

MATEMATICKO-FYZIKÁLNÍ FAKULTA  
PRAHA

**DISKURZNÍ VZTAHY V ČEŠTINĚ A JEJICH  
ZACHYCENÍ V PRAŽSKÉM ZÁVISLOSTNÍM  
KORPUSU 2.0**

LUCIE MLADOVÁ

úfal/ckl technical report

TR-2008-40



UNIVERSITAS CAROLINA PRAGENSIS

DISKURZNÍ VZTAHY V ČEŠTINĚ  
A JEJICH ZACHYCENÍ  
V PRAŽSKÉM ZÁVISLOSTNÍM KORPUSU 2.0

Lucie Mladová

ÚFAL/CKL Technická zpráva TR-2008-40  
ISSN 1214-5521

Prosinec 2008

Copies of ÚFAL/CKL Technical Reports can be ordered from:

Institute of Formal and Applied Linguistics (ÚFAL MFF UK)

Faculty of Mathematics and Physics, Charles University

Malostranské nám. 25, CZ-11800 Prague 1

Czech Republic

or can be obtained via the Web: <http://ufal.mff.cuni.cz>

## Acknowledgement

Dokument vznikl za finanční podpory grantů Grantové agentury České republiky: GA 405/06/0589, GA 405/09/0729; grantu Ministerstva školství, mládeže a tělovýchovy: MŠMT ČR LC536 a grantu projektu Companions ISTFP6-034434<sup>1</sup>.

---

<sup>1</sup> This work was funded in part by the Companions project ([www.companionsproject.org](http://www.companionsproject.org)) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number ISTFP6-034434.

# Obsah

Acknowledgement.....	2
Obsah.....	3
Úvod.....	5
Kapitola 1. Teoretická východiska.....	8
Kapitola 2. Teorie, metody a terminologie.....	9
2.1 Závislostní a frázový syntaktický popis.....	9
2.1.1 Pražský závislostní korpus (PDT).....	9
2.1.1.1 Funkční generativní popis.....	9
2.1.1.2 Tektogramatická reprezentace v Pražském závislostním korpusu.....	10
2.1.2 Frázový syntaktický popis v Penn Treebanku (PTB).....	12
2.1.3 Penn Discourse Treebank a jeho anotace.....	14
2.1.3.1 Discourse connectives v PDTB a jejich klasifikace.....	15
2.1.3.2 Argumenty discourse connectives, abstraktní objekty.....	17
2.2 Textová lingvistika a diskurz.....	19
2.2.1 Zahraniční a české uvažování o diskurzu (textu).....	19
2.2.1.1 Diskurz a text v české lingvistice.....	19
2.2.1.2 Diskurz a text ve světové lingvistice.....	20
2.2.1.3 Pojmový aparát – textová lingvistika (textová syntax, hypersyntax, nadvětná syntax) versus discourse analysis.....	24
2.3 Textová koherence a prostředky jejího vyjádření – přehled.....	25
2.3.1 Aktuální členění, tematické posloupnosti.....	25
2.3.2 Koreference (anaforické řetězce, izotopické řetězce).....	26
2.3.3 Syntaktická struktura textu.....	26
2.3.4 Rétorické vztahy (obsahově sémantické vztahy, rematické vztahy).....	26
2.3.5 Temporální souvztažnosti.....	27
2.3.6 Grafické a zvukové členění textu.....	27
2.3.7 Komunikační a pragmatické faktory.....	27
2.4 Pojetí diskurzu ve Funkčním generativním popisu.....	28
2.4.1 Jaké vztahy zachycovat v popisu diskurzu?.....	28
2.4.2 Předmět zkoumání.....	29
2.4.3 Termíny vztahující se k popisu diskurzu.....	30
Kapitola 3. Rétorické vztahy v diskurzu.....	32
3.1 Segmented Discourse Representation Theory.....	32
3.2 Pojetí rétorických vztahů v české lingvistice.....	39
3.3 Příklady rétorických diskurzních vztahů mezi větami z Pražského závislostního korpusu.....	41
3.4 Shrnutí.....	43
Kapitola 4. Diskurzní vztahy zachycené v PDT.....	45
4.1 Syntaktická závislost.....	45
4.2 Souřadnost.....	47
4.3 Diskurzní vztahy signalizované funktorem PREC.....	49
4.4 Diskurzní vztahy uvnitř věty a přes hranice věty.....	49
Kapitola 5. Diskurzní konektory.....	51
5.1 Explicitní a implicitní vyjádření diskurzních vztahů.....	52
5.2 Tektogramatický funktor PREC.....	53
5.2.1 Základní charakteristika.....	53

5.2.2 Korpusové výzkumy – Analýza výrazů s funktorem PREC v PDT 2.0.....	55
5.2.3 K slovnědruhové charakteristice výrazů s funkcí PREC a diskurzních konektorů obecně.....	57
5.2.4 Binarita.....	61
5.2.5 „Antecedenty“ výrazů s funkcí PREC.....	63
5.2.6 Povaha „druhého“ argumentu.....	71
5.2.6.1 PREC bez funkce diskurzního konektoru – připojování členské.....	72
5.2.6.2 Gramatická elipsa řídicího slovesa v klauzi s výrazem s funktorem PREC... ..	77
5.2.6.3 PREC jako potomek koordinačního uzlu.....	77
5.2.6.4 Nepravé závislé klauze vztažné jako reprezentant problémových okruhů.....	78
5.2.6.5 Kataforické konektory.....	80
5.2.6.6 Shrnutí.....	80
5.2.7 Hraniční a problémové případy.....	81
5.3 Modifikátory souřadných spojení a jejich vztah k diskurzním konektorům.....	83
5.4 Rematizátor ve funkci diskurzního konektoru.....	88
5.5 Výrazy vyjadřující postoj mluvčího ve funkci diskurzního konektoru.....	92
5.6 Temporální doplnění slovesa jako diskurzní konektory – několik poznámek k temporálním vztahům v diskurzu.....	95
5.7 Interpunkční znaménka ve funkci diskurzních konektorů.....	98
5.8 Další výrazy.....	100
5.9 Shrnutí.....	105
Kapitola 6. Závěr.....	107
Kapitola 7. Otázky, problémy, budoucí kroky.....	109
Summary.....	111
Přílohy.....	112
Příloha 1: Seznam explicitních discourse connectives v PDTB 1.0.....	112
Příloha 2 : A. Seznam anglických výrazů s funktorem PREC pro anotaci PEDT.....	114
B. Discourse connectives z PDTB bez funkce PREC v anotaci PEDT.....	116
Příloha 3: Seznam českých výrazů s funktorem PREC v PDT 2.0.....	117
Příloha 4: Seznam českých výrazů s funktorem CM v PDT 2.0 .....	120
Příloha 5: Seznam příslovcí s funktorem ATT .....	122
Seznam užívaných funktorů.....	123
Seznam zkratk.....	125
Literatura a prameny.....	126

# Úvod

Tato příručka je výstupem z korpusového výzkumu věnovanému vzájemnému vztahu syntakticko-sémantické struktury věty a struktury diskurzu (textu) zpracovanému v diplomové práci Lucie Mladové (Ústav českého jazyka a teorie komunikace FF UK, 2008). Z rozsáhlé problematiky popisu diskurzu v jeho různých aspektech (např. koreferenční vztahy, tematicko-rematická výstavba diskurzu atd.) se zabývá především syntakticky motivovanými vztahy, tj. otázkou, do **jaké míry lze ze syntaktického a sémantického popisu věty vyčíst informace o diskurzních vztazích** a jakého charakteru tyto informace jsou. Stěžejní část dokumentu je věnována lexikálním prostředkům vyjadřujícím spojování či navazování klauzí a vět, nazýváme je **diskurzními konektory**. Do této kategorie zahrnujeme jednak většinu spojek a spojovacích výrazů vyjadřujících druh syntaktického vztahu mezi klauzemi uvnitř věty<sup>2</sup> a jednak převážně adverbialní a částicové výrazy, které signalizují takové spojování, respektive připojování, přes hranice jedné věty. Z takto vymezené kategorie diskurzních konektorů se soustředíme především na posledně jmenované prostředky připojovací typu:

(1) *Díky přístupu firmy neztratil lékař ani den. Technici totiž zvládli výměnu zařízení ordinace za víkend.* PDT<sup>3</sup>

Výzkum shrnutý v tomto dokumentu se opírá o dva hlavní teoretické přístupy a zároveň zdroje lingvistických dat: zdrojem českého jazykového materiálu je Pražský závislostní korpus (PDT) a především jeho tzv. tektogramatická (hloubkově syntaktická a sémantická) reprezentace, v jejímž rámci byly výrazy vyjadřující návaznost klauze, v níž stojí, na předchozí kontext označeny sémantickou značkou PREC (reference to PREceding Context).

Druhým reflektovaným přístupem je pojetí a zachycení diskurzních vztahů ve filadelfském textovém korpusu Penn Discourse Treebank (PDTB), který se zabývá problematikou diskurzních vztahů v angličtině.

---

2 Větou rozumíme útvar od tečky k tečce, tedy i souvětí, zatímco klauzí máme na mysli větu jednoduchou, viz též kapitola 2.4.3.

3 Zkratkou PDT jsou označeny příkladové věty z Pražského závislostního korpusu 2.0, vyhledané pomocí nástrojů Netgraph či TrEd. Jiné zdroje jsou vždy uváděny bezprostředně za příkladem.

V dokumentu jsou pojednány následující body:

- představení tektogramatické reprezentace v PDT a filadelfského přístupu k popisu diskurzu jakožto dvou základních teoretických východisek výzkumu
- vymezení pojetí diskurzu a diskurzních vztahů v teoretickém rámci funkčního generativního popisu (FGP)
- vymezení a popsání skupiny diskurzních konektorů v češtině na základě hloubkově syntaktické anotace v Pražském závislostním korpusu
- příspěvek ke vzniku korpusu zaměřeného na popis diskurzních vztahů v češtině, tedy v jazyce strukturně poměrně odlišném od angličtiny.<sup>4</sup>

Studie je rozdělena do šesti hlavních kapitol. První tři jsou věnovány teoretickým poznatkům o diskurzu a diskurzních vztazích, následující tři jsou praktického zaměření a opírají se o konkrétní korpusové výzkumy.

