničí může být těch důvodů více .*
- (3) *Stačí správně odpovédět na soutěžní otázku/Obj*
- (4) *Byl by vděčný za pomoc/Obj , kterou jsme schválili*
- (5) *Dostala ji od Goebbelsovy manželky/Obj jako svatební dar .*
- (6) *Pevnost byla Němci/Obj dobyta ztečí v únoru*
- (7) *a stěžovala si na problémy způsobené těžaři/Obj*

### 1.5 Přísllovečné určení (Adv)

Přísllovečné určení je větný člen, který rozvíjí sloveso, adjektivum nebo adverbium. Jeho tvar není určen řídicím členem, vyjadřuje okolnosti (čas, místo, míru, způsob aj.). Některé případy obvykle považované za přísllovečná určení se v PDT řadí mezi předměty, naopak některé tradiční předměty se řadí mezi přísllovečná určení (viz výše, podrobněji také u jednotlivých kombinací funkcí a pádů). Přísllovečné určení může být vyjádřeno prostým (1) a (2) i předložkovým (3) a (4) pádem substantiva.

Mezi přísllovečným určením a předmětem je, zvláště v případě předložkových pádů, často velmi nejasná hranice, určitý tvar substantiva závislý na stejném slovese lze totiž i podle významu interpretovat jednou jako přísllovečné určení (5), jindy jako předmět (6). Přenesené významy se řadí spíše k předmětu. Kromě rozhodování mezi předmětem a přísllovečným určením je leckdy problematické i rozlišování mezi neshodným přívlastkem a přísllovečným určením, protože není jasné, na kterém větném členu má záviset předložková fráze.

- (1) *Příštího dne/Adv dorazilo čtyři sta mužů Čestné dělostřelecké roty .*
- (2) *Dům šachu a dámy , touto dobou/Adv obsazený převážně mužskou populací .*
- (3) *Ten se v roce/Adv 1935 stal letním sídlem spisovatele Karla Čapka ,*
- (4) *Britské síly , postupující podél řeky/Adv Tigris , se příliš roztáhly*
- (5) *Najednou Ajla ucítla , jak jí něco narazilo na paži/Adv .*
- (6) *Narazili jsme však na nepochopení/Obj .*

### 1.6 Elipsa (ExD)

Označení elipsa se v PDT používá pro několik různých typů syntaktických jevů, primárně ale pro větné členy, jejichž řídicí člen byl vypuštěn (aktuální elipsy). Typickými příklady

elipsy v PDT jsou řídicí větné členy ve výpovědích (větách) bez slovesa, a to jak lexikalizované obraty (1), tak jiné výpovědi bez slovesa (2) a samostatně stojící aktuální elipsy (3). Dále se takto označují větné členy v koordinovaných větách, kde je vypuštěno opakující se sloveso (4 a 5). Řadí se sem i vokativ, a to jak samostatně stojící (mimo větu), tak vokativ použitý ve větě jako samostatné oslovení (6). Za elipsu se považují i srovnávací výrazy se spojkami *jako* a *než* (7).

- (1) *Dobrou **noc/ExD** !*
- (2) *Na **Rýn/ExD** , **Rýn/ExD** , německý **Rýn/ExD** !*
- (3) *kde ji vyměnili za kávu , sůl a cukr . A za **peníze/ExD** .*
- (4) *Slyším jak nás volají , všechny **syny/ExD** svobody .*
- (5) *Hořelo tam totiž nespočetně svíček , jedna **za** každého **člověka/ExD** , který kdy žil na zemi .*
- (6) *Komu to říkáš , **otrapo/ExD** ?*
- (7) *Mé srdce je **tvrdé** jako **pecka/ExD** z broskve .*

## 2. Pády substantiv

Syntaktické funkce substantiv vztahujeme v této práci k jejich pádům (prostým nebo předložkovým). Stručně tedy představíme pojetí pádu, z něhož vycházíme, a způsob, jak budeme dále s pády substantiv zacházet a jak je budeme označovat.

### 2.1 Pojetí pádu

Pád substantiv považujeme za výrazový prostředek, který umožňuje zapojení substantiv do větných struktur, vyjádření jejich vztahu k ostatním větným členům a v některých případech také doplnění vlastního významu substantiva o další významové prvky. Pro účely této práce nepovažujeme striktní rozdělení mezi tzv. syntaktickými a sémantickými pády za vhodné pro popis jazykového systému. Jako užitečnější se jeví rozdělení pádů, které nabízí J. Kuryłowicz (Kuryłowicz 1949, s. 29). Rozděluje pády na pády syntaktické, jež mají primárně funkci syntaktickou, ale sekundárně mohou mít i funkci sémantickou, a na pády sémantické, které jsou primárně sémantické, ale sekundárně mohou nabývat i funkcí syntaktických. Toto rozdělení se jasně odráží ve frekvenci použití prostých a předložkových pádů v češtině, např. prostého genitivu nebo předložky *v* s lokálem. V obou případech lze najít doklady syntaktického i sémantického využití pádu (či kombinace předložky a pádu), ale u genitivu je

přínejmenším z frekvenčního hlediska centrální funkce syntaktická, zatímco u předložky v a lokálu je centrální funkce sémantická.

Rozdělení na syntaktické a sémantické funkce pádů v této práci dále nepoužíváme. Lze ho částečně vztáhnout k syntaktickým funkcím (*Obj* oproti *Adv*), ale ani pády v příslovečných určeních (primárně předložkové pády) nelze obecně prohlásit za ryze sémantické.

Vztah předložky a pádu považujeme za nedělitelný. Předložka není řídicím slovem substantiva, ale součástí komplexního morfému, který spolu s pádem substantiva vyjadřuje syntaktické i sémantické zapojení substantiva do větné struktury. „*La préposition n'est pas le regens de la forme casuelle, mais elle est un sous-morphème, bien que principal, du morphème composé II (consistant de préposition + désinence casuelle)*“ (Kuryłowicz 1949, s. 24). Podobně např. (Sgall 1967, s. 51): „*Spojení předložky s pádem je tedy komplexní jednotka morfologické roviny*“; ve shodě s tímto pojetím důsledně rozdělujeme pády prosté a pády předložkové.

Kvůli přehlednosti se často omezujeme jen na toto rozlišení (prostý nebo předložkový pád), ale vždy, když používáme pojem „předložkový pád“, znamená to spojení konkrétních předložek s pády substantiv. Tam, kde je to možné, pracujeme s jednotlivými spojeními konkrétních předložek a pádů substantiv. V PDT se vztah předložky a substantiva vyjadřuje technickou závislostí substantiva na jeho předložce, i v našem výkladu tak může zaznít, že substantivum je „závislé“ na předložce. Tato závislost je však technická, neznámá to, že bychom pojímali vztah předložky a substantiva jako determinační.

## 2.2 Prosté a předložkové pády substantiv

V výkladu vztahu mezi funkcí a tvarem substantiv budeme tedy rozlišovat mezi prostými a předložkovými pády. Za předložkový pád považujeme tvar substantiva, který určuje nebo spoluurčuje předložka, na níž je substantivum formálně závislé (včetně substantiv, které jsou součástí koordinace či apozice, jež je formálně závislá na předložce). Za předložkový pád nepovažujeme případy, kdy substantivum rozvíjí jiné syntaktické substantivum v předložkovém pádě, a to ani v případech shodných přívlastků nebo doplňků.

U genitivu, dativu, akuzativu a instrumentálu uvažujeme tedy dvě varianty: prostý a předložkový pád. Označujeme je samotným číslem u prostého pádu (**2, 3, 4, 7**), písmenem **p** s číslem u předložkového pádu (**p2, p3 ...**). Nominativ uvažujeme pouze jako prostý pád (**1**), předložkový nominativ u skutečně českých slov (tedy ne v cizích vlastních jménech) se vyskytuje natolik zřídka (méně než 0,01 % ze všech substantiv), že je možné ho zanedbat. Vokativ (**5**) se vyskytuje vždy jako pád prostý. Lokál se naproti tomu vyskytuje vždy jako

předložkový (**p6**), i když ve zcela výjimečných případech nemusí stát jeho předložka v téže větě. Shodné přívlasky v lokálu řadíme pro zjednodušení k předložkovému lokálu (na rozdíl od ostatních shodných přívlasků, které přiřazujeme z technických důvodů k prostým pádům).

Realizace syntaktické funkce pádem substantiva (kombinace syntaktické funkce a pádu) označujeme, když je třeba, spojením zkratky funkce a čísla pádu: **Obj\_p6**, **Atr\_2**, **Adv\_7** aj.

Pojmem *pád* budeme nadále rozumět buď předložkový, nebo prostý pád (například předložkový genitiv nebo prostý instrumentál), celkem tedy pracujeme s jedenácti pády. Bylo by samozřejmě přesnější vždy rozlišovat, která předložka byla použita, ale počet kombinací by byl příliš vysoký a údaje nepřehledné. V komentářích realizací syntaktických funkcí předložkovými pády budou podány podrobnější informace o jednotlivých předložkách a jejich frekvencích.

### 3. Funkční styly

Ve výkladu v následující kapitole se příležitostně zmíníme o funkčních stylech (žánrech) v korpusu SYN2005, tématem třetí kapitoly pak bude srovnání funkčních stylů v korpusu z hlediska syntaktických funkcí a jejich realizací pády substantiv. V korpusu SYN2005 jsou definovány tři subkorpuse podle žánrů: subkorpus beletrie **BEL**, subkorpus odborných textů **ODB** a subkorpus publicistický **PUB**. Podrobnější popis složení jednotlivých subkorpusů je uveden na <http://ucnk.ff.cuni.cz/syn2005.php>.

Mezi žánry není a ani nemůže být nijak ostrá hranice: k beletrii se například řadí i texty publicistického charakteru (mj. několik ročníků časopisu *Lidé a země*) nebo texty charakteru populárně naučného (literatura faktu, např. *Kulturní Chrudim minulosti a současnosti*). Subkorpus **ODB** obsahuje například i zájmové a profesní časopisy (*Chatař a chalupář*). Subkorpus **PUB** obsahuje sice pouze texty novin a neoborných časopisů, přesto jsou některé texty v něm obsažené blíže beletrii, jiné odborné literatuře. Ve třetí kapitole si však ukážeme, že se žánry přes tyto neostré hranice jasně liší, mimo jiné frekvencí syntaktických funkcí a pádů substantiv.

Statistické údaje o korpusu SYN2005 a jeho subkorpusech uvádíme v tabulce. V prvních třech sloupcích se nacházejí údaje převzaté ze statistické publikace o korpusu SYN2005 (Bartoň et al. 2009, s. 16 a s. 130): celkový počet pozic (včetně interpunkce), podíl žánrů v korpusu vypočtený na základě pozic a podíl substantiv v jednotlivých subkorpusech (podíl v rámci slovních druhů nepočítaje interpunkci). V posledních dvou sloupcích představujeme

údaje, s nimiž pracujeme v této práci: počet substantiv a podíl substantiv z jednotlivých žánrových subkorpusů na všech substantivech v korpusu. Uvádíme pouze „započítaná“ substantiva, tedy substantiva, která mají určený pád (ne zkratky) a jež nebyla z údajů vyloučena kvůli málo časté syntaktické funkci nebo nesmyslné kombinaci syntaktické funkce a pádu (viz dále).

Tabulka funkčních stylů v korpusu SYN2005

	pozic	podíl žánrů v korpusu (pozice)	podíl substantiv v (sub)korpusu	započítaných substantiv	podíl žánrů v korpusu (substantiva)
SYN2005	122 419 948	100 %	30,2 %	28 655 821	100 %
BEL	50 011 887	40,9 %	24,3 %	9 257 251	32,3 %
ODB	32 692 293	26,7 %	34,5 %	8 696 437	30,4 %
PUB	39 715 768	32,4 %	33,8 %	10 702 133	37,3 %

## 4. Zpracování dat pro statistické účely

Jak bylo řečeno v první části práce, automatická syntaktická anotace (včetně opravného modulu) je dosud relativně málo spolehlivá. Údaje o frekvencích realizace syntaktických funkcí pády substantiv tak bylo nutno manuálně ověřit na vzorcích a upravit podle výsledků. Spolehlivost údajů je nejnižší u velmi málo frekventovaných kombinací funkce a pádu, pro přehlednost a pro vyšší přesnost byly tedy nejméně časté údaje zcela zanedbány.

### 4.1 Ověření a korekce údajů

Pro ověření údajů získaných ze statisticky anotovaného korpusu jsme využili manuální kontrolu všech kombinací pádu a syntaktické funkce na vzorcích z anotovaného korpusu. Vzorky byly náhodně vybrány z různých textů, a to tak, aby byl zachován stejný poměr žánrů (funkčních stylů) jako v celém korpusu. Velikost vzorků odpovídala frekvenci dané kombinace funkce a pádu, od 100 pro nejméně časté, po 500 u několika nejčastějších kombinací. U každého vzorku jsme zjišťovali, jaká část byla určena správně a u jaké části byl chybně určen pád nebo syntaktická funkce.

Jako příklad můžeme uvést příslovečná určení vyjádřená předložkovým instrumentálem. Cca 74 % ze vzorku bylo určeno správně (pád i funkce), dalších 25 % mělo správně určený pád, ale chybně syntaktickou funkci (12 % *Atr*, 9 % *ExD*, 4 % *Obj*). Jedno procento substantiv ve vzorku mělo chybně určen jak pád, tak funkci (chybné koordinace), správně

měly být substantivům přiřazeny funkce **Sb** a nominativ.

Při manuálním ověřování výsledků se ukázalo, že chyby parseru jsou víceméně konzistentní: rozbor dvou nebo více dostatečně velkých vzorků (100 a více) téže kombinace dával vždy velmi podobné výsledky. Nejfrekventovanější kombinace syntaktické funkce a pádu (subjekt vyjádřený nominativem, přívlastek vyjádřený prostým genitivem, předmět vyjádřený prostým akuzativem) měly spolehlivost přes 90 %, nejméně časté kombinace byly z velké části nebo vždy chybné.

Zjištěná procenta byla poté použita k přepočítání údajů, od údajů pro každou kombinaci funkce a pádu byly odečten podíl chyb a naopak k nim byly přičteny podíly, které jim přináležely od jiných kombinací. Díky tomu jsme snížili dopad chybovosti statistického parseru na kvalitu získaných údajů.

## 4.2 Zjednodušení dat

Pro lepší přehlednost v tabulkách, které budeme uvádět, a především proto, že u málo častých kombinací funkcí a pádů je spolehlivost anotace velmi nízká, byly ze statistických údajů vyloučeny nejméně početné údaje: v žádné tabulce neuvádíme kombinace funkce a pádu, které představují méně než 0,1 % celku (celkem jsou v tabulkách všechna substantiva, všechna substantiva určitého pádu nebo všechna substantiva s určitou syntaktickou funkcí).

Jak již bylo řečeno výše, zcela byly také vyloučeny některé syntaktické funkce substantiv, jejichž podíl mezi všemi substantivy byl příliš nízký: všechny kombinované funkce (**AtrAdv**, **ObjAtr** aj.), doplněk (**Atv**, **AtvV**), součást víceslovné předložky (**AuxP**) nebo spojky (**AuxY**).

## Kapitola II

### Syntaktické funkce substantiv realizované prostými a předložkovými pády

V této kapitole představíme nejprve hrubé statistické údaje o syntaktických funkcích substantiv realizovaných prostými a předložkovými pády, a to formou přehledné tabulky. Data sama o sobě bez podrobnějšího představení a vyčíslení toho, jaké jevy v sobě jednotlivé kategorie zahrnují, nemají velkou informační hodnotu, ale poslouží pro lepší orientaci v následujícím výkladu. Po souhrnné prezentaci údajů představíme postupně všechny relevantní kombinace syntaktické funkce a pádu, u kombinací, které lze dále třídit, vyčíslíme podíl důležitých podkategorií (např. v rámci funkce **Obj** vyjádřené prostým instrumentálem vyčíslíme, jaký podíl na této funkci mají substantiva s funkcí „původce děje“ u opisného pasiva a jaký podíl mají valenční předměty aktivních sloves).

#### 1. Tabulka frekvence syntaktických funkcí substantiv a jejich realizace pády v korpusu SYN2005

V tabulce představují sloupce syntaktické funkce a řádky jednotlivé pády (prosté označené číslem, předložkové označené písmenem **p** a číslem). Uvnitř tabulky jsou uvedeny údaje v procentech z celkového počtu substantiv, přesněji z celkového počtu započtených substantiv, kombinace pádu a funkce, které byly zanedbány, se do celku nepočítají. Hodnoty byly zaokrouhleny na desetiny procenta, nižší než 0,05 % byly zanedbány.

## Frekvence syntaktických funkcí substantiv ve vztahu k jejich pádu

SYN2005	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	20,3	1,1	2,6	0,1	0	4,2	28,4
2	0	0	17,9	1,0	0,2	0,2	19,3
p2	0	0	2,4	0,5	4,6	0,1	7,6
3	0	0	0,1	1,6	0	0	1,8
p3	0	0	0,4	0,6	1,0	0	2,1
4	0	0	0,1	13,0	0,7	0,3	14,1
p4	0	0	1,5	1,7	2,8	0,1	6,1
5	0	0	0	0	0	0,4	0,4
p6	0	0	3,3	0,7	7,4	0,2	11,6
7	0	0,9	0,4	1,1	1,8	0,1	4,3
p7	0	0	1,8	0,6	1,6	0,2	4,2
SUM	20,4	2,0	30,5	20,9	20,2	5,9	100 %

Z tabulky je zřejmé, že nejčastější syntaktickou funkcí substantiv v korpusu je přívlastek (**Atr**), další tři funkce jsou z hlediska frekvence velmi vyrovnané: **Sb**, **Obj** a **Adv**. Nejčastější realizací funkce pádem je **Sb** v nominativu, dále **Atr** v prostém genitivu, **Obj** v prostém akuzativu a **Adv** v předložkovém lokálu (**červeně** zvýrazněno). V následující kapitole, v níž budeme srovnávat funkční styly v korpusu na základě obdobných tabulek, uvidíme, že mezi funkčními styly je poměrně vysoká variabilita jak v celkové frekvenci jednotlivých funkcí, tak v jejich realizaci jednotlivými pády. Než se ale dostaneme k tomuto srovnání, představíme podrobně jednotlivé relevantní kombinace syntaktických funkcí a pádů.

## 2. Popis jednotlivých realizací syntaktických funkcí pády

V této části postupně představíme jednotlivé realizace syntaktických funkcí pády. Prezentovány budou všechny kombinace, jejichž podíl v rámci syntaktické funkce je alespoň 0,1 %. Okrajově zmíníme i některé další méně frekventované, ale syntakticky smysluplné kombinace (například subjekt v předložkovém pádu). U některých kombinací není rozsáhlý komentář potřeba, zvláště u frekventovaných realizací funkce a pádu, které nelze dále rozčlenit na syntaktické podkategorie (například u subjektu v nominativu či předmětu v prostém akuzativu). U jiných je nutné pečlivé vyčíslení, ze kterých jevů a v jaké míře se daná kombinace skládá (přívlastek v prostém genitivu). Popis realizací funkcí postupuje podle syntaktických funkcí, v rámci syntaktických funkcí podle pádů. Uvádíme-li v analýze dílčí frekvence, jsou vždy přepočítané na 1 000 000 substantiv v korpusu (nepočítaje zanedbané a

zcela nesmyslné kombinace funkce a pádu), značeno jako **ppm** (parts per million, 10 000 ppm je jedno procento).

## 2.1 Podmět (Sb)

„Podmět je člen predikační dvojice, jemuž se přísudkem přisuzuje nějaký znak“ (Šmilauer 1966, s. 126). Mezi šesti funkcemi vyjádřenými substantivy, s nimiž pracujeme, je podíl podmětu 20,4 %. Substantivní podmět je v korpusu SYN2005 realizován z 99,8 % prostým nominativem, ve zbývajících 0,2 % prostým genitivem. Vzácně se objevuje i podmět v předložkových pádech, zde ho sice také představíme jako možnou kombinaci, ve statistických údajích však byl zanedbán.

### 2.1.1 Podmět v prostém nominativu (Sb\_1)

Podmět v prostém nominativu je nejčastější kombinace syntaktické funkce a pádu v korpusu SYN2005 i ve dvou ze tří funkčních stylů v korpusu. Kombinace je syntakticky koherentní, nemá smysl ji uměle rozdělovat na podkategorie. Ze sémantického hlediska se mezi nejčastější podměty řadí označení osoby: *člověk, muž, žena, dítě, otec, matka* (1); označení instituce: *společnost, firma, vláda, město* (2) a substantivní kvantifikátory: *většina, část, řada* (3).

- (1) *Snažil se , aby si lidé/Sb mysleli , že nemá strach .*
- (2) *Také proto vláda/Sb rozhodla , že správcem se stane některý ze státních podniků .*
- (3) *Čeká mě spousta/Sb další práce .*

### 2.1.2 Podmět v prostém genitivu (Sb\_2)

Podmět v prostém genitivu má v rámci podmětu podíl jen 0,15 %, v rámci všech kombinací 314 ppm. Mezi podměty je nejčastější genitivový záporový u slovesa *být*, cca 250 ppm, často v ustálených, lexikalizovaných obrazech (*není ... divu, pochyb, stopy, sporu, důvodu, úniku, památky, návratu*), jako v příkladu (1) nebo v podmínkových větách s jinými subjekty (2) a (3).

- (1) *v domě není živé duše/Sb , matka spí .*
- (2) *Kdyby nebylo Ajly/Sb , zemřel bych s ním .*
- (3) *Možná by byla trvala mnohem déle , nebýt vnějšího zásahu/Sb carevny Alžběty ,*

Méně častý je genitiv partitivní, především se slovesy *přibýt, přibývat, ubýt, ubývat*: cca 35 ppm (4) nebo *dostat se, dostávat se*: 10 ppm (5).

- (4) *Loni nám sice znovu přibýlo dopravních nehod/Sb , méně ale je smrtelných zranění*
- (5) *věděli velmi dobře , že státu se nedostávalo prostředků/Sb na sociální pomoc ,*

V PDT se za subjekt považují i výrazy závislé na predikativních spojeních *být třeba, být zapotřebí*. Subjektem je tu nejčastěji infinitiv nebo vedlejší věta, substantivum v prostém genitivu jen zřídka, takových subjektů je mezi substantivy v korpusu asi 30 ppm (6) a (7).

- (6) *Bylo třeba dalších osmi let/Sb namáhavého úsilí , aby ze záblesku inspirace vznikla ucelená teorie ,*
- (7) *Lépe by bylo ji zničit , k tomu je však zapotřebí dračích ohňů/Sb .*

Většina ostatních předmětů v prostém genitivu (cca 5 ppm) jsou subjekty pasiva u sloves s genitivní valencí: *dosáhnout, dosahovat, použít, používat*, aj. (8) a (9). V anotačním manuálu ani v datech PDT (tj. ani ve trénovacích datech pro parser) se takové subjekty nenacházejí, automaticky však byly takto anotovány. Z hlediska syntaktické teorie je to neobvyklá, ale ne zcela nesmyslná možnost interpretace takové struktury. Frekvence takto anotovaných subjektů je zanedbatelná.

- (8) *Četných výpravných efektů/Sb se dosáhlo použitím početného komparsu v tancích , bitvách a průvodech .*
- (9) *Nevím , jakého kopírovacího procesu/Sb bylo použito .*

### 2.1.3 Podmět v předložkových pádech (Sb\_p2, Sb\_p3, Sb\_p6)

V tradiční gramatice se za podmět považují i některé předložkové fráze vyjadřující podílnost a přibližnost. Formalismus PDT s takovým podmětem také počítá, ale jde o velmi málo častý jev (v PDT cca 50 ppm, nepovažujeme-li slova *tisíc* a *sto* za substantiva, ale za číslovky; jinak cca 90 ppm). V automaticky anotovaném korpusu jsou podměty realizované substantivem v předložkovém pádu zastoupeny přibližně stejně (52 ppm), ale jen zlomek takových automaticky nalezených podmětů je identifikován správně. V korpusu je možné na základě závislosti předložkové fráze, morfologické značky slovesa (neutrum singuláru

v případě přičestí, u přítomného času singulár) a dalších indicií, jako je kvantifikace, vyhledat potenciální subjekty v předložkové frázi a mezi nimi manuálně vyhledat vhodné příklady (1) a (2), na automatické značkování se však není možné spolehnout. Podmět v předložkových pádech u substantiv se vyskytuje pouze v genitivu (s předložkami *kolem*, *okolo*), v dativu (předložka *k*) a v lokálu (*po*). V akuzativu, po předložkách *přes* a *na*, je formálním podmětem obvykle číslovka, ne substantivum (3).

- (1) *Na postranních křídlech je po dvou výjevech/Sb z biblického života .*
- (2) *Uskutečnilo se kolem 80 pohovorů/Sb a schůzek/Sb s manažery ČS*
- (3) *Tuto službu využívá již přes 100000/Sb klientů .*

## 2.2 Jmenná část verbonominálního přísudku (*Pnom*)

Funkce *Pnom* je nejméně frekventovanou funkcí, kterou zahrnujeme do statistických údajů, tuto funkci má 2,0 % substantiv. Příisudek je tvořen sponou *být* (výjimečně *bývat*, cca v 0,5 % případů) a substantivem v prostém nominativu. S nominativem se střídá instrumentál, zřídka i genitiv; předložkové pády závislé na slovese *být* se ve formalismu PDT značkují vždy jako příslovečná určení. V celém korpusu SYN2005 je funkce *Pnom* realizována častěji nominativem (55,4 %) než instrumentálem (44,1 %), tento poměr se u jednotlivých funkčních stylů výrazně liší. Podíl prostého genitivu ve funkci *Pnom* je v korpusu jen cca 0,5 %.

### 2.2.1 Jmenná část verbonominálního přísudku v prostém nominativu (*Pnom\_1*)

Nejčastějším lexikálním obsazením funkce *Pnom* v nominativu jsou abstraktní, obecná jména, která jsou v textu obvykle dále doplňována (vedlejší větou či přívlasky): *pravda*, *věc*, *otázka*, *problém*, *škoda*, *záležitost*, *doba*, *důvod*, *místo*, *případ*, *způsob* (1) a (2). Dalšími častými nominálními predikáty jsou obecná označení osoby: *člověk*, *muž*, *žena* (3) a (4).

- (1) *To je pravý důvod/Pnom toho , že 8 . srpen byl nazván " černým dnem " ,*
- (2) *Potíže byly ta poslední věc/Pnom , kterou tu čekal .*
- (3) *Marcela je jediná žena/Pnom , kterou fakt miluji .*
- (4) *Víte , kolego , jsem člověk/Pnom nadaný svatou trpělivostí ,*

### 2.2.2 Jmenná část verbonominálního přísudku v prostém genitivu (*Pnom\_2*)

Funkce **Pnom** je realizována prostým genitivem (genitivem vlastnosti) jen velmi zřídka (95 ppm, 0,5 % ze všech nominálních predikátů). Prostý genitiv se se sponou *být* používá pro označení vlastnosti, často jde o víceméně ustálené obraty. Vlastnosti většinou charakterizují člověka: *je ... názoru, mínění; povahy, charakteru; původu, rodu, národnosti, krve; postavy, věku, pohlaví, vyznání* aj. (1) a (2), ale mohou charakterizovat i jiné předměty: *je ... druhu, rázu, typu, rozsahu, data, barvy* aj. (3) a (4). Podstatné jméno v genitivu je vždy nějak rozvíto, obvykle přívlastkem, někdy vedlejší větou.

- (1) *Byli toho **názoru/Pnom**, že se mám po dlouhém věznění pořádně krmit*
- (2) *Stejně tak lze mít za to, že jeho syn **nebyl** české **národnosti/Pnom**.*
- (3) *Menší jímka **je rozměru/Pnom** 18,7 x 18,7 m*
- (4) *Většina uvedených citací **je** staršího **data/Pnom**.*

### 2.2.3 Jmenná část verbonominálního přísudku v prostém instrumentálu (*Pnom\_7*)

Nejčastější substantiva ve funkci jmenové části přísudku se sponou v prostém instrumentálu vyjadřují kauzální vztahy: *je ... důvodem, příčinou, podmínkou, důsledkem, výsledkem, předpokladem* (1) a (2), frekventovaná jsou i další abstrakta: *je ... součástí, cílem, úkolem, předmětem, problémem, zdrojem, základem, záležitostí, otázkou* (3). Mezi nejčastějšími substantivy se objevuje výrazně méně možných označení osob než v případě nominativu: *členem, autorem, svědkem* (4).

- (1) ***Podmínkou/Pnom je**, aby pohyb " sbalení " byl shodný se směrem získané točivosti .*
- (2) *Infekce **jsou** hlavním **důvodem/Pnom** úmrtí nemocných po transplantaci jater*
- (3) *Jednota církve, překonání rozkolů, **je** trvalým **úkolem/Pnom***
- (4) *Nejnešťastnějším **členem/Pnom** rodiny **byl** však Einsteinův mladší syn Eduard,*

## 2.3 Přívlastek (*Atr*)

Substantivní přívlastek je větný člen rozvíjející syntaktické substantivum ve větě, v korpusu SYN2005 je to s 30,5 % nejčastější syntaktická funkce. Funkce **Atr** zahrnuje jednak neshodné přívlastky (vyjádřené prostým či předložkovým pádem), jednak tzv. přívlastek shodný. Jako shodný přívlastek se v PDT označují součásti souřadného

(adordinačního) spojení typu *pan/Atr Novák* nebo *řeka Vltava/Atr*. Shodné přívlastky představují 4,1 % všech přívlastků v korpusu, poměr se u jednotlivých pádů výrazně liší. Shodné přívlastky mají stejné vlastnosti napříč pády, proto je popíšeme v samostatném odstavci, na nějž budeme v další analýze přívlastků odkazovat.