Úvodní kapitola osvětluje teoretická východiska práce se syntakticky anotovaným korpusem. Druhá, rozsáhlejší kapitola přibližuje teorii a metodologii, se kterými zde pracujeme. Nejprve jsou představeny korpusy PDT a PDTB. Dále se věnujeme problematice užívání pojmů *diskurz* a *text* ve světové a v české lingvistice a podáváme i stručný exkurz do dějin analýzy diskurzu a textové lingvistiky. Dále poukazujeme na aspekty diskurzu, kterým se věnuje současná lingvistika, a vymezujeme tak konkrétně vlastní téma této studie – syntakticky motivované, konektivní diskurzní vztahy. V závěru druhé kapitoly pak definujeme pojetí diskurzu ve Funkčním generativním popisu a tedy teprve na tomto místě vysvětlujeme užívané termíny z oblasti syntaxe a diskurzu. Třetí kapitola je odbočkou: představením jednoho ze současných formálních modelů diskurzu upozorňuje na existenci konektivních diskurzních vztahů, které nejsou primárně založené na větné syntaxi. Jsou to tzv. rétorické vztahy, neboli vztahy kompoziční. Ve čtvrté kapitole již pracujeme s PDT. Ukazujeme v ní, jakými způsoby je možné z tektogramatické reprezentace věty abstrahovat diskurzní vztahy, a zjišťujeme, co všechno syntakticky anotovaný korpus může vypovídat o vztazích v textu. Pátá, stěžejní kapitola je pak věnována rozboru diskurzních konektorů jakožto explicitních lexikálních prostředků vyjádření konektivních diskurzních vztahů. Vychází z rozsáhlé analýzy korpusových dat v PDT, tj. ze sémantické anotace konektorů

---

4 Tím jsou míněny zejména vysoký stupeň flektivnosti a tzv. volný slovosled v češtině.



a výrazů, jež s nimi hraničí. Šestá kapitola shrnuje dosažená zjištění o diskurzních vztazích a zejména o diskurzních konektorech. Sedmá kapitola je kapitolou doplňující, zmiňujeme v ní některé otázky a problematické jevy spojené s analýzou diskurzních vztahů a zamýšlíme se nad dalšími cestami zkoumání a popisu diskurzu.

## Kapitola 1. Teoretická východiska

Syntaktické a sémantické vztahy uvnitř větné struktury zároveň obsahují informaci o vztazích „vyšších jednotek“, vztazích textových či diskurzních. Předpokládáme, že diskurzního charakteru jsou jednak některé druhy hypotaktického spojování, konkrétně vztah mezi větou hlavní a mezi vedlejší větou adverbiální, jednak některé vztahy vyjadřované paratakticky. Také se domníváme, že ukazateli diskurzního vztahu mohou být některé lexikální jednotky, v jejichž sémantice převládá význam podobný významu syntaktických spojek, tedy význam spojování či navazování na předchozí kontext v nejširším slova smyslu. Těmito výrazy ovšem nemáme na mysli výrazy deiktické, u nichž je třeba striktně rozlišovat, o jaký druh odkazování v tom kterém případě jde – ne vždy implikují vztah textový. Máme na mysli výrazy částicového a adverbiálního charakteru, které se chovají podobně jako výrazy spojovací.

V Pražském závislostním korpusu byla na tektogramatické rovině provedena anotace hloubkových syntaktických vztahů a rozpracována jejich sémantická klasifikace. Domníváme se, že velkou část této syntaktické anotace bude možno převzít a adaptovat pro budoucí anotaci diskurzu. Zároveň ale předpokládáme, že pro potřeby komplexního a zároveň co nejekonomičtějšího zpracování vztahů v diskurzu bude třeba některé zásady platné pro anotaci tektogramatické roviny přeformulovat, modifikovat některá teoretická východiska a provést novou sémantickou klasifikaci budoucích „diskurzních spojovacích výrazů“.

Tento dokument vychází při popisu diskurzních konektorů kromě jejich zachycení v Pražském závislostním korpusu i z vymezení tzv. *discourse connectives* ve filadelfském Penn Discourse Treebanku 1.0. Domníváme se, že anotace diskurzu angličtiny, jak ji provádí Penn Discourse Treebank, bude inspirativním zdrojem pro uvažování o zachycení diskurzu v českém korpusu. Odlišnosti dle našeho názoru vyvstanou spíše na rovině teoretického přístupu a anotačních schémat než tím, že by diskurzní vztahy byly jazykově specifické. Očekáváme, že angličtina a čeština se v popisu diskurzních vztahů navzájem neliší více, než je tomu v jejich popisu syntaktickém<sup>5</sup> a že jazykové odlišnosti se budou v naší analýze projevat zejména jednotlivě, v rámci porovnávání chování odpovídajících jazykových jednotek.

---

5 Srov. Šindlerová, Mladová, Toman a Cinková (2007).

## Kapitola 2. Teorie, metody a terminologie

### 2.1 Závislostní a frázový syntaktický popis

#### 2.1.1 Pražský závislostní korpus (PDT)

Pražský závislostní korpus (Prague Dependency Treebank, PDT) je projekt lingvistů z Ústavu formální a aplikované lingvistiky MFF UK započatý v roce 1995. Jedná se o databázi českých publicistických textů<sup>6</sup> o velikosti přibližně 2 milióny slovních jednotek, texty jsou přežaté z Českého národního korpusu. Oproti ČNK jsou opatřeny kromě morfologické také syntaktickou a sémantickou anotací, zachyceny jsou ale i referenční vztahy a aktuální členění.

##### 2.1.1.1 Funkční generativní popis

Závislostní syntaktický popis, který využívá analytická a tektogramatická rovina v PDT, je součástí pražského Funkčního generativního popisu (FGP, Functional Generative Description). Tento teoretický rámec, zformulovaný v 60. letech Petrem Sgallem a jeho spolupracovníky (Sgall 1967, Sgall et al.1969), vymezuje pravidla pro všestranný formální popis přirozeného jazyka především za účelem strojového učení, strojového překlada a pro další lingvistické i počítačové experimenty.

V závislostním syntaktickém popisu FGP je za centrum (či vrchol) věty považováno sloveso ve funkci predikátu a slovesná valence. Všechny ostatní členy včetně subjektu jsou přímo nebo nepřímo závislé na slovese.

V souladu se stratifikačním přístupem k jazyku ve FGP existují tři roviny jazykového popisu, tj. tři roviny anotace textů v PDT: Postupujeme-li směrem nahoru od jazykových forem k jejich funkci (či významu), označujeme za nejnižší anotační vrstvu tzv. *rovinu morfologickou*, na které jsou všechny slovní jednotky opatřeny morfologickým značkováním. Prostřední *analytická rovina* zachycuje povrchově syntaktické (větněčlenské) vztahy a má již v korpusu jiné grafické zpracování: Věty na analytické rovině mají formu stromové struktury. Nejvyšší rovina, zachycující stromovým diagramem hloubkově syntaktické a sémantické

---

6 Lidové Noviny, Mladá fronta Dnes, Českomoravský profit, Vesmír

vztahy, se nazývá *tektogramatická*. Za čtvrtou, neanotační rovinu pak bývá považována rovina čistého základového textu – *slovní rovina*. Roviny anotace jsou vzájemně propojeny, z vyšší vedou odkazy do nižších a zároveň je ve vyšší rovině zpravidla uchována lingvistická informace přidělená „níže“, například i v tektogramatickém stromě jsou zobrazeny morfologické kategorie přiřazené určité slovní jednotce na rovině morfologické. Omezíme se pouze na popis tektogramatické reprezentace.

### 2.1.1.2 Tektogramatická reprezentace v Pražském závislostním korpusu

Hlubková významová struktura věty je v Pražském závislostním korpusu znázorněna na tektogramatické rovině pomocí projekivní stromové struktury. Tektogramatický strom se skládá z různých druhů uzlů a hran, uzlům jsou navíc přiděleny různé atributy.

*Tektogramatický uzel* buď reprezentuje jazykový výraz existující v povrchové stavbě věty, nebo je uměle vytvořen a přidán do stromu. Jedním z případů, kdy se uzel přidává do stromu, jsou v povrchové struktuře věty nepřítomná osobní zájmena. V hloubkové struktuře je reprezentuje nově vytvořený uzel se zástupným t-lematem *#PersPron*. (viz obr. 1). Naopak některé jazykové jednotky v tektogramatické reprezentaci svůj vlastní uzel nemají (např. předložky, podřadící spojky, pomocná slovesa).

Jak již bylo zmíněno, vrcholem (tj. uzlem, který je nejvýše) neboli kořenem tektogramatického stromu je sloveso.<sup>7</sup> Slovesná doplnění včetně subjektu, tedy členy na slovese závislé, jsou pak ve stromě umístěny níže. Syntaktická závislost mezi dvěma uzly v tektogramatickém stromě je znázorněna *hranami*, rovnými spojnicemi mezi dvěma uzly. Hrany ovšem ne vždy vyjadřují syntaktickou závislost, mohou mít i jiné funkce (naznačují koordinační spojení či další specifické syntaktické vztahy nebo mají pomocný technický charakter).

*Atributy* přiřazené uzlům popisují jejich syntaktické a sémantické rysy. Jeden z nejdůležitějších atributů, tektogramatický *funktor*, ale pouze zdánlivě patří přímo uzlu: Je sémantickou charakteristikou syntaktického vztahu mezi uzlem závislým a řídícím, mezi dcerou a matkou, je tedy fakticky atributem hrany spíše než uzlu. Zmíníme alespoň několik základních druhů funktorů: Dle pojetí valence v FGP se slovesná doplnění dělí na aktanty (neboli vnitřní participanty), které jsou reprezentovány funktoři:

---

<sup>7</sup> Výjimku tvoří souřadící výraz, který spojuje klauze a tedy predikáty. Viz obr. 1.

ACT – aktor, kognitivní role půvoce/nositele děje,

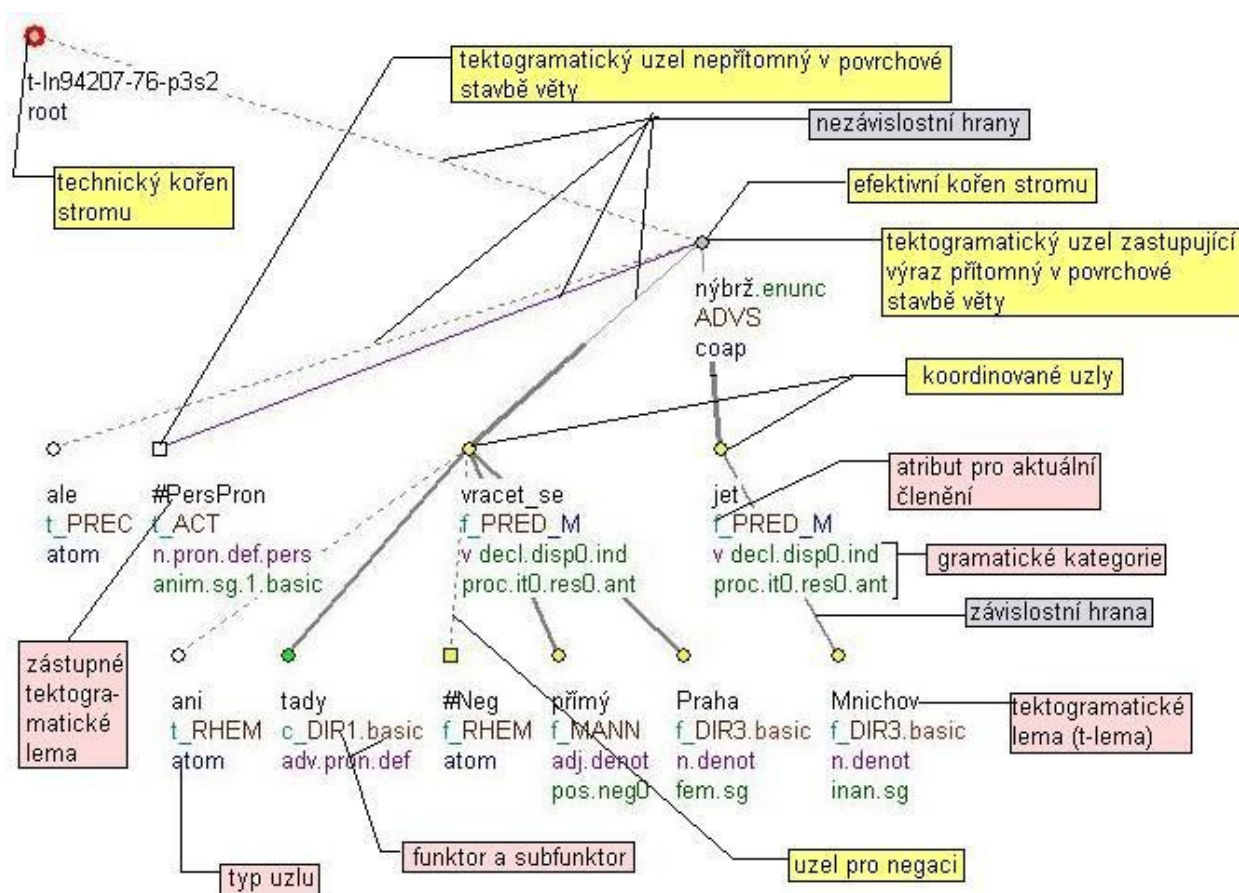
PAT – patiens, předmět dějem zasažený,

EFF – efekt, výsledek děje,

ADDR – adresát, příjemce děje,

ORIG – origo, původ děje;

a na volná slovesná doplnění, jejichž sémantická škála je široká (funktory časové, místní a směrové, kauzální, různá vyjádření způsobu atd.), např. TWHEN, LOC, DIR, MANN, CAUS, CPR, ACMP.<sup>8</sup>



*Ale ani odtud jsem se přímo nevracel do Prahy, nýbrž jsem jel do Mnichova.* PDT

**Obr. 1** Příkladový tectogrammatický strom s popisem

<sup>8</sup> Vysvětlení zkratk všech užívaných funktorů viz Seznam používaných funktorů v závěru této práce.

Dále existuje skupina funktorů pro významy souřadných spojení (tedy nepopisující druh závislosti, ale vztah mezi koordinovanými členy), funktory pro výrazy vyjadřující např. postoj mluvčího k ději (ATT), modální charakteristiku (MOD) či návaznost na předcházející kontext (PREC) atd.<sup>9</sup>

Na příkladové větě (obr. 1) si ukážeme, jaké informace lze z tektogramatického stromu vyčíst. Informace o uzlech jsou ve žlutých rámečcích, o jejich attributech v růžových a o hranách v šedých. Pro úplnost ještě nutno dodat, že na obrázku 1 není viditelné existující propojení korpusu, respektive jeho slovesných uzlů s valenčním slovníkem ValLex<sup>10</sup>, ze kterého je každému slovesu přiřazen valenční rámec odpovídající příslušnému slovesnému významu v té které větě. Dále také chybí grafické znázornění koreferenčních vztahů, které ale tektogramatická reprezentace také zachycuje. Ty jsou ve stromech znázorněny barevnými šipkami mezi koreferujícími a koreferovanými uzly, barvy šipek se liší dle typu koreferenčního vztahu.

Pro účely této studie má z tektogramatické analýzy největší význam pojetí syntaktické závislosti a souřadnosti, dále konkrétní uzly reprezentující výrazy potenciálně relevantní pro analýzu diskurzu (i ty, které jsou v t-rovině skryté!), a v neposlední řadě sémantické značení, tedy některé z funktorů.

### 2.1.2 Frázový syntaktický popis v Penn Treebanku (PTB)

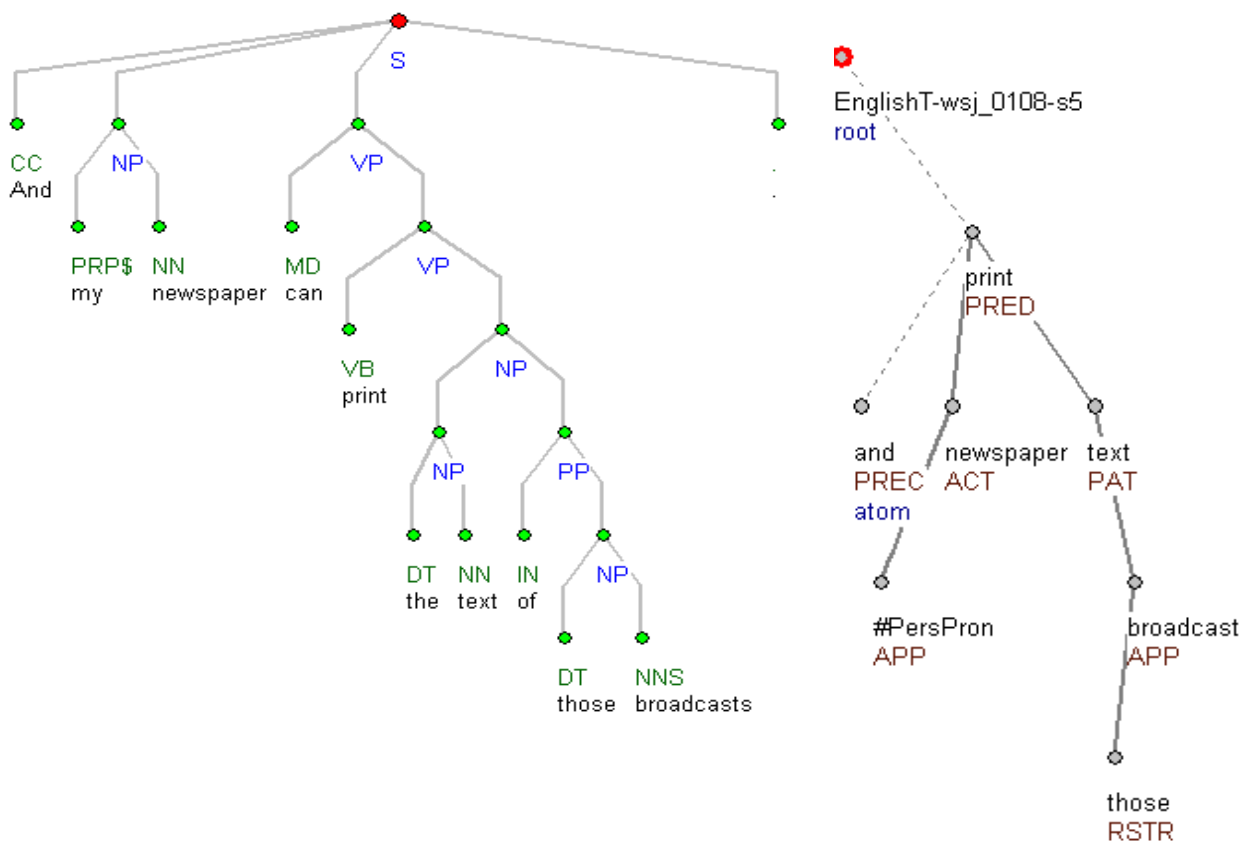
Korpus zachycující diskurzní vztahy, filadelfský Penn Discourse Treebank (PDTB), je nástavbou či pokračováním projektu Penn Treebank (PTB), anglického syntaktického korpusu anotovaného pomocí rozboru na bezprostřední složky (neboli pomocí frázových stromových diagramů). Texty obou korpusů jsou tedy stejné, jde o přibližně 49 000 vět z ekonomického časopisu Wall Street Journal (WSJ). Ačkoli anotace diskurzních vztahů v PDTB proběhla na holých textech a ne na syntaktických stromech PTB, je v syntaktických stromech rovněž zachycena pomocí odkazů. V následující kapitole představíme nejprve stručně Penn Treebank a zvláště teoretická východiska jeho anotačního schématu, teprve poté se budeme zabývat diskurzním korpusem.

Strukturní lingvistika nabízí ve své tradici dva základní modely syntaktického popisu.

<sup>9</sup> Detailní třídění funktorů v *Manuálu pro tektogramatickou anotaci* (Mikulová et al. 2005: kapitola 6, Funktory a subfunktory) a v Seznamu užívaných funktorů (kap. 11 v této práci).

<sup>10</sup> ValLex – Valency Lexicon of Czech Verbs: <http://ufal.mff.cuni.cz/vallex/2.0/doc/home.html>

Zatímco v Evropě se spolu se strukturalismem vyvíjel závislostní přístup (Becker, Tesnière), v Americe se od třicátých let uplatňuje bloomfieldovská analýza na bezprostřední složky (Bloomfield 1933),<sup>11</sup> kterou dále rozpracoval Z. Harris (1951) a uplatňuje se i v chomskyánské generativní gramatice. Těchto tendencí se drží i současné lingvistické projekty. Hlavní rozdíly spočívají zejména v následujícím: Oproti závislostnímu popisu jazyka v pražském FGP, kde je každý uzel terminální, tj. reprezentuje konkrétní jazykový výraz, existují v americkém frázovém (složkovém) popisu neterminální uzly – fráze. Frázový strom tak bude obsahovat vždy více uzlů, než obsahuje strom závislostní. Ve frázové analýze se věta (sentence) primárně rozpadá na část verbální (verbal phrase) a část nominální (noun phrase)  $S \rightarrow VP + NP$  (neboli subjektovou a predikátovou), zatímco v závislostním popisu je subjekt chápán jako doplnění slovesa a je rovnocenný s ostatními aktanty.