### 2.3.1 Shodný přívlastek (*Atr*)

Pro zařazení souřadných jmenných skupin typu *pan/Atr Arne/Atr Novák* nebo *řeka Vltava/Atr* do závislostní struktury bylo v PDT zvoleno toto řešení: u označení osob se za řídicí člen spojení považuje poslední označení osoby (zde jméno *Novák*), u předmětů (institucí, krajinných útvarů aj.) je řídicí první člen (zde apellativum *řeka*). Ostatní členy spojení jsou označeny jako přívlastky (*Atr*) závislé na řídicím členu. Označení „shodný“ přívlastek je dáno především shodou v pádě v rámci celé souřadné skupiny. Shoda v rodě ani v čísle není povinná.

Když označují osobu, mohou se souřadná spojení tohoto typu skládat ze dvou nebo více substantiv, často tří až čtyř, výjimečně i více (1). Shodné přívlastky jsou relativně málo frekvencované, jen 4,1 % všech přívlastků, 1,2 % ze všech substantiv.

Označení osob (apellativa i propria) jsou mnohem častější (96,8 %) než označení předmětů (3,2 %). Většinou označují muže (87,8 % shodných přívlastků označujících osoby je mužského životného rodu), což odráží genderově nevyrovnané zaměření textů. Tento nevyrovnaný poměr se projevuje ve všech žánrech (**BEL** 87,3 %, **ODB** 91,1 %, **PUB** 87,4 %).

Celkem 57,9 % shodných přívlastků označujících osoby jsou propria, poněkud nejčastěji křestní jména; nejčastější jsou: *Jan, Václav, Jiří, Petr, Pavel*. Nejčastější ženská jména v těchto spojeních jsou: *Jana, Eva, Marie, Hana, Helena*, ovšem nejčastější *Jana* je mezi všemi křestními jmény až na 28. místě. Mezi nejčastějšími apellativy (42,1 % ze shodných přívlastků) jsou „honorifika“ (*pan/pán, paní, panna, slečna, soudruh*), označení politické funkce (*ministr, prezident, premiér, starosta, král, císař*), vedoucí funkce (*ředitel, předseda, šéf*), označení dosaženého vzdělání (*profesor, doktor, inženýr*) a vojenské funkce (*kapitán, generál*). Nejčastější souřadná spojení v SYN2005 jsou *Václav Havel* a *Václav Klaus*, v případě dvou apellativ to jsou spojení *pan doktor, Pán Bůh* (2) a (3). U feminin jsou nejčastější spojení *paní Mooshabrová* (všechna pocházejí z jednoho textu) a *Panna Maria* (4). Výjimečně se souřadné skupiny neshodují v rodě (5).

- (1) mimořádně vzdělaný a osvícený pedagogický **spisovatel/Atr** a **reformátor/Atr hrabě/Atr František/Atr Josef/Atr Kinský** , ctitel Komenského .
- (2) se zástupci Občanského fóra v čele s **Václavem/Atr Havlem** .
- (3) Možná že po Tobě chci moc , **Pane/Atr Bože** , možná po Tobě chci moc
- (4) řekl hlasem , který teď **paní/Atr Mooshabrové** už nepřišel tak strohý
- (5) Poprvé se fanouškům v dresu Slavie představila i největší **posila/Atr** pražského klubu **záložník/Atr Martin/Atr Lukeš** .

Pro neživé objekty se souřadná spojení (shodné přívlasky) používají zřídka, pouze ve 3,2 % shodných přívlasků, cca 380 ppm v rámci všech substantiv. Funkce shodný přívlastek se připisuje propriu, názvu předmětu, které následuje po apelativu vyjadřujícím jeho druh. U označení předmětů v nominativu nelze určit, zda je proprium shodné, nebo zda je v nominativu jmenovacím. U apelativ, u kterých je v korpusu doloženo alespoň 10 shodných tvarů v jiném pádě, považujeme přívlasky v nominativu za shodné, přestože se u většiny těchto apelativ vyskytují i doklady nominativu jmenovacího (apelativum, které není v nominativu, je rozvíto následujícím propriem v nominativu).

Apelativa obvykle označují typy lidských sídel: *město, obec, městečko, metropole, hrad, zámek, klášter* (6) a (7) nebo krajinných útvarů: *řeka, hora, ostrov, jezero, říčka* (8) a (9).

Výjimečně může být apelativum označující neživý předmět rozvíto následujícím shodným apelativem, nejčastěji je to apelativum *měsíc* a název měsíce, 10 ppm v korpusu (10), velmi vzácně i jiná spojení (11).

- (6) V potoce Pšovka pod **hradem Kokořínem/Atr** jsem ji chytil poprvé asi před devíti lety
- (7) v nichž se dnes nachází například hlavní **město Praha/Atr** i jihomoravská **metropole Brno/Atr**
- (8) U Svatého Jana překročíte **potok Libochovku/Atr** , na jehož březích prý roste *pěřovník pštrosí* .
- (9) Na **vrchu Bohdalci/Atr** u Prahy . . .
- (10) Islámští věřící dodnes respektují postní **měsíc ramadán/Atr** ,
- (11) Na Zemi je rozšířen hlavně v **nerostech fosforečnanech/Atr** ( *fosfátech* ) .

Souřadné přívlasky (označení osob i označení předmětů) se nejčastěji používají v nominativu (65,8 %) a v genitivu (20,5 %).

### 2.3.2 Přívlástek v prostém nominativu (Atr\_1)

Přívlástek v prostém nominativu zahrnuje jednak menší část neshodných přívlástků včetně obrátů citátového charakteru (32,1 %), jednak shodné přívlastky, které byly představeny výše (67,9 %). Část neshodných přívlástků v nominativu je velmi blízká shodným přívlástkům označujícím předměty: neshodné přívlastky jsou vlastní jména rozvíjející předcházející apelativum, které označuje druh označeného předmětu. Počet různých apelativ, která jsou v korpusu rozvíjena neshodným přívlástkem, je ale výrazně vyšší než u předmětů označených souřadnými skupinami, kde jich je cca 40 (s alespoň 10 doloženými shodnými tvary mimo nominativ). Zde je různých apelativ několik set (je obtížné určit přesně počet, protože u méně frekventovaných případů rychle narůstá chybovost určení závislosti a funkce). Apelativa označují typy sídel (*město, sídliště, klášter*), krajinné útvary (*řeka, údolí, sopka*), instituce (*divadlo, společnost, firma*), umělecká díla (*film, román, opera*), výrobky (*vůz, boty, tyčinka*) apod. Propria v nominativu, jež následují po řídicím apelativu, jsou pojmenováním příslušného předmětu (1), (2) a (3). Název je většinou jednoduchý (jedno proprium, případně rozvítené adjektivem), ale v některých případech, zvláště u pojmenování uměleckých děl, může mít charakter citátu (4).

- (1) *v létech 1942 - 1945 byla učitelkou dětí v koncentračním táboře Terezín/Atr .*
- (2) *Uprostřed pralesa , na březích řeky Ogooné/Atr , ve vesnici Lambaréné/Atr , založil misijní zdravotnickou stanici .*
- (3) *Jako je možné dělat reklamu na boty Adidas/Atr , je možné dělat reklamu na normální , lidské chování*
- (4) *je možné nalézt například ve filmu Cesta/Atr ke slávě ( The Road to Glory , 1936 )*

Zvláštním případem v těchto spojeních jsou apelativa *jméno, slovo, název, pojem, pojmenování, označení* ap. rozvítené neshodným přívlástkem v nominativu (propriem či apelativem). Patří k relativně frekventovaným apelativům v těchto spojeních. Funkce těchto apelativ je odlišná od předcházejících, neoznačují nějaký mimotextový předmět, ale zpředměťují pojmenování v textu (5), (6) a (7).

- (5) *Ono prosté slovo déšť/Atr , které neznamena skoro nic pro civilistu se střechem nad hlavou*
- (6) *tvoru , jemuž by mělo být odebráno označení žena/Atr a nahrazeno titulem bestie/Atr*

(7) sídlo s libozvučným **jménem Bertramka/Atr** , která se tak jmenuje po svém původním majiteli

Méně časté než obecná apelativa jsou ve spojení s neshodným nominativním přívlastkem řídicím slovem vlastní jména. Většinou je neshodný přívlastek součástí víceslovného názvu, přičemž upřesňuje místo působení nebo původu instituce (např. firmy či klubu: *Škoda Plzeň, Tesla Pardubice, Viktoria Žižkov*), zaměření instituce (*Poldi Ocel, Škoda Energo*), podtyp značky (např. automobilu: *Ford Focus, Škoda Favorit*) aj. (8) a (9). Výjimečně stojí neshodný přívlastek před řídicím jménem (10).

(8) *Zájem o dodávky mají Čína a Ruská federace , favoritem však je opět **Škoda Plzeň/Atr** ,*

(9) *Program fotbalistů Liberce v **Poháru Intertoto/Atr** nabral rychlý spád .*

(10) *možnost se zevrubně seznámit se stavbou víceúčelové haly **Sazka/Atr Aréna** .*

### 2.3.3 Přívlastek v prostém genitivu (*Atr\_2*)

Přívlastek v prostém genitivu je druhá nejčastější realizace syntaktické funkce pádem v korpusu SYN2005 (po podmětu v nominativu) se 17,9 % ze substantiv. Jen velmi malý podíl připadá na shodné přívlastky typu *Václav/Atr Havel* (1,4 %), jež byly popsány výše, zbývajících 98,6 % jsou přívlastky neshodné.

Neshodné genitivní přívlastky lze částečně roztrdit na základě syntaktických vlastností. Celkem 14,1 % genitivních přívlastků (vč. shodných přívlastků) připadá na přívlastek rozvíjející kvantifikátor, a to v 7,8 % vyjádřený číslovkou (určitou nebo neurčitou, popř. číslicí): *několik, mnoho, pár, deset, sto, tisíc, milión, miliarda, pět, šest...* (1), (2) a (3). Dalších 6,3 % genitivních přívlastků závisí na substantivním kvantifikátoru: *část, řada, většina, počet, polovina, množství, spousta, procento, desítka, tucet, litr, kilogram, hejno, záplava* aj. (4, 5 a 6). Přesné určení frekvence přívlastků závislých na substantivních kvantifikátorech není možné: jednak je obtížné stanovit, která substantiva jsou substantivními kvantifikátory, jednak ani genitivní přívlastky závislé na potenciálních substantivních kvantifikátorech nejsou vždy vyjádřením počítaného předmětu, srov.: *sklenice vody/Atr* oproti *ze sklenic pohřebního bratrstva/Atr*; *řeky krve/Atr* oproti *nejdelší řeka Evropy/Atr*).

(1) *Pojď , už jsi tu nebyl dobrých **deset let/Atr** .*

- (2) Michaelovi na **pár vteřin/Atr** klesla hlava až na hrudník .
- (3) Bohužel bude také propuštěno na **420 lidí/Atr** , tedy asi **čtvrtina pracovníků/Atr** .
- (4) Ale jako **většina Řeků/Atr** i on vyšel z moře .
- (5) Prokousala se **hromadou místních kronik/Atr** z farností kolem jezera .
- (6) Alkohol - více než **litr piva/Atr** nebo půl **litru vína/Atr** denně poškozuje játra

Verbální substantiva odvozená z trpných příčestí na *-ní/tí* (včetně substantiv, jež prošla dalším sémantickým vývojem a nepoužívají se primárně jako názvy dějů, např. *oddělení*) jsou řídicími větnými členy dalších 12,6 % genitivních přívlasků. Mezi nejčastějšími jsou např.: *vedení, řešení, zvýšení, řízení, využití, použití, snížení, oddělení, rozhodnutí, zajištění* (7) a (8). Přívlasky nejčastěji vyjadřují předmět slovesa před nominalizací: *řešení problémů, plnění úkolů, odnětí svobody* (7) nebo podmět slovesa před nominalizací: *rozhodnutí soudu, narození dítěte, usnesení vlády* (8). Došlo-li k posunu významu od původního pojmenování děje, nelze u přívlasku funkce subjekt/objekt před nominalizací rozlišovat: *hnutí odporu, oddělení policie, postavení firmy* (9).

- (7) Adekvátním **řešením** naznačených **problémů/Atr** by podle našeho názoru bylo ,
- (8) Tímto způsobem se brzy po **vypuknutí války/Atr** dostali do vlády socialisté .
- (9) Od brooklynského **oddělení vražd/Atr** se možná takové spolupráce nedočkáte .

Ostatní přívlasky (71,9 % přívlasků v prostém genitivu) je obtížné sémanticky nebo syntakticky třídit, jev je totiž velmi rozmanitý: v korpusu SYN2005 je cca 3600 různých lemmat řídicích větných členů, která nespádají do žádné z předešlých kategorií a jsou alespoň stokrát rozvíta genitivním přívlaskem. Mezi nejčastější patří další deverbativa: *výroba, ochrana, rozvoj, podpora, vznik*; různá abstrakta: *konec, možnost, výsledek, úroveň, důvod*; označení osoby (funkce): *člen, ministr, předseda, šéf, mluvčí* a podobně. Z hlediska významu přívlasku je přesné třídění také problematické, lze uvažovat například o široce přívlastňovacím významu: *majetek společnosti, břeh řeky, centrum Prahy* (10), o subjektivním genitivu: *růst cen, rychlost světla, bolest hlavy* (11) nebo třeba o genitivu cíle, zaměření či předmětu: *nabídka služeb, ministr financí, úřad práce* (12), ale přesnou hranici mezi jednotlivými typy nelze vytyčit, tím méně pak vyčíslit jejich zastoupení v korpusu.

- (10) V jižní části zbyl **kostelík** sv. **Klimenta/Atr** ( *odtud jméno ostrova/Atr* ) , původně románský

(11) *Tuzemský rum bude možné najít na pultech prodejen pod tradičním názvem až do konce roku/Atr 2003 .*

(12) *Nárok na náhradu škody/Atr uplatnilo zatím sedm poškozených či pozůstalých .*

Přívlastek obvykle následuje po řídicím substantivu a je od něj oddělen pouze jinými přívlastky (shodnými či neshodnými). Zvláště u přívlastků závislých na kvantifikátorech však není výjimkou ani opačné pořadí a přívlastek, mezi nímž a jeho řídicím větným členem může stát i sloveso (13) a (14).

(13) *Takových historek/Atr byla spousta , však o tom bylo natočeno několik televizních pořadů .*

(14) *No samozřejmě , žen/Atr poznal mnoho , ale nikdy takovou , která by ho naučila opravdu milovat*

#### 2.3.4 Přívlastek v předložkovém genitivu (Atr\_p2)

Přívlastky v předložkových pádech jsou vždy neshodné (shodné přívlastky řadíme k prostým pádům). V případě předložkového genitivu je důležitou skupinou typ, kdy předložková fráze s předložkou z rozvíjí přídavné jméno, zájmeno či číslovku ve funkci syntaktického substantiva: *jedna z dívek*. Tento typ se na neshodných přívlastcích v předložkovém genitivu v korpusu podílí 11,2 %. Nejčastějšími řídicími větnými členy jsou číslovky základní: *jeden, dva, tři*; číslovky řadové: *první, druhý, třetí*; zájmena: *některý, někdo, nikdo, ten, který*; přídavná jména blízká zájmenům: *mnohý, další* aj. (1) a (2).

(1) *Invaze německých armád do Francie roku 1914 byla prvním ze čtyř vojenských hazardů/Atr ,*

(2) *I když mnozí z českých pánů/Atr či rytířů/Atr zachytili své soukromí v denících i v korespondenci ,*

Mezi ostatními přívlastky převažují předložkové fráze s předložkami *z, do, od* a *u*, celkem 76,4 % přívlastků (47,4 % s předložkou *z*, 15,9 % s předložkou *do*, 7,1 % s předložkou *od*, 6,0 % s předložkou *u*). Řídicím substantivem přívlastků v předložkovém genitivu je v 8,0 % deverbativní substantivum (pojmenování děje s koncovkou *-ní/tí*, včetně substantiv, jejichž význam se od původního pojmenování děje dalším vývojem odchýlil). Nejčastějšími

slovesnými řídicími substantivy jsou: *obvinění, podezření, zapojení*.

Přívlastky mají nejčastěji význam lokální: *cesta do práce, odchod ze scény* (3); časový: *noc ze soboty na neděli, období od ledna do listopadu* (4); nebo vyjadřují předmět zasažený dějem (přívlastek „valenčních“ substantiv): *strach z odplaty, zásah do organismu, obvinění z vraždy, dopisy od přátel, osvobození od daně* (5). Dále se vyskytují přívlastky vyjadřující účel, původ aj.: *prášek do pečiva, vejce od slepic* (6).

(3) *Invaze německých armád do Francie/Atr* roku 1914 byla prvním ze čtyř vojenských hazardů ,

(4) *aby přednesl informaci o vývoji bezpečnostní situace od posledního zasedání/Atr štábu .*

(5) *Dnes si manželé užívají radosti z třinácti vnoučat/Atr a dvanácti pravnoučat/Atr .*

(6) *Byl viditelně odvozen od slavného středověkého dřevorezu Sv. Jiří na koni od Albrechta Dürera/Atr ( 1471 – 1528 ) .*

Ostatních 12,2 % neshodných přívlastků v předložkovém genitivu jsou předložkové fráze s jinými předložkami s genitivem (*bez, včetně, kolem, podle, za* a další): *píseň beze slov, roztoči včetně klíšťat, cesta kolem světa, evangelium podle Matouše, lisování za tepla, cena okolo čtyřiceti korun, vztahy uvnitř rodiny, cesta podél pobřeží, komunikace prostřednictvím Internetu, spor ohledně kompetencí* apod. (7), (8) a (9).

(7) *Torpédový útok ponořené ponorky bez předchozího varování/Atr se v praxi prosazoval váhavě*

(8) *Absence hmotněprávních ustanovení ohledně doby/Atr ochrany rovněž působily potíže při poskytování vzájemné ochrany .*

(9) *Bouře odtrhnou mořské řasy ze skal podél pobřeží/Atr a vy najdete v jejich chuchvalcích těla medúz a krabů .*

### 2.3.5 Přívlastek v prostém dativu (Atr\_3)

Přívlastky v prostém dativu se v korpusu vyskytují zřídka, jen 1292 ppm (0,13 %) všech substantiv. Většina z nich jsou shodné přívlastky, součásti souřadných skupin typu *pan Novák* (cca 60 %<sup>3</sup>). Mezi neshodnými přívlastky menší část (cca 25 % z neshodných přívlastků, tj.

3 Vzhledem k vysoké chybovosti automatické syntaktické anotace u této kombinace (a obecně u všech málo frekventovaných kombinací pádu a funkce) uvádíme jen přibližné údaje.

cca 120 ppm) rozvíjí deverbativní přívlastky (přímá pojmenování děje odvozená od sloves sufixy -ní/tí): *poskytování, porozumění, přizpůsobení, předání, podřízení, doporučení, poskytnutí, udělení, přiblížení, zabránění* aj. (1), (2) a (3).

- (1) *v Praze bylo běžnou praxí poskytování služeb matkám/Atr , které si přály slehnout potají .*
- (2) *Zabránění domluvě/Atr agresorů na křivé výpovědi .*
- (3) *c ) doručením rezignace prezidentu/Atr republiky , d ) právní mocí rozsudku*

Ostatní neshodné přívlastky rozvíjejí jiná, široce „valenční“ substantiva. Tyto přívlastky rozvíjejí řídicí substantiva v několika významových okruzích, mezi něž patří například pomoc včetně finanční asistence: *pomoc, podpora, dotace* (4). Dále (spíše pozitivní) mezilidské vztahy: *vděčnost, věrnost, oddanost, důvěra* (5). Písemné zprávy: *dopis, list, zpráva* (6). Jiná komunikace: *výzva, otázka, ultimátum* (7). Pozitivní ocenění: *pocta, sláva, chvála* (8). Transakce, nejen obchodní: *prodej, nabídka, půjčka* (9) a jiné.

- (4) *Tím nemají čas na pomoc samičce/Atr , která začne stavět hnízdo*
- (5) *Taková věrnost americkým ideálům/Atr se oceňuje !*
- (6) *Při četbě listu Římanům/Atr . . . se mu náhle vyjasnilo ,*
- (7) *OK , ale jako vzkaz příchozímu hostu/Atr to působí trochu nečekaně*
- (8) *Na řadě je pocta mrtvým kamarádům/Atr :*
- (9) *společný prodej podílů v ČEZ a v šesti REAS ( dále ČEZ + 6 REAS ) ve výběrovém řízení strategickému investorovi/Atr*

Výjimečně může být neshodný přívlastek od řídicího členu oddělen slovesem a dokonce mu předcházet (10).

- (10) *Spojili jsme se a mým úvahám/Atr byl konec .*

### 2.3.6 Přívlastek v předložkovém dativu (Atr\_p3)

Neshodné přívlastky v předložkovém dativu jsou v korpusu přibližně třikrát častější než neshodné přívlastky v dativu prostém, i tak je to ale málo frekventovaná kombinace, přívlastek realizovaný předložkovým dativem představuje pouze 4049 ppm (0,4 %) všech

substantiv. Celkem 76,5 % těchto přívlastků je s předložkou *k*, 14,6 % s předložkou *proti*, 4,6 % s předložkou *vůči*, 2,4 % s předložkou *kvůli*. Zbývající 1,9 % procenta těchto přívlastků jsou předložkové fráze s předložkami *oproti*, *díky*, *navzdory*, *naproti*, *vstříc* aj.

Celkem 8,2 % neshodných přívlastků s předložkovým dativem rozvíjí deverbativní substantiva (-ní/tí), nejčastěji: *povolení*, *opatření*, *připojení*, *přihlédnutí*, *tažení* (1 a 2). Význam těchto substantiv se často posunul a nejde již o přímá pojmenování děje.

- (1) *Přijmout účinná **opatření k ochraně/Atr** československé ekonomiky před mimořádnými událostmi*
- (2) *Je totiž otázkou , zda je **tažení proti** Evropské **unii/Atr** v této chvíli na místě .*

Z hlediska vztahu k řídicímu substantivu vyjadřují neshodné přívlastky v předložkovém dativu nejčastěji předmět zasažený dějem či vlastností, jež vyjadřuje řídicí substantivum: *úcta k životu*, *vztah k přírodě*, *protest proti válce*, *boj proti terorismu*, *výzva k úhradě*, *přístup k informacím*, *imunita vůči chorobám* aj. Sémantika řídicích substantiv těchto přívlastků je poměrně pestrá, většinou jsou to abstrakta nebo původně konkrétní substantiva s přeneseným významem; označují různé vztahy: *nenávisť*, *úcta*, *důvod*, *přístup*, *sklon ... k* (3); různé aspekty konfliktu: *válka*, *boj*, *obrana*, *imunita*, *protest ... proti* (4); souhlas nebo výzvu autority: *souhlas*, *právo*, *povolení*, *pokyn*, *rozkaz ... k* (5) a jiné. Předložky jsou částečně variabilní, předložku *proti* a někdy i *k* nahrazuje předložka *vůči*.

- (3) *Budete mít **tendenci k** přílišné **agresivitě/Atr** .*
- (4) *Pokračuje tu starý **zápas proti hlouposti/Atr** ( stultitia ) ,*
- (5) ***Oprávnění k** této **činnosti/Atr** prokazují průkazem státního dozoru nad kapitálovým trhem .*

Výrazně méně časté jsou přívlastky s významem okolnostního určení. Přívlastky vyjadřují místo: *cesta k moři*, *postup k pobřeží*, *výlet k vodopádům* a účel: *klíč k bytu*, *dárek k Vánocům*, *výstava k 100. výročí*. Hranice mezi okolnostním určením a předmětem (valenčním přívlastkem) není vždy jasná, abstrakta se spíše řadí k (široce pojatým) valenčním přívlastkům, konkrétní spíše vyjadřují okolnosti (6) a (7), (8) a (9).

- (6) *Vycítila jeho nesnáze a ukázala mu **cestu k odpovědi/Atr** .*
- (7) *Zahnula za roh na **cestu k** našemu **domu/Atr** .*

- (8) *Jinak řečeno , diferenciace je klíč k úspěchu/Atr .*
- (9) *přečetl si možná nějaký váš dopis , našel klíč k mému stolu/Atr . . .*

### 2.3.7 Přívlastek v prostém akuzativu (Atr\_4)

Přívlastek v prostém akuzativu je také málo častá realizace syntaktické funkce pádem, mezi všemi substantivy je takových kombinací jen 1225 ppm (0,12 %). Až na výjimky, které uvedeme dále, jsou všechny tyto přívlastky shodné (součásti souřadných skupin, jež byly popsány výše). Prostý akuzativ jako primárně předmětný pád v zásadě neumožňuje, aby byl využit jako přívlastek. Přesto se akuzativní přívlastek může objevit, a to v případě nominalizace slovesa s okolnostním určením v prostém akuzativu (určení času nebo míry). V korpusu SYN2005 jsme našli pouze dva takové příklady (tři, počítáme-li oba členy koordinace samostatně) neshodného přívlastku v akuzativu, v obou má akuzativ funkci určení času (1) a (2).

- (1) *Ideální je mírné cvičení hodinu/Atr obden nebo aspoň půlhodinu/Atr denně .*
- (2) *Anamnéza plošné bolesti s trváním nejméně tři měsíce/Atr .*

### 2.3.8 Přívlastek v předložkovém akuzativu (Atr\_p4)

Neshodné přívlastky v předložkovém akuzativu jsou přibližně desetkrát častější než přívlastky v předložkovém dativu, jejich podíl na všech substantivech je 1,5 % (4,9 % z přívlastků). Podíl čtyř nejčastějších předložek v rámci těchto přívlastků je 93 %: *na* (46,8 %), *pro* (28,6 %), *o* (9,6 %) a *za* (8,1 %), např. *odpověď na otázku, smysl pro humor, pokus o převrat, boj za nezávislost*. Kromě těchto čtyř jsou neshodné přívlastky řízeny dalšími akuzativními předložkami: *přes, v, po, pod, mimo, nad, mezi, před, skrz*, např. *most přes řeku, víra v Boha, rána pod pás* aj.

Mezi řídicími substantivy těchto přívlastků je podíl deverbativních substantiv (-ní/tí) poměrně nízký (6,7 %), často jde navíc o substantiva, jejichž význam je od prostého pojmenování děje vzdálený (*zařízení, sdružení, hnutí, označení, pojištění*). Jen malá část deverbativních substantiv si zachovala původní charakter i valenci slovesa (*čekání na Godota, volání o pomoc*).

Většinu řídicích substantiv neshodných přívlastků lze považovat za široce valenční, přívlastky mají funkci předmětu zasaženého dějem (cíl, téma aj.). Podrobnější významové

třídění těchto „valenčních“ substantiv je obtížné, kategorie víceméně odpovídají sémantickému třídění valenčních sloves s předložkami *na*, *o*, *pro* a *za*, a celkem by jich mohlo být i několik desítek. Převládají abstrakta, často dějová; řídicí substantiva vyjadřují např. působení: *vliv, působení, dopad, účinek ... na* (3), úsilí: *snaha, úsilí, pokus, usilování ... o* (4) nebo strach: *strach, starost, obava, bázeň ... o* (5). Ani jednotlivá valenční substantiva nejsou sémanticky jednotná, srov. *zájem o práci, zájem o koupi* (cíl je předmět) oproti *zájem o kulturu, zájem o problematiku raně novověkých dějin* (cíl je informace o předmětu).

- (3) *Prověrka **vlivu** elektrárny **na** životní **prostředí/Atr** měla být podle nich uzavřena až po vzájemných konzultacích*
- (4) *Záleží pak **na uchazeči o práci/Atr** a zaměstnavateli , aby si domluvili vznik pracovního poměru .*
- (5) *V **obavách o život/Atr** se pak přestala bránit .*

Kromě častějších neshodných přívlastků závislých na „valenčních“ substantivech je mezi přívlastky s předložkovým akuzativem i nezanedbatelná část okolnostních určení. Přívlastky nejčastěji určují místo: *vstup na stránky; lanovka na Sněžku; most přes Labe* (6); účel: *salva na počest; těsto na knedlíky, prostor pro cestující* (7) a důvod nebo příčinu: *penále za nedodržení (smluvních podmínek), žaloba pro zanedbání (povinné péče), stížnost pro podjatost* (8).

- (6) *Nejlepší basketbalistka Evropy roku 1996 Eva Němcová zvažuje **návrat pod** bezedné **koše/Atr** .*
- (7) *jaké to je , když vás **paní na hlídání/Atr** nechá na holičkách .*
- (8) *Obnovuje se trestní **stíhání** generála Lorence **pro zneužití/Atr** pravomoci veřejného činitele .*

I neshodné přívlastky v předložkovém akuzativu výjimečně předcházejí řídicímu substantivu (9) a (10).

- (9) ***O problematiku/Atr** bydlení je ze strany médií velký **zájem** ,*
- (10) *abych si mohla **na věc/Atr** vytvořit vlastní **názor** ,*

Taková spojení lze také interpretovat jako předmět složeného slovesně-jmenného přísudku (*je zájem ... o problematiku/Obj*), v souladu s formalismem PDT zde ovšem zachováváme interpretaci přívlaskovou. Automaticky bylo takto označeno jen 0,5 % přívlasků v předložkovém akuzativu.

### 2.5.9 Přívlasek v prostém vokativu (Atr\_5)

Přívlasků v prostém vokativu je v korpusu jen 459 ppm (0,05 %), je to nejméně častá realizace přívlasku. Nízká frekvence odpovídá i celkové nízké frekvenci pádu (0,4 % ze všech substantiv). Až na cca 70 výjimek (2 ppm) jsou všechny tyto přívlasky shodné, součásti souřadných spojení typu *pan Novák (pane/Atr továrníku/Atr Bierhanzle)*. Ojedinelé výskyty neshodných přívlasků ve vokativu jsou přívlasky citátového charakteru závislé na slově *oslovení* (1) až (3) nebo, v jednom případě, *volání* (4). Citátový charakter je ve většině případů zvýrazněn použitím uvozovek (3) a (4). Podobný citátový charakter mají vokativy v názvech písní a modliteb: *chorál Svatý Václave, modlitba Otče náš* (5) a (6).