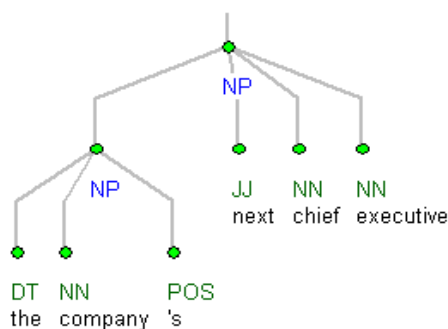
**Obr. 2** Srovnání frázového a tektogramatického stromu pro větu (2)

<sup>11</sup> Též IC analýza, Immediate-Constituent Analysis.

Nejlépe uvidíme rozdíly ve struktuře na ukázkových stromech. Vlevo je strom frázový převzatý z PTB, vpravo strom tektogramatický (tedy závislostní) z PEDT. Oba stromy reprezentují jednu a tu samou větu (2).

(2) *And my newspaper can print the text of those broadcasts.* WSJ

Další rozdíl ve frázovém a závislostním popisu je v počtu potomků uzlů ve stromových strukturách. Zatímco frázový strom je striktně binární, tj. fráze se skládá právě ze dvou složek, uzel v závislostním popisu může mít různý počet potomků. Kvůli ekonomickému zobrazení se však PTB při zachycení jmenných frází důsledně binarity vzdal, viz obr. 3:



**Obr. 3** Zachycení rozvíte jmenné fráze v PTB

### 2.1.3 Penn Discourse Treebank<sup>12</sup> a jeho anotace

Penn Discourse Treebank je projekt, jehož cílem je popsat diskurzní vztahy na textech Wall Street Journalu. Přestože se zkoumání diskurzu, či chceme-li, „úrovně textu“ nepochybně odvíjí od předchozí syntaktické a sémantické analýzy na poli větném, vyžaduje odlišnost jazykových „úrovní“ i odlišné uvažování o diskurzních vztazích a jejich popisu. (Závislost syntaktická je jiného druhu než „závislost“ v diskurzním korpusu, a stejně tak i pojmy predikát a jeho argumenty se v závislostní syntaxi podstatně liší od těch, se kterými pracuje PDTB.) Přístup k anotaci diskurzu se tedy nutně liší od obou popsaných syntaktických přístupů.

<sup>12</sup> První verze korpusu, materiály k anotaci i technické nástroje viz <http://www.seas.upenn.edu/~pdtb/pdtb-corpus-1.0/> [k 15. 9. 2007]; vydání druhé verze je plánováno na prosinec 2007.



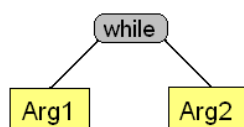
V PDTB je nejprve vymezena a podrobně charakterizována skupina tzv. *discourse connectives*, což jsou výrazy určitým způsobem spojující různě rozsáhlé textové úseky. *Discourse connectives* – nazvěme je v češtině „konektory diskurzu“ (diskurzí konektory)<sup>13</sup> – jsou pak hodnoceny jako predikáty binárních diskurzí vztahů a přijímají tedy po právě dvou argumentech. Jako argument 2 *Arg2* se označuje ten z obou textových úseků, který obsahuje konektor, zbývající argument je označen *Arg1*. Graficky se první argument označuje kurzívou, druhý tučně. Diskurzí konektor bývá podtržen. Uveďme hned z kraje příklady:

(3) *John eats porridge for breakfast, while Mary eats muesli.* PDTB

(4) Since McDonald's menu prices rose this year, the actual decline may have been more. PDTB

(5) *But*, says Mr. Dinkins, he did get an office. So he shouldn't complain. PDTB

Jiné grafické znázornění věty (3) by mohlo vypadat takto:



**Obr. 4** Příklad zachycení diskurzího vztahu

V následujících kapitolách se budeme nejprve blíže věnovat diskurzí konektorům, poté jejich argumentům.

### 2.1.3.1 *Discourse connectives* v PDTB a jejich klasifikace

Anotační schéma PDTB rozlišuje tzv. explicitní a implicitní diskurzí konektory. Explicitní jsou povrchově přítomné, implicitní jsou takové, které sice povrchově přítomné nejsou, ale dají se jednoznačně doplnit mezi dva textové úseky, mezi kterými existuje diskurzí vztah.

<sup>13</sup> Termín „konektor“ lze nalézt ve významu jazykového prostředku ve funkci pojítka v textu v několika základních gramatických příručkách češtiny. Proto ho přebíráme i pro účely této práce, přidáváme k němu adjektivum „diskurzí“.

Pravidla pro anotaci implicitních diskurzních konektorů jsou podrobně popsána v anotačním manuálu pro PDTB (Prasad et al. 2006). V prvním vydání PDTB proběhla anotace implicitních diskurzních konektorů zatím pouze na třech sekcích korpusu, bylo pro ni však vyvinuto sémantické třídění, které zatím anotace explicitních diskurzních konektorů v PDTB nemá (plánuje se pro druhé vydání v prosinci 2007).

Diskurzní konektory v PDTB jsou, co se týče jejich gramatické klasifikace, heterogenní skupina výrazů. Každý diskurzní konektor obsahuje informaci o sémantickém vztahu dvou syntaktických entit. (viz Příloha 1 – Seznam explicitních *discourse connectives* v PDTB 1.0):

Jsou to:

1, Souřadící spojky jako např.: *and, but, or, nor*

Anotovány jsou pouze souřadící spojky mezi větami hlavními. Souřadně spojené vedlejší věty (či větné členy) nejsou považovány za diskurzní vztah (Prasad et al. 2006: 5).

2, Podřadící spojky, např.: *because, when, since, although, as soon as*

3, Tzv. *discourse adverbials*, diskurzní adverbia, mezi která patří adverbia a předložkové fráze (ADVP a PP), např.: *instead, therefore, however, otherwise, then, as a result, for examle, on the other hand* atd.

Ne vždy ale mají takové výrazy v PDTB funkci diskurzního konektoru. Někdy mohou mít primárně jinou funkci: *well, anyway, in fact* spíše souvisí tematicko-rematickými vztahy, někdy se nevztahují ke dvěma argumentům, ale pouze k jednomu: *frankly, strangely*; někdy jsou tyto výrazy homonymní s výrazy v jiné funkci: *and* spojující dvě jmenné fráze; *before* jako předložka i jako podřadící spojka s temporálním významem; *for example* jako diskurzní konektor i jako rematizátor<sup>14</sup>.

Dále se v PDTB poukazuje na skupinu modifikovaných diskurzních konektorů: *only if, just because, even though...* Protože jsou tyto formy velmi produktivní, řadí se při anotaci vždy ke svým základovým formám. Modifikace je zde chápána velice široce, například u diskurzního konektoru *after* lze nalézt dlouhý seznam modifikací typu *two weeks after*.

Další podskupinou jsou tzv. paralelní diskurzní konektory. Tak jsou označeny párové spojovací výrazy, které fungují jako jeden konektor a přijímají stejné argumenty: *on one hand – one the other hand, if – then, either – or*.

---

14 U rematizátorů je to složitější: Americké schéma s některými z nich (tak, jak jsou definovány FGP) počítá jako s diskurzními konektory, zatímco český přístup vede poměrně jasnou hranici mezi výrazy primárně rematizujícími a výrazy primárně vyjadřujícími návaznost na předchozí klauzi. K tomu blíže v kapitole 5.4.

### 2.1.3.2 Argumenty *discourse connectives*, abstraktní objekty

PDTB definuje argument dle Nicholase Ashera (1993) jako abstraktní objekt (*Abstract Object, AO*). Hierarchii abstraktních objektů zachycuje obr. 5.



**Obr. 5** Hierarchie abstraktních objektů (Asher 1993)

Jako abstraktní objekty jsou interpretovány klauze, mohou tak být ale interpretovány také další syntaktické struktury, například nominalizace, deiktické výrazy, sekvence o více větách atd. PDTB se zatím soustřeďuje pouze na anotaci vztahů mezi klauzemi a mezivětných vztahů, a tudíž nechává nominalizace a deixi zatím stranou.<sup>15</sup>

Jak je vidět z příkladových vět (3) - (5), pořadí argumentů může být různé v závislosti na typu diskurzního konektoru, argument nemusí být spojitý (5), tj. může být přerušen textem, který do argumentu nepatří, a též postavení diskurzního konektoru vzhledem k jeho dvěma argumentům může být různé.

Zatímco příkladové věty (3) a (4) ukazují diskurzni vztah uvnitř souvětí, příklad (5) ukazuje anotaci mezivětných diskurzni vztahů, tedy „přes tečku“. Mezivětné vztahy byly anotovány u všech explicitních diskurzni konektorů, implicitní konektory se doplňovaly (anotovaly) pouze v rámci stejného odstavce.

Argumenty dvou různých diskurzni konektorů se mohou různě překrývat, mohou být

15 S anotací nominalizací typu

(1) *instead of featuring a major East Coast team against a West Coast team, it pitted the Los Angeles Dodgers against the losina Oakland A's.*

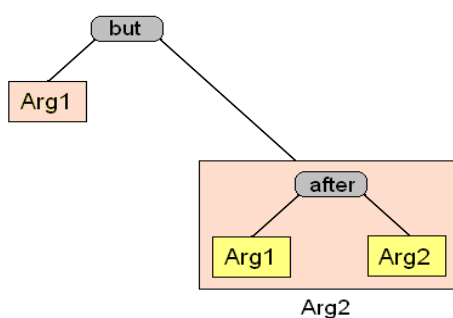
počítá PDTB ve svých budoucích vydáních, v prvních dvou však ještě nebudou takové případy zachyceny.

(částečně) sdílené dvěma diskurzními konektory atd. Uvedme příklad, kdy druhý argument diskurzního vztahu A plně obsahuje diskurzní vztah B, tj. diskurzní konektor i oba jeho argumenty.<sup>16</sup>

(6a) *The drop in earnings had been anticipated by most Wall Street analysts, **but the results were reported after the market closed.*** (A) PDTB

(6b) *The drop in earnings had been anticipated by most Wall Street analysts, but *the results were reported* **after the market closed.*** (B) PDTB

Jiné znázornění:



**Obr. 6** Příklad diskurzního vztahu: Vnořená relace

Nejobtížnější částí anotace diskurzních vztahů bylo v PDTB určení rozsahu jednotlivých argumentů. V tomto bodě se anotátoři často lišili (různé hodnocení parentezí, závislých klauzí apod.). Proto byl zaveden tzv. minimální princip (*Minimality Principle*), (Prasad et al. 2006: 12), který vymezuje argument jako „minimální“ textový úsek potřebný k interpretaci vztahu vyjádřeného diskurzním konektorem. Ostatní části relevantní pro daný diskurzní vztah se označují SUP 1 (*supplementary to Arg 1*) či SUP 2.

Jak již bylo zmíněno, nepoužívá PDTB k anotaci diskurzu frázové stromy Penn Treebanku, neboť syntaktické struktury se ne vždy překrývají s diskurzními (Webber et al. 2005: 5). Anotace sice proběhla na holých textech, ale data v PTB a v PDTB jsou propojena odkazy a ve vyhledávacích nástrojích lze snadno získat informace např. o syntaktické struktuře argumentů.

<sup>16</sup> K tomu podrobněji viz Lee et al. (2006).

## **2.2 Textová lingvistika a diskurz**

### **2.2.1 Zahraníční a české uvažování o diskurzu (textu)**

Zkoumání jazykových projevů či užívání jazyka má dlouhou historii. Už od starověku se tím, co bychom dnes mohli nazvat text či jazykový projev, zabývala rétorika. Od vzniku moderní strukturní lingvistiky, tedy od dob Saussurových, je všeobecnou tendencí postupovat při popisu jazyka od nižších jazykových rovin a jednotek k vyšším. A tak spolu s poznatkem, že věta (či výpověď) vyňatá z jazykového projevu (psaného textu, rozhovoru atd.) ve velké většině případů není plně srozumitelná, ale téměř vždy vyžaduje k jasné interpretaci předcházející a/nebo situační kontext, byl v 70. letech minulého století v lingvistice vznesen požadavek, že za základní jednotku jazyka je třeba považovat text (Petöfi, van Dijk). Vznikla nová oblast lingvistického bádání, věnující se vztahům „nad hranicí věty“ a zdůrazňující jejich důležitost pro plné pochopení informací obsažených v textu.

#### **2.2.1.1 Diskurz a text v české lingvistice**

V českém prostředí začali věnovat více pozornosti jazykovým projevům, jejich obsahu či tématu, nadvětným vztahům a textu někteří lingvisté Pražské školy, a to ještě před vznikem textové lingvistiky jakožto samostatné disciplíny. (zejména Skalička 1937, 1948; Mathesius 1939, 1943 atd). Významným příspěvkem mezinárodního dosahu jsou Mathesiovy kontrastivní studie o slovosledu a aktuálním členění větném, které je určováno jednak syntaktickými vlastnostmi daného jazyka, jednak vztahy kontextovými. Na Mathesia a Pražský lingvistický kroužek zde navázali zejména Jan Firbas v Brně (funkční perspektiva větná) a Petr Sgall a Eva Hajičová v Praze (topic-focus articulation, salience – stupně aktivovanosti textu).

Z jiného hlediska se nadvětným jazykovým útvarům věnovala a věnuje stylistika, která nabízí zejména typologii jazykových projevů, dále úvahy o jejich výstavbě a funkci v komunikaci. Stylistikou a stavbou textů se zabýval především Karel Hausenblas (1964, 1971). Systematicky se (mimo jiné) jevy na pomezí syntaxe a textových vztahů zabývá František Daneš (zejména 1985). Z českých lingvistů dále k tomuto tématu různými způsoby přispěli a přispívají například Jan Kořenský (1992), Josef Hrbáček (1994), Jarmila Tárnyiková (2002),

Jana Hoffmannová (1983, 1984, 1997) a další.

V souvislosti s komunikačně-pragmatickým obratem v lingvistice v 70. letech minulého století se dostaly do popředí lingvistického zájmu okolnosti vzniku a recepcce jazykových projevů, mluvená komunikace, účel komunikace atd. Textová lingvistika se tedy jakožto jazykovědná disciplína začala formovat dvojitým způsobem: (i) jako další stupeň strukturního popisu jazyka, stupeň „nad“ syntaxí, (kdy na rozdíl od stylistiky, která popisuje především rozdíly, jsou v textu hledány shodné, systémové jevy); (ii) jako nauka o fungování jazykového projevu v komunikaci, tedy s přesahem do pragmatiky a stylistiky. Tento přístup převažuje i v současném lingvistickém bádání.

Saussurovské *langue a parole* se začíná problematizovat. Snaha o vysledování a popsání systematických jevů v užívání jazyka je pokusem svědčícím o stejném přístupu lingvistiky k jevům parolovým jako k jevům jazykového systému. Nelze než přisvědčit, že i pro zkoumání jazyka v jeho užívání je systematický popis relevantní, neboť textové vztahy nejsou náhodné. Nelze však říci, že textové vztahy jsou systémové<sup>17</sup>; proto se jeví jako nejvhodnější prozatím opustit snahu o přesné zařazení textové lingvistiky v rámci didakticky určené distinkce *langue a parole*, dodržované v důsledku především v terminologii: rozlišuje se *věta a výpověď* jako jednotky systémové a komunikační, výklad pojmů jako promluva a text se komplikuje i tím, že se lingvistické teorie a modely množí a interdisciplinarizují.

### 2.2.1.2 Diskurz a text ve světové lingvistice

Pojem *discourse analysis* v americké lingvistice poprvé použil Z. Harris (1952), a ačkoli sám nepodal ucelený výklad tohoto pojmu a spíše na text aplikoval své teze o transformacích a ekvivalenci, dokládají jeho studie, že text vykazuje jistý stupeň opakování a paralelismy v syntaktických strukturách. Samotné slovo *discourse*, v českém překladu doslovně *promluva*, dále také (*jazykový*) *projev* či *text*, jako termín užívá jako jeden z prvních holandský vědec Teun A. van Dijk, který bývá považován za jednoho ze zakladatelů textové lingvistiky. Van Dijk se v sedmdesátých letech zabývá problematikou explicitního popisu (gramatických) struktur v textu, který postrádá v dosavadním lingvistickém bádání, a vyčítá chomskyánské

---

17 „It is much more straightforward to decide what constitutes a grammatical or acceptable sentence than what constitutes a grammatical or acceptable sentence sequence, paragraph, text, or discourse. When we move beyond the sentence boundary, we enter a domain characterized by greater freedom of selection or variation and lesser conformity with established rules.“ (Dressler/Beaugrande 1981)

generativní gramaticy, že není uzpůsobena pro zkoumání textových struktur. Jeho *text grammars* je generativní model zaměřený na sémantiku textu. Od van Dijka též pochází dnes již vžitý pojem *makrostruktury*, hloubkové obsahové struktury textu. Van Dijk se však později odklání od lingvistiky přes pragmatiku k sociálním bádáním a zakládá v devadesátých letech vědní obor *discourse studies*, „a new cross-discipline that comprises the theory and analysis of text and talk in virtually all disciplines of humanities and social sciences“ (van Dijk 1997: XI). Čistě lingvistický přístup k „text and talk“ je pouze jednou z jím nabízených možností analýzy diskurzu. Další jsou pragmatické, sociální, sémiotické, kognitivní, zabývají se politikou, kulturou, genderovou problematikou atd., navíc má každý přístup svůj vlastní teoretický rámec. Spojení *discourse analysis* se tedy problematizuje, je v tomto chápání široce pojatým hyperonymem a textová lingvistika by jakožto malá podmnožina *discourse analysis* nejlépe odpovídala tomu, co van Dijk v současnosti nazývá zkoumání *discourse as verbal structure*.<sup>18</sup>

Analýze diskurzu v užším, lingvistickém slova smyslu, tedy textové lingvistice, se ve světě od sedmdesátých let věnuje mnoho pozornosti v rámci různých přístupů či modelů jazyka. Pokoušet se o jejich ucelený přehled by bylo nad rámec této studie, zmíníme zde proto pouze několik významných. Deborah Schiffrin (1994) v úvodu své knihy *Approaches to Discourse* vyjmenovává šest převládajících konceptů diskurzu v rámci lingvistiky: teorie řečových aktů, pragmatika, etnometodologie, interakční sociolingvistika, konverzační analýza a teorie variace. Zároveň dělí přístupy k diskurzu dle jejich zaměřenosti na jazykový systém či na užití jazyka na strukturní (formální) a funkční. Například formální přístupy, dle Schiffrinové (1994: 21) např. Chomsky, považují jazyk primárně za mentální jev, zatímco funkční přístupy (Halliday) jej považují za jev sociální.

---

18 verbal = jazykový, ne slovesný!

"Structural"	"Functional"
Structure of language (code) as grammar	Structure of speech (act, event) as ways of speaking
Use merely implements, perhaps limits, may correlate with, what is analyzed as code; analysis of code prior to analysis of use	Analysis of use prior to analysis of code; organization of use discloses additional features and relations; shows code and use in integral (dialectical) relation
Referential function, fully semanticized uses as norm	Gamut of stylistic or social functions
Elements and structures analytically arbitrary (in cross-cultural or historical perspective), or universal (in theoretical perspective)	Elements and structures as ethnographically appropriate ("psychiatrically" in Sapir's sense)
Functional (adaptive) equivalence of languages; all languages essentially (potentially) equal	Functional (adaptive) differentiation of languages, varieties, styles; these being existentially (actually) not necessarily equivalent
Single homogeneous code and community ("replication of uniformity")	Speech community as matrix of code-repertoires, or speech styles ("organization of diversity")
Fundamental concepts, such as speech community, speech act, fluent speaker, functions of speech and of languages, taken for granted or arbitrarily postulated	Fundamental concepts taken as problematic and to be investigated

**Obr. 7** Formální (strukturní) a funkční přístupy k diskurzu (Schiffirin 1994)

Paradigma rysů formálních a funkčních přístupů k diskurzu dle Schiffrinové (1994: 21) velmi dobře ukazuje na to, že jasnou hranici mezi nimi vést nelze. Spíše bychom mohli tvrdit, že textová lingvistika (pokud ji vnímáme jako „hypersyntax“, jako disciplínu zkoumající spojitost a sémantiku textu) zůstává v oblasti *langue* a disciplíny jako pragmatika, stylistika, sociolingvistika, jejichž předmětem je také text, patří do oblasti *parole*. Jak už ale bylo zmíněno výše, toto vymezení je nanejvýš problematické.