- (1) *Neoslovujte uchazeče mladá paní , chlapče , slečno , ale používejte oslovení slečno/Atr , paní/Atr , pane/Atr + příjmení .*
- (2) *Oslovení Vážený pane řediteli/Atr je po příjmení a titulu bráno jako nejdůležitější ,*
- (3) *Ježíšovo oslovení " Otče/Atr " vyjadřuje něco , co se nekonečně ukazuje na celém pozemském rodičovství .*
- (4) *Slyšel jsem volání " Conchito/Atr " ,*
- (5) *obraz člověka sklíčeného utrpením , jenž hledá útěchu v modlitbě Otče/Atr náš .*
- (6) *výslovně citoval jeho výklad i estetické hodnocení písně Hospodine/Atr , pomiluj ny ,*

### 2.5.10 Přívlasek v předložkovém lokálu (Atr\_p6)

Přívlasek v předložkovém lokálu je kromě genitivního přívlasku nejčastější realizací přívlasku (10,8 % z přívlasků, 3,3 % ze všech substantiv). Mezi přívlasky v předložkovém lokálu byly výjimečně zařazeny i shodné přívlasky typu *pan Novák* či *řeka Vltava* (ty se u ostatních pádů počítají jako přívlasky s prostým pádem). Tyto shodné přívlasky představují 9,1 % přívlasků v lokálu. Mezi neshodnými přívlasky je poměr různých předložek: 59,9 % předložka *v*; 21,3 % *na*; 11,5 % *o*; 4,3 % *po*; 3,0 % *při*.

Podíl deverbativních substantiv (*-ní/tí*) na řídicích substantivech neshodných přívlasků v předložkovém lokálu je 8,1 %; i zde se význam těchto substantiv často posunul od prostého

popisu děje, např. *zasedání MMF v Praze, zastupování státu ve věcech majetkových* (1) a (2).

- (1) Ovšem před **použitím** těchto metod **v** široké klinické **praxi/Atr** musí proběhnout velmi náročné klinické zkoušení .
- (2) Šrejber je podezřelý ze **zneužívání** informací **v** obchodním **styku/Atr** .

Z hlediska významového a funkčního vztahu přívlastků a řídicích substantiv představuje cca 70 % neshodných přívlastků určení místa (dohromady u slovesných i neslovesných substantiv), a to obvykle s předložkou *v* (3), zřídka i s dalšími předložkami (4) a (5).

- (3) *jak armádní špičky v Pchjongjangu hodnotily po válce v Zálivu/Atr porážku Saddáma Husajna :*
- (4) *odtud lze dorazit i na pláži u Kalami při sv. cípu/Atr ostrova ,*
- (5) *např. kliky vstojí oporou o stěnu nebo ručkování po podlaze/Atr ze stoje dopředu a zpět .*

Dalších 12,5 % neshodných přívlastků rozvíjí „valenční“ substantiva, přívlastky mají význam předmětu zasaženého dějem (téma, cíl aj.). Většina (57,3 %) neshodných přívlastků je na řídicím substantivu závislá prostřednictvím předložky *o*, přívlastky vyjadřují široce pojaté téma: *zákon, smlouva, ustanovení; informace, představa, úvaha; diskuse, řeč, vyprávění; kniha, článek, film ... o* (6).

Druhé nejčastější jsou předložkové fráze s předložkou *v* s 21,9 % v rámci „valenčních“ substantiv. Zde je hranice mezi místním určením a předmětem velmi nejasná, řídicí substantiva, jež k „valenčním“ substantivům řadíme (*změna, rozdíl, úspěch, pokrok, podíl, výhra ... v*), mají sice přenesený význam a neshodné přívlastky na nich závislé jsou často abstraktní, ale význam určení místa je u nich nadále znatelný, s výjimkou slova *pokračování* (7). U všech těchto „valenčních“ substantiv jsou častá i místní doplnění s touž předložkou (8).

Další „valenční“ substantiva mají rekci s předložkou *na* (12,9 %): *podíl, účast, závislost, škoda, ublížení, zájem, práce ... na* (9) a s předložkou *po* (7,9 %): *touha, stesk, poptávka, pátrání, stopa, vdova ... po* (10).

- (6) *V bezprostřední reakci na zprávy o zemětřesení/Atr* Voltaire napsal svému bankéři ,
- (7) *Jiným příkladem je pokračování v blokádě/Atr* Německa Dohodou po válce .

- (8) rakousko-uherskou monarchií , jejímž **pokračováním v** bývalém **Zalitavsku/Atr** byla Maďarská republika .
- (9) Výkon tohoto práva slouží občanům k jejich **účasti na** politickém **životě/Atr** společnosti ,
- (10) že **volání po svobodě/Atr** sílí každým dnem ,

Kromě místních určení a „valenčních“ přívlastků vyjadřují neshodné přívlastky v lokálu také časová určení, často s předložkami *po* a *v*, řídčeji *při* a *o*, s předložkou *na* pouze v několika typických obratech: *na začátku*, *na konci*, *na přelomu*, *na jaře* (11), (12) a (13).

- (11) *Pak s námi společně pronesla birkat ha-mazon* , **požehnání po jídle/Atr** .
- (12) *Je samozřejmé* , že bez Kristova **vzkříšení o** oněch slavných prvních **Velikonocích/Atr** křesťanské éry by nic z toho možné nebylo ,
- (13) *Po invazi do Tibetu v roce/Atr* 1950 zde Číňané vybudovali obrovské středisko jaderného výzkumu ,

Přívlastky v lokálu mohou vyjadřovat i jiná okolnostní určení, například určení způsobu (14) a (15). Méně často, ale nijak vzácně vyjadřují přívlastky i jiné vlastnosti řídicího substantiva, které nelze přiřadit k určení okolnosti (16) a (17).

- (14) *Možnosti jsou v zásadě dvě* : nová dražba , nebo **prodej** podniku **po částech/Atr** .
- (15) *Odlehčení končetiny* **chůzí o berlič/Atr** je nutné *po dobu* několika měsíců .
- (16) *Turci doposud zrušili kontrakt* francouzského podniku Alcatel **ve výši/Atr** 200 miliónů amerických dolarů
- (17) *Prezidentské* **apartmá o rozloze/Atr** 300 metrů čtverečních ( *pro Karla Brücknera* ) .

### 2.5.11 Přívlastek v prostém instrumentálu (Atr\_7)

Prostý instrumentál se k vyjádření přívlastku nepoužívá často. Podíl realizací přívlastku prostým instrumentálem mezi přívlastky je 1,2 % , mezi všemi substantivy 0,4 % (3688 ppm). 17,6 % přívlastků realizovaných prostým instrumentálem jsou shodné přívlastky (součásti souřadných spojení).

Řídící substantiva neshodných přívlastků jsou ze 42,5 % tvořeny deverbativními substantivy (-ní/tí), ostatní jsou dějová substantiva (až na drobné výjimky, cca 1,3 % , které

budou popsány níže). Tato dějová substantiva jsou zčásti odvozená od sloves: *procházka, jízda, platba*, zčásti jsou jiného původu, ale v jejich významu je tak či onak děj obsažen: *cesta, smrt, infekce*.

Význam neshodných přívlastků v prostém instrumentálu lze rozčlenit do několika skupin, jejichž podíl není možné bez podrobného rozboru významů přívlastků přesně vyčíslit. Malá část těchto přívlastků vyjadřuje podmět děje vyjádřeného substantivem: *schválení parlamentem, bodnutí hmyzem, osvobození města americkou armádou* (1) a (2). Významnější část představují předměty zasažené dějem: *hod oštěpem, plýtvání penězi, obchod obilím, pokrčení rameny, jmenování ministrem, nákaza antraxem* (3) a (4).

- (1) Tato bezpříkladná **potupa** " **domorodci/Atr** " ochromila poněkud italské kolonialistické ambice ,
- (2) Také pozdní doba kvetení zde vylučuje **poškození** květů nízkými **teplotami/Atr** .
- (3) zlepšení životního prostředí , zvýšení bezpečnosti **zásobování palivy/Atr** ,
- (4) Pokutě padesát tisíc korun či stíhání pro **pohrdání soudem/Atr** s hrozbou až dvouleté basy se včera Maxa vyhnul na poslední chvíli .

Ostatní přívlastky mají význam různých okolnostních určení, především určení místa: *pochod Prahou, cesta lesem, průlet atmosférou, průvodkyně pořadem* (5) a (6) a určení způsobu a nástroje: *bodnutí nožem, cesta vlakem, práce rukama, platba kartou* (7) a (8). Ostatní okolnostní určení substantiv jsou méně častá, a to jak určení času: *učednice druhým rokem, vykonatelnost rozhodnutí okamžikem doručení* (9), tak určení příčiny: *opotřebením otěrem, smrt udušením* (10). Určení příčiny je často obtížné odlišit od subjektového přívlastku, srov. příklad (2), nebo od určení způsobu.

- (5) **Průtok** krve **ledvinami/Atr** se omezuje při svalové práci .
- (6) Na **řezu** kostěným **kanálkem/Atr** má přibližně trojúhelníkovitý tvar .
- (7) a proto **křest ohněm/Atr** zažijí pouze před soupeřovými diváky .
- (8) Při odvetném **úderu pěstí/Atr** utrpí jen jeden člověk - útočník .
- (9) zákoník práce , ve znění pozdějších předpisů ( **účinnost dnem/Atr** vyhlášení )
- (10) Proud je velmi nebezpečný a může způsobit **smrt utonutím/Atr** při bleskovém stoupaní hladiny .

Výjimku, kdy řídicí substantivum nemusí být slovesné ani dějové, jsou přívlastky blíže

určující vlastnost řídicího substantiva: *jménem*, *původem*, *tělem* i *duší* apod. (11, 12 a 13). Tyto přívlasky se oproti výše uvedeným typům objevují zřídka, celkem představují cca 1,3 % neshodných přívlasků vyjádřených prostým instrumentálem.

- (11) *Muž z Kladna jménem/Atr Eduard vyházel z kontejneru odpadky proto , aby si pohodlněji ustlal .*
- (12) *kteří jsou pevně spjati s nadacemi původem/Atr z Maroka a Saúdské Arábie .*
- (13) *Opravdový sportovec tělem/Atr i duší/Atr Jiří Králík spolu s podnikatelem Oldřichem Karpíškem .*

### 2.5.12 Přívlasek v předložkovém instrumentálu (Atr\_p7)

Neshodné přívlasky realizované předložkovým instrumentálem mají mezi všemi přívlasky podíl 5,8 %, mezi substantivy 1,8 %. Mezi neshodné přívlasky počítáme i nepravé přívlasky s předložkou s vyjadřující slučovací poměr (*hrabě s chotí*). Nejčastěji používanou předložkou v těchto neshodných přívlascích je předložka *s* (69,3 %). Následují předložky *mezi* (12,8 %), *nad* (6,6 %), *před* (5,0 %), *pod* (3,7 %), *za* (2,6 %) a *napříč* (0,1 %). Podíl deverbativních substantiv (*-ní/tí*) na řídicích substantivech těchto přívlasků je 8,0 %, nejčastější jsou *spojení*, *setkání*, *jednání ... s/mezi*.

Z hlediska významu neshodných přívlasků a jejich řídicích substantiv jsou nejvíce zastoupena určení průvodní okolnosti (provázající osoby, vlastnosti aj.), představují cca 45 % přívlasků v nepředložkovém instrumentálu (1), (2) a (3). Řadíme sem také hypotaktické vyjádření slučovacího vztahu: je poměrně málo časté, osoby jsou takto označeny pouze v 180 ppm, tj. 1 % přívlasků v předložkovém instrumentálu (4) a (5).

- (1) *Stavba domu s pečovatelskou službou/Atr je tak minimálně na několik dní zastavena .*
- (2) *Před stánkem s pečivem/Atr na Palmovce se krátce před polednem vytvořila souvislá fronta netrpělivých zákazníků .*
- (3) *Vedle ní byl muž s placatou čepicí/Atr na hlavě .*
- (4) *Když Honzík s Mařenkou/Atr vystoupili z lodi , byli příjemně překvapeni ,*
- (5) *Pak přijela matka s otcem/Atr a převzali velení .*

Druhým nejčastějším typem neshodných přívlasků v předložkovém instrumentálu jsou přívlasky „předmětné“ (zasažené dějem či vlastností řídicího substantiva), které rozvíjejí

slovesná, dějová a další široce „valenční“ substantiva: *problém, potíž, obchod, manipulace, nakládání ... s; vztah, smlouva, kontakt, spor, srážka ... s/mezi; kontrola, dohled, vítězství, rozhořčení, rozpaky ... nad; honba, hon, pohledávka, ohlédnutí, zaostávání ... za* aj. (6, 7 a 8). Mnoho substantiv, jež jsou často rozvity přívlastky s lokálním významem, lze v případě abstraktních přívlastků řadit také mezi „valenční“ substantiva: *cesta za úspěchem, přechod mezi realismem a romantismem* (9).

(6) *Hádka mezi* dvěma *mladíky/Atr* a *mužem/Atr* v Orelské ulici v Praze 10 vyústila málem ve vraždu .

(7) *Jak bude podle zákona č. 458/2000 Sb. postupně otevírán* *trh s elektřinou/Atr* ?

(8) *Tu je na místě menší filosofické* *zamyšlení nad pojmem/Atr* hodnota .

(9) *Propast mezi lidmi/Atr* s nízkými a vysokými příjmy se podle loňských statistik stále více prohlubuje .

Třetím častým významem neshodných přívlastků v předložkovém instrumentálu je místní a časové určení. Místní určení může rozvíjet dějové substantivum: *putování napříč Antarktidou, přesuny mezi budovami školy*, častěji ale rozvíjí ostatní, nedějová substantiva: *střecha nad hlavou, silnice mezi Děčínem a Prahou* (10) a (11). Za místní určení považujeme i složené místní názvy: *Ždár nad Sázavou, Frenštát pod Radhoštěm* (12). V místních a časových určeních typu *metr pod povrchem* nebo *den před Vánoci* se v PDT považuje za řídicí substantivum v akuzativu, na kterém je závislá předložková fráze. Přestože by bylo možné najít argumenty i pro opačné řešení, tj. řídicí je předložková fráze a substantivum v akuzativu udává míru (jak hluboko *pod povrchem*, jak dlouho *před Vánoci*), řešení použité v PDT přebíráme. Mezi místními určeními je takových obrátů poměrně málo (13), naopak většina neshodných přívlastků v předložkovém instrumentálu, které určují čas, má právě tento charakter (14). V korpusu se ale vyskytují i jiné „časové“ přívlastky v předložkovém instrumentálu (15).

(10) *Lávka tak pěším umožní* *přechod mezi* oběma *břehy/Atr* Vltavy .

(11) *Já je viděl , ty dva blázny , z* *ohybu cesty za kolejemi/Atr* , kam teď chodíme se *stavebním komandem* ,

(12) *Již v posledním domácím zápase s* *Ústím nad Labem/Atr* jsme naznačili zlepšení .

(13) *Byla to mobilní nemocnice , umístěná několik* *kilometrů za liniemi/Atr*

(14) *V tom posledním ,* *den před popravou/Atr* , uvedl , že svého činu nelituje

(15) *obdobím jeho života ,* *rok mezi říjnem/Atr* 1912 a 1913 byl naopak dobou

## 2.6 Předmět (*Obj*)

Předmět je větný člen závislý na slovese, adjektivu, výjimečně na adverbium, jeho pád (případně i předložka) je určen řídicím členem. V rámci všech funkcí realizovaných substantivy má předmět podíl 20,9 %. Nejčastěji je realizován prostým akuzativem (62,3 %). Funkci **Obj** má i původce děje v prostém instrumentálu nebo určení původu a výsledku děje (vše tradičně **Adv**).

### 2.6.1 Předmět v prostém nominativu (*Obj\_1*)

Na rozdíl od tradičního řazení k příslovečnému určení způsobu se ve formalismu PDT za předmět považují názvy v nominativu, které jsou závislé na slovesech *nazvat*, *nazývat*, či na adjektivech *nazvaný*, *zvaný*, dále výrazy připojené spojkou *jako* u sloves, u kterých takové obraty patří k valenci slovesa (*označit jako*) a výrazy v nominativu závislé na slově *viz*. Automatické značkování korpusu bylo v tomto případě zcela nespolehlivé, bylo nutné samostatně vyhledat relevantní příklady bez ohledu na značkování. Předměty realizované prostým nominativem jsou velmi málo frekventované, představují méně než 0,1 % všech substantiv (553 ppm).

Předměty v nominativu patří do čtyř přibližně stejně frekventovaných skupin. Jsou to výrazy po spojce *jako* závislé na slovesech *označit* a *označovat*, pojmenování závislá na slovesech *nazvat* či *nazývat* (2), dále pojmenování závislá na přídavných jménech *nazvaný*, *nazývaný*, *zvaný* (3) a odkazy závislé na pokleslém, gramatikalizovaném imperativu *viz* (4).

- (1) *Jako populární hudba/Obj se přibližně od roku 1960 označuje směs white blues , rocku a písň*
- (2) *Toto předivo potravních vztahů , tedy kdo se kým živí , se nazývá potravní síť/Obj .*
- (3) *Zatím se dívkám šťastně podařilo vyjít z tajné chodby v místě zvaném Kobylí hlava/Obj .*
- (4) *Hluboký zážitek z účinků barevných oken lze zažít v sakrálním prostoru některé z katedrál - viz v Praze katedrála/Obj sv. Víta nebo v Chartres ve Francii*

### 2.6.2 Předmět v prostém genitivu (*Obj\_2*)

Prostým genitivem se realizuje 4,8 % všech předmětů, tato kombinace pádu a funkce představuje 1,0 % všech substantiv. Většina z nich rozvíjí slovesa s genitivní valencí, jichž jsou v češtině desítky, jen malá část je závislá na přídavných jménech (4,2 %), vzácně se

v předmětné funkci užívá genitivu záporového (0,6 %) nebo partitivního (0,1 %).

Slovesa s genitivní valencí jsou často zvrtná, většinou s reflexivem *se*. Význam těchto sloves je pestrý, bylo by možné vytvořit až několik desítek sémantických okruhů, do nichž bychom slovesa rozřídili. Místo toho uvedeme jen několik příkladů. Mezi valenční slovesa s genitivem patří slovesa vyjadřující dotyk a cíl: *dotknout se*, *dosáhnout*, *chopit se* (1) a slovesa užívání, u nichž se genitiv střídá s akuzativem: *použít*, *využívat* (2). Genitivní předmět mají slovesa vyjadřující obavy: *bát se*, *obávat se* (3), otázku: *ptát se*, *tázat se* (4); odluku: *zbavit se*, *zprostit* (5) a další.

Uvádíme seznam dvaceti nejčastějších sloves s genitivním předmětem (v závorce je uveden počet substantivních předmětů v genitivu, přepočítáno na milion slov v korpusu): *dosáhnout* (425 ppm), *týkat se* (271 ppm), *využít* (167 ppm), *zbavit se* (149 ppm), *zúčastnit se* (112 ppm), *dotknout se* (91 ppm), *dosahovat* (89 ppm), *využívat* (89 ppm), *bát se* (87 ppm), *dočkat se* (82 ppm), *vzdát se* (82 ppm), *dostat se* (77 ppm), *zeptat se* (76 ppm), *všimnout si* (75 ppm), *účastnit se* (71 ppm), *držet se* (64 ppm), *nabýt* (59 ppm), *použít* (56 ppm), *ujmout se* (56 ppm), *používat* (54 ppm).

- (1) *Naše celebrity dlouho nečekaly a chopily se taky příležitosti/Obj.*
- (2) *Nemůžu zneužít jeho důvěry/Obj.*
- (3) *kdo se neumí bránit, lekne se křiku/Obj, i když má bajonety*
- (4) *Ale musel jsem se taky optat přítelkyně/Obj, co ona by ráda.*
- (5) *Nicméně ani tehdy se nevzdal naděje/Obj.*

Na přídavných jménech je závislá jen malá část, 4,2 % genitivních předmětů (420 ppm v rámci substantiv v korpusu); pouze pět nejčastějších adjektiv má 10 či více substantivních předmětů v genitivu na milion slov v korpusu. Genitivní předměty mohou být závislé na deverbativních adjektivech odvozených od sloves s genitivní valencí (6), ale také na původních přídavných jménech, a to na jejich obvyklých, adjektivních tvarech (7) i na krátkých, predikativních formách (8).

Dvacet nejčastějších přídavných jmen s genitivní valencí (u pěti nejčastějších uvádíme i frekvenci substantivních předmětů v genitivu): *plný* (196 ppm), *týkající* (99 ppm), *schopný* (39 ppm), *vědomý* (31 ppm), *hodný* (15 ppm), *využívající*, *zbavený*, *prostý*, *poslušný*, *znalý*, *mocný*, *dotýkající se*, *účastný*, *účastníci se*, *dosahující*, *blízký*, *daleký*, *pamětlivý*, *užívající*, *důstojný*.

- (6) *Odmítá odpovídat na otázky **týkající se smrti/Obj** .*
- (7) *Kdyby neměla oči **plné slz/Obj** , mohl by si myslet , že hraje komedii nebo že žertuje .*
- (8) *Byla **si** dojemně **vědoma** svého **stavu/Obj** , byla jím zkroušena*

Záporový genitiv u sloves s akuzativní valencí (prostý genitiv nahrazuje prostý akuzativ v případech negace slovesa) se v korpusu SYN2005 vyskytuje jen zřídka, podíl těchto případů na substantivech v korpusu je cca 60 ppm (0,6 % genitivních předmětů). Záporový genitiv se téměř vždy objevuje v lexikalizovaných obratech, nejčastěji se slovesem *mít* (9) a (10), dále například se slovesy *brát*, *znát*, *snést* (11). Několik nejčastějších kombinací se záporovým genitivem: *nemít ... obdoby, potuchy, konce, trvání, smyslu, pochyb; nemít ... kouska citu, studu, odvahy; nevědět si rady; neznat ... mezí, hranic; nebrat konce; nezamhouřit oka; nesnést odkladu* aj.

- (9) *Románek však **neměl** dlouhého **trvání/Obj** .*
- (10) *Kam se cpeš mezi vládnoucí elitu , když **nemáš** **kouska/Obj** strategického smyslu ?*
- (11) *Ale vášeň , která **nezná** **mezí/Obj** , nezná ani žádné omezování .*

Partitivního genitivu se u sloves s akuzativní valencí užívá ještě vzácněji než genitivu záporového, cca 10 ppm v rámci substantiv v korpusu (0,1 % genitivních předmětů). Obvykle se používá v několika lexikalizovaných obratech, např. *nalít čistého vína* (12) a (13), výjimečně i v jiných spojeních, kde použití genitivu signalizuje větší množství (14).

- (12) *Možná by bylo lepší **nalít** si čistého **vína/Obj** .*
- (13) *Bratříčku Serváci , nemohl bys mně **podat vody/Obj***
- (14) ***Nakoupil** různých **dárků/Obj** a chystal se na cestu .*

Za předmět považujeme také substantiva v genitivu závislá na přísudku složeném ze slovesa *být* a predikativa *líto* (15) a (16). Těchto předmětů je mezi substantivy v korpusu cca 5 ppm.

- (15) ***Je** nám **líto** vašeho **utrpení/Obj** a želíme vašich zmarněných životů .*
- (16) *Toho **člověka/Obj** **je** mi opravdu **líto** . . .*

### 2.6.3 Předmět v předložkovém genitivu (*Obj\_p2*)

V korpusu se substantivy v předložkovém genitivu realizuje předmět poměrně zřídka, tato kombinace funkce a pádu je pouze u 0,5 % substantiv v korpusu (2,6 % substantivních předmětů). Za předmět se v PDT, a tedy i v syntakticky anotovaném korpusu SYN2005 považuje také určení původu (látky), výsledku a původce děje, jak podrobněji předvedeme dále.

Jako u jiných předložkových frází je rozlišení mezi předmětem a příslovečným určením často obtížné, stejné sloveso může být rozvito předložkovou frází se stejnou předložkou jednou s významem okolnostním, podruhé s významem předmětu zasaženého dějem. Tato nejednoznačnost se negativně projevuje i na spolehlivosti automatického značkování: celkové statistické údaje jsou sice díky opravám a manuálním sondám do korpusu poměrně spolehlivé, relativní frekvence jednotlivých sloves a předložek ale neuvádíme.

Předmět v předložkovém genitivu se pojí se čtyřmi předložkami: *z* (44,4 %), *do* (39,8 %), *od* (14,4 %) a *bez* (1,4 %). Předložkové fráze s ostatními předložkami s genitivem mají většinou funkci *Adv*, zřídka i *Atr* či *ExD*, ale ne *Obj*. Předmět rozvíjí v 93,0 % sloveso, v 6,5 % rozvíjí přídavné jméno a v 0,5 % rozvíjí příslovce.

S předložkou *z* se nejčastěji pojí slovesa: *obvinit, nařknout ... z; radovat se, těšit se ... z; vyplývat, vycházet ... z; vzpamatovat se, probrat se ... z; těžít, profitovat ... z* a další (1) a (2). Předložkovými frázemi s předložkou *z* se také často vyjadřují určení původu/látky (3) a (4).

- (1) *Potrvá dva až tři roky , než se Mongolsko **vzpamatuje z** kruté **zimy/Obj** ,*
- (2) *Policie mladíka **podezřívala z porušování/Obj** domovní svobody , **krádeže/Obj** a **poškození/Obj** cizí věci .*
- (3) *Každý se **skládá ze záhlaví/Obj** a datové **části/Obj** .*
- (4) ***Vyráběny** byly převážně **z jasanového dřeva/Obj** , které pro svoji pružnost nejlépe odolávalo zatížení .*

S předložkou *do* a s genitivem jsou nejčastější slovesa: *dát se, pustit se ... do; zamilovat se, zakoukat se ... do; zabrat se, pohroužit se ... do; vcítit se, vžít se ... do; nutit do* aj. (5) a (6). Zřídka se používá předložkových frází s předložkou *do* k určení výsledku děje, a to se slovesy *přejít, přecházet, přerůst* aj. (7) a (8), častější jsou ale předložky *v* nebo *na* s akuzativem. Určení výsledku děje (*Obj*) lze často jen těžko odlišit od (metaforického) určení místa (*Adv*).

- (5) Potom **se pouští do otvírání/Obj** okna .
- (6) aby dívku **do sňatku/Obj** nenutil ,
- (7) Svah **přecházel do roviny/Obj** .
- (8) ale žena začala hystericky ječet , až její skřek **přešel do vzlykotu/Obj** .

S předložkou *od* jsou nejčastější slovesa: *lišit, odlišovat ... od; osvobodit, oprostít ... od; izolovat, oddělit ... od; upustit, odstoupit ... od* aj. (9) a (10). Předložkové fráze s předložkou *od* také vyjadřují původ (zdroj), nejčastěji se slovesy *dostat, přijmout, získat, vyžadovat, očekávat ... od* (11). Původce děje (hloubkový subjekt, subjekt infinitivu) vyjadřují tyto předložkové fráze v obrazech *nechat* s infinitivem plnovýznamového slovesa (12).

Jen jedno sloveso má předmět s předložkou *bez*: *obejít se ... bez* (13).

- (9) Těsně před hlasováním se poslance pokusil **od jejich záměru/Obj** odradit senátor Martin Mejstřík
- (10) Názor ekonomů **se od názoru/Obj** odborníků na životní prostředí **liší** .
- (11) Proto nyní **očekávají od Američanů/Obj** lepší odpověď než tento nepřítel přátelský vzkaz
- (12) To jsem si ale nechal **rozmluvit od vězeňské psycholožky/Obj** .
- (13) Místní farmáři **se bez laciné pracovní síly/Obj** neobejdou ,

Přídavná jména, na nichž jsou závislé předměty v předložkovém genitivu, jsou až na výjimky (*vinný/vinen, odlišný*) odvozená od sloves (z přítomných, trpných a minulých příčestí) a zachovávají si vazbu původního slovesa. Nejčastější jsou adjektiva s předložkou *z*, předměty mají často význam původu/látky: *složený, vyrobený, vyplývající, podezřelý, obviněný ... z*; dále adjektiva s předložkou *od*: *odlišný, odvozený, oddělující, osvobozený ... od*; nejméně častá jsou přídavná jména, jež mají předmět s předložkou *do*: *zamilovaný, zasahující, zapletený ... do*.

- (14) Ten byl uznán **vinným z pokusu/Obj** vraždy a vyfasoval deset let .
- (15) Pavouci mají hlavohrud' **oddělenou od zadečku/Obj** stopkou a osm jednoduchých oček .
- (16) Byla jste hodně **zasvěcená do jeho záležitostí/Obj** ?

Příslovce jsou rozvíta předměty v předložkovém genitivu jen výjimečně, v 0,5 % těchto

předmětů. Většina je závislá na adverbium *daleko* (17), některá adverbia jsou také odvozena od (slovesných) valenčních adjektiv (18) a (19).

- (17) *Původ jezevčíka sahá nepochybně daleko do minulosti/Obj* ,
- (18) *vězeň byl dočasně umístěn odděleně od ostatních odsouzených/Obj* .
- (19) *některé vztahy upravují pouze rámcově , odlišně od znění/Obj zákona , či vůbec*

#### 2.6.4 Předmět v prostém dativu (*Obj\_3*)

Předměty realizované prostým dativem představují v korpusu cca 1,6 % všech substantiv (7,7 % předmětů). Zahrnují jednak valenční dativ, jednak zčásti tzv. volný dativ. Volný dativ (tedy dativ, který se volně pojí se slovesy, jež nemají obligatorní dativní valenci) se k předmětu řadí jednak v případě subjektového dativu (*člověku/Obj se chce zvracet*), jednak v případě fakultativního valenčního dativu (*bicykl prodal neznámému muži/Obj*). Naopak dativ prospěchový (*můžete ho natrhat mamince/Adv do vázy*), zřetelový (*Majiteli/Adv způsobil škodu 50 tisíc*) a přivlastňovací (*Anně/Adv se zavíraly oči*) se řadí k příslovečným určení. Automatické rozlišení mezi syntaktickými funkcemi je v těchto případech velmi nespolehlivé, dílčí statistické údaje v těchto kategoriích jsou jen přibližné, celkové údaje o frekvenci realizace předmětu prostým dativem byly však ověřeny na vzorcích z korpusu a jsou přibližně správné.