Ranou sférou zájmu textové lingvistiky bylo zejména zkoumání textové soudržnosti (*koherence*, *koheze*) a prostředků jejího vyjádření. Michael A. K. Halliday a Ruqaiya Hasanová (1976) se věnují zejména lexikálním prostředkům jakožto spojovacím prvkům



k vyjádření textové soudržnosti a zároveň popírají existenci struktury textu, tak, jak chápeme syntaktickou strukturu věty: „Whatever relation there is among the parts of a text – the sentences, the paragraphs, or turns in a dialogue – it is not the same as structure in the usual sense, the relation which links the parts of a sentence or a clause.“ (Halliday, Hasan 1976: 6); „Between sentences, there are no structural relations“ (1976: 27). Organizaci textu nazývají *textura* a míní tím všechny druhy textové koheze (dle Hallidaye jich je pět: *conjunctions, reference, substitution, ellipsis, lexical cohesion*). Text je podle nich vytvářen jako „neatly woven texture“, neboli je propojen sítí (heterogenních) kohezních vazeb.

Dalšími světově významnými příspěvateli jsou Wolfgang Dressler a Robert-Alain de Beaugrande. Společně přeložili Dresslerův příspěvek *Einführung in die Textlinguistik* z roku 1972, rychlým rozvojem této vědní oblasti v průběhu sedmdesátých let však byli donuceni značně zrevidovat a rozšířit svá východiska. Roku 1981 pak vydali rozsáhlou shrnující studii o textové lingvistice *Introduction to Text Linguistics*, ve které reflektují vzniklou pluralitu v přístupech k textu (či diskurzu)<sup>19</sup> a vymezují textově lingvistická bádání jako hledání určitých tendencí oproti hledání pravidel v dosavadních formálních přístupech<sup>20</sup>. Dle jejich pohledu je pak text komunikační událostí, která splňuje sedm kritérií *textovosti* (communicative occurrence which meets seven standards of textuality) (1981), která jsou: koheze, koherence, intencionalita, akceptabilita (přijatelnost pro adresáta), informativnost, situativnost a intertextualita.

Současné teorie diskurzu zkoumají především mechanismy textové koherence, vliv kontextu na užití různých druhů jazykových prostředků, podmínky porozumění textu apod. Velká pozornost se věnuje dialogu a jeho modelování, např. problematice tzv. „question answering“. Tak nacházejí poznatky textové lingvistiky či analýzy diskurzu využití v oblasti výzkumu umělé inteligence, zejména při vytváření dialogových systémů.<sup>21</sup>

---

19, „Hardly an aspect of human thought, action, and interaction is not permeated to some degree by language. We cannot escape being reductive in our theories and models. Yet we must bear in mind that reductions are temporary, undesirable conditions to be removed as soon as it is feasible. We may even find that an integrated, comprehensive approach actually leads to a simpler account of language overall than a fragmented, restricted one: preoccupation with exactness of detail in isolated domains can block our vision for sweeping correlations across the whole spectrum.“ (Dressler/Beaugrande 1981)

20, „We should look to discover *regularities, strategies, motivations, preferences, and defaults* rather than *rules and laws*. *Dominances* can offer more realistic classifications than can *strict categories*. *Acceptability and appropriateness* are more crucial standards for texts than *grammaticality* and *well-formedness*. *Human reasoning processes* are more essential to using and conveying knowledge in texts than are *logical proofs*. It is the task of science to systemize the *fuzziness* of its objects of inquiry, not to ignore it or argue it away.“ (Dressler, Beaugrande 1981)

21 viz např. *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue* (2007).

### 2.2.1.3 Pojmový aparát – textová lingvistika (textová syntax, hypersyntax, nadvětná syntax) versus *discourse analysis*

V předchozí kapitole byly termíny *text*, *discourse/diskurz* a další užívány vždy tak, jak s nimi zacházely jednotlivé školy. Jak jsme zjistili, i v užším rámci čistě lingvistických přístupů k diskurzu je termín *analýza diskurzu (discourse analysis)* širší než pojetí oboru *textová lingvistika*. K terminologickému zmatení dochází nejen v souvislosti s rozvojem rozličných lingvistických i nelingvistických „diskurzních“ disciplín, ale také s nejednoznačností zařazení toho či onoho aspektu do oblasti systému jazyka či do oblasti komunikace. V neposlední řadě platí, co škola, to jiné užívání téhož termínu.

Problematické terminologického vymezení pojmů diskurz (*discourse*) a text se věnuje v úvodu své anglické monografie *From text to texture* Jarmila Tárnyiková (2002: 19); poukazuje zde na nejčastější rozdíly v chápání těchto dvou pojmů v různých lingvistických konceptech, v Evropě i v Americe. V různých pojetích se zpravidla pojmy *text* a *diskurz* liší jedním z uvedených aspektů:

<u>text</u>	<u>diskurz</u>
psaná komunikace	mluvená komunikace
abstraktní konstrukt (funkce <i>langue</i> )	manifestace <i>parole</i>
evropská lingvistická tradice	americká lingvistická tradice
kombinace vět	užití kombinace vět
	text + situační kontext
statický koncept (produkt)	dynamický koncept (proces vytváření textu)

V české lingvistické tradici je vžitý spíše pojem *text*, termín *diskurz* je spíše okrajový, částečně kvůli problematickému překladu do češtiny, částečně také kvůli své polysémantické povaze. Karel Hausenblas překládá *discourse* jako *jazykový projev*, objevuje se i přejaté slovo *diskurs/diskurz*, častěji však zatím jako pojem literární vědy či sociologických disciplín (Foucault, Bachtin aj.), odkud proniká především do stylistiky. Pojem textu se obecně v lingvistice v poslední době otevřel, text je vnímán jako proces a i jako produkt a přesahuje původní hranici psaného jazykového projevu. Diskurz však v nejširším pojetí zahrnuje sdělení (text), mluvčího (autora), adresáta sdělení, situační kontext. Nabízí se tedy dvojí chápání:

buďto lze diskurz chápat jako obecnější jev než text, nebo lze diskurz a text považovat za synonyma ve smyslu *jazykového projevu*, přičemž zdůrazníme to, co je pro tyto pojmy v zásadě společné a naopak zanedbáme rozdíly v jejich interpretacích dané především různě kladeným důrazem v lingvistickém zkoumání.

## 2.3 Textová koherence a prostředky jejího vyjádření – přehled

Jednou ze základních vlastností textu a také jednou z podmínek jeho srozumitelnosti je koherence textových jednotek. Pojmem *koherence* máme na mysli spojitost, soudržnost jazykového projevu v nejširším pojetí, tedy soudržnost po stránce tematické, významové i formální. Na koherenci jazykového projevu, textu či diskurzu se podílí řada faktorů. V této kapitole podáme stručný přehled jevů a vztahů, které jsou obecně pokládány za vztahy či jevy textové, a zmíníme základní (jazykové) prostředky jejich vyjádření. Každá textová jednotka koherentního textu obsahuje alespoň jeden takový prvek, a ten ji spojuje s textovým okolím (srov. MČ III: 685). Často se jazykové prostředky vyjádření textové koherence podílejí na více typech vztahů najednou, textové vztahy různého typu jsou jimi provázané.<sup>22</sup>

Tento (nikoli vyčerpávající) přehled má za cíl ukázat, kde v rámci textové lingvistiky se nachází předmět našeho zkoumání.

### 2.3.1 Aktuální členění, tematické posloupnosti

Aktuální členění dělí obsah výpovědi na jednotlivé části podle jejich specifických funkcí v komunikačním procesu, zejména podle „novosti“ respektive „známosti“ informace, kterou obsahují. Zda je informace nová či známá, vyplývá především z kontextu. Aktuální členění tímto způsobem může zachytit, jakým způsobem z hlediska informační struktury je text vystavěn, nabízí různé druhy tzv. tematických posloupností.

Nejběžnější prostředky vyjádření aktuálního členění v češtině jsou povrchový slovosled, intonace, aktualizační částice (rematizátory), gramatické prostředky (vytýkání, elipsa atd.) (MČ III: 559).

---

<sup>22</sup> Například užití ukazovacího zájmena *ten* signalizuje jednak kontextovou zapojenost výrazu z hlediska aktuálního členění a jednak koreferenční vztah anaforického odkazování.

### 2.3.2 Koreference (anaforické řetězce, izotopické řetězce)

Koreference je vzájemný vztah dvou jazykových výrazů, které odkazují k témuž denotátu (předmětu řeči). Nemusí jít vždy o odkazování uvnitř textu, ale i o odkazování exoforické, tj. mimotextové. Koreferenční vztahy (nejčastěji anaforické) identifikují či diferencují v textu stejné denotáty nejčastěji opakováním nebo nahrazováním jazykových výrazů. V textu tak vznikají tzv. anaforické (izotopické) řetězce. Textovým vztahem je pouze koreference textová, která využívá nejrůznějších odkazovacích prostředků, u koreference gramatické je možné určit koreferovaný člen na základě gramatických pravidel (zejména Panevová 1992).

Vztahy textové koreference jsou signalizovány zejména pronominalizací, gramatickou shodou a užitím opakování, synonym, parafrází, hyponym/hyperonym v rámci lexikální koheze. Nejčastějšími gramatickými prostředky vyjádření koreference jsou osobní a přivlastňovací zájmena, ukazovací zájmena, zájmenná příslovce atd.

### 2.3.3 Syntaktická struktura textu

Syntaktické vztahy hypotaktické a parataktické, jak jsou známé z větné syntaxe, překračují často hranice věty. Stejně jako ve větě jsou tyto vztahy i v textu nejsnadněji identifikovatelné formálně – pomocí spojovacích výrazů. Tyto výrazy (v rámci věty se jim říká spojovací, za hranicí věty zpravidla připojovací, obecně *konektory*) plní většinou dvojí funkci. Jednak formálně spojují/připojují textové jednotky (spojková koheze, MČ III: 663), jednak plní i funkci vyjadřovat určitý sémantický vztah mezi nimi. V případě, že mezi textovými jednotkami žádný konektor není povrchově přítomný, sémantický vztah je implicitní, tj. není explicitně vyjádřen lexikálním prostředkem. Syntaktické vztahy mohou v textu tvořit hierarchické struktury, podobně, jako je tomu na rovině věty.