Slovesa s dativní valencí (s jediným dativním předmětem i s tzv. nepřímým, druhým dativním předmětem) mají mnoho různých významů, uvedeme jen několik typických příkladů. Mezi nejčastější významy sloves patří komunikace: *říci, slíbit, vyčítat, oznámit, odpovídat* (1); dávání, předávání: *dát, věnovat, předat, poskytnout, prodat* (2); podobnost a shoda: *podobat se, rovnat se, odpovídat* (3); příznivé i nepřiznivé (mezilidské) vztahy: *důvěřovat, vděčit, pomoci, závidět, křivdit* (4); odpor a obrana: *vyhnout se, zabránit, odolat, vzepřít se, předcházet* (5) a jiné.

Uvádíme seznam dvaceti nejčastějších sloves s dativní valencí (v závorce počet dativních předmětů závislých na tomto slovese na milion substantiv v korpusu): *říci* (442 ppm), *věnovat* (437 ppm), *dát* (380 ppm), *odpovídat* (237 ppm), *dávat* (204 ppm), *podat se* (183 ppm), *patřit* (173 ppm), *pomoci* (169 ppm), *říkat* (144 ppm), *vyhnout se* (140 ppm), *zabránit* (120 ppm), *podobat se* (114 ppm), *poskytnout* (107 ppm), *nabídnout* (104 ppm), *sloužit* (101 ppm), *předat* (101 ppm), *poskytovat* (97 ppm), *pomáhat* (95 ppm), *líbit se* (89 ppm), *bránit* (89 ppm).

- (1) **Cizincům/Obj** na to **odpovídám** , že se nescházíme , protože se nesnášíme .
- (2) Vzpomněl jsem si , jak jsem před lety **půjčil Pepovi/Obj** Věru Linhartovou Meziprůzkum nejbliž uplynulého .
- (3) Strop nemocenské **odpovídá platu/Obj** nad 18000 Kč .
- (4) **Závidím** pražským **studentům/Obj** , jak jsou povzneseni nad naše předsudky
- (5) **Čelíc** značným zásobovacím **problémům/Obj** , s malými zálohami , které ještě mohla povolát , nebyla německá armáda schopna zastavit spojenecké protiútoky v srpnu 1918 .

Subjektový dativ (který je obtížné automaticky identifikovat) se vyskytuje zřídka, především se slovesem *dostat se*, cca 40 ppm (6), a u výrazu *nechce se* s infinitivem, cca 10 ppm (7). Ostatní subjektové dativy vyjádřené substantivy jsou ještě vzácnější (8).

- (6) **Ženám/Obj se dostalo** tohoto práva ve Velké Británii až po první světové válce ,
- (7) **Otcům/Obj se nechce** platit na své děti .
- (8) Po spolknutí lžičky šťávy **se miminkům/Obj** lépe **spí** .

Dativní předměty závislé na přídavných jménech jsou mezi substantivy **Obj\_3** zastoupeny přibližně čtyřmi procenty (650 ppm mezi substantivy v korpusu). Přídavná jména s dativní valencí jsou z 55 % původní adjektiva: *blízký*, *milý*, *cizí* a adjektiva, jejichž původní vztah ke slovesům je synchronně zastřený: *škodlivý*, *srozumitelný*, *dostupný* (9), (10) a (11). Ze 45 % jsou to adjektiva slovesná, odvozená od přítomných, trpných a minulých příčestí: *podléhající*, *vyhrazený*, *uvyklý* (12), (13) a (14). Významy předmětů adjektiv víceméně odpovídají významům u sloves, přičemž například adjektivum *bezpečný* má dativní valenci jen jako negativní (zřejmě by tedy mělo být lemmatizováno jako *nebezpečný*).

Dvacet nejčastějších adjektiv se substantivními předměty v prostém dativu (s frekvencí v závorce u adjektiv častějších než 10 ppm v korpusu): *podobný* (75 ppm), *věnovaný* (42 ppm), *odpovídající* (32 ppm), *blízký* (29 ppm), *určený* (17 ppm), *přístupný* (17 ppm), *úměrný* (15 ppm), *věrný* (13 ppm), *známý* (13 ppm), *vlastní* (12 ppm), *rovný*, *patřící*, *sloužící*, *adresovaný*, *přítomný*, *vzdálený*, *vystavený*, *zasvěcený*, *podřízený*, *poskytovaný*.

- (9) Moře **podobné** temnému **přízraku/Obj** poplácávalo bezděky písek
- (10) A následoval páteční večer - **milý Bohu/Obj** i **lidem/Obj** .
- (11) Jeho zranění nejsou **životu/Obj nebezpečná** , ale zanechají následky v podobě jizev ,
- (12) vypadají jako odsouzení **nastavující** krk **katu/Obj** . . .

- (13) Panel **věnovaný otázkám/Obj** opatrnostní regulace ve vztahu k práci  
(14) kde se scházejí mladí lidé **propadlí drogám/Obj** .

Zřídka (cca 20 ppm) jsou předměty v prostém dativu závislé i na příslovcích, a to na slovech *blízko, úměrně a přiměřeně* (15), (16) a (17).

- (15) Ve fotbalové Lize mistrů se tak posunul o krok **blíže** svému vítěznému **snu/Obj** .  
(16) Později zvyšovaly noviny plat obou bratří **úměrně** stoupajícím **cenám/Obj** životních potřeb a též jejich stoupající **oblíbenosti/Obj** .  
(17) ale vždycky byl oblečený slušně a **přiměřeně místu/Obj** , kde se nacházel .

### 2.6.5 Předmět v předložkovém dativu (Obj\_p3)

Pro vyjádření předmětné funkce se předložkový dativ používá ve 3,1 % předmětů (0,6 % substantiv v korpusu). Hranici mezi předmětem a okolnostním určením (zvl. určením místa a účelu) není ani u tohoto předložkového pádu možné jasně vymezit, zvláště pak ne automaticky (rozhodnutí je často závislé na významu slov), dílčí frekvence v rámci této kategorie jsou tedy jen přibližné, zakládají se na korekci automaticky anotovaných dat s využitím manuálních sond do korpusu.

Problematiku rozlišování mezi syntaktickými funkcemi lze dobře ukázat na substantivech v předložkovém dativu závislých prostřednictvím předložky *k* na slovese *dojít*. Z 22 858 takových substantiv (bez ohledu na automaticky přiřazenou syntaktickou funkci) jich má přibližně 80 % význam hloubkového subjektu, události, která nastala či nastane; sloveso je ve třetí osobě singuláru, přičestí je středního rodu (1). Cca 10 % těchto předložkových frází má význam předmětu/cíle, jehož někdo dosahuje; nejčastější jsou zde obraty *dojít k ... závěru, názoru, řešení, rozhodnutí, poznání* (2). Cca 10 % připadá na skutečně místní určení, lokální cíl, kam se někdo přemísťuje (3). Takovou analýzu lze provést pro jedno sloveso, ne však pro stovky sloves, na nichž jsou substantiva v předložkovém dativu závislá (v korpusu je 400 různých sloves, na nichž jsou závislá substantiva v předložkovém dativu stokrát nebo vícekrát).

- (1) **Ke** druhému **incidentu/Obj** **došlo** při jeho cestě vlakem do Pretorie ,  
(2) *potřebovala jsem dospět k nějakému konci* , zatím jsem ještě **nedošla k** žádnému **řešení/Obj** .

- (3) **Došel** jsem **k** recepčnímu **pultu/Adv** a postavil se na konec krátké fronty .

Většina předmětů, 94,6 %, je řízena předložkou *k* (*odsoudit k smrti/Obj*); 5,1 % řídí předložka *proti*: (*odvolat se proti rozsudku/Obj*); a 0,3 % předložka *vůči* (*odolný vůči nízkým teplotám/Obj*). Většina předmětů v předložkovém dativu rozvíjí slovesa, jen 1,2 % rozvíjí adjektiva a 0,6 % adverbia.

S předložkou *k* se pojí slovesa vyjadřující cíl, podíl, vztah, událost, modalitu aj.: *patřit, náležet, připojit se, přispět ... k; dospět, přimět, nutit, odsoudit ... k; dojít, docházet, schylovat se ... k; svolit, hodit se, stačit ... k* a další (4), (5) a (6).

- (4) že **se schyluje k** občanské **válce/Obj** v Irsku .  
(5) Nakonec **dospěl k závěru/Obj** , že by měli odjet co nejdřív ,  
(6) V historii dluhových krizí **se k** tomuto radikálnímu **řešení/Obj** **uchýlilo** jen málo zemí ,

S předložkou *proti* se pojí slovesa s významem (široce pojatého) konfliktu nebo ochrany: *bojovat, protestovat, stavět se, postavit se, vystoupit, bránit, chránit, očkovat, pojistit ... proti* (7) a (8). Předložka *vůči* často nahrazuje předložky *k* a *proti* v jejich valenčních rámcích (*chovat se ... k/vůči; provinit se proti/vůči*), s některými slovesy se předložka *vůči* používá primárně: *uplatnit, uplatňovat, postupovat ... vůči* (9). Předmět s předložkou *vůči* plynule přechází k určení zřetele.

- (7) Co mohou **namítat proti Regině/Obj** ?  
(8) tak jako druhý syn jsem **se vzbouřil proti** svému **lidu/Obj** .  
(9) Žijeme v čase a nemůžeme zapomenout , že věčnost **se vůči času/Obj** **vymezuje** i negativně ;

Přídavná jména představují cca 1,2 % řídicích větných členů předmětů v předložkovém dativu. Ze 71 % jsou slovesná, odvozená od přítomných a trpných příčestí: *opravňující, vyzývající, odsouzený ... k*. Menší část, 29 % přídavných jmen je neodvozených nebo odvozených jiným způsobem: *potřebný, lhostejný ... k; odolný ... proti/vůči*. Nejčastější (85 %) jsou adjektiva s předložkou *k* a dativem: *vztahující, vedoucí, určený, nutný, směřující ... k* (10); méně častá (10 %) jsou adjektiva s předložkou *proti*: *odolný, namířený,*

zaměřený, bojující, protestující ... proti (11); nejméně často jsou adjektiva rozvíjena předměty s předložkou vůči: odolný, imunní, nepřátelský, citlivý, kritický ... vůči (12).

(10) Proč je tak citlivý , tak **náchylný k návalům/Obj** nevysvětlitelné lítosti ?

(11) Víš určitě , že je **očkována proti tetanu/Obj** ?

(12) A soudci **neloajální vůči státu/Obj** mohou jen těžko pomáhat s vytvářením loajality u svých spoluobčanů .

Adverbia jsou řídicími větnými členy předmětů v předložkovém dativu pouze v 0,6 % případech. Většinou je řídicím adverbium slovo *blízko* (13), zřídka jsou předmětem rozvíta i jiná adverbia (14) a (15).

(13) i přesto měli domácí hodně **blízko ke vstřelení/Obj** gólu .

(14) - **analogicky k** psychologicky pojaté **podobizně/Obj** mladého muže .

(15) přehněte na zkoušku všechny proužky **kolmo k ploše/Obj** kolečka .

#### 2.6.6 Předmět v prostém akuzativu (Obj\_4)

Syntaktická funkce předmět se prostým akuzativem substantiv vyjadřuje velmi často, v 62,3 % předmětů, tuto kombinaci funkce a pádu má 13,0 % substantiv v korpusu, je to tedy třetí nejčastější kombinace syntaktické funkce a pádu u substantiv. Řídicím větným členem je téměř vždy sloveso, jen u 0,6 % akuzativních předmětů je to adjektivum (příslovce není rozvíto prostým akuzativem nikdy).

Předměty v prostém akuzativu jsou primárně předměty zasažené dějem slovesa, i když intenzita zásahu je velmi variabilní (*zničit* vs. *pozorovat*). Sémantické třídění sloves s akuzativní valencí je značně problematické, možných sémantických kategorií by bylo mnoho, rozdělení byť jen frekventovaných sloves by bylo velmi obtížné: v korpusu je více než 3000 různých sloves s akuzativní valencí, které mají sto nebo více substantivních předmětů v akuzativu. Uvádíme jen několik příkladů frekventovaných sémantických okruhů. Mezi nejčastějšími slovesy s akuzativní valencí jsou například slovesa vyjadřující vlastnictví a jeho změny: *mít, dostat, dát, získat, koupit, ztratit, vzít* (1); slovesa vnímání a uvažování: *vidět, slyšet, číst, znát, chápat, představit si* (2); slovesa typická pro akuzativ, jež přímo zasahují předmět: *dělat, udělat, tvořit, vytvořit, založit* (3); slovesa vyjadřující význam a reprezentaci: *znamenat, představovat, reprezentovat, projevit, ukazovat* (4).

- (1) Vystudovala Právnickou fakultu Univerzity Karlovy , kde získala titul/Obj magistra v oblasti práv .
- (2) Nevím jen , jestli si budoucnost/Obj představovala takhle .
- (3) Pro skutečné apokryfy si vytvořme systém/Obj .
- (4) První kolo může taky představovat ztrátu/Obj pět až osm miliónů korun .

Na tranzitivních slovesech je obvykle závislý pouze jeden (nekoordinovaný) substantivní předmět v akuzativu. V jazykovém systému však dva akuzativní předměty rozvíjející jedno sloveso přípustné jsou a v korpusu se skutečně vyskytuje několik desítek takových případů: ve 40 případech (1,5 ppm) jsou předměty závislé na slovese učit nebo na slovese odvozeném od slovesa učit (5), (6) a (7), v cca 200 případech (7 ppm) jsou předměty závislé na slovese stát, jeden vyjadřuje hloubkový subjekt (osobu, která platí/ztrácí), druhý míru (8) a (9).

- (5) Děvčata/Obj učil jeden z bratrů benediktinů matematiku/Obj .
- (6) Tímto způsobem lze ptáka/Obj odnaučit opakovaný křik/Obj .
- (7) Naši generaci/Obj tyto věci/Obj nikdo neučil .
- (8) Jeden náboj do kalašnikova stojí palestinské teroristy/Obj tři dolary/Obj
- (9) Tehdejšího ředitele Pavla Holuba/Obj to stálo místo/Obj .

Přídavná jména jsou řídicími členy akuzativních předmětů jen relativně zřídka (0,6 % předmětů v prostém akuzativu, cca 800 ppm). Až na jednu výjimku: dlužný/dlužen (10) jsou to vždy slovesná adjektiva utvořená z přítomného přičestí (11), (12) a (13).

Nejčastější adjektiva jsou: připomínající (49 ppm), obsahující (37 ppm), umožňující (29 ppm), zajišťující (17 ppm), představující (16 ppm), podporující (14 ppm), spojující (14 ppm), zahrnující (12 ppm), tvořící (12 ppm), mající (12 ppm), ovlivňující (12 ppm), poskytující (10 ppm), vyžadující, upravující, nesoucí, nabízející, přesahující, využívající, vyrábějící, zachycující.

- (10) Jsem ti dlužen omluvu/Obj za středu .
- (11) Ta je pátým největším producentem v rámci Organizace zemí vyvážejících ropu/Obj .
- (12) Události , mající operativní charakter/Obj , zasílají správy StB na II . správu SNB .
- (13) elektronických přístrojů zvládajících život/Obj ohrožující poruchy/Obj srdečního rytmu

### 2.6.7 Předmět v předložkovém akuzativu (*Obj\_p4*)

Předložkovým akuzativem se funkce **Obj** realizuje v 8,0 %, podíl kombinace **Obj\_p4** na všech substantivech je 1,7 %. Předměty jsou obvykle závislé na slovesech, jen v 1,8 % jsou řízeny adjektivy. Substantivní předměty v předložkovém akuzativu se v korpusu vyskytují s předložkami: *na* (45,1 %), *o* (28,8 %), *za* (19,9 %), *v* (4,3 %), *pro* (1,8 %); výjimečně i s předložkami *přes* a *nad*.

Slovesa s valencí *na* s akuzativem lze rozdělit do mnoha sémantických podkategorií, z nich nejčastější jsou slovesa vnímání apod.: *dívat se, podívat se, pohlédnout, ukázat ... na* (1); slovesa vyjadřující mentální procesy: *myslet, zapomenout, vzpomenout si, pamatovat, soustředit se ... na* (2); slovesa s významem reakce: *reagovat, odpovědět ... na* (3); slovesa očekávání: *čekat, těšit se ... na* a mnoho dalších, např.: *chystat se, útočit, stěžovat si ... na*.

- (1) **Pohlédl na hodinky/Obj** : teprve poledne !
- (2) Moc rád **vzpomínám na dětství/Obj** a samozřejmě **Vánoce/Obj** a **dárky/Obj** .
- (3) **Na tyto změny/Obj** jsme **reagovali** velmi pružně ,

Mezi slovesy s valencí *o* s akuzativem jsou nejčastější slovesa s vyprázdněným obsahem, tematizující předmět: *jít, jednat se ... o* (4). Další časté významové skupiny jsou slovesa vyjadřující ztrátu předmětu: *přijít, připravit, obrat ... o* (5); slovesa zájmu a snahy: *požádat, usilovat, snažit se, zajímat se, bojovat ... o* (6); slovesa vyjadřující péči: *starat se, postarat se, pečovat ... o* a jiné, např. *bát se*.

- (4) **Jedná se o** velmi vzácný **případ/Obj** , trojčata jsem nezažil za posledních 40 let ,
- (5) Rasová diskriminace jej **připravila o možnost/Obj** studia na dramatické konzervatoři ,
- (6) Židovské náboženství bylo **monoteistické a usilovalo o očištění/Obj** od uctívání pohanských bohů .

Nejčastějšími významovými skupinami sloves, na nichž jsou v korpusu závislé předměty s akuzativem a předložkou *za*, jsou slovesa s jedním předmětem v prostém akuzativu a druhým předmětem, který vyjadřuje vlastnost připisovanou předmětu: *považovat, pokládat, označit, označovat, vydávat ... za* (7). Mezi dalšími častá slovesa patří směnné transakce (placení): *zaplatit, platit, vyměnit ... za* (8); slovesa vyjadřující vděčnost: *děkovat, vděčit, poděkovat ... za* (9); a další, např. *provdat, odpovídat, bojovat ... za*.

- (7) **Považuji** ale **za** svou **povinnost/Obj** ještě jednou vám připomenout , jakému riziku čelíte
- (8) Ovšem příliš to nevdává , protože **zaplaceno za** své **služby/Obj** stejně nedostanou .
- (9) ale já musím říci , že lékařům **vděčím za život/Obj** .

Slovesa s předložkou *v* s akuzativem nejčastěji vyjadřují proces (změnu), předmět je výsledkem tohoto procesu: *proměnit, změnit, přetvořit, vyústit, přerůst ... v* (10). Dále sem patří především slovesa víry a důvěry: *věřit, doufat, uvěřit ... v* (11); a některá sémanticky vyprázdňená slovesa, která se používají v ustálených obratech, kde substantivum považujeme za předmět: *brát, vzít, přijít ... v*, např. *brát v potaz, přicházet v úvahu* aj. (12).

- (10) nejprve ji **transformuje v energii/Obj** chemickou a teprve potom **v další formy/Obj** energie .
- (11) Tak vzniká církev - společenství všech , kdo **v Krista/Obj uvěří** a žijí jako jeho přátelé .
- (12) Je třeba **brát v potaz/Obj** , že současně může dojít k poklesu cen energií .

Předložkové fráze s předložkou *pro* mají obvykle význam zřetele, účelu či prospěchu, za valenční považujeme jen menší část sloves, např. *znamenat, rozhodnout se, udělat, pracovat, hlasovat ... pro* (13) a (14). Za sloveso valenční s předložkou *nad* a akuzativem považujeme jediné sloveso: *povznést se nad* (15).

- (13) Zatímco **pro zákazníky/Obj** to **znamená** nižší ceny , zelináři mají problémy .
- (14) Kdy jste se **pro salto/Obj** vůbec **rozhodl** ?
- (15) nedokážu se **povznést nad** jednotlivý **osud/Obj** , zvláště jde - li o osud můj

V korpusu se také vyskytují předložkové fráze s předložkami *přes* a *na* s významem přibližnosti, které jsou předměty tranzitivních sloves (16), (17) a (18). Případů, kdy je zde předmětem substantivum, je málo, obvykle je totiž předmětem číslovka. Mezi substantivy může být předmětem jen jiný kvantifikátor: *tuna, metrák; desítka, stovka* (16) a (17) nebo musí být substantivum kvantifikováno číslovkou *dva až čtyři* (18). Tyto příklady je obtížné vyhledat automaticky, celkem jsme v korpusu našli jen desítku příkladů.

- (16) Z dobře obaleného stromu může **získat** i **přes metrák/Obj** jablek ,

- (17) Během 14 dnů **zkontrolovali na stovku/Obj** provozoven na zpracování masa a mléka .  
 (18) vždyť hned sedm volejbalistů Bělgorodu **má přes dva metry/Obj** .

Adjektiva jsou řídicím větným členem předmětů v předložkovém akuzativu poměrně zřídka (1,8 % **Obj\_p4**). Přibližně polovina předmětů (asi 53 %) je závislých na slovesných adjektivech odvozených ze slovesných příčestí: *usilující, považovaný, zvyklý*. Neslovesná adjektiva jsou jen o málo méně frekventovaná (cca 47 %): *odpovědný, hrdý, vděčný*. Nejčastější adjektiva s valencí s předložkovým akuzativem jsou: *zaměřený, zvyklý, bohatý, citlivý, hrdý ... na* (19); *odpovědný, považovaný, vděčný ... za* (20); *usilující, pečující, bojující ... o* (21).

- (19) Oslík byl tak **pyšný na** svůj honácký **úspěch/Obj** , že se rozhodl , že poběží .  
 (20) Myslíte tu , co byla **vdaná za** toho **spisovatele/Obj** ?  
 (21) Roku 1419 s pomocí Zikmunda porazil Turky , **usilující o ovládnutí/Obj** pevnosti Severin .

Substantivní předmět v předložkovém akuzativu závislý na příslovci nebyl v korpusu nalezen. Spojení *vzhledem na* (cca 5 ppm) je spíše nepravá předložka, předložková fráze u příslovce *nezvykle* je určením zřetele (22), na příslovcích odvozených z valenčních adjektiv *hrdě* a *pyšně* jsou v korpusu závislá jen demonstrativa uvozující vedlejší větu (23) a (24).

- (22) , působí tento " drobeček " **na** evropské **poměry/Adv** přinejmenším **nezvykle** .  
 (23) Lidé tu žijí a vypadají spokojeně a trochu **hrdě na to/Obj** , co dokázali .  
 (24) Vypadal tak **pyšně na to/Obj** , že ho něco takového napadlo ,

### 2.6.8 Předmět v předložkovém lokálu (**Obj\_p6**)

Předmět se realizuje předložkovým lokálem ve 3,2 %, tj. 0,7 % všech substantiv v korpusu jsou předměty realizované předložkovým lokálem. Řídicím větným členem jsou z 95,8 % slovesa, ze 3,8 % adjektiva a výjimečně, v 0,4 %, i příslovce. Nejčastější předložkou u předmětů v předložkovém lokálu je *o* (60,2 %), dále předložka *na* (16,6 %), předložka *v* (11,8 %), nejméně častá je u předmětů předložka *po* (11,4 %).

Automaticky rozlišit mezi předmětem a okolnostním určením je i u těchto předložkových frází velmi obtížné, jak dokládají příklady (1) až (6). Můžeme sice vycházet z lemmat

valenčních sloves (*pokračovat v, pracovat na, medítovat o*) a u frekventovaných adverbialních sousloví se naopak opřít o tvary v předložkových frázích (v *Itálii, na principu, o Vánocích, o samotě*), ale ani v jednom směru není určení spolehlivé. Například substantiva v lokálu s předložkou *na*, které rozvíjejí sloveso *pracovat*, jsou přibližně ze dvou třetin předměty, z jedné třetiny okolnostní určení, většinou místní (2), (3) a (4).

- (1) **V Itálii/Adv a v Salcburku/Adv** pak Mozart dále **pokračuje v** konvenční symfonické **produkci/Obj**
- (2) *stravování dělníků , kteří **pracovali na kostele/Obj** sv. Antonína ,*
- (3) *přes den **pracoval na stavbě/Adv** , v rukavičkárně nebo papírně i jako podomní prodavač vysavačů .*
- (4) *Přístroj **pracuje na principu/Adv** rychlého a přesného vývinu tepla magnetickou indukci .*
- (5) *s **předsedou Poslanecké sněmovny Václavem Klausem jsme meditovali o Vánocích/Obj** .*

Substantivní předměty v lokálu s předložkou *o* mají užší význam než většina jiných skupin předmětů, vyjadřují (široce pojaté) téma či obsah, s nímž zachází valenční sloveso. V korpusu je 70 různých sloves s valencí s předložkou *o* a s lokálem, na nichž je závislých sto nebo více substantivních předmětů. Mezi nejčastější patří slovesa komunikace (slovesa dicendi): *mluvit, hovořit, informovat, vyprávět, psát, říci, zmínit se, říkat, bavit se, diskutovat ... o* (6), slovesa rozhodování a jednání: *rozhodnout, rozhodovat, jednat, vyjednávat, hlasovat ... o* (7) a slovesa vyjadřující mentální procesy (úvahy): *vědět, uvažovat, dozvědět se, přemýšlet, pochybovat, snít ... o* (8).

- (6) **O obtížné cestě/Obj** Řecka do ES a posléze do EMU **promluvil** Isaac Sabethai ,
- (7) *Dohlédací úřad má rovněž **rozhodnout o osudu/Obj** čtyř státních obchodních bank ,*
- (8) *Neměl ani představivost k tomu , aby **zapřemýšlel o nových metodách/Obj** prodeje .*

Předměty s předložkou *na* a s lokálem závisí nejčastěji na slovesech, která vyjadřují závislost: *záviset, záležet ... na* (9); podíl: *podílet se, účastnit se, spolupracovat ... na* (10); dohodu: *shodnout se, dohodnout se, usnést se ... na* (11) a jiné, z častějších například *pracovat, trvat* či *pochutnat si ... na*.

- (9) *Na pánvi opékejte o něco déle , doba opékání závisí na velikosti/Obj rajčat .*
- (10) *Na vzniku/Obj encefalopatie se podílí hypoxie , hypoglykémie , krvácení do CNS a i přímý účinek pertusového toxinu .*
- (11) *Na základě některých významných argumentů jsme se shodli na společném postupu/Obj .*

Slovesa, na nichž jsou závislé substantivní předměty v lokálu s předložkou *v*, jsou z hlediska významu nesourodá. K nejčastějším patří: *pokračovat, bránit, zabránit, vyznat se, angažovat se, tkvít, spočívat, libovat si ... v (12), (13) a (14).*

- (12) *Jeho nástupce Ludvík XII . Orleánský v této válce/Obj pokračoval ,*
- (13) *Vy se v ženách/Obj moc nevyznáte , vidíte ?*
- (14) *Problém tkví v časovém postupu/Obj .*

Předměty v předložkovém lokálu s předložkou *po* nejčastěji vyjadřují hledání (často vizuální): *pátrat, rozhlédnout se, rozhlížet se, podívat se, ptát se ... po (15)* a touhu či nedostatek: *toužit, volat, stýskat se, lapat, prahnout ... po (16)*. Mezi ostatními jsou nejčastější: *zdědit, převzít, chtít, zbýt ... po (17)*.

- (15) *Zastesklo se mu po parním mandlu/Obj , po liščatech/Obj , po domově/Obj .*
- (16) *Lernere , podívej se po nějakých elektrických svítilnách/Obj .*
- (17) *Po zastupitelích/Obj proto chtěl , aby na odpovědné pracovníky podali trestní oznámení .*

Přídavná jména, na nichž jsou závislé předměty v předložkovém lokálu, jsou převážně slovesná, odvozená od přítomných, trpných a minulých příčestí sloves: *toužící, přesvědčený, závislý*. I ostatní valenční přídavná jména jsou původem odvozená od sloves, ale tento vztah je synchronně méně zřetelný (*shodný, úspěšný*). Samotné adjektivum *závislý* je řídicím větným členem pro celých 62 % předmětů v lokálu závislých na adjektivech (18). Mezi další častá valenční adjektiva s lokálem patří: *pojdnávající, svědčící, přesvědčený ... o (19); toužící, pojmenovaný, následující ... po (20); spočívající, úspěšný, shodný ... v; založený, podílející se, spáchaný ... na.*

- (18) *Toto řešení je samozřejmě závislé na dohodě/Obj s nájemníky ,*

- (19) *Tehdy mu totiž car - naprosto nepochybující o jeho loajalitě/Obj - oznámil ,*  
(20) *Z Prahy často zaznívá imperativ volající po jednotě/Obj díla a života tvůrce , po stvrzení/Obj díla životem ,*

V korpusu jsme našli pouze jedno příslovce rozvítené předložkovým lokálem: (ne)závisle: 24 výskytů na milion substantiv v korpusu, až na zcela zanedbatelné výjimky *nezávisle* na (21) a (22). Pozitivní tvar příslovce má substantivní předmět v lokálu pouze v 10 výskytech v korpusu, tj. 0,35 ppm (23).