### 2.3.4 Rétorické vztahy (obsahově sémantické vztahy, rematické vztahy)

Rétorické vztahy v textu jsou takové, jejichž sémantika není motivována syntakticky, nýbrž kompoziční výstavbou textu. Určují především, jakým způsobem obsah výpovědi navazuje na obsah celého předchozího úseku, zda jej rozvádí, shrnuje, udává příklad atd. Podrobněji

rétorickým vztahům věnujeme kapitolu 3.

### **2.3.5 Temporální souvztažnosti**

Obligatorností vyjádření gramatické kategorie slovesného času v indikativních strukturách je dána poměrně svébytná temporální struktura diskurzu, na rozdíl od obvykle méně viditelných vztahů prostorových. Shoda časových forem významně přispívá i ke koherenci textu. Temporalita bývá kromě slovesného času vyjádřena adverbialními doplněními s časovým významem a, jak chceme ukážeme v kapitole 5.6, i některými konektory.

### **2.3.6 Grafické a zvukové členění textu**

K srozumitelnosti textu přispívají i prostředky formálního členění textu: U mluvených projevů je zásadním členícím prostředkem intonace, u psaných interpunkce. Intonace i interpunkce ovšem mají schopnost nejen formálně členit text, ale mohou být také významotvorné. Mohou mít vztah k aktuálnímu členění, některé druhy interpunkce mohou mít v tektogramatické reprezentaci funkci konektoru (viz kapitola 5.7).

### **2.3.7 Komunikační a pragmatické faktory**

Koherence pragmatická je založena na mimotextových jevech, je opřena o společné znalosti a zkušenosti účastníků komunikace, které nejsou předmětem sdělení. Pragmatická koherence textu tak není objektivní vlastností textu, nýbrž je textu připisována recipiency, a tedy může být pro každého recipienta jiná. (Hrbáček 1994: 27) V této oblasti se zkoumají například presupozice, situační kontext (world knowledge či domain knowledge – Asher 1993, Asher a Lascarides 2003), asociační vztahy, exofory, bridging anafory a salience – hierarchie stupňů aktivovanosti prvků společné zásoby znalostí komunikačních partnerů (Hajičová, Vrbová 1982 aj.).

## 2.4 Pojetí diskurzu ve Funkčním generativním popisu

Na základě předchozích kapitol nyní vymezíme pojetí diskurzu ve Funkčním generativním popisu. Diskurzem se v FGP rozumí **užívání jazyka jako systému v procesu komunikace**, diskurz je tedy chápán jako posloupnost výpovědních událostí. Smysl výpovědní události je tvořen významem věty jako jednotky jazyka a přiřazením reference referujícím výrazům (zejm. Sgall, Hajičová, Panevová 1986). Z toho plyne, že nelze chápat diskurzní vztahy jako gramatiku textu, ale že jde o pojem širší: diskurz je provázaný sítí syntaktických, sémantických a pragmatických vztahů.

V souvislosti s otázkou, zda zařadit diskurz do sféry langue či parole (viz kapitola 2.2.1.1), vyvstává otázka, zda mluvit v rámci FGP o další rovině jazykového popisu, o rovině „diskurzní“, která by byla nástavbou nad rovinou hloubkově syntaktickou (tektogramatickou), a jako taková by byla v rámci anotačního schématu PDT i anotována. „Sometimes, a level of *discourse* or ”hypersyntax“ in the language system is postulated, too, representing one of the early branches of the first attempts at text linguistics in Prague.“ (Hajičová, Čermák 2003: 7). Jistě by bylo možné chápat klauze na rovině syntaktického popisu jako diskurzní jednotky, jejichž funkce se projevuje na rovině diskurzní. Roviny popisu by tak zůstaly provázané. Problematické ovšem je, že pro správnou interpretaci diskurzu a tedy i jeho jednotlivých částí je třeba více informací, než poskytuje detailní popis a provázání jazykových rovin. Je nutné zohlednit mimojazykové skutečnosti, kontext nejen jazykový, ale i situační, je nutné dívat se na diskurz v jeho aktuálních souvislostech. Je tedy třeba udělat „úrok stranou“ od systémového stratifikačního popisu jazyka a vyjít z předpokladu, že některé vztahy v diskurzu se blíží systematickému popisu jazyka a lze v nich vyzkoumat určité zákonitosti a pravidelnosti, konečné porozumění smyslu jazykového projevu je však závislé na jeho zakotvení v komunikačním procesu, který je sám o sobě jedinečný.

### 2.4.1. Jaké vztahy zachycovat v popisu diskurzu?

Vedle zachycení hloubkové syntaxe, včetně aktuálního členění, zahrnují tzv. tektogramatické struktury v Pražském závislostním korpusu některé druhy gramatické a textové koreference. Pro zachycení některých aspektů diskurzu v rámci PDT se počítá s převzetím anotace vztahů

**aktuálního členění**, lépe řečeno s uchováním této informace v diskurzu, dále by se přejala anotace **textové koreference**. Dále by se využilo anotace diskurzních konektorů v tektogramatické rovině (především tektogramatický funktor PREC), od kterých by se vycházelo při popisování, sémantické klasifikaci a konečně při anotaci **syntakticky motivovaných vztahů** v diskurzu. V souvislosti s aktuálním členěním se diskutuje možnost anotace tzv. „**topics**“ – **obsahových témat** v diskurzu a taktéž anotace **stupňů aktivace společných znalostí** mluvčího a posluchače. Otázkou zůstává, zda v diskurzu zachycovat vztahy kompoziční, čili **rétorické**, a zda se věnovat **temporalitě diskurzu**. Zodpovězení této otázky vyžaduje ještě detailní průzkum těchto typů diskurzních vztahů.

Nezodpovězenou zatím zůstala také otázka, proč vůbec tvořit diskurzní rovinu v PDT a proč zmíněné vztahy anotovat. Kromě rozšířených možností lingvistických bádání v oblasti diskurzu, která by anotovaná data umožňovala, může rozvinutí poznatků v této oblasti umožnit vyvození určitých algoritmů pro počítačové modelování diskurzu či určitých vztahů v něm. V neposlední řadě se zkoumáním komplexnějšího jevu (diskurzu) můžeme zpětně dozvědět něco nového o jevech méně komplexních (o struktuře a fungování věty apod.).

Syntakticky anotovaný korpus češtiny umožňuje založit lingvistický výzkum na předzpracovaném jazykovém materiálu. Jeho nevýhodou však je, že jazykový materiál Pražského závislostního korpusu pochází z poměrně úzce stylisticky vymezeného okruhu publicistických textů. Výzkum je tak nutně ovlivněn konvencí užívání jazyka v publicistice.

#### **2.4.2 Předmět zkoumání**

Předmětem této studie jsou zejména ty diskurzní vztahy, které se zakládají na syntaxi, tj. jsou z velké části již zachycené v tektogramatické rovině PDT. Nezabýváme se zde aktuálním členěním, vztahy koreferenčními (s výjimkou toho, kdy jsou diskurzní konektory chápány jako anaforické), nebudeme se zabývat ani pragmatickými faktory a zvukovým či grafickým členěním diskurzu. Vztahům rétorickým a temporálním věnujeme samostatné kapitoly zejména z toho důvodu, že často mají jako svůj identifikátor určitý typ diskurzního konektoru.

Není však prvotním účelem studie podat podrobnou sémantickou klasifikaci diskurzních vztahů založených na syntaxi. Tento návrh již byl učiněn na základě porovnání závislostních

vztahů tektogramatické roviny PDT 2.0 a připravované druhé verze Penn Discourse Treebanku (Zikánová 2007). Vyžaduje samozřejmě další prozkoumání a rozvedení a v této práci se k němu často vztahujeme. Zkoumání dalších možností sémantické hierarchie těchto vztahů však samo o sobě předmětem této studie není, přibližujeme se mu ovšem z jiných stran: (i) přes lexikální prostředky vyjádření těchto vztahů; (ii) přes náš návrh vymezení syntaktických vztahů v diskurzu vůči vztahům na hranici syntaxe a rétorické kompozice textu, tj. vůči vztahům, které jsou jiného charakteru, než mají vztahy zachycené závislostní hranou na tektogramatické rovině PDT.

Tato studie se tedy věnuje především diskurzním konektorům jako lexikálním prostředkům zachycení těchto diskurzních vztahů. Vychází z rozboru jazykových výrazů, které v tektogramatické rovině PDT dostaly funktor PREC (reference to preceding context), a tedy je jejich hlavní funkcí referovat k předchozímu kontextu. Jsou to zejména výrazy typu *však, ovšem, přitom, navíc, tedy, totiž* atp. Dále se zabývá těmi výrazy, které mají v TR jinou sémantickou značku, ale jejich funkce se zdá být podobná – tedy spojovací, připojovací, odkazující k předcházejícímu kontextu.

#### 2.4.3 Termíny vztahující se k popisu diskurzu

Termínů *diskurz* a *text* jsme v této práci doposud užívali tak, jak s nimi pracovaly jednotlivé školy. V následující části práce, které pojednává již především o přístupu FDP, budeme používat termínu *diskurz* pro jazykové projevy obecně. Slova *text* budeme užívat ve smyslu psaného dokumentu, artefaktu, nejčastěji ve významu konkrétního textu z PDT, na kterém provádíme analýzu. Zároveň ale zůstaneme u již tradičního užívání slova *text* u pojmů jako *textová koherence, textová koreference, textová syntax* atd. Synonymně užíváme výrazy jako *segment textu/diskurzu, textová/diskurzní jednotka, textové/diskurzní vztahy* atd.

Termínů *klauze* a *věta* budeme užívat v souladu s Manuálem pro anotaci na tektogramatické rovině PDT (Mikulová et. al 2005), *klauze* je tedy jednoduchá věta s jednou predikací, *věta* je hyperonymní označení jak pro klauzi, tak pro souvětí a i pro výpověď, často je pro naše účely relevantní zejména charakteristika věty jako útvaru „od tečky k tečce“.