- (21) *Tyto složky deformace vznikají nezávisle na působícím napětí/Obj .*  
(22) *Jinou debatou je ale budoucnost světa existujícího nezávisle na člověku/Obj .*  
(23) *Oběti začínají vypovídat závisle na agresorech/Obj , zcela proti pravdě a svému zájmu .*

### 2.6.9 Předmět v prostém instrumentálu (Obj\_7)

Podíl prostého instrumentálu na substantivních předmětech v korpusu je 5,3 %, kombinace funkce a pádu se na všech substantivech podílí 1,1 %. Předmět v instrumentálu zahrnuje kromě substantiv závislých na valenčních slovesech i původce děje u opisného pasiva a adjektiv odvozených od trpného přičestí, celkem 39 % předmětů v prostém instrumentálu. Valenční předměty sloves představují přibližně 60 % ze všech předmětů v prostém instrumentálu, původce děje u trpného rodu sloves asi 25 %, původce děje u slovesných adjektiv (-ný/tý/lý) cca 14 %, valenční předměty adjektiv asi 1,6 %.

Hranice mezi předmětem v prostém instrumentálu a určením prostředku, způsobu nebo místa je velmi nejasná. V případě posunu významu od konkrétního, např. místního určení, k obecnějšímu významu u slovesa, na němž jsou často závislá substantiva v instrumentálu, volíme funkci **Obj**, zůstává-li okolnostní určení zřetelné, volíme funkci **Adv** (1) a (2). Automatické rozlišení mezi funkcemi **Adv** a **Obj** je nespolehlivé, lze se sice opřít o časté adverbialní instrumentály jako *způsobem*, *začátkem* či *septem*, ale jinak je pro rozlišení nutné porozumění významu věty. Určení dílčích frekvencí je jen přibližné, s využitím manuálních sond v korpusu.

- (1) *a jikry musejí projít stádiem/Obj částečného vyschnutí substrátu .*  
(2) *aby prošli o několik měsíců později další krvavou lázní/Adv při ofenzivě na Chemin des Dames .*

Valenční slovesa s prostým instrumentálem jsou syntakticky i sémanticky pestrá, uvádíme jen příklady nejvíce zastoupených skupin. Řadíme sem slovesa sponového charakteru, u nichž má předmět význam vlastnosti: *stát se, stávat se, zůstat, zůstávat* (3). Patří sem slovesa fázová: *začít, začínat, končit, počínat*, (4); slovesa pohybu, zvláště pohybu částmi těla: *zavrtět, pokrčit, mávnout, pohybovat, pohnout, potřást* (5); slovesa s významem manipulace s předmětem: *zabývat se, disponovat, obchodovat* (6); nebo slovesa s významem hojnosti: *naplnit, plnit, oplývat, vyznačovat se* (7). Dvacet nejčastějších sloves s instrumentálovou valencí (v závorce počet substantivních předmětů v instrumentálu na milión substantiv v korpusu): *stát se* (1418 ppm), *zabývat se* (361 ppm), *stávat se* (223 ppm), *trpět* (147 ppm), *zůstávat* (130 ppm), *zavrtět* (123 ppm), *skončit* (101 ppm), *pokrčit* (99 ppm), *zůstat* (90 ppm), *začít* (88 ppm), *projít* (84 ppm), *vyznačovat se* (80 ppm), *nazývat* (73 ppm), *počínat* (70 ppm), *končit* (70 ppm), *nazvat* (59 ppm), *živit se* (58 ppm), *začínat* (58 ppm), *naplnit* (55 ppm), *disponovat* (54 ppm).

- (3) *Navíc ani asijská krize nezůstala izolovaným případem/Obj.*
- (4) *Začínáváme tangem/Obj a končíme kvapíkem/Obj.*
- (5) *Vrtíš nechápavě hlavou/Obj.*
- (6) *Zabývejme se odpovědností/Obj vedení.*
- (7) *hlavně proto, že neoplývá dostatečnou láskou/Obj, tolerancí/Obj a pokorou/Obj, zvláště tou k Václavu Klausovi a ODS.*

V korpusu je cca 2300 ppm substantiv v prostém instrumentálu, které vyjadřují původce děje u opisného trpného rodu sloves; zde tuto syntaktickou funkci řadíme mezi předměty. Původce děje lze snadno rozlišit u podstatných jmen označujících osoby (8). U neživých objektů nelze automaticky určit, zda substantivum v instrumentálu vyjadřuje původce děje nebo nástroj, někdy je toto rozlišení problematické i s porozuměním významu věty – v příkladu (9) lze ve spojení *chráněn zákonem* považovat slovo *zákonem* za hloubkový subjekt (a tedy předmět), nebo za nástroj. Ve většině případů (10) ale funkci manuálně určit lze, z manuálních sond v korpusu vyplývá, že poměr mezi kategoriemi původce děje a nástroje u pasiva je přibližně 3 : 1.

- (8) *Celé bývalé vedení banky bylo svého času jedním vyšetřovatelem/Obj obviněno, ten ale pak od policie odešel.*
- (9) *Ačkoli je vlk chráněn zákonem/Obj, bývá ročně pytláky či vesničany uloveno asi 20*

kusů .

(10) Nejčastější typ průduchu **je tvořen** svěracími **buňkami/Obj** ledvinovitého tvaru

Přídavná jména, na nichž jsou závislé předměty v prostém instrumentálu, jsou z velké části (cca 90 %, 1550 ppm) odvozená od trpných příčestí, řidčeji minulých příčestí, přičemž předmět vyjadřuje původce děje. Nejčastější adjektiva odvozená od trpných příčestí s původcem děje v instrumentálu jsou: *způsobený, stanovený, vedený, tvořený, vyvolaný* (11) a (12). Od minulých příčestí jsou odvozená adjektiva: *porostlý, posedlý, nasáklý, zarostlý* (13). Většina ostatních adjektiv je také slovesných, ovšem s valenčním instrumentálem: *zabývající se, trpící, vyznačující se, vonící, naplněný, proslulý* (14); malá část adjektiv se substantivním předmětem v instrumentálu není (alespoň synchronně) odvozena od sloves: *známý, jistý, podobný, charakteristický, pověstný* (15).

(11) *protože se přece jen trochu styděl za své předešlé excesy , **vyvolané touhou/Obj** po poznání .*

(12) *A navštěvovati nemocné je i **Písmem/Obj daná** křesťanská ctnost .*

(13) *Většina jeho postav je **posedlá mocí/Obj** či **penězi/Obj** a usiluje o ně manipulací druhých .*

(14) *ve svém díle , **vyznačujícím se elegancí/Obj** kresby a citlivou **představivostí/Obj** , však mířila k dřevorytu a ilustracím .*

(15) *Ani tím **ředitelem/Obj si** nejsem **jist** .*

#### 2.6.10 Předmět v předložkovém instrumentálu (Obj\_p7)

Mezi předměty vyjádřenými substantivy má předložkový instrumentál podíl 2,7 %, mezi všemi substantivy má tato kombinace 0,6 %. Řídícím větným členem jsou obvykle slovesa (87,9 %), méně často adjektiva (11,6 %), adverbia jen ojedinele (0,5 %). Většina předmětů v předložkovém instrumentálu, celkem 88,6 %, je s předložkou *s*, všechny výskyty předmětů s ostatními předložkami dohromady jsou osmkrát méně časté. V korpusu jsou substantivní předměty s předložkami: *před* 5,1 %, *nad* 2,2 %, *mezi* 2,0 %, *za* 1,6 % a *pod* 0,5 %.

Předměty v instrumentálu s předložkou *s* závislé na slovesech jsou v mnoha případech obtížně odlišitelné od „sociativního“ příslovečného určení, tj. určení provázející osoby či okolnosti (1), (2) a (3). S využitím manuálních sond do korpusu byly odhadnuty poměry

okolnostních určení a předmětů u nejméně frekventovaných slov a podle toho byly upraveny údaje. Substantivní předměty s předložkou s nejčastěji rozvíjejí slovesa komunikace: *mluvit, jednat, hovořit ... s* (4); slovesa vyjadřující vztahy mezi osobami: *seznámit (se), spolupracovat, setkat se ... s* (5) i jiné vztahy: *souviset, spojit, spojovat ... s* (6). Řídící slovesa také často vyjadřují zacházení s předmětem: *počítat, pracovat, zacházet ... s* (7). S předložkou s a instrumentálem mají předměty i fázová slovesa: *začít, začínat, skončit ... s* (8).

- (1) České sklo **nepřišlo s** novými **konceptemi/Obj** , ty zůstaly vyhrazeny Lotrinčanu Emilu Gallému
- (2) Ten muž , co **přišel s** vrchním inspektorem **Tavernerem/Adv** , je taky detektiv , že ?
- (3) Zní to hezky , ale možná pátrači **přijdou s křížkem/Adv** po funuse .
- (4) Ovšem ten , kdo ho zažil **bavit se s přáteli/Obj** , velice rychle pozná , že žádný morous není .
- (5) **Seznámil** jsem tě **s** tvou **ženou/Obj** .
- (6) Může **souviset s** trestnou **činností/Obj** a vylovenými **mrtvolami/Obj** .
- (7) Je tomu tak i tam , kde **pracuje s fantazií/Obj** .
- (8) Na místo , kde vloni **začali se stavbou/Obj** , se už nevrátili .

Substantivní předměty s předložkou před a instrumentálem rozvíjejí nejčastěji valenční slovesa: *chránit, zachránit, varovat, bránit, utéci, prchat, stát ... před* (9) a (10); předměty s předložkou nad rozvíjejí slovesa: *zvíťezit, vyhrát, zamyslet se, žasnout, převažovat, plakat, bdít ... nad* (11) a (12).

- (9) V prvním týdnu po přesazení ji **chráníme před** silným **sluncem/Obj** .
- (10) Unie přitom v dané oblasti **stojí před řešením/Obj** dalších problémů ,
- (11) **Užasli nad** tou **krásou/Obj** a vynutili si exkurzi .
- (12) Jsou lidé , kteří ještě dnes dokážou **plakat nad osudem/Obj** Anny Kareninové .

Předměty v předložkovém instrumentálu s předložkami mezi, za a pod rozvíjejí nejčastěji tato slovesa: *rozlišovat, vybrat, volit, rozhodovat, patřit ... mezi* (13) a (14); *zaostávat, zaostat, honit, stát ... za* (15) a (16); *představit si, představovat si, rozumět, chápat ... pod* (17) a (18).

- (13) Kareta obrovská je želva , patří mezi plazy/Obj žijící v teplých mořích .
- (14) Proto bych mezi dvěma zly/Obj volil raději Rusy .
- (15) Celkově však Uhry za vyspělejšími zeměmi/Obj Evropy viditelně zaostávaly .
- (16) Někteří experti míní , že za útokem/Obj stojí ruská nebo arabská mafie .
- (17) Pod pojmem/Obj větev se rozumí propojení uzlů větvemi .
- (18) Co si však lze pod takovým označením/Obj představit ?

Přídavná jména jsou předměty v předložkovém instrumentálu rozvíta jen zřídka, řídicími adjektivy jsou v cca 75 % adjektiva slovesná (související, spojený, užaslý), ve 25 % jiná adjektiva (shodný, známý, bezpečný). Adjektiva mají nejčastěji valenci s předložkou s: spojený, spokojený, srovnatelný ... s (19); řidčeji s předložkami pod a před: známý pod, chráněný, chránící, prchající ... před (20) a (21); výjimečně s dalšími předložkami, např. bdící nad, váhající mezi, zaostávající za.

- (19) Právě Súdán je s Rwandou/Obj srovnatelný , přesněji řečeno srovnatelný s výchozí situací/Obj .
- (20) Tisíciletá období chladu střídaná teplejšími mezičasy jsou většinou známa pod lidovým označením/Obj ledové doby .
- (21) Závěj se ukládá v oblasti za překážkou chráněné před větrem/Obj .

Příslovce jsou v korpusu rozvíta předmětem v předložkovém instrumentále jen ojedinele. Závislé předměty jsou v instrumentálu s předložkou s, příslovce jsou zčásti utvořené z valenčních adjektiv: shodně, srovnatelně ... s (22). Mezi další příslovce patří např. zároveň nebo pryč (23) a (24).

- (22) Producentem filmu je Vratislav Šlajer , který shodně s oběma autory/Obj tvrdí , že realizaci nechtějí uspěchat ,
- (23) Zároveň s dohodou/Obj se snad blíží také konec směňování lékařů
- (24) kde skandovali " Pryč s komunisty/Obj " ,

## 2.7 Přísllovečné určení (Adv)

Přísllovečné určení je větný člen, který rozvíjí slovesa, přídavná jména a příslovce (popř. předložkové fráze). Tvar substantiva (jeho pád, případně předložka) s funkcí Adv neurčuje

vazba jeho řídicího větného členu, ale především význam okolnostního určení. Přísllovečné určení je většinou realizováno předložkovým pádem substantiva (86,8 %). Mezi přísllovečná určení se řadí také nevazebný dativ, především dativ přivlastňovací a zřetelový (1) a předložkové fráze závislé na sponě (2).

V PDT se přísllovečné určení dále sémanticky netřídí, i když v anotačním manuálu se sémantická klasifikace používá, umožňuje totiž přehlednější formální rozdělení přísllovečných určení a uvádění příkladů. Ani v automaticky syntakticky anotovaném korpusu se přísllovečná určení sémanticky netřídí, základní sémantické klasifikace využijeme pouze pro exemplifikaci, když budeme v jednotlivých odstavcích představovat realizace přísllovečného určení jednotlivými pády substantiv. Pracujeme se základním určením **místa**, **času**, **způsobu** (včetně míry, nástroje a zřetele) a **kauzality** (příčiny, důsledku, účelu aj.), toto hrubé rozdělení přebíráme ze základního rozdělení (Šmilauer 1966, s. 257), s drobnými rozdíly danými především odlišnou interpretací v PDT. Hranice mezi jednotlivými podtypy jsou leckdy nejasné, zvláště u abstraktních substantiv.

(1) **Mužům/Adv** je asi představa tábora , ve kterém jsou týrány ženy , obzvláště nesnesitelná .

(2) Jediným zbožím , které **nebylo na lístky/Adv** , byla zábava .

### 2.7.1 Přísllovečné určení v prostém nominativu (*Adv\_1*)

Substantiva v nominativu se v automaticky syntakticky anotovaném korpusu považují za přísllovečné určení pouze tehdy, když jsou závislá na slovesech *jmenovat se*, *pojmenovat*, popř. na adjektivu *pojmenovaný* (1) a (2). Celkový počet výskytů této kombinace je cca 120 ppm. Interpretace se zde liší od pojmenování v nominativu závislých na slovesech *nazvat*, *nazývat* (*Obj*), příčinou rozdílu je ale pravděpodobně chyba v manuálním značkování, ne záměrné rozdělení podobného jevu do dvou různých kategorií. V celkových statistických údajích je frekvence jak předmětu v nominativu, tak přísllovečného určení v nominativu zanedbatelná.

(1) A zatím poslední příspěvek z tohoto soudku **se jmenuje Zrcadlo/Adv** a vychází opět v *MF DNES* ;

(2) strojeným hlasem uvádí adaptaci dosud nikdy nerealizovaného scénáře Václava Havla **pojmenovaného Osudná noc/Adv** .

### 2.7.2 Přísllovečné určení v prostém genitivu (Adv\_2)

Substantiva v genitivu jsou ve funkci **Adv** poměrně zřídka (0,9 % přísllovečných určení, 0,2 % všech substantiv). Tato přísllovečná určení až na drobné výjimky určují čas. Nejčastější v této funkci jsou tvary *dne* (cca 450 ppm) a *roku* (cca 350 ppm). Substantiva jsou vždy nějak rozvita: předcházejícím shodným přívlastkem: *toho dne, jednoho krásného dne, každého třetího roku* aj. (1) a (2); nebo následujícím neshodným přívlastkem: *dne 4. září 2002, roku 550 před Kristem* (3) a (4). Celkem asi 1000 ppm mají v korpusu genitivní tvary měsíců v roce s předcházející číslovkou: *1. září, posledního května* (5) a (6). Ostatní časová určení mají výrazně nižší frekvenci, např. *oné noci, příštího večera, léta Páně* apod.

- (1) *co jeho nejmladší syn Jan zvedl onoho památného dne/Adv na Staroměstské radnici v Praze meč*
- (2) *Toho roku/Adv byla prý nejchladnější zima v polských dějinách .*
- (3) *založená králem Karlem IV. dne/Adv 7 . dubna 1348 v Praze .*
- (4) *Kachličky z Kantonu , zhruba třicet na třicet , které roku/Adv 1100 po Kr. dovezl Ezechiel Rabi , pokrývaly podlahy , stěny a strop té malé synagogy .*
- (5) *Zloděje , kteří sedmého března/Adv ukradli v Roudnici nad Labem obraz Máří Magdaleny od Jana Zrzavého , policie již dopadla .*
- (6) *Thérésa , převezena 7 . thermidoru/Adv do věznice Justičního paláce , pochopila , že se octla před revolučním tribunálem .*

Kromě časového určení se prostého genitivu používá i pro vyjádření srovnávaného předmětu u komparativu adjektiv: *starší padesáti let/Adv*. V korpusu je cca 30 ppm takových obrátů, především u komparativů *starší* a *mladší* (7), výjimečně u jiných (8).

- (7) *takže prosím , aby osoby mladší osmnácti let/Adv následující tři odstavce raději vynechaly .*
- (8) *z městečka samotného je ke spatření dvacet tři štítů vyšších 5000 metrů/Adv .*

### 2.7.3 Přísllovečné určení v předložkovém genitivu (Adv\_p2)

Přísllovečná určení realizovaná předložkovým genitivem představují 22,9 % všech

přísllovečných určení (4,6 % všech substantiv v korpusu). Předložkový genitiv je druhý nejčastější pád mezi substantivy ve funkci **Adv**. Předložkové fráze ve funkci **Adv** jsou řízeny více než třiceti různými primárními předložkami a dalšími asi padesáti nevlastními předložkami.

Nejčastější genitivní předložky, které řídí předložkové fráze ve funkci **Adv**, jsou (v závorce je procento zastoupení předložky v rámci **Adv\_p2**): *do* (35,7 %), *z* (22,9 %), *od* (10,0 %), *u* (7,6 %), *podle* (6,8 %), *bez*, *za*, *během*, *kolem*, *kromě*, *vedle*, *místo*, *pomocí*, *prostřednictvím*, *uprostřed*, *včetně*, *podél*, *dle*, *okolo*, *uvnitř*.

Za nevlastní předložky se považují jednak víceslovné předložkové obraty, jednak slova, která jsou dosud morfologicky značkována jako substantiva, ale mají ustálenou předložkovou funkci (v PDT mají syntaktickou funkci **AuxP**). Mezi nejčastější nevlastní předložky s genitivem patří: *v rámci*, *na základě*, *v případě*, *v průběhu*, *koncem*, *z hlediska*, *v důsledku*, *počátkem*, *pod vlivem*, *začátkem*.

Substantiva v předložkovém genitivu ve funkci **Adv** jsou nejčastěji určením místa (1), (2), (3) a času (4), (5), (6).

- (1) *Mladšího syna prodá do otroctví/Adv nebo na vojnu a peníze přidá k rodinnému jmění .*
- (2) *Když se Svoboda vrátil , Kočka seděl u stolu/Adv , hlavu položenou na složených pažích ,*
- (3) *Potom obdaroval maminku velkým karafiátem , který vytáhl zpoza ucha/Adv .*
- (4) *jak se případně zlepšil stav popsanych věcí od loňského října/Adv ,*
- (5) *že během tří let/Adv toto číslo přesáhne 25 milionů .*
- (6) *Za vlády/Adv Napoleona byl P. dočasně připojen k Francii ,*

Přísllovečná určení způsobu (7) a (8) vyjádřená předložkovým genitivem substantiv jsou méně častá než určení místa a času, určení kauzality jsou mezi **Adv** v předložkovém genitivu zastoupena nejméně (9) a (10).

- (7) *Bach vstřelil , přetvořil a do poslední kapky/Adv využil všeho , co bylo v této formě v minulosti napsáno .*
- (8) *Podle jejího názoru/Adv byly záruky , kterých se Malé dohodě dostalo od Francie a dalších signatářů paktu čtyř , dostatečným důvodem*
- (9) *Z dosud neznámé příčiny/Adv se vzňal odpadkový koš , od něho obložení kuchyně*

(10) *Má se brzy **dostavit za účelem sepsání/Adv** protokolu .*

Nejčastější spojení předložky a substantiva v genitivu ve funkci **Adv** (substantiva jsou většinou dále rozvíta přivlastky) jsou: *od roku* (320 ppm), *do konce* (280 ppm), *od doby* (250 ppm), *do míry* (200 ppm), *z důvodů* (200 ppm).

#### 2.7.4 Přísluvečné určení v prostém dativu (**Adv\_3**)

Prostý dativ substantiv se pro vyjádření tradičně pojatého přísluvečného určení nepoužívá, funkci **Adv** mají v PDT pouze volné, nevazební dativy: dativ zřetelový (1) a (2), přivlastňovací (3) a (4) a prospěchový (5) a (6). Automaticky lze nevazební dativ od vazebního rozlišit jen velmi obtížně, údaje o frekvenci jevu se zakládají na odhadu s využitím manuálních sond do korpusu. Podle tohoto odhadu je celkový počet výskytů **Adv** v prostém dativu cca 250 ppm, což je frekvence zanedbatelná jak v rámci všech substantiv, tak v rámci substantiv ve funkci **Adv** (0,1 %).

- (1) *Pobertové **způsobili majitelům/Adv** škodu za 275000 korun .*
- (2) ***Kapele/Adv** to **bylo** buřt , a pokud ne , nedávali to na sobě znát .*
- (3) *Den před Silvestrem se jeho **snaše/Adv** **narodila** vnučka Petra .*
- (4) *Zpitý jak Dán **rozbil policistům/Adv** židle , stůl , okna a jednomu dokonce i ústa .*
- (5) *Brzy vstávat , **najít dětem/Adv** nové kamarády a přistavět barák*
- (6) *Páteční dohoda s Evropskou unií **uvolnila Číně/Adv** cestu na světový trh .*

#### 2.7.5 Přísluvečné určení v předložkovém dativu (**Adv\_p3**)

Substantivem v předložkovém dativu se realizuje 5,0 % substantivních přísluvečných určení (1,0 % všech substantiv). Nejčastější předložkou u těchto **Adv** je předložka *k* (74,1 %); další předložky jsou: *proti* (8,1 %), *kvůli* (7,3 %), *díky* (5,6 %), *vůči* (1,7 %), *navzdory*, *oproti*, *vstříc*, *naproti*, *dík*, *vzdor*, *napospas*.

Přísluvečná určení v prostém dativu nejčastěji vyjadřují místo (1), (2) a (3) a kauzalitu: účel, podmínku, přípustku aj. (4), (5) a (6).

- (1) *Kdybyste je **posadili k televizi/Adv** , at' se dívají na pohádku*
- (2) *Budova zastupitelství **leží** přímo **naproti** obří **mešitě/Adv** ,*

- (3) V kalném listopadovém svítání **jedeme** po dálnici **vstříc** rudému **slunci/Adv** .
- (4) když zrudlý moula **předváděl k obveselení/Adv** dvora improvizovaný taneček .
- (5) Byla to tichá skupina , unavená a **chladná navzdory ohni/Adv** v krbu ,
- (6) **Díky laskavosti/Adv** nakladatelů stále **cestuji** ,

Méně častá jsou příslovečná určení způsobu (míry, zřetele aj.) v předložkovém dativu (7) a (8), určení času se objevují pouze s předložkou *k* (9) a (10), a to poměrně zřídka.

- (7) Jeho už tak dost velký nos se **proti** normálnímu **stavu/Adv** **zvětšil** přinejmenším dvakrát
- (8) Budu **mluvit** stručně a **k věci/Adv** .
- (9) Dohoda o tom byla **uzavřena** s Rumunskem již **k lednu/Adv** 1935 .
- (10) Když **k ránu/Adv** **přišla** domů a procházela zahradou , necítila žádný žal .

Nejčastější spojení předložky a substantiva v dativu ve funkci **Adv** jsou: *k dispozici* (260 ppm), *ke dveřím* (85 ppm), *ke konci* (75 ppm), *ke stolu* (70 ppm), *k zemi* (65 ppm).

### 2.7.6 Přííslovečné určení v prostém akuzativu (Adv\_4)

Prostým akuzativem substantiv se funkce **Adv** realizuje v 3,3 % (0,7 % všech substantiv). Prostý akuzativ se používá především pro určení času, kromě toho také určuje míru u rozměrů či vzdáleností. V souslovích s opakujícím se substantivem a předložkou (*noc za noci*) také určuje čas nebo způsob.

Akuzativ jako příslovečné určení nejčastěji, ve více než 90 %, vyjadřuje časové určení: umístění v čase (*kdy?*), délku trvání (*jak dlouho?*) a frekvenci (*jak často?*). Umístění v čase i frekvence jsou vždy vyjádřeny substantivem s významem času, které je rozvito přívlastkem, většinou shodným: *onen den, hodinu po poledni, každou neděli* (1). Určení délky trvání nemusí být rozvito vždy: *hodinu, rok*; většinou je však také rozvito: *pouhou minutu, dvě hodiny, celou cestu* (2). Délka trvání může být také vyjádřena spojením kvantifikátoru a časového údaje, v případě substantiv má funkci **Adv** substantivní kvantifikátor: *řadu týdnů, spoustu let* (3).

Nejčastější substantiva v prostém akuzativu ve funkci určení času jsou (v pořadí podle frekvence): *den, týden, doba, rok, noc, večer, chvíle, měsíc, ráno, hodina, odpoledne, okamžik, dopoledne, chvilka, minuta*. Mezi nejčastějšími časovými údaji v akuzativu jsou

také dny v týdnu, nejčastější (jen s malým rozdílem) jsou: *pátek, sobota, neděle*; následují (také podle frekvence): *čtvrtek, pondělí, úterý, středa*.

Substantiva jsou nejčastěji rozvíta shodnými přívlastky (také podle frekvence): *každý, celý, ten, druhý, příští, minulý, jeden, tento, první, další, třetí, poslední, následující, týž, tehle, jediný, čtvrtý, pátý, předminulý, pouhý*.

- (1) *Teplejší počasí už druhý měsíc/Adv po sobě pomáhá nezaměstnaným Čechům najít si alespoň sezónní práci ,*
- (2) *takže celou tu druhou noc/Adv slyšel v třeskající tmě dunění mořských peřejí o vlnolam*
- (3) *I když řadu/Adv měsíců vévodila žebříčkům volebních odhadů ODS , těsně před datem voleb ji předstihli socialisté*

Určení míry (vzdálenosti) v prostém akuzativu se používá výrazně méně než určení času, představuje cca 4 % příslovečných určení v prostém akuzativu (4). Substantiva v akuzativu často rozvíjejí adjektiva: *vysoký, dlouhý* apod. (5). Řadí se sem i obraty typu *metr před vápnem*, v nichž po substantivu v akuzativu následuje předložková fráze (6). V PDT se v takových spojeních považuje první substantivum za řídicí, druhé za závislé. Toto řešení přebíráme i zde, i když by bylo možné najít argumenty i pro opačnou závislost (první substantivum vyjadřuje míru vzdálenosti vyjádřené předložkovou frází, např. výše uvedené spojení by bylo odpovědí na otázku: *jak daleko před vápnem?*).

- (4) *Příčnou tyčku uvažte asi 3 centimetry/Adv od jednoho konce podélného dřívka .*
- (5) *Tehdy tu byl skoro čtyři metry/Adv hluboký rybník !*
- (6) *A teď je krok/Adv od prohry .*

Přibližně 3 % substantiv v prostém akuzativu ve funkci **Adv** tvoří víceméně ustálené slovní obraty složené ze substantiva v akuzativu, předložky a téhož substantiva v pádu daném předložkou (*den ode dne*). Tato spojení nejčastěji určují čas: *čas od času, rok od roku, minutu po minutě* a způsob: *krok za krokem, metr po metru, kousek po kousku, bok po boku, případ od případu, ráz na ráz* (7), (8) a (9).

- (7) *Rok/Adv za rokem vysílá Anglie čerstvé odvedence do první bitevní linie ,*
- (8) *Rozsáhlé výčty živočichů , které je obývají , se oblast/Adv od oblasti různé ,*
- (9) *Vzpomínali jsme slovo/Adv od slova , děj/Adv od děje , přidávali jsme své osobní vzpomínky .*

### 2.7.7 Přísluvečné určení v předložkovém akuzativu (Adv\_p4)

Substantiva v předložkovém akuzativu se na přísluvečných určeních v korpusu podílejí 14,0 %, na všech substantivech mají podíl 2,8 %. Předložkové fráze s funkcí **Adv** jsou řízeny předložkami: *na* (48,5 %), *pro* (18,4 %), *v* (9,8 %), *za* (7,4 %), *přes* (4,8 %), *o*, *po*, *mezi*, *pod*, *mimo*, *před*, *nad*, *skrz*.

Substantiva v předložkovém akuzativu s funkcí Adv nejčastěji určují místo (1), (2) a (3), jen o málo méně často určují čas (4), (5) a (6).

- (1) *Maigret mu strčil pod nos/Adv svůj odznak .*
- (2) *přišla řeč i na národní rýsy/Adv firemních kultur .*
- (3) *Rádi bychom se dostali přes laťku/Adv kapel , které jsou známé jen v regionu*
- (4) *Zdá se však , že odešel v pravý čas/Adv , na vrcholu tvůrčích sil .*
- (5) *Zastavil se tam na mžik/Adv a zatajil dech .*
- (6) *Odešla a za notnou chvíli/Adv se vrátila v domácím oděvu ,*

Určení způsobu v předložkovém akuzativu (určení míry, zřetele aj.) jsou v korpusu méně častá než určení času, ale vyskytují se i tak hojně (7), (8) a (9). Nejméně častým přísluvečným určením v předložkovém akuzativu je určení kauzality, především příčiny a účelu (10), (11) a (12) .

- (7) *Torpédo nás minulo snad o metr/Adv .*
- (8) *Byl rád , že nebude na obtíž/Adv .*
- (9) *Se svým bratrem vyřizovala pro matku/Adv mnoho pochůzek .*
- (10) *Řízení bylo zahájeno na žádost/Adv Ředitelství silnic a dálnic .*
- (11) *Jejich snaha však skončí pro nedostatek/Adv záloh a zásob .*
- (12) *Za nedodržení/Adv nově stanoveného termínu k provedení nápravy je možné zvýšit původně určenou pokutu až o 100 procent .*

Mezi nejčastější přísluvečná určení v předložkovém akuzativu patří spojení předložky *v* se jmény dnů v týdnu (*v sobotu*). Kromě těchto spojení jsou velmi frekventovaná další časová určení: *ve chvíli*, *po dobu*; místní určení: *na stranu*, *na zem* a určení způsobu: *na pohled* (substantiva ve všech uvedených spojeních jsou často rozvíta přívlastky).

### 2.7.8 Přísluvečné určení v prostém vokativu (Adv\_5)

Přísluvečné určení může být vyjádřeno i substantivem ve vokativu, takových případů je však v korpusu jen velmi málo: 23 ppm. Přísluvečná určení jsou závislá na slovesech říci, říkat, oslovit, oslovovat (1) až (5).

- (1) Mně jednou **řekli Šéfe/Adv** !
- (2) " Liborovi **říkáme boubelko/Adv** , " upozorňuje další aktérka Dana Morávková .
- (3) Jinak by mu přece **neříkaly " táto/Adv "** .
- (4) Urazil se poté , když ho jeden z novinářů **oslovil Honzo/Adv** .
- (5) Ani za nic prý nebude svou bývalou komornou **oslovovat ' sestro/Adv '** .

### 2.7.9 Přísluvečné určení v předložkovém lokálu (Adv\_p6)

Předložkový lokál je nejčastějším pádem substantiv ve funkci přísluvečného určení, substantiva v tomto pádu se na všech **Adv** podílejí 36,9 %. Kombinace **Adv\_p6** je tak s podílem 7,4 % na všech substantivech čtvrtou nejčastější realizací syntaktické funkce pádem. Většina přísluvečných určení je s předložkou *v* (63,2 %). Ostatní předložky se používají méně: *na* (20,5 %), *po* (10,0 %), *při* (5,6 %). Předložka *o* s lokálem se pro realizaci funkce **Adv** používá jen výjimečně (0,7 %), většinou se používá k vyjádření předmětu.

Substantiva s funkcí **Adv** v předložkovém lokálu jsou nejčastěji určením místa (1), (2) a (3), velmi často také určením času (4), (5) a (6).

- (1) Připadal jsem si jako klouček **nachytaný na švestkách/Adv** .
- (2) **V** některých **zemích/Adv** již **zakázali** šamanům přístup na hřiště .
- (3) Vrátil se do haly , pomalu se **vlekl po schodech/Adv** a nahoře zamířil do ložnice .
- (4) **Ve třicátých letech/Adv** **podlehly** svodům fašismu miliony lidí .
- (5) Poté však **po** dlouhá **desetiletí/Adv** k žádné podstatné technologické změně **nedošlo** .
- (6) ( **nenapadne** - li **o Vánocích/Adv** sníh , bude **sněžit o Velikonocích/Adv** )

Určení způsobu (vlastního způsobu, dopravního prostředku, průvodních okolností aj.) v předložkovém lokálu (7), (8) a (9) je znatelně méně časté, i tak ale dostatečně zastoupené v korpusu. Naopak určení kauzality (přípustka, účel apod.) v lokálu se vyskytuje jen zřídka (10) a (11).

- (7) v upomínku na to , že jím v roce 1779 **projel na koni/Adv** Josef II .
- (8) Posadme se a **v klidu/Adv** si o tom **promluvm** .
- (9) Potřetí se **vrátil** s pláčem a **po kolenou/Adv** a úpěnlivě žadonil o občanství NDR .
- (10) Že **při** těchto **schopnostech/Adv** **nevyužila** své tíživosti , ale ochotně se omezila na prostor ženám vlastní , zvyšovalo kouzlo její osobnosti .
- (11) **Byl** jsem s barmanem **na rybách/Adv** .

Spojení předložky a substantiva v lokálu s funkcí **Adv** jsou v korpusu SYN2005 nejčastěji určeni času: v roce, v letech, v době, v noci, ve chvíli; dále určeni místa: na místě, na straně, v Praze; a jiné: v případě, ve skutečnosti.

### 2.7.10 Příslůvečné určení v prostém instrumentálu (**Adv\_7**)

Mezi substantivy s funkcí **Adv** je podíl prostého instrumentálu 8,8 %, kombinace **Adv\_7** představuje 1,8 % všech substantiv. Mezi příslůvečná určení neřadíme určení původce děje u opisného pasiva, tato funkce se v analyzovaném korpusu považuje za **Obj**.

Prostého instrumentálu substantiv se v češtině užívá většinou pro vyjádření způsobu: nástroje, dopravního prostředku, zřetele, vlastního způsobu aj. (1), (2) a (3). Výrazně méně časté je určení místa, především jako odpověď na otázku *kudy?* (4), (5) a (6).

- (1) Ze západního obzoru se **rychlostí/Adv** blesku **rozvine** hejno havranů ,
- (2) Všechny podrobnosti **dostává** elektronickou **poštou/Adv** .
- (3) Jelikož byla zvířata **střílena** jednotlivě **kulkou/Adv** , měla nejlepší šance na přežití ta , která se chovala plaše
- (4) **Místností/Adv** ozval se řev :
- (5) sborový zpěv , jenž **se** zlehka **nesl soumrakem/Adv** .
- (6) **Prodírali se spleti/Adv** kořenů v prastaré chodbě k cíli ,

Časové určení v prostém instrumentálu je v korpusu jen o málo méně časté než určení místa. V této funkci se může vyskytovat několik nerozvitých substantiv v instrumentálu, např. *časem, chvílemi, dnem i nocí* a dále substantiva rozvitá přívlastky podobně jako časová určení v prostém akuzativu: *tou dobou, tímto dnem, každým rokem* (7), (8) a (9). Ustálené tvary rozvitě genitivem jiného substantiva považujeme za nevlastní předložky: *začátkem, počátkem, koncem* apod., za substantivní příslůvečná určení je tedy nepovažujeme. Určení

kauzality se v prostém instrumentálu vyskytuje zřídka, většinou jako určení příčiny: vykřiknout hrůzou, výjimečně i jako určení účelu aj. (10), (11) a (12).

- (7) Západočeši **chvílemi/Adv připomínali** přezrálé hrušky ,
- (8) **Pracoval** neúnavně **dnem/Adv** i **nocí/Adv** .
- (9) *Trápí vás , že každým **rokem/Adv zanikne** několik českých pivovarů ,*
- (10) *Mnoho jich cestou **zahyne hladem/Adv** a **žízní/Adv** ,*
- (11) *Ted' jsem se jen **chvěl radostí/Adv** a mlčel .*
- (12) *Je vedro a zemědělci začínají naříkat , že nic neroste a za chvíli prý **půjdeme žebrotou/Adv** .*

Nejčastější substantiva v prostém instrumentálu ve funkci **Adv** jsou: *způsobem, cestou, směrem, časem, formou, rychlostí, zákonem, jménem*. Mezi nejčastější příslovečná určení realizovaná prostým instrumentálem patří také názvy částí těla (nepočítaje samozřejmě předměty v instrumentálu, které jsou také časté, např. *vrtět hlavou* či *mávnout rukou*): *hlavou (blesknout hlavou), rukou (chytit rukou), rukama (projít rukama), očima* aj.

#### 2.7.11 Příslopečné určení v předložkovém instrumentálu (**Adv\_p7**)

Substantiva v předložkovém instrumentálu s funkcí **Adv** představují 8,1 % substantiv s funkcí **Adv**, 1,6 % všech substantiv. Předložkové fráze **Adv\_p7** jsou řízeny sedmi předložkami: *s* (47,6 %), *před* (21,3 %), *pod* (9,7 %), *mezi* (9,1 %), *za* (8,3 %), *nad* (3,9 %), *napříč* (0,1 %). Malá část substantiv **Adv\_p7** (2,8 %) je řízena víceslovnými předložkovými výrazy, např. *společně s, v souvislosti s, v souladu s, ve srovnání s* aj.

Substantiva **Adv\_p7** jsou nejčastěji příslovečným určením způsobu, včetně průvodních okolností či provázející osoby (1), (2) a (3). Příslopečné určení místa vyjádřené předložkovým instrumentálem má přibližně poloviční frekvenci (4), (5) a (6).

- (1) **Objevila** se **s podnosem/Adv** plným sklenic a lahví , který mi připadal pro ni příliš těžký ,
- (2) *Obě země se ale připravují na vstup do unie a **pod tlakem/Adv** Bruselu prostředí pro podnikání **vylepšují** .*
- (3) *Polovinu brandy **nahřejeme** v naběračce **nad plamenem/Adv** , zapálíme a hořící nalijeme na zeleninu .*
- (4) *Po několika hodinách veslování se **nad horizontem/Adv** začala **rýsovat** temná silueta*

vzdáleného ostrova .

- (5) Dodávka s rakvemi umně **kličkuje mezi zákony/Adv** obou zemí .
- (6) Po odeznění obrozenského romantismu 19 . století jí už musela **být za pecí/Adv** venkovské chalupy dlouhá chvíle .

Přísllovečná určení času v předložkovém instrumentálu jsou jen o málo méně časté než určení místa, také s přibližně poloviční frekvencí oproti určení způsobu (7), (8) a (9). Naproti tomu určení kauzality (především účelu, výjimečně i podmínky) se předložkovým instrumentálem realizují jen velmi zřídka (10), (11) a (12).

- (7) **Před válkou/Adv** jste **byl** uznávaným odborníkem .
- (8) Dva podezřelé muže **zatkla nad ránem/Adv** v Rousínově na Vyškovsku policejní hlídka .
- (9) **Mezi** jedenáctou a čtrnáctou **hodinou/Adv** , kdy panuje největší žár , pak **nevycházíme** s dětmi vůbec ven ,
- (10) Vyplývá to z plánů americké armády , které ale byly možná **zveřejněny s** nějakým jiným **cílem/Adv** .
- (11) Znovu opakuji , má žena **jela** do Řecka **za** konkrétním **účelem/Adv** ,
- (12) Už není nevolník a **s** tvým **souhlasem/Adv** bychom mu mohli **dát** kus země na Greenfieldu

Mezi nejčastější přísllovečná určení v předložkovém instrumentálu patří jednak časová určení: *před lety, před rokem, před časem, před (několika) dny*, jednak určení způsobu: *s úsměvem, s pomocí, se (zavřenýma) očima, s (velkou) pravděpodobností, za zády, pod názvem, za dveřmi*.

## 2.8 Aktuální elipsy a řídicí větné členy výpovědi bez slovesa (ExD)

Syntaktická funkce **ExD** v PDT zahrnuje několik dílčích kategorií. Řadí se sem především slova, jejichž řídicí člen byl aktuálně vypuštěn (1), nepatří sem tedy lexikalizované elipsy. Za konstrukce s elidovaným predikátem se považují i srovnávací spojení se spojky *jak, jako, než* (2). Funkce **ExD** se také přiřazuje řídicím větným členům ve výpovědích bez slovesa (3), v názvech kapitol apod. (4). Syntaktickou funkci **ExD** má i vokativ ve funkci oslovení, ať stojí samostatně, nebo se používá v rámci komplexní výpovědi (5).

- (1) *Neboť jsem přišel postavit **syna/Obj proti** jeho **otci/Obj** , **dceru/ExD proti matce/ExD** , **snachu/ExD proti tchyni/ExD** ;*
- (2) *plet' žlutavě snědou , vlasy černé jako **tuš/ExD** a tak hojné , že by stačily na dva účesy*  
.
- (3) *Šla nejdřív zhasnout světla . Všude zhasíná světla . **V kuchyni/ExD** . **Na schodišti/ExD** do sklepa . **Ve sklepě/ExD** . **V prádelně/ExD** .*
- (4) ***Objevitel/ExD** radioaktivity . Sto padesáté **výročí/ExD** narození . V pondělí 20 . ledna 1896 na pravidelné schůzi pařížské Akademie věd nechal jeden z členů kolovat snímky*
- (5) *Nic mi neříkejte , **sestro/ExD** , já vám a zdejšími doktorům prostě věřím ,*

Aktuální elipsy uvnitř souvětí je velmi obtížné automaticky odlišit od koordinačních a apozičních vztahů, určení funkce **ExD** je v důsledku toho nespolehlivé, pro získání spolehlivých statistických údajů bylo nutné vycházet z manuálních sond v korpusu.

Funkci **ExD** má 5,9 % substantiv v korpusu, kromě prostého nominativu a vokativu se však vyskytuje jen zřídka (kombinace s ostatními pády mají frekvenci mezi 300 a 2300 ppm, podíl všech pádů vyjma nominativu a vokativu na substantivech v korpusu je 1,3 %). Proto samostatně představíme jen realizaci funkce **ExD** nominativem a vokativem, potom v jednom odstavci shrneme příklady realizací této funkce všemi dalšími pády, předložkovými i nepředložkovými.

### 2.8.1 Elipsa v prostém nominativu (**ExD\_1**)

V prostém nominativu je 71,6 % substantiv se syntaktickou funkcí **ExD**, kombinace **ExD\_1** má na všech substantivech v korpusu podíl 4,2 %, je to šestá nejčastější realizace syntaktické funkce pádem. Kvůli značné nespolehlivosti automatického značkování aktuální elipsy, tj. elipsy řídicího členu substantiva ve větě či v souvětí, kde se nachází sloveso, jsou údaje o dílčích frekvencích jen velmi přibližné.

Nejčastěji (přibližně v 70 %) se v korpusu substantiva v nominativu s funkcí **ExD** vyskytují ve větách/výpovědích bez slovesa v určitém tvaru. Jsou to především názvy děl, oddílů, kapitol, informace o autorech, různé podtitulky aj., zvláště často v publicistickém stylu (1) a (2). Méně často jsou tyto nominativy samostatně stojícími výpověďmi, které navazují na předcházející text (3).

- (1) ***Vývoj/ExD** papírových platidel od roku 1918*  
**VLADIMÍR KUDLÍK/ExD**

- (2) zahájena v jedenáct hodin . **BRNO/ExD** - Masarykova univerzita chce do roku 2005
- (3) ( začal ... jednu po druhé strkat důstojníkovi do rukou ...) Lord **Byron/ExD** , edinburský tisk . **Milton/ExD** . **Shelley/ExD** a **Keats/ExD** . A tohle je komentované vydání Chaucera .

Asi 20 % substantiv **ExD\_1** se nachází ve větách se slovesem v určitém tvaru, ve kterých nedošlo k elizi opakujícího se predikátu (příklady k této elizi níže). Do těchto 20 % spadá mnoho různých typů konstrukcí (4), (5) a (6).

- (4) **Únava/ExD** - sotva se držím na lavici , natož abych se přeplazila ke kbelíku studené vody a občerstvila ručník na hlavě .
- (5) **Pravda/ExD** , později by možná pochyboval , dokonce i litoval .
- (6) Otevřeno **úterý/ExD** až **neděle/ExD** 10.00 až 12.00 , 13.00 až 16.00 hodin

Srovnávací obraty se spojky jako a než se na substantivech v nominativu s funkcí **ExD** podílejí přibližně sedmi procenty (7), (8) a (9).

- (7) hned by byla zdravá jako **rybička/ExD** .
- (8) Termitiště se tyčí jako pískovcové **sloupy/ExD** , často vyšší než dospělý **člověk/ExD** .
- (9) kdo zastává jiný názor než Václav **Klaus/ExD** , je potenciální terorista

Elipsa opakovaného predikátu byla mezi substantivy automaticky označenými jako **ExD** nalezena jen zřídka, představuje pouze cca 3 % substantiv **ExD\_1** (10), (11) a (12).

- (10) Jenomže Christian je daleko a **Paříž/ExD** plná svodů . . .
- (11) Richelieu **zemřel** roku 1642 a **Ludvík/ExD** XIII . roku 1643 .
- (12) Tato smlouva je sepsána ve 4 vyhotoveních stejné právní závaznosti , z nichž nájemce **obdrží 1 a pronajímatel/ExD** 3 vyhotovení .

### 2.8.2 Elipsa v prostém vokativu (**Exd\_5**)

Na rozdíl od elipsy v nominativu je realizace funkce **ExD** prostým vokativem zcela přehledná: funkci **ExD** má substantivum ve vokativu použité jako oslovení; tuto funkci má většina vokativů (88,7 %), ostatní vokativy jsou téměř vždy shodnými přívlastky jiných vokativů s funkcí **ExD**.

Podíl substantiv ve vokativu s funkcí **ExD** na všech (započítaných) substantivech v korpusu je 0,4 % (3793 ppm). Přibližně v jedné čtvrtině výskytů tvoří vokativy samostatnou výpověď nebo se nacházejí ve složitější výpovědi, ale bez slovesa (1) až (5).

- (1) Milý **příteli/ExD** !
- (2) Můj ty **bože/ExD** !
- (3) Vy komunistický **svině/ExD** !
- (4) Ahoj , **děvče/ExD** !
- (5) Ahoj , **Alfréde/ExD** , tady máma .

Častěji (cca 75 %) jsou vokativy součástí komplexních výpovědí se slovesem v určitém tvaru, sloveso je často ve druhé osobě (6) až (10).

- (6) **Adame/ExD** , máš mne rád ?
- (7) " Pane **prezidente/ExD** , udělám všechno , co si přejete , " řekla a chystala se odejít .
- (8) " Nech toho , **maminko/ExD** , " prohlásila jsem na prahu , " dej těm pánům klidně všechno .
- (9) " Ztraceně , můj **poklade/ExD** , jsi to ty ? " zvolal doktor Angelino užasle .
- (10) " Obžalovaný vinu přiznává , ctihodný **soude/ExD** , " řekl .

### 2.8.3 Substantiva se syntaktickou funkcí **ExD** v jiných pádech než v nominativu a vokativu

Substantiva s funkcí **ExD** se objevují v prostých i předložkových pádech, a to přibližně ve stejné frekvenci. Mezi prostými pády je nejčastější akuzativ (3041 ppm), dále následují genitiv (1673 ppm), instrumentál (1291 ppm) a dativ (476 ppm). Prosté pády stojí nejčastěji v samostatných výpovědích bez slovesa (1) a (2), zřídka ve strukturách s elidovaným opakovaným predikátem (3) a (4) a zřídka se také nacházejí ve srovnávacích obrazech se spojkami *jako* a *než* (5).

- (1) *Děšť promlouvá ke mně skrz megafony okapů . Nikde ani živáčka/ExD . Večerní temnota ovíjí svými šálami kmeny stromů .*
- (2) *Toho jsme my stromy ušetřeny . Bohu/ExD díky/ExD . A co teprve vysvobození z pomíjivosti času ,*
- (3) *Richelieu zemřel roku 1642 a Ludvík XIII . roku/ExD 1643 .*

- (4) Tato smlouva je sepsána ve 4 vyhotoveních stejné právní závaznosti , z nichž nájemce obdrží 1 a pronajímatel 3 **vyhotovení/ExD** .
- (5) vaše připomínka je mi právě tak nejasná jako slečně **Mitnickové/ExD** .

Mezi předložkovými pády s funkcí ExD jsou také nejčastější samostatné výpovědi bez slovesa (6), (7) a (8), předložkové fráze ve strukturách s vypuštěným opakovaným predikátem se vyskytují vzácně (9). Jen relativně zřídka se také vyskytují srovnávací obraty se spojkami *jako* a *než*, po kterých následuje předložková fráze (9) a (10).

- (6) **Na Bělehrad/ExD** !
- (7) V zoo se mi nejvíce líbily malé žirafy . **Kvůli jejich bezbrannosti/ExD a čistotě/ExD** .
- (8) / **Od našeho zvláštního zpravodaje/ExD** /
- (9) Za měsíc zaplatíme pět stovek a **za jízdné/ExD** 350 korun , což je ještě nejméně o tři stovky méně než **v Opavě/ExD** , " konstatoval Tomáš ( 20 ) z Brna .
- (10) U řady pacientů jsou na předním konci žeber změny stejné jako **u rachitidy/ExD** .

## 2.9 Závěr

V této kapitole byly podrobně představeny realizace syntaktických funkcí různými pády substantiv. Kromě vlastní informační hodnoty a utříděných korpusových dokladů mají tyto odstavce sloužit jako referenční údaje, které umožní lepší pochopení celkových statistických údajů i srovnání žánrů v korpusu, jež představíme v následující kapitole.

## Kapitola III

# Frekvence syntaktických funkcí a jejich realizací pády substantiv v korpusu SYN2005

Na následujících stránkách představíme frekvence syntaktických funkcí a jejich realizací pády v korpusu SYN2005 a v jeho třech subkorpusech: beletrie (**BEL**), odborné literatury (**ODB**) a publicistiky (**PUB**). Frekvence se prezentují vždy ve čtveřicích tabulek, jedna tabulka pro celý korpus (úsporněji označovaný **SYN05**), tři tabulky pro subkorpusey.

### 1. Tabulky frekvence syntaktických funkcí a pádů substantiv

V každé tabulce jsou ve sloupcích uvedeny syntaktické funkce: **Sb**, **Pnom**, **Atr**, **Obj**, **Adv** a **ExD**. V řádcích jsou uvedeny pády substantiv, odděleně prosté a předložkové: **1**, **2**, **p2**, **3**, **p3** atd. Uvnitř tabulky jsou pak číslicemi vyjádřené údaje o relativních frekvencích jednotlivých kombinací syntaktické funkce a pádu.

Celkem zde prezentujeme čtyři čtveřice tabulek. V prvních třech čtveřicích tabulek jsou frekvenční údaje v tabulce vyjádřením podílů v procentech (zaokrouhleno na desetiny procenta), ve čtvrté čtveřici tabulek jsou údaje podrobnější, vyjádřené v ppm (počet výskytů na 1 milión substantiv). Znak procent je kvůli přehlednosti uveden pouze u celku (100 %).

Základem pro výpočet procent v první čtveřici tabulek je počet výskytů všech (započítaných) substantiv v korpusu či subkorpusu, např. údaj 15,5 u kombinace **Obj\_4** v subkorpusu **BEL** znamená, že substantiva v prostém akuzativu s funkcí **Obj** představují 15,5 % ze všech substantiv v daném subkorpusu.

Ve druhé čtveřici tabulek je základem pro výpočet celkový počet substantiv s určitou syntaktickou funkcí, údaje v tabulkách vyjadřují, jak často se daná syntaktická funkce realizuje jednotlivými pády. Například údaj 60,4 u kombinace **Obj\_4** v subkorpusu **ODB** vyjadřuje, že v tomto subkorpusu je 60,4 % substantivních předmětů v prostém akuzativu.

Ve třetí čtveřici tabulek vychází výpočet procent ze základu, kterým je počet substantiv daného pádu. Údaje v tabulkách tak ukazují, jak často se daný pád využívá k realizaci určité syntaktické funkce. Například číslo 92,0 u kombinace **Obj\_4** v subkorpusu **PUB** udává, že se v tomto subkorpusu používá prostý akuzativ z 92,0 % k realizaci syntaktické funkce **Obj** (předmět).

Poslední čtveřice tabulek má za základ počet výskytů všech substantiv v korpusu či

subkorpusu, stejně jako první čtveřice. Údaje však nejsou vyjádřeny v procentech (zaokrouhlené na desetiny procent), ale v ppm, tj. v počtu výskytů na 1 milión substantiv. V údajích je tak vyjádřeno o tři desetinná místa více a v tabulkách se díky tomu objeví i některé málo frekventované kombinace, např. **Atr\_5**. Např. u kombinace **Obj\_4** v subkorpusu BEL je v první sadě tabulek uvedeno 15,5 %, ve čtvrté čtveřici tabulek je na stejném místě 155 247 ppm. Protože údaje přepočítáváme na základě manuálních sond v korpusu, kterými jsme korigovali nespolehlivé výsledky automatického syntaktického značkování, nelze považovat údaj 155 247 za spolehlivý o nic více než zaokrouhlená procenta odpovídající číslu 155 000. Zaokrouhlení je však v tomto případě zbytečné.

Tyto podrobnější údaje lze také použít jako základ pro nové výpočty, jimiž by si čtenář mohl chtít přizpůsobit tabulku svým potřebám: je možné využít podrobnější údaje uvedené u jednotlivých kombinací v předcházející kapitole a tabulku přepočítat, popř. z ní vyloučit některé části (např. by bylo možné oddělit shodné přívlastky od neshodných, počítat původce děje v prostém instrumentálu jako příslovečné určení aj.).

## 1.1 Podíl realizací syntaktických funkcí pády na všech substantivech

V první čtveřici tabulek představujeme podíly kombinací syntaktické funkce a pádu na celkovém počtu výskytů substantiv v subkorpusech. Údaje jsou v procentech, zaokrouhlené na desetiny procenta.

<b>SYN05</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	20,4	1,1	2,6	0,1	0	4,2	28,4
2	0	0	17,9	1,0	0,2	0,2	19,3
p2	0	0	2,4	0,5	4,6	0,1	7,6
3	0	0	0,1	1,6	0	0	1,8
p3	0	0	0,4	0,6	1,0	0	2,1
4	0	0	0,1	13,0	0,7	0,3	14,1
p4	0	0	1,5	1,7	2,8	0,1	6,1
5	0	0	0	0	0	0,4	0,4
p6	0	0	3,3	0,7	7,4	0,2	11,6
7	0	0,9	0,4	1,1	1,8	0,1	4,3
p7	0	0	1,8	0,6	1,6	0,2	4,2
SUM	20,4	2,0	30,5	20,9	20,2	5,9	100 %

<b>BEL</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	21,4	1,7	1,8	0,1	0	4,0	29,0
2	0	0	11,8	1,0	0,1	0,1	13,1
p2	0	0	2,2	0,7	5,6	0,1	8,6
3	0	0	0,1	1,8	0	0,1	2,0
p3	0	0	0,4	0,6	1,1	0	2,2
4	0	0	0,1	15,5	0,9	0,4	16,9
p4	0	0	1,2	1,8	3,4	0,1	6,4
5	0	0	0,1	0	0	0,7	0,9
p6	0	0	2,6	0,7	7,8	0,2	11,4
7	0	0,7	0,4	1,3	2,1	0,2	4,7
p7	0	0	1,8	0,7	2,1	0,3	4,9
SUM	21,5	2,4	22,5	24,1	23,2	6,3	100 %

Nejvíce se liší subkorpusey **BEL** a **ODB**, a to nejčastější kombinací (**Sb\_1** oproti **Atr\_2**), nejčastější syntaktickou funkcí celkem (**Obj** oproti **Atr**); liší se i v dalších oblastech (např. realizace **Pnom**).

**Červeně** je zvýrazněna nejvyšší frekvence kombinace funkce a pádu v každé z tabulek a nejvyšší celková frekvence syntaktické funkce. **Zeleně** jsou zvýrazněny oblasti, které budou předmětem samostatné analýzy. **Modře** jsou zvýrazněny dvě další oblasti, v nichž se žánry odlišují (*Obj\_4* a *ExD\_5*).

<b>ODB</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	18,4	<b>0,7</b>	1,8	0,1	0	4,4	25,3
2	0	0	<b>23,2</b>	1,1	0,3	0,2	24,9
p2	0	0	2,4	0,5	4,0	0,1	6,9
3	0	0	0,1	1,4	0	0	1,6
p3	0	0	0,5	0,7	1,1	0	2,4
4	0	0	0,1	<b>11,9</b>	0,4	0,3	12,8
p4	0	0	1,7	1,6	2,3	0,1	5,7
5	0	0	0	0	0	<b>0,2</b>	0,2
p6	0	0	3,5	0,6	7,0	0,2	11,3
7	0	<b>1,2</b>	0,4	<b>1,2</b>	1,9	0,1	5,0
p7	0	0	1,8	0,5	1,3	0,2	3,9
SUM	18,4	1,9	<b>35,6</b>	19,7	18,5	5,8	100 %

<b>PUB</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	<b>21,0</b>	<b>0,9</b>	4,0	0,1	0	4,3	30,4
2	0	0	<b>18,9</b>	0,9	0,2	0,2	20,2
p2	0	0	2,5	0,5	4,3	0,1	7,4
3	0	0	0,1	1,6	0	0	1,8
p3	0	0	0,4	0,6	0,9	0	1,8
4	0	0	0,1	<b>11,8</b>	0,7	0,2	12,9
p4	0	0	1,6	1,6	2,7	0,1	6,0
5	0	0	0	0	0	<b>0,2</b>	0,2
p6	0	0	3,7	0,7	7,5	0,2	12,0
7	0	<b>0,8</b>	0,3	<b>0,8</b>	1,3	0,1	3,4
p7	0	0	1,7	0,5	1,5	0,2	3,9
SUM	21,0	1,8	<b>33,4</b>	19,2	19,0	5,7	100 %

Mezi subkorpora **BEL** a **PUB** jsou větší rozdíly než mezi **BEL** a **ODB** například v celkové frekvenci funkcí *Pnom* a *Obj*. V naprosté většině údajů jsou si subkorpora **PUB** a **ODB** bližší než **BEL** a **PUB** nebo **BEL** a **ODB**.

## 1.2 Podíl pádů na syntaktických funkcích substantiv

Druhá čtveřice tabulek ukazuje, jak často se syntaktické funkce realizují určitým pádem, přesněji řečeno, jaký je podíl substantiv v jednotlivých pádech na všech substantivech s danou syntaktickou funkcí. Údaje jsou v procentech, zaokrouhlené na desetiny procenta.

<b>SYN05</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	99,8	55,4	8,7	0,3	0,1	71,6	28,4
2	0,2	0,5	58,6	4,8	0,9	2,8	19,3
p2	0	0	7,8	2,6	22,8	1,7	7,6
3	0	0	0,4	7,7	0,1	0,8	1,8
p3	0	0	1,3	3,1	5,0	0,6	2,1
4	0	0	0,4	62,3	3,3	5,1	14,1
p4	0	0	4,9	8,0	14,0	1,2	6,1
5	0	0	0,2	0	0	6,4	0,4
p6	0	0	10,8	3,2	36,8	3,7	11,6
7	0	44,1	1,2	5,3	8,8	2,2	4,3
p7	0	0	5,8	2,7	8,1	3,8	4,2
SUM	100 %	100 %	100 %	100 %	100 %	100 %	100 %

<b>BEL</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	99,8	71,2	8,0	0,2	0,1	63,8	29,0
2	0,2	0,5	52,4	4,0	0,6	2,2	13,1
p2	0	0	9,9	2,7	23,9	2,0	8,6
3	0	0	0,6	7,4	0,1	1,0	2,0
p3	0	0	1,6	2,6	4,9	0,6	2,2
4	0	0	0,5	64,4	3,8	6,3	16,9
p4	0	0	5,2	7,4	14,6	1,4	6,4
5	0	0	0,5	0	0	11,9	0,9
p6	0	0	11,6	3,0	33,7	3,6	11,4
7	0	28,3	1,6	5,5	9,2	2,4	4,7
p7	0	0	8,0	2,7	9,1	4,7	4,9
SUM	100 %	100 %	100 %	100 %	100 %	100 %	100 %

V těchto tabulkách jsou některé rozdíly mezi žánry zřetelnější než v tabulkách předchozích, např. poměr realizace funkce **Pnom** nominativem a instrumentálem substantiv v subkorpusech **BEL** a **ODB**.

**Zeleně** jsou zvýrazněny oblasti, které budou předmětem samostatné analýzy a zde je rozdíl mezi žánry v dané oblasti dostatečně zřetelný. **Modře** je zvýrazněna další oblast, v níž se žánry odlišují (*ExD\_5*).

<b>ODB</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	99,8	<b>36,9</b>	5,1	0,3	0	74,6	25,3
2	0,2	0,5	<b>65,2</b>	5,8	1,5	3,7	24,9
p2	0	0	6,7	2,3	21,9	1,4	6,9
3	0	0	0,3	7,3	0,1	0,6	1,6
p3	0	0	1,4	3,8	6,0	0,8	2,4
4	0	0	0,3	60,4	2,4	5,0	12,8
p4	0	0	4,8	8,2	12,6	1,0	5,7
5	0	0	0	0	0	<b>3,6</b>	0,2
p6	0	0	9,8	3,0	37,9	3,5	11,3
7	0	<b>62,6</b>	1,2	<b>6,3</b>	10,5	2,6	5,0
p7	0	0	5,1	2,6	7,1	3,2	3,9
SUM	100 %	100 %	100 %	100 %	100 %	100 %	100 %

<b>PUB</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	99,9	<b>53,1</b>	12,1	0,3	0,1	76,5	30,4
2	0,1	0,4	<b>56,6</b>	4,9	0,9	2,7	20,2
p2	0	0	7,6	2,7	22,5	1,6	7,4
3	0	0	0,4	8,4	0,1	0,7	1,8
p3	0	0	1,1	3,0	4,5	0,6	1,8
4	0	0	0,4	61,7	3,5	4,1	12,9
p4	0	0	4,7	8,6	14,4	1,2	6,0
5	0	0	0	0	0	<b>3,5</b>	0,2
p6	0	0	11,1	3,4	39,3	3,9	12,0
7	0	<b>46,5</b>	1,0	<b>4,3</b>	7,0	1,6	3,4
p7	0	0	5,0	2,7	7,9	3,6	3,9
SUM	100 %	100 %	100 %	100 %	100 %	100 %	100 %

Modře zvýrazněný je podíl substantiv ve vokativu na funkci *ExD*. V subkorpusu **BEL** je více než třikrát častější než v korpusu **PUB** nebo **ODB**.

### 1.3 Podíl syntaktických funkcí na pádech substantiv

Tabulky ve třetí čtveřici obsahují procentuální podíly syntaktických funkcí na jednotlivých pádech substantiv. Údaj v políčku vyjadřuje, jak často se v rámci jednoho pádu vyskytuje určitá syntaktická funkce. Údaje jsou v procentech, zaokrouhleny na desetiny procenta.

<b>SYN05</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	71,6	3,9	<b>9,3</b>	0,2	0	14,9	100 %
2	0,2	0	92,7	5,2	1,0	0,9	100 %
p2	0	0	<b>31,2</b>	7,1	<b>60,4</b>	1,3	100 %
3	0	0	7,1	89,5	0,7	2,6	100 %
p3	0	0	19,2	30,6	48,4	1,8	100 %
4	0	0	0,9	92,3	4,7	2,2	100 %
p4	0	0	<b>24,5</b>	27,7	<b>46,6</b>	1,2	100 %
5	0	0	10,7	0	0,5	88,7	100 %
p6	0	0	<b>28,3</b>	5,7	<b>64,1</b>	1,9	100 %
7	0	20,8	8,6	26,0	41,5	3,0	100 %
p7	0	0	41,9	13,5	39,2	5,4	100 %
SUM	20,4	2,0	30,5	20,9	20,2	5,9	100 %

<b>BEL</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	73,8	5,9	<b>6,2</b>	0,2	0,1	13,8	100 %
2	0,3	0,1	90,2	7,3	1,0	1,1	100 %
p2	0	0	<b>26,0</b>	7,6	<b>64,9</b>	1,5	100 %
3	0	0	6,7	89,2	0,8	3,2	100 %
p3	0	0	16,9	29,4	52,1	1,7	100 %
4	0	0	0,7	91,7	5,2	2,3	100 %
p4	0	0	<b>18,0</b>	27,9	<b>52,7</b>	1,4	100 %
5	0	0	12,8	0	0,7	86,5	100 %
p6	0	0	<b>22,9</b>	6,3	<b>68,7</b>	2,0	100 %
7	0	14,6	7,7	28,5	45,8	3,3	100 %
p7	0	0	37,0	13,4	43,6	6,0	100 %
SUM	21,5	2,4	22,5	24,1	23,2	6,3	100 %

V tabulkách je zvýrazněn podíl nominativu použitého jako přívlastek (**Atr\_1**): v subkorpusu **PUB** se prostý nominativ používá pro tuto funkci více než dvakrát častěji než v subkorpusu **BEL**.

V těchto tabulkách jsou **modře** zvýrazněny další oblasti, v nichž se žánry odlišují (stručný komentář uvádíme pod tabulkami).

<b>ODB</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	72,6	2,8	<b>7,1</b>	0,2	0	17,2	100 %
2	0,1	0	93,3	4,6	1,1	0,9	100 %
p2	0	0	<b>34,1</b>	6,6	<b>58,1</b>	1,2	100 %
3	0	0	6,7	90,3	0,6	2,4	100 %
p3	0	0	21,1	31,0	46,0	1,9	100 %
4	0	0	0,9	93,3	3,5	2,3	100 %
p4	0	0	<b>30,1</b>	28,2	<b>40,7</b>	1,0	100 %
5	0	0	5,8	0	0,1	94,0	100 %
p6	0	0	<b>31,0</b>	5,2	<b>61,9</b>	1,8	100 %
7	0	24,2	8,4	25,1	39,2	3,0	100 %
p7	0	0	47,6	13,4	34,2	4,8	100 %
SUM	18,4	1,9	35,6	19,7	18,5	5,8	100 %

<b>PUB</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	69,1	3,1	<b>13,3</b>	0,2	0	14,2	100 %
2	0,1	0	93,5	4,7	0,9	0,7	100 %
p2	0	0	<b>34,1</b>	6,9	<b>57,7</b>	1,2	100 %
3	0	0	7,8	89,3	0,6	2,2	100 %
p3	0	0	19,6	31,5	47,2	1,8	100 %
4	0	0	1,1	92,0	5,1	1,8	100 %
p4	0	0	<b>26,3</b>	27,2	<b>45,4</b>	1,1	100 %
5	0	0	7,6	0	0,4	92,0	100 %
p6	0	0	<b>30,6</b>	5,5	<b>62,0</b>	1,9	100 %
7	0	24,3	9,9	24,1	39,0	2,7	100 %
p7	0	0	42,7	13,6	38,5	5,2	100 %
SUM	21,0	1,8	33,4	19,2	19,0	5,7	100 %

Zvýrazněny jsou také dvojice použití předložkových pádů (**p2**, **p4** a **p6**) jako **Atr** a jako **Obj**. V subkorpusech **PUB** a **ODB** realizují tyto předložkové pády mnohem častěji neshodné přívlastky než v subkorpusech **BEL**, kde se tyto pády používají více jako příslovečná určení.

#### 1.4 Podíl realizací syntaktických funkcí pády na všech substantivech (ppm)

V poslední čtveřici statistických tabulek uvádíme podrobnější údaje vyjádřené v počtu výskytů kombinací funkce a pády substantiva na milión substantiv v korpusu. Údaje z větší části nejsou tak přesné, jak jsou zde představeny, ale zaokrouhlení by situaci jen zhoršilo.

<b>SYN05</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	203 488	11 193	26 435	553	119	42 380	284 171
2	314	95	179 094	10 093	1 894	1 673	193 165
p2	0	0	23 828	5 406	46 166	1 001	76 404
3	0	0	1 291	16 184	125	476	18 078
p3	0	0	4 048	6 443	10 190	371	21 053
4	0	0	1 225	130 455	6 678	3 041	141 400
p4	0	0	14 892	16 823	28 245	712	60 674
5	0	0	459	0	23	3 793	4 275
p6	0	0	32 863	6 595	74 442	2 192	116 093
7	0	8 909	3 688	11 135	17 763	1 291	42 788
p7	0	0	17 561	5 639	16 419	2 273	41 893
SUM	203 803	20 198	305 389	209 330	202 069	59 208	ppm

<b>BEL</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	214 181	17 179	18 036	549	210	40 181	290 339
2	367	126	117 895	9 599	1 340	1 406	130 736
p2	0	0	22 316	6 531	55 648	1 259	85 754
3	0	0	1 347	17 817	161	649	19 975
p3	0	0	3 660	6 369	11 293	359	21 683
4	0	0	1 155	155 247	8 882	3 954	169 239
p4	0	0	11 584	17 894	33 845	875	64 200
5	0	0	1 111	0	59	7 489	8 660
p6	0	0	26 122	7 226	78 341	2 290	113 980
7	0	6 820	3 620	13 341	21 433	1 539	46 755
p7	0	0	18 000	6 527	21 206	2 939	48 674
SUM	214 549	24 126	224 851	241 105	232 422	62 944	ppm

K datům zde nepřipojujeme další komentář, v této podobě jsou uvedena, aby mohla sloužit k případnému vytváření vlastních statistických tabulek, které by si čtenář upravil na základě podrobných rozborů, jež jsme uvedli v předcházející kapitole.

Jak bylo řečeno výše, u velmi málo frekventovaných kombinací je automatické syntaktické značkování velmi nespolehlivé. U některých takových kombinací (*Adv\_3*, *Adv\_5*) byly statistické údaje v tabulkách po manuálním přepočítání a ověření v korpusu zachovány, u jiných (např. *Sb\_p2*, *Sb\_p6*), kde se nepodařilo spolehlivě zjistit frekvence, byly údaje zanedbány (místo velmi nepřesného čísla v řádu desítek je uvedena nula).

<b>ODB</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	184 022	7 073	18 072	568	48	43 606	253 391
2	315	96	232 142	11 394	2 678	2 152	248 780
p2	0	0	23 698	4 567	40 379	835	69 480
3	0	0	1 074	14 462	103	378	16 018
p3	0	0	5 080	7 445	11 054	446	24 027
4	0	0	1 136	119 077	4 477	2 922	127 613
p4	0	0	17 230	16 188	23 322	591	57 333
5	0	0	130	0	2	2 102	2 235
p6	0	0	35 057	5 911	69 973	2 038	112 981
7	0	12 000	4 178	12 458	19 465	1 498	49 601
p7	0	0	18 333	5 157	13 178	1 865	38 535
SUM	184 338	19 169	356 135	197 232	184 684	58 438	ppm

<b>PUB</b>	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
1	210 056	9 362	40 496	545	99	43 287	303 847
2	267	69	188 923	9 462	1 735	1 514	201 973
p2	0	0	25 242	5 116	42 667	914	73 941
3	0	0	1 420	16 171	112	405	18 110
p3	0	0	3 545	5 693	8 533	320	18 092
4	0	0	1 358	118 255	6 562	2 348	128 524
p4	0	0	15 855	16 411	27 402	670	60 339
5	0	0	162	0	8	1 970	2 141
p6	0	0	36 911	6 604	74 700	2 233	120 449
7	0	8 204	3 348	8 152	13 206	909	33 821
p7	0	0	16 552	5 262	14 911	2 030	38 757
SUM	210 324	17 636	333 818	191 676	189 940	56 603	ppm

## 2. Srovnání subkorpusů podle syntaktických funkcí a jejich realizace pády

V tomto oddíle doplníme podrobnější komentář k rozdílům mezi žánrovými subkorpusy. V uvedených tabulkách frekvencí syntaktických funkcí a jejich realizací pády substantiv byly zvýrazněny některé oblasti, v nichž se subkorpusy zvláště lišily. Tyto rozdíly se pokusíme vysvětlit v následujících odstavcích, podle potřeby doplníme i další údaje o syntaxi substantiv v žánrových subkorpusech. Srovnáme žánry podle celkové frekvence syntaktických funkcí, ukážeme, jak se žánry liší realizací verbonominálního predikátu, představíme rozdíly v použití přívlastku v prostém genitivu a v použití syntaktické funkce **Obj** v prostém instrumentálu. V posledním pododdíle shrneme ostatní zajímavé rozdíly mezi žánry.

### 2.1 Srovnání žánrů podle frekvence syntaktických funkcí

V tabulkách výše již byly celkové frekvence syntaktických funkcí uvedeny, pro přehlednost je však uvádíme ještě jednou v jednodušší tabulce: frekvence syntaktických funkcí substantiv v subkorpusech **BEL**, **ODB** a **PUB** a v korpusu **SYN2005**. **Červeně** je zvýrazněna nejčastější syntaktická funkce.

Frekvence syntaktických funkcí substantiv v korpusu SYN2005 a jeho subkorpusech

	Sb	Pnom	Atr	Obj	Adv	ExD	SUM
<b>BEL</b>	21,5	2,4	22,5	<b>24,1</b>	23,2	6,3	100 %
<b>ODB</b>	18,4	1,9	<b>35,6</b>	19,7	18,5	5,8	100 %
<b>PUB</b>	21,0	1,8	<b>33,4</b>	19,2	19,0	5,7	100 %
<b>SYN05</b>	20,4	2,0	<b>30,5</b>	20,9	20,2	5,9	100 %

Největší rozdíl mezi žánry je ve frekvenci substantivních přívlastků. Zastoupení funkce **Atr** v subkorpusu **ODB** je o 58 % vyšší než v subkorpusu **BEL**. To odpovídá větší tendenci k nominalizaci, kterou pozorujeme u odborné literatury a jen o málo méně i v publicistice (viz dále statistiku výskytu rozsáhlých nominálních skupin). Naopak v subkorpusu **BEL** je výrazně vyšší podíl předmětů a příslovečných určení.

Vyšší podíl větných členů rozvíjejících slovesa (**Sb**, **Obj**, **Adv**) mezi substantivy v **BEL** oproti **ODB** a **PUB** vyplývá z poměru slovních druhů v subkorpusech, který představujeme v následující tabulce. V subkorpusu **ODB** je více než dvojnásobný poměr substantiv ku slovesům než v korpusu **BEL**, subkorpus **PUB** je v tomto ohledu bližší **ODB** než **BEL**. Údaje přebíráme z publikace Statistiky češtiny (Bartoň a kol. 2009, s. 130). V subkorpusu

**BEL** jsou tedy substantiva samozřejmě mnohem častěji závislá na slovesech než v druhých dvou subkorpusech.

Podíl slovních druhů (sloves a substantiv) v korpusu SYN2005 a jeho subkorpusech

	substantiva	slovesa	ostatní	celkem	poměr subst.: slovesa
<b>BEL</b>	24,3	21,2	54,5	100 %	1,15 : 1
<b>ODB</b>	34,5	13,9	51,6	100 %	2,48 : 1
<b>PUB</b>	33,8	16,0	50,2	100 %	2,11 : 1
<b>SYN05</b>	30,2	17,5	52,3	100 %	1,73 : 1

V subkorpusu **BEL** je tedy více sloves a jsou tu jednodušší nominální skupiny, v **PUB** je méně sloves a nominální skupiny jsou složitější, v **ODB** je sloves nejméně a nejvíce se zde vyskytují složité nominální skupiny. Typické věty v jednotlivých subkorpusech by (velmi zjednodušeně) mohly vypadat takto (**zeleně** vyznačena slovesa a reflexivum, **modře** a **červeně** jednotlivé nominální skupiny):

**BEL:** *Mladý muž miloval svou Bělu z celé duše .*

**ODB:** *Průkaz účasti složek imunitního systému v patogenezi akutních koronárních příhod a chronické ischemické choroby srdeční byl hlavním podnětem pro studium jejich prognostického významu [ 27 ] .*

**PUB:** *Podle nejnovějších informací norské policie se počet obětí úterní srážky dvou vlaků v Norsku zvýšil na 19 .*

Největší rozdíl v zastoupení syntaktických funkcí mezi subkorpusem **ODB** a **PUB** je v podílu substantiv s funkcí **Sb**. V **PUB** je o 14 % více substantiv s funkcí **Sb** než v **ODB** (21,0 % substantiv v **PUB** oproti 18,4 % v **ODB**). Příčinu tohoto rozdílu je třeba hledat v charakteru textů: subkorpus **PUB** se skládá z většího množství velmi krátkých textů – jednotlivých novinových článků, v nichž je třeba vždy znovu představit subjekt a několikrát ho opakovat. Subkorpus **ODB** se skládá z textů delších, kde tak časté použití podmětu není nutné.

## 2.2 Nominální skupiny v subkorpusech BEL, ODB a PUB

V předcházejícím výkladu bylo zmíněno, že v subkorpusu **ODB** je výrazně vyšší tendence k nominalizaci než v subkorpusu **BEL**. Toto tvrzení můžeme doložit i statistikou výskytu nominálních skupin s vysokým počtem substantiv. V korpusu jsme vyhledali jmenné fráze

složené z jednoho substantiva následovaného shodnými přívlastky (syntaktickými adjektivy), substantivy v genitivu a předložkovými frázemi se substantivem, neoddělené žádným slovesem, spojkou ani jinou potenciální hranicí nominální fráze. Vyhledané skupiny byly rozříděny podle počtu substantiv v nich obsažených. Požadavky byly restriktivní, nebyly tak zdaleka vyhledány všechny nominální fráze s vyšším počtem substantiv. Některé vyhledané nominální fráze naopak ve skutečnosti netvoří jeden celek, pro srovnání počtu rozsáhlých nominálních skupin v žánrových subkorpusech však tento postup vyhovuje. V tabulce uvádíme počet výskytů nominálních skupin o daném počtu substantiv přepočítaný na 1 milión textových pozic (tokenů).

Počet nominálních skupin o daném počtu substantiv na 1 milión pozic

substantiv	10	9	8	7	6	5
<b>BEL</b>	0,1	0,3	0,6	3,2	17,3	87,0
<b>ODB</b>	1,3	5,5	14,1	49,3	167,2	591,8
<b>PUB</b>	0,8	2,8	10,4	36,9	135,6	484,3
<b>SYN05</b>	0,7	2,5	7,4	26,4	95,7	350,6

Z tabulky je zřejmé, že nejvyšší počet rozsáhlých nominálních skupin je v subkorpusu **ODB**, v subkorpusu **PUB** jich je méně, ale rozdíl je mnohem menší než rozdíl oproti subkorpusu **BEL**, kde se nominální skupiny s vyšším počtem substantiv vyskytují jen vzácně.

### 2.3 Srovnání žánrů v korpusu podle realizace verbonominálního predikátu (*Pnom*)

Syntaktická funkce verbonominální predikát (*Pnom*) je v korpusu nejméně frekventovanou funkcí (z funkcí zahrnutých do zpracování). Mezi žánry se použití substantiv s funkcí *Pnom* značně liší, a to jak v celkové frekvenci, tak poměrem mezi pády: nominativem a instrumentálem. Přehled použití funkce *Pnom* představujeme opět v přehledné tabulce.

V tabulce jsou uvedeny údaje o realizaci funkce *Pnom* pády substantiv v jednotlivých žánrech. V prvních čtyřech sloupcích jsou relativní frekvence realizace syntaktické funkce pádem (počet výskytů v subkorpusu přepočítaný na 1 milión substantiv: ppm). V druhých čtyřech sloupcích jsou poměry realizací funkce *Pnom* pády v procentech (všechny výskyty substantiv s funkcí *Pnom* v daném subkorpusu odpovídají 100 %). **Červeně** jsou zvýrazněny nejvyšší hodnoty, **modře** hodnoty nejnižší.

## Realizace funkce **Pnom** pády substantiv

	BEL	ODB	PUB	SYN05	BEL	ODB	PUB	SYN05
<b>Pnom_1</b>	<b>17 179</b>	<b>7 073</b>	9 362	11 193	<b>71,2</b>	<b>36,9</b>	53,1	55,4
<b>Pnom_2</b>	126	96	69	95	0,5	0,5	0,4	0,5
<b>Pnom_7</b>	<b>6 820</b>	<b>12 000</b>	8 204	8 909	<b>28,3</b>	<b>62,6</b>	46,5	44,1
	<b>24 126</b>	19 169	<b>17 636</b>	20 198	100 %	100 %	100 %	100 %

Nejvíce se substantiva s funkcí **Pnom** používají v subkorpusu beletrie, o 37 % více než v subkorpusu publicistiky. V subkorpusu **BEL** je funkce **Pnom** realizována většinou (71,2 %) nominativem, v subkorpusu **ODB** většinou instrumentálem (62,6 %), v subkorpusu **PUB** je poměr víceméně vyrovnaný, mírně převažuje realizace nominativem (53,1 %).

Jmenné části verbonominálního přísudku se mezi žánry liší i lexikálně, proto uvádíme i dva seznamy nejčastějších tvarů substantiv s funkcí **Pnom** v subkorpusech. První seznam představuje deset nejčastějších tvarů substantiv s funkcí **Pnom** v nominativu v každém z žánrových subkorpusů, druhý seznam uvádí deset nejčastějších tvarů substantiv s funkcí **Pnom** v instrumentálu. Číselné údaje uvádějí počet výskytů daného tvaru v této funkci na 1 milión substantiv v korpusu (ppm). **Modře** jsou zvýrazněny tvary, které se mezi prvními deseti vyskytují jen v jednom žánru.

Deset nejčastějších tvarů substantiv v nominativu s funkcí **Pnom** v subkorpusech

**BEL:** *pravda* (658), *člověk* (409), *věc* (273), ***žena*** (118), ***muž*** (114), *otázka* (108), ***nápad*** (97), ***práce*** (91), ***místo*** (91), *lidé* (89).

**ODB:** *pravda* (136), *otázka* (71), *člověk* (60), *problém* (57), *věc* (55), ***způsob*** (44), ***proces*** (39), *doba* (35), *záležitost* (32), *lidé* (32).

**PUB:** *pravda* (273), *věc* (145), *problém* (131), *otázka* (124), *člověk* (109), ***škoda*** (70), *záležitost* (61), ***důvod*** (49), *lidé* (48), *doba* (47).

Deset nejčastějších tvarů substantiv v instrumentálu s funkcí **Pnom** v subkorpusech

**BEL:** *součástí* (140), *cílem* (81), *výsledkem* (66), *příčinou* (62), ***svědkem*** (60), ***zvykem*** (54), ***dílem*** (52), *členem* (51), *důvodem* (49), ***úkolem*** (45).

**ODB:** *součástí* (553), *cílem* (416), *výsledkem* (314), *příkladem* (227), *příčinou* (223), ***základem*** (190), ***předmětem*** (186), ***zdrojem*** (178), *důvodem* (150), *problémem* (138).

**PUB:** *součástí* (363), *důvodem* (232), *cílem* (232), *příčinou* (133), *výsledkem* (131), *členem* (106), *problémem* (100), ***otázkou*** (92), *příkladem* (84), ***výjimkou*** (79).

## 2.4 Srovnání žánrů v korpusu podle užití přívlastku v prostém genitivu (*Atr\_2*)

Další oblastí, v níž se subkorporusy značně liší, jsou substantiva s funkcí *Atr* v prostém genitivu. Odlišná je jak celková frekvence substantiv *Atr\_2* (v subkorpusu **ODB** je téměř dvojnásobné zastoupení substantiv *Atr\_2* oproti **BEL**), tak jejich zapojení do větných struktur, tj. řídicí větné členy těchto přívlastků.

Mezi substantivy s funkcí *Atr* v prostém genitivu dokážeme automaticky rozlišit čtyři typy zapojení přívlastku do větné struktury. Frekvenci těchto typů uvádíme v přehledné tabulce níže. Automaticky rozlišujeme substantiva s funkcí shodného přívlastku: *pana/Atr Nováka* či *řeky Vltavy/Atr* (v tabulce pod názvem „shodný *Atr*“), substantiva, která jsou jako neshodný přívlastek závislá na číslovce: *120 druhů/Atr* (v tabulce pod názvem „číslovka“) nebo na substantivním kvantifikátoru: *spousta přátel/Atr* („**N kvant.**“) a substantiva závislá na slovesném substantivu: *vyplnění dotazníku/Atr* („**sloves. N**“). Přestože tyto čtyři skupiny představují ve všech žánrech jen menší část substantiv *Atr\_2* (okolo 30 %), je statistika těchto podtypů genitivního přívlastku pro srovnání žánrů užitečná. V prvních čtyřech sloupcích uvádíme frekvence přepočítané na 1 milión substantiv (ppm), v dalších čtyřech sloupcích jsou podíly jednotlivých typů na všech genitivních přívlastcích v procentech (základem je počet výskytů substantiv *Atr\_2* v subkorpusu). **Červeně** jsou zvýrazněny nejvyšší hodnoty, **modře** hodnoty nejnižší.

Frekvence substantivních přívlastků v prostém genitivu podle jejich řídicích větných členů

	BEL	ODB	PUB	SYN05	BEL	ODB	PUB	SYN05
shodný <i>Atr</i>	1 569	<b>1 124</b>	<b>4 561</b>	2 542	1,3	<b>0,5</b>	<b>2,4</b>	1,4
číslovka	10 351	<b>9 588</b>	<b>20 906</b>	14 057	8,8	<b>4,1</b>	<b>11,1</b>	7,8
<b>N kvant.</b>	<b>8 535</b>	<b>14 445</b>	10 997	11 242	<b>7,2</b>	6,2	<b>5,8</b>	6,3
sloves. <b>N</b>	<b>8 442</b>	<b>40 340</b>	20 174	22 496	<b>7,2</b>	<b>17,4</b>	10,7	12,6
ostatní	<b>88 998</b>	<b>166 645</b>	132 285	128 757	<b>75,5</b>	71,8	<b>70,0</b>	71,9
celkem	<b>117 895</b>	<b>232 142</b>	188 923	179 094	100 %	100 %	100 %	100 %

Jak bylo řečeno výše, podíl substantiv *Atr\_2* na substantivech v **BEL** je téměř dvakrát nižší než v **ODB**.

Shodné přívlastky (často křestní jména závislá na následujícím příjmení: *Václava/Atr Havla*) mají nejvyšší frekvenci i nejvyšší podíl na přívlastcích v subkorpusu **PUB**. V subkorpusu **PUB** jsou také nejvíce zastoupeny neshodné přívlastky v genitivu plurálu závislé na číslovkách.

Neshodné přívlastky rozvíjející substantivní kvantifikátory (*většina, řada; litr, kilogram;*

hejno, sklenice) jsou nejčastější v **ODB**. Kromě přibližných substantivních kvantifikátorů (většina, řada, množství) a měr (kilogram, litr, m<sup>3</sup>) jsou v subkorpusu **ODB** časté i „kuchyňské“ kvantifikátory (sklenice, lžíce, špetka aj.), protože do tohoto subkorpusu patří i texty kuchařek. Nejvyšší poměr substantivních kvantifikátorů v rámci genitivních přívlastků je v **BEL**.

V subkorpusu **ODB** jsou výrazně frekventovanější neshodné přívlastky rozvíjející slovesná substantiva (na -ní/tí): mezi všemi substantivy jsou v **ODB** téměř pětkrát častější než v **BEL** a dvakrát častější než v **PUB**, vyšší je i jejich podíl na řídicích větných členech **Atr\_2**. Častější užívání slovesných substantiv v **ODB** je dáno snahou o hutný, kondenzovaný výklad. Nejčastějšími slovesnými substantivy v **ODB** jsou: řešení, řízení, využití, použití, zvýšení.

Nejčastějšími řídicími větnými členy přívlastků, které patří do kategorie „ostatní“, jsou převážně abstraktní substantiva (v **ODB** a **BEL**), v **PUB** navíc pojmenování funkcí (ministr, ředitel, mluvčí). Subkorpusy se i v této oblasti značně liší, proto uvádíme seznam deseti nejčastějších takových substantiv v každém ze subkorpusů. Průnik mezi žánry je poměrně malý, slova, která se vyskytují mezi nejčastějšími u dvou subkorpusů, zvýrazňujeme **červeně**.

Seznam nejčastějších řídicích substantiv neshodných **Atr\_2** (mimo kvantifikátory a deverbativa)

**BEL:** **konec** (995), strana (567), **doba** (527), pocit (491), den (468), **člen** (427), dveře (393), jméno (371), okraj (362), místo (338).

**ODB:** oblast (1847), způsob (1264), **doba** (1199), systém (1151), typ (1105), možnost (1074), druh (1053), forma (1050), změna (1004), vývoj (916).

**PUB:** ministr (1128), **konec** (1116), ředitel (998), ministerstvo (975), **člen** (911), mluvčí (906), cena (901), šéf (834), předseda (747), rok (731).

## 2.5 Předmět v prostém instrumentálu (**Obj\_7**)

Syntaktická funkce **Obj** v prostém instrumentálu zahrnuje dva typy funkce: valenční předmět a původce děje u opisného pasiva. I v této oblasti se žánrové subkorpusy značně různí. Rozdíly představujeme opět v tabulce, kde jsou substantiva **Obj\_7** v každém subkorpusu rozděleny do čtyř kategorií. První kategorie zahrnuje valenční předměty sloves, např. stát se, zabývat se, trpět, pohnout, vrtět aj. (v tabulce „**val. Obj V**“). Ve druhé kategorii jsou valenční předměty adjektiv, např. disponující, známý, páchnoucí, trpící, zabývající se („**val. Obj A**“).

Třetí kategorie obsahuje původce děje u opisného pasiva, např. *byl odsouzen soudem/Obj* („pův. děje V“), čtvrtá kategorie zahrnuje původce děje u adjektiv odvozených od trpných příčestí, např. *vydaný nakladatelstvím/Obj* („pův. děje A“). V prvních čtyřech sloupcích jsou uvedené počty výskytů přepočítané na 1 milión substantiv v subkorpusu (ppm), ve druhých čtyřech sloupcích jsou procentuální údaje (základem je počet substantiv *Obj\_7* v subkorpusu). **Červeně** jsou zvýrazněny nejvyšší hodnoty, **modře** hodnoty nejnižší.

Frekvence substantiv s funkcí *Obj* v prostém instrumentálu podle podtypů funkce

	BEL	ODB	PUB	SYN05	BEL	ODB	PUB	SYN05
val. <i>Obj</i> V	<b>9 169</b>	5 669	<b>5 431</b>	6 673	<b>68,7</b>	<b>45,5</b>	66,6	59,9
val. <i>Obj</i> A	147	247	144	178	1,1	2,0	1,8	1,6
pův. děje V	2 288	<b>4 638</b>	<b>1 562</b>	2 730	<b>17,2</b>	<b>37,2</b>	19,2	24,5
pův. děje A	1 737	<b>1 904</b>	<b>1 015</b>	1 554	13,0	<b>15,3</b>	<b>12,5</b>	14,0
celkem	<b>13 341</b>	12 458	<b>8 152</b>	11 135	100 %	100 %	100 %	100 %

Z tabulky vidíme, že substantiva *Obj\_7* mají celkově nejvyšší podíl mezi substantivy v subkorpusu **BEL**, a to především jako valenční předměty sloves, zatímco v **PUB** se používají relativně málo. V subkorpusu **ODB** vyjadřuje nadpoloviční většina substantiv *Obj\_7*, celkem 52,5 %, původce děje u pasiva (u sloves v trpném rodě i u slovesných adjektiv). Tento poměr je dvakrát vyšší než u druhých dvou subkorpusů.

## 2.6 Další rozdíly mezi žánrovými subkorpusy

Na závěr srovnání žánrových subkorpusů stručně uvedeme několik dalších rozdílů, jež byly zvýrazněny v tabulkách syntaktických funkcí. Předně je to velký rozdíl ve frekvenci substantivních předmětů v prostém akuzativu (*Obj\_4*): 15,5 % v **BEL**, 11,9 % v **ODB** a 11,8 % v **PUB**. Rozdíl vyplývá z většího podílu sloves v **BEL**: kromě *Sb\_1* je *Obj\_4* nejčastější realizace syntaktické funkce závislé na slovesech. V **PUB** je ještě o desetinu procenta méně *Obj\_4* než v **ODB** především kvůli vyšší frekvenci podmětu v **PUB** (důvod uveden výše).

Velký rozdíl mezi žánry je i v použití substantiv ve vokativu (celkově i jako *ExD\_5*). V korpusu **BEL** je přibližně čtyřikrát vyšší podíl substantiv ve vokativu na všech substantivech oproti **ODB** a **PUB**. Poměr vokativu je díky dialogům v beletrii obecně vyšší, nejvíce je pak zastoupen v divadelních hrách.

Mezi přívlastky je v subkorpusu **PUB** výrazně vyšší procento nominativu: 13,3 % přívlastků v **PUB** oproti 6,2 % v **BEL** a 7,1 % v **PUB**. Tento rozdíl je dán jak vyšším

poměrem shodných přívlasků (*ministr/Atr vnitra Gross*), tak vyšším poměrem nominativů jmenovacích (*divadlo Semafor/Atr*) v subkorpusu **PUB**.

Posledním výrazným rozdílem mezi žánry, který jsme zvýraznili v tabulkách syntaktických funkcí, jsou poměry využití některých předložkových pádů (**p2**, **p4**, **p6**) pro realizaci přívlasku a příslovečného určení. V **BEL** výrazně převažují příslovečná určení, poměr **Adv : Atr** u těchto tří předložkových pádů celkem je 2,8 : 1; v **ODB** a **PUB** je poměr vyrovnanější (1,8 : 1, resp. 1,9 : 1). Tento rozdíl opět vyplývá z většího zastoupení složitých nominálních skupin v subkorpusech **ODB** a **PUB**.

## 2.7 Závěr

Texty v korpusu SYN2005 jsou rozděleny do tří žánrových subkorpuseů beletrie, odborné literatury a publicistiky. Toto rozdělení nemá ostré hranice, v subkorpusech **BEL** i **ODB** se například objevují i texty publicistického charakteru (např. odborné, zájmové, cestopisné časopisy). Přesto se rozdíly mezi žánry jasně projevují v rozdílech ve frekvenci syntaktických funkcí a jejich realizací prostými a předložkovými pády substantiv. I v rámci jednotlivých kombinací syntaktické funkce a pádu (jako je přívlask v prostém genitivu) jsou značné rozdíly mezi žánry v tom, jakým způsobem se tato substantiva začleňují do větných struktur.

## 3. Srovnání frekvencí syntaktických substantiv v korpusu SYN2005 a v Kvantitativních charakteristikách současné češtiny

Tato práce by nebyla úplná, kdyby se alespoň stručně nevyrovnala se starší prací na stejné téma: s příslušnými pasážemi publikace Marie Těšitelové a kolektivu: *Kvantitativní charakteristiky současné češtiny* (Těšitelová, 1985), dále jen **KCh**. Představíme tuto práci a korpus, z něhož vychází, a srovnáme tabulky syntaktických funkcí substantiv se zřetelem k pádu, které jsme uvedli výše, s tabulkami z **KCh**.

### 3.1 Kvantitativní charakteristiky současné češtiny

*Kvantitativní charakteristiky současné češtiny* je publikace, která vyšla v roce 1985 jako statistický doplněk Mluvnice češtiny. Obsahuje soubor komentovaných statistických údajů o českém jazyce: o jeho fonologické, grafematické, lexikální a gramatické (morfologické a syntaktické) rovině. Většina statistických údajů pochází z manuálně anotovaného korpusu věcného stylu o rozsahu 540 000 slov. Tento korpus byl vytvořen ze 180 úryvků po 3000 slovech, z toho 56 % textů pocházelo z odborného stylu, 33 % z publicistického stylu a 11 %

ze stylu administrativního. Lexikum a morfologie byly zčásti zpracovány na základě staršího materiálu použitého pro frekvenční slovník češtiny (Jelínek a kol. 1961). Pro popis jazyka na (povrchově) syntaktické rovině vycházeli autoři ze syntaktické koncepce Vl. Šmilauera (Šmilauer 1966).

Syntaktickým funkcím větných členů se věnuje výklad na s. 100–115. Syntaktické funkce substantiv se zřetelem k jejich pádu pak autoři představují ve dvou tabulkách (č. 35 a 36). Údaje jsou v procentech, základem pro výpočet procent je v první tabulce počet výskytů substantiv v určitém pádu, v druhé je základem počet výskytů substantiv s určitou syntaktickou funkcí. Syntaktické funkce v **KCh.** jsou: **S**: subjekt; **Pvn**: jmenná část sponového predikátu; **O**: objekt; **Ad**: adverbiale; **At**: atribut **Ap**: apozice; **Atv**: doplněk; **Z**: základ věty jednočlenné a **jiné**: ostatní syntaktické funkce, např. oslovení u vokativu. **KCh.** neobsahují podrobný popis použitých syntaktických funkcí (odvolávají se na koncepci Vl. Šmilauera), u některých kombinací se lze jen dohadovat (např. podmět v instrumentálu: pravděpodobně typ „*Pantáta s panímámou/S pracují ve fabrice.*“).

Autoři se rozhodli zanedbat v tabulkách rozdíl mezi prostými a předložkovými pády, přestože přinejmenším pro část textů byly údaje zpracovány i se zahrnutím tohoto rozdílu: pro některé funkce jsou dílčí statistické údaje uvedeny v dřívější, přípravné publikaci (Těšitelová et al. 1983, s. 78–79).

Tabulky uvádíme s drobným rozdílem v pořadí (**Ap** jsme posunuli o dva sloupce dále). Tučně jsou zvýrazněny syntaktické funkce, které budou předmětem srovnání v dalším oddíle.

Kvantitativní charakteristiky současné češtiny (tab. č. 35)

Frekvence syntaktických funkcí u substantiv (v %)										
pád	<b>S</b>	<b>Pvn</b>	<b>At</b>	<b>O</b>	<b>Ad</b>	Ap	Atv	Z	jiné	$\Sigma$
<b>N</b>	<b>71,83</b>	<b>4,41</b>	<b>7,05</b>	<b>0,02</b>	<b>1,50</b>	9,09	0,54	4,02	1,54	<b>100,00</b>
<b>G</b>	<b>0,19</b>	<b>0,26</b>	<b>80,56</b>	<b>3,12</b>	<b>12,85</b>	2,60	0,25	0,06	0,11	<b>100,00</b>
<b>D</b>	-	<b>0,84</b>	<b>21,14</b>	<b>55,15</b>	<b>19,05</b>	3,42	0,04	0,27	0,09	<b>100,00</b>
<b>A</b>	-	<b>0,09</b>	<b>12,19</b>	<b>70,66</b>	<b>12,17</b>	3,51	0,61	0,50	0,27	<b>100,00</b>
<b>V</b>	-	-	-	-	-	20,20	-	3,54	76,26	<b>100,00</b>
<b>L</b>	<b>0,02</b>	<b>1,34</b>	<b>29,24</b>	<b>5,92</b>	<b>57,60</b>	4,43	0,90	0,08	0,47	<b>100,00</b>
<b>I</b>	<b>0,39</b>	<b>13,66</b>	<b>23,60</b>	<b>18,79</b>	<b>36,60</b>	3,71	2,75	0,18	0,32	<b>100,00</b>
<b><math>\Sigma</math></b>	<b>17,1</b>	<b>2,48</b>	<b>37,82</b>	<b>17,68</b>	<b>17,92</b>	4,66	0,66	1,10	0,62	<b>100,00</b>

Frekvence syntaktických funkcí u substantiv (v %)										
pád	S	Pvn	At	O	Ad	Ap	Atv	Z	jiné	Σ
N	99,41	41,96	4,40	0,03	1,98	46,01	19,20	86,79	58,33	23,61
G	0,39	3,62	72,94	6,03	24,55	19,07	12,94	1,82	6,32	34,24
D	-	1,25	2,08	11,58	3,95	2,72	0,25	0,91	0,53	3,71
A	-	0,65	5,59	69,29	11,77	13,05	16,03	7,85	7,57	17,34
V	-	-	-	-	-	0,48	-	0,35	13,45	0,11
L	0,01	6,88	9,82	4,26	40,83	12,08	17,19	0,86	9,53	12,70
I	0,19	45,64	5,17	8,81	16,92	6,59	34,39	1,42	4,27	8,29
Σ	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

### 3.2 Srovnání frekvencí syntaktických funkcí substantiv v SYN2005 a v KCh.

Pro přehlednost uvádíme srovnání statistik z *Kvantitativních charakteristik současné češtiny* a statistik uvedených v této práci ve dvou společných tabulkách. Srovnáváme jen pět syntaktických funkcí, které jsou zpracované podobně v obou korpusech: **S**, **Pvn**, **At**, **O** a **Ad**.

#### 3.2.1 Přizpůsobení údajů o syntaktických funkcích substantiv z této práce Kvantitativním charakteristikám

Pro snazší porovnání jsme alespoň částečně přizpůsobili data v tabulkách z korpusu SYN2005: shodné přívlastky typu *pan/Atr Novák* byly z přívlastku vyřazeny (předpokládáme, že se v KCh. řadí mezi Ap nebo „jiné“), frekvence určení původce děje se místo v kombinaci **Obj\_7** počítají v **Adv\_7**. Substantiva **Obj\_1** a **Adv\_1** počítáme společně jako **Adv\_1**.

Jiné koncepční rozdíly ale odstranit nelze, např. není možné převést předložkové fráze závislé na sponě *být* z funkce **Adv** do **Pnom**, protože chybějí data, která by umožnila automaticky rozlišit mezi okolnostním určením a skutečnými součástmi verbonominálního predikátu (*Byli jsme venku jen na chvíli/Adv* oproti *Doutníky byly na dosah/Pnom*). Zásadním koncepčním rozdílem je také to, že zachováváme rozdělení pádů na prosté a předložkové. Při srovnání je tedy třeba porovnávat jeden údaj z **KCh.** se dvěma údaji (nebo jejich součtem) ze syntakticky anotovaného korpusu SYN2005.

Údaje jsme zaokrouhlili na jedno desetinné místo. Přesnost na dvě desetinná místa není v syntakticky značkováném jazykovém korpusu relevantní a tabulky jsou díky tomu přehlednější.

### 3.2.2 Společné tabulky syntaktických funkcí substantiv ze SYN2005 a z KCh

V tabulkách uvádíme ve sloupcích vedle sebe syntaktické funkce z korpusu **SYN2005** (**modře** zvýrazněno) a syntaktické funkce z **KCh.** (zvýrazněno **červeně**); zkratky syntaktických funkcí ponecháváme beze změny. V řádcích jsou uvedeny pády (označeny číslicí, ne zkratkou pádu; rozlišujeme pády prosté a předložkové), ve sloupcích z **KCh.** je pro prostý i předložkový pád dohromady jedno číslo.

V první tabulce je základem pro výpočet procent celkový počet výskytů substantiv s daným pádem (u SYN2005 dohromady prostým i předložkovým pádem). Sloupce z **KCh.** obsahují vybrané údaje z tabulky č. 35 v **KCh.** **Červeně** jsou zvýrazněny rozdíly vyplývající z rozdílné koncepce, **modře** největší rozdíly v údajích. Vzhledem k tomu, že některé syntaktické funkce byly z tabulky vyřazeny, není číslo v pravém sloupci součtem údajů v tabulce, ale součtem všech výskytů daného pádu substantiv v analyzovaném korpusu.

Syntaktické funkce substantiv v **SYN2005** a v **KCh.** (základ je výskyt pádu)

	<b>Sb</b>	<b>S</b>	<b>Pnom</b>	<b>Pvn</b>	<b>Atr</b>	<b>At</b>	<b>Obj</b>	<b>O</b>	<b>Adv</b>	<b>Ad</b>	<b>Σ</b>	<b>Σ</b>
1	76,4	71,8	4,2	4,4	<b>3,3</b>	<b>7,1</b>	0	0	<b>0,24</b>	<b>1,5</b>	100	100
2	0,1	0,2	0	0,3	66,1	80,6	<b>3,8</b>	<b>3,1</b>	<b>0,71</b>	<b>12,9</b>	71,7	100
p2	0		0		8,9		<b>2,0</b>		<b>17,3</b>		28,3	
3	0	0	0	<b>0,8</b>	<b>1,4</b>	<b>21,1</b>	42,2	55,2	0,33	19,1	46,2	100
p3	0		0		<b>10,6</b>		16,8		26,6		53,8	
4	0	0	0	<b>0,1</b>	0	12,2	65,0	70,7	3,32	12,2	70,0	100
p4	0		0		7,4		8,4		14,1		30,0	
5	0	0	0	0	0,2	0	0	0	0,60	0	100	100
p6	0	0	0	<b>1,3</b>	26,4	29,2	5,8	5,9	65,8	57,6	100	100
7	0	<b>0,4</b>	10,6	13,7	3,6	23,6	8,2	18,8	<b>26,2</b>	<b>36,6</b>	50,5	100
p7	0		0		20,9		6,7		<b>19,5</b>		49,5	
<b>Σ</b>	20,9	17,1	2,1	2,5	28,7	37,8	21,1	17,7	21,2	17,9	100	100

V druhé tabulce je základem pro výpočet procent celkový počet výskytů substantiv s danou syntaktickou funkcí. Sloupce z **KCh.** obsahují vybrané údaje z tabulky č. 36 v Kvantitativních charakteristikách současné češtiny. **Červeně** jsou zvýrazněny rozdíly vyplývající z rozdílné koncepce, **modře** největší rozdíly v údajích.

Syntaktické funkce substantiv v **SYN2005** a v **KCh.** (základ je výskyt syntaktické funkce)

	<b>Sb</b>	<b>S</b>	<b>Pnom</b>	<b>Pvn</b>	<b>Atr</b>	<b>At</b>	<b>Obj</b>	<b>O</b>	<b>Adv</b>	<b>Ad</b>	<b>Σ</b>	<b>Σ</b>
1	99,9	99,4	<b>55,4</b>	<b>42,0</b>	3,1	4,4	0,0	0	<b>0,3</b>	<b>2,0</b>	27,4	23,6
2	0,2	0,4	0,5	<b>3,6</b>	63,3	72,9	4,8	6,0	0,9	24,6	19,6	34,2
p2	0		0		8,5		2,6		22,8		7,9	
3	0	0	0	<b>1,3</b>	0,2	2,1	7,7	11,6	<b>0,1</b>	<b>4,0</b>	1,8	3,7
p3	0		0		1,5		3,1		<b>5,0</b>		2,2	
4	0	0	0	<b>0,7</b>	0	5,6	62,3	69,3	<b>3,3</b>	<b>11,8</b>	14,4	17,3
p4	0		0		5,3		8,0		<b>14,0</b>		6,2	
5	0	0	0	<b>0</b>	0	0	0,0	0	0,0	0	0,4	0,1
p6	0	0	0	<b>6,9</b>	10,7	9,8	3,2	4,3	36,8	40,8	11,6	12,7
7	0	<b>0,2</b>	44,1	45,6	1,1	5,2	3,3	8,8	10,9	16,9	4,3	8,3
p7	0		0		6,3		2,7		8,1		4,3	
<b>Σ</b>	100	100	100	100	100	100	100	100	100	100	100	100

Přes odlišné koncepty značkování syntaktických funkcí se obě tabulky ve většině oblastí víceméně shodují. Určitým nedostatkem **KCh.** je nerozlišování prostých a předložkových pádů, není jasné, co autory vedlo k zanedbání této důležité informace. Tabulky s rozlišením prostého a předložkového pádu nejsou o mnoho složitější nebo méně přehledné a potřebné údaje měli autoři k dispozici.

Některé závěry v **KCh.** se jeví ve světle rozsáhlejších analýz jako poněkud ukvapené, např. (Těšitelová 1985, s. 63): „Z hlediska syntaktických funkcí vzato je v české větě relativně nejčastější syntaktická funkce atributu (v 38 %): nejčastěji ji pak zastává **G** (v 81 %)“. Jak jsme ukázali výše, toto tvrzení může přibližně platit pro tzv. „věcný styl“, v beletrii je podíl přívlastku výrazně nižší (22,5 %), nejčastější syntaktickou funkcí je předmět (24,1 %). Závislost na charakteru analyzovaného textu si autoři uvědomují: „Rozdíly ve frekvenci pádů u substantiv a v syntaktických funkcích, které zastávají, jsou v češtině citlivé na faktory objektivní i subjektivní, tj. na funkční styl a druh projevu, ale i na individuální styl autora, na tematiku apod.“ (Těšitelová 1985, s. 63), ale výše uvedený výrok měli formulovat opatrněji.

#### 4. Závěr

Ve druhé části práce jsme představili syntaktické funkce substantiv v korpusu SYN2005 a jeho žánrových subkorpusech. Podrobně jsme analyzovali jednotlivé realizace syntaktických funkcí prostými a předložkovými pády substantiv, kde to bylo užitečné, byly v rámci těchto kombinací vymezeny syntakticky nebo sémanticky zajímavé podtypy. Vše bylo doloženo bohatými příkladovým materiálem z korpusu. Dále jsme představili tabulky, v nichž byly

představeny frekvence syntaktických substantiv se zřetelem k pádům v celém korpusu i v jeho žánrových subkorpusech. Důležité rozdíly mezi subkorpusy jsme analyzovali podrobněji a zdůvodnili. Nakonec jsme srovnali výsledky práce se starší publikací: *Kvantitativními charakteristikami současné češtiny*.

## Závěr

V české literatuře dosud chyběly statistické údaje o frekvenci využití pádu substantiv k realizaci syntaktických funkcí, které by se zakládaly na rozsáhlém textovém korpusu. Rozhodli jsme se proto tento nedostatek napravit: získat z korpusu SYN2005 data o tom, jaké funkce realizují v češtině prosté a předložkové pády substantiv.

K tomuto účelu bylo nutné opatřit korpus syntaktickou anotací. Zvolili jsme formalismus analytické roviny Pražského závislostního korpusu a s pomocí stochastického MST parseru jsme každému textovému slovu v korpusu přiřadili řídicí větný člen a syntaktickou funkci. Výstup MST parseru však nebyl dostatečně spolehlivý, obsahoval příliš vysoké procento chyb. Na základě podrobného rozboru chyb parseru (manuálního i automatického) jsme zjistili, že chyby vykazují určitou pravidelnost a že značnou část těchto chyb by mělo být možné automaticky opravit.

Vytvořili jsme automatický opravný modul, který koriguje jednotlivé typy chybných struktur pomocí samostatných pravidel. Tato pravidla využívají rozsáhlé databáze znalostí o českém jazyce, které jsme získali z textových korpusů ČNK, vyhledávají a opravují chyby v určení závislostí a syntaktických funkcí. Pravidla zasahují v různých oblastech, opravují například základní strukturu hlavních a vedlejších vět v souvětí, závislosti a syntaktické funkce předložkových frází aj. Po implementaci dvaceti šesti pravidel však zlepšení výstupu zaostává za očekáváním, výstup parseru zlepšuje jen o 1,33 procentního bodu, tj. snižuje počet chyb pouze o cca 6 %. Opravný modul lze ovšem dále rozvíjet: je možné doplňovat další pravidla, rozšiřovat valenční seznamy i jiné seznamy vlastností slov a podobně. Opravný modul lze aplikovat na výsledky kteréhokoli závislostního parseru, který vychází z dat Pražského závislostního korpusu, a zlepšit tak jeho výsledky. Kromě vlastního vývoje opravného modulu jsme navrhli několik dalších postupů vhodných ke zlepšení kvality syntaktické anotace.

Korpus SYN2005, který jsme opatřili syntaktickou anotací s pomocí MST parseru a opravného modulu, posloužil jako základ k výzkumu vztahu formy a funkce substantiv. V korpusu jsme zkoumali frekvence syntaktických funkcí a jejich realizací prostými a předložkovými pády substantiv. Pro omezení vlivu chybovosti syntaktické anotace jsme použili rozsáhlé manuální sondy do korpusu, na jejichž základě jsme údaje korigovali.

Získali jsme poměrně spolehlivé údaje o frekvencích syntaktických substantiv ve vztahu k jejich prostému či předložkovému pádu. Podrobně jsme popsali jednotlivé realizace syntaktických funkcí prostými a předložkovými pády, vyčíslili jsme zastoupení jevů, které se

na každé kombinaci funkce a pádu podílejí. Ve čtyřech čtveřicích přehledných tabulek jsme představili frekvence syntaktických funkcí substantiv se zřetelem k jejich pádu, a to v celém korpusu SYN2005 a v jeho třech žánrových subkorpusech. Odlišný způsob zpracování dat v každém souboru tabulek umožnil poukázat na některé nuance ve vztahu mezi funkcemi a pády substantiv i na různé typy rozdílů mezi subkorpusem.

Syntaktická anotace korpusu nám také umožnila podrobněji porovnat funkční styly v korpusu SYN2005. Korpus se dělí na tři žánrové subkorpuse: beletrie, odborné literatury a publicistiky. Tyto subkorpuse jsme srovnali podle celkové frekvence syntaktických funkcí substantiv i podle některých dílčích jevů: realizace jmenné části verbonominálního predikátu nominativem nebo instrumentálem; způsob využití neshodných přívlastků v prostém genitivu; poměr valenčních předmětů v prostém instrumentálu a určení původce děje u opisného pasiva. Přestože hranice mezi subkorpusem v SYN2005 nejsou ostré, zaznamenali jsme značné rozdíly mezi subkorpusem ve všech analyzovaných jevech.

# Literatura

Bartoň Tomáš, Václav Cvrček, František Čermák, Tomáš Jelínek, Vladimír Petkevič (2009). *Statistiky češtiny*. Lidové noviny, Praha.

Český národní korpus – SYN2005. Ústav Českého národního korpusu FF UK, Praha 2005. Dostupný z WWW: <<http://www.korpus.cz>>.

Čechová Marie a kol. (2000). *Čeština – řeč a jazyk*. ISV, Praha.

Hajič Jan (2004a). *Complex Corpus Annotation: The Prague Dependency Treebank*. Jazykovedný ústav Ľ. Štúra, SAV, Bratislava.

Hajič Jan (2004b). *Disambiguation of Rich Inflection*. Karolinum, Praha.

Hajič Jan, Alena Böhmová, Eva Hajičová, Barbora Vidová Hladká (2003). The Prague Dependency Treebank: A Three-Level Annotation Scenario. In: *Treebanks: Building and Using Parsed Corpora*, Kluwer Academic Publishers, Amsterdam, s. 103–127.

Hajič Jan, Eva Hajičová, Jarmila Panevová, Petr Sgall (1998). Syntax v Českém národním korpusu. In: *Slovo a slovesnost*, 59, s. 168–177.

Hajič Jan, Jarmila Panevová, Eva Buráňová, Zdeňka Urešová, Alla Bémová (1999). *Anotace Pražského závislostního korpusu na analytické rovině: pokyny pro anotátory (html)*.

Hajič Jan, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, Magda Ševčíková Razímová (2006). *Prague Dependency Treebank 2.0, CD-ROM*. Linguistic Data Consortium, Philadelphia.

Holan Tomáš, Zdeněk Žabokrtský (2006). Combining Czech Dependency Parsers. In: *Lecture Notes in Computer Science, Proceedings of the 9th International Conference, TSD 2006*, Springer Verlag, Berlin/Heidelberg, s. 95–102.

Jelínek Jaroslav, Josef Václav Bečka, Marie Těšitelová (1961). *Frekvence slov, slovních druhů a tvarů v českém jazyce*. SPN, Praha.

Jelínek Tomáš (2008). Nové značkování v Českém národním korpusu. In: *Naše řeč*, 91, s. 13–20.

Jelínek Tomáš, Vladimír Petkevič (2011). Systém jazykového značkování současné psané češtiny. In: *Korpusová lingvistika Praha 2011, sv. 3: Gramatika a značkování korpusů*. Nakladatelství Lidové noviny / Ústav českého národního korpusu, Praha, s. 154–170.

Jelínek Tomáš (2012). Automatic Rule-Based Correction of Stochastic Syntactic Annotation of Czech. In: *Slavic Languages in Formal Grammar. Proceedings of FDSL 8.5*. Peter Lang, Frankfurt am Main, s. 65–81.

- Kuryłowicz Jerzy (1949). Le problème du classement des cas. In: *Biuletyn polskiego towarzystwa językoznawczego 9*, Kraków 1949, s. 20–43.
- Květoň Pavel (2006). *Rule-based Morphological Disambiguation*. Disertační práce, MFF UK, Praha.
- McDonald Ryan, Fernando Pereira, Kiril Ribarov, Jan Hajič (2005). Non-Projective Dependency Parsing using Spanning Tree Algorithms. In: *Proceedings of HLT-EMNLP 2005*. ACL, Vancouver, s. 523–530.
- Mluvnice češtiny 2 (1986). Academia, Praha.
- Mluvnice češtiny 3 (1987). Academia, Praha.
- Nivre Joakim, Johan Hall, Jens Nilsson (2006), MaltParser: A Data-Driven Parser-Generator for Dependency Parsing. In *Proceedings of LREC 2006*. ELRA, Genova, s. 2216–2219.
- Novák Václav, Zdeněk Žabokrtský (2007). Feature Engineering in Maximum Spanning Tree Dependency Parser. In: *Lecture Notes in Computer Science, Proceedings of the 10th International Conference TSD 2007*, Springer Verlag, Berlin/Heidelberg, s. 92–98.
- Panevová Jarmila (1980). *Formy a funkce ve stavbě české věty*. Academia, Praha.
- Sgall Petr (1967). *Generativní popis jazyka a česká deklinace*. Academia, Praha.
- Skalička Vladimír (1950). Poznámky k teorii pádů. In.: *Slovo a slovesnost*, 12, s. 134–152.
- Spoustová Johanka (2007). *Kombinované statisticko-pravidlové metody značkování češtiny*. Disertační práce, MFF UK, Praha.
- Šmilauer Vladimír (1966). *Novočeská skladba*. SPN, Praha.
- Těšitelová Marie a kol. (1983). *Kvantitativní charakteristiky současné odborné češtiny (v rámci věcného stylu)*. ÚJČ ČSAV, Praha.
- Těšitelová Marie a kol. (1985). *Kvantitativní charakteristiky současné češtiny*. Academia, Praha.
- Uličný Oldřich (2000). *Instrumentál v struktuře české věty*. Karolinum, Praha.
- Zeman Daniel, Zdeněk Žabokrtský (2005). Improving Parsing Accuracy by Combining Diverse Dependency Parsers. In: *Proceedings of the 9th IWPT*. ACL, Vancouver, s. 171–178.
- Žabokrtský Zdeněk, Jan Ptáček, Petr Pajas (2008). TectoMT: Highly Modular MT System with Tectogrammatcs Used as Transfer Layer. In *Proceedings of WMT'08*. Columbus, s. 167–170.