*Spojku* rozumíme tradiční slovnědruhovou kategorií, *spojovacím výrazem* jakýkoli jazykový výraz, který má spojovací funkci na rovině větného popisu, (*diskurzním*)



*konektorem* spojovací či připojovací výrazy, které mají tuto funkci na rovině popisu diskurzu. U tohoto pojmu vycházíme jednak z tradice popisu v českých lingvistických příručkách (nevymezujeme ovšem opozici junktor – konektor, naopak chápeme junktory, čili spojovací prostředky mezi klauzemi, jako součást diskurzních konektorů), jednak z překladu anglického výrazu *discourse connective*. Rozdíl mezi funkcí spojovací a funkcí připojovací chápeme tak, že spojování se uskutečňuje v rámci věty mezi klauzemi, zatímco připojování mezi větami. Označení *discourse marker* patří těm jazykovým výrazům, které signalizují určitý diskurzní vztah, nemusí mít pouze funkci konektivní (tedy spojit či připojit dva textové segmenty). V češtině tyto výrazy pojmenováváme obecně jako *diskurzní částice*.

Označení *diskurzní/textové vztahy* (preferujeme označení *diskurzní vztahy*) je problematické. V přehledových částech studie označuje všechny typy vztahů, které se v diskurzu vyskytují, jejich stručný přehled podává kapitola 2.3. V kapitole 5 je však tohoto označení záměrně užito ve významu užším – chápeme tak diskurzní vztahy založené na syntaxi, tj. ty diskurzní vztahy, ve kterých se zpravidla signalizuje propojenost sousedních klauzí nebo vět určitým operátorem, diskurzním konektorem. Jsme si vědomi této homonymie, nicméně jsme nenašli žádné vhodné označení pro užší skupinu těchto vztahů a zároveň předpokládáme, že správná interpretace těchto termínů je snadno odvoditelná z kontextu.

Termíny *koherence* a *koheze* bývají v textové lingvistice často používány nejednotně. Někdy se těchto pojmů užívá synonymně (např. Hrbáček 1994: 9), někdy se koherence váže na významovou spojitost textu a koheze je projev koherence na výrazové, povrchové rovině. (MČ III: 633), někdy jsou tyto pojmy chápány naopak (Černý 1998: 213). Zde se držíme druhého přístupu, koherence je tedy významová spojitost a soudržnost textu a je jedním ze základních předpokladů pro srozumitelnost textu.

Pojmů *predikát* a jeho *argumenty se* v FGP užívá v syntaxi k popisu slovesa a jeho doplnění. Na rovině popisu diskurzu se za predikát (binárního) vztahu považuje diskurzní konektor, který přijímá argumenty v podobě určitých segmentů diskurzu, nejčastěji klauzí.

## Kapitola 3. Rétorické vztahy v diskurzu

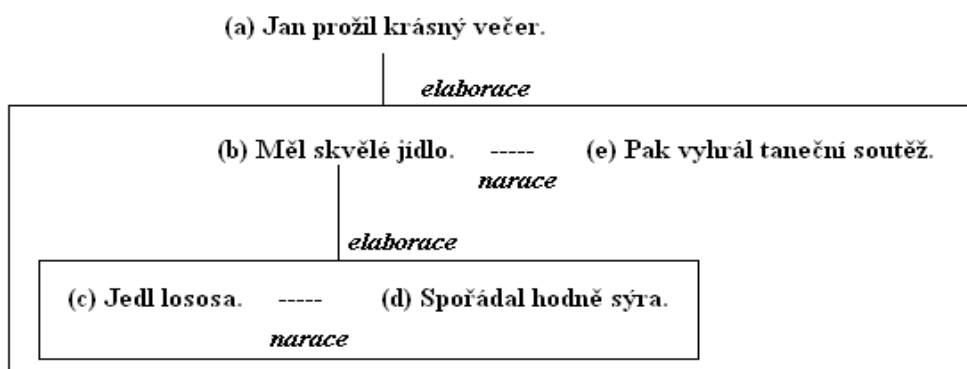
V návaznosti na dřívější teorie zabývající se rétorickými vztahy v diskurzu, zejména v návaznosti na Rhetorical Structure Theory (RST) (Mann a Thompson 1986), se od devadesátých let vyvíjejí přístupy, které usilují jednak o interpretaci diskurzu a jednak o vytvoření formálně precizní teorie pro modelování struktury diskurzu jinak než na základě syntaktických vztahů. Přiblížíme si model diskurzu, který stojí na pomezí sémantiky a pragmatiky, abychom si objasnili pojem *rétorických vztahů* v diskurzu a zamysleli se nad jejich využitelností pro náš popis diskurzních vztahů.

### 3.1 Segmented Discourse Representation Theory

Nicholas Asher (1993) ve své teorii diskurzu chápe diskurzí vztahy jako pojící prostředek pro vytváření komplexních jednotek z jednotek jednodušších. Rozlišuje mezi vztahy rétorickými a koherenčními. Koherenční vztahy segmentují diskurz na základě vztahů mezi *eventualities* (děje a stavy), které jsou popisovány jednotlivými konstituenty diskurzu, zatímco vztahy rétorické segmentují diskurz dle rétorické funkce té propozice, která je ve vztahu s propozicemi již začleněnými v diskurzu. (1993: 264) Některé z těchto vztahů závisí na struktuře konstituentů, některé na jejich sémantice. Rozdíl mezi rétorickými a koherenčními vztahy je dle Ashera sémantický: rétorické vztahy obsahují informaci o pravdivostních podmínkách v rámci diskurzu (např. *explanation* – vysvětlení), koherenční vztahy obvykle přímo vypovídají o pravdivostních podmínkách samotných konstituentů (např. příčina, která zčásti odpovídá rétorickému vztahu *explanation*). Asher uvádí, že povaha rétorických vztahů neumožňuje jejich úplný výčet, často jsou pro jiné druhy diskurzů relevantní jiné „soubory“, nicméně podává výčet těch užitečných, „of the useful ones“: *elaboration, explanation, instance, purpose, continuation, consequence, generalization a contrast*. Tyto rétorické vztahy pak definuje s využitím prostředků formální logiky.

V novější práci (Asher a Lascarides 2003) v rámci své teorie diskurzu (Segmented Discourse Representation Theory, SDRT) od popisu koherenčních vztahů upouští a pracuje se sedmi základními rétorickými vztahy, které jsou již synonymem pro vztahy diskurzí: *Narration* (časová následnost propozic), *Elaboration* (uvedení detailu či výčtu vzhledem

k obsahu předchozí propozice), *Explanation* (vysvětlení, předčasnost propozic), *Result* (výsledek/následek), *Background* (děj v pozadí) *Contrast* (kontrast proti předchozímu obsahu, v širokém pojetí), *Parallel* (současnost propozicí). Diskurz je koherentní, tedy dává smysl („makes sense“) tehdy, (i) pokud jsou všechny propozice rétoricky spojeny s jinou propozicí (částí informace) v diskurzu; (ii) pokud správně interpretujeme všechny anafory. Asher tvrdí, že rétorická struktura diskurzu je hierarchická; například vztah *Narration* – narace – implikuje souřadnost, zatímco *Elaboration* – elaborace, rozvinutí – podřadnost. Hierarchická rétorická struktura diskurzu může dle Ashera vypovídat o dalších vztazích: rétorické vztahy určují hranice pro srozumitelnost anafory (na základě její vzdálenosti od antecedentu, tj. místa zapojení v rétorické struktuře) a jejich sémantika zahrnuje informace o temporálních vztazích v diskurzu. Pojetí koordinace a subordinace rétorických vztahů v diskurzu je ale značně odlišné od pojetí syntaktického. Popisuje totiž především možnosti a hranice navazování propozic v diskurzu, a ukazuje souvislost i s možnostmi pronominální reference. Zjednodušeně lze říci, že aby byl diskurz koherentní, lze navázat další klauzi pouze na pravé hranici rétorické struktury (*right frontier constraint*, 2003:10), tedy na propozici v předchozí klauzi či na jakoukoli propozici, která předchází klauzi dominuje. Podrobněji si to ukážeme na následujícím příkladu (věty jsou přejaté z Asherovy analýzy a přeložené do češtiny):



**Obr. 8** Ukázka rétorické struktury diskurzu dle SDRT

(7)

(a) *Jan prožil krásný večer.*

(b) *Měl skvělé jídlo.*

(c) *Jedl lososa.*

(d) *Spořádal hodně sýra.*

(e) *Pak vyhrál taneční soutěž.*

Grafické znázornění rétorické struktury tohoto diskurzu, jak jej podává Asher, ukazuje, že (a) dominuje všechny následující věty vztahem elaborace, stejně jako (b) dominuje věty (c) a (d). (c) a (d) jsou v souřadném vztahu narace (čili časové následnosti) a stejně je tomu tak i mezi (b) a (e). Asher chce ukázat, že hierarchie rétorických vztahů vylučuje, aby věta

(f) *Měl krásnou růžovou barvu.*

v diskurzu následovala, tj. navazovala na (e), neboť se vztahuje k (c), k lososovi.

Zároveň je možné, aby věta

(g) *Pak šel spokojený domů.*

v diskurzu následovala, neboť se může vztahovat jak k větě poslední, tak hypertématu v (a), které jsou obě na pravé hranici struktury. SDRT tuto možnost dvojí interpretace nevylučuje.<sup>23</sup>

V pojetí SDRT se může (na rozdíl od dřívější Rhetorical Structure Theory a od Penn Discourse Treebanku) mezi dvěma segmenty diskurzu existovat dvojí vztah. Pokud mezi dvěma segmenty existuje rétorický vztah narace, čili propozice následují v čase za sebou, mohou být rétoricky propojeny dalším vztahem. V příkladovém diskurzu, historice o dvou mužích, kteří ve snaze se vzájemně vyhnout do sebe nakonec vrazí (2003: 34), jsou první a poslední věta spojeny vztahem narace a zároveň vztahem výsledku (*result*). Také kvalita (nebo „čitelnost“, explicitnost) rétorických vztahů může být různá, pohybuje se na jakési stupnici. Asher zde uplatňuje tzv. princip maximální koherence (MDC – *Maximizing*

---

23 Věta (g) může navazovat na každou větu diskurzu, protože je v každé zachován anaforický řetězec ke slovu Jan. Zařazením věty (g) se však zruší možnost stejného pokračování diskurzu, a tedy se změní i rétorická struktura. Věta (f) by oproti tomu na pokračování diskurzu neměla vliv.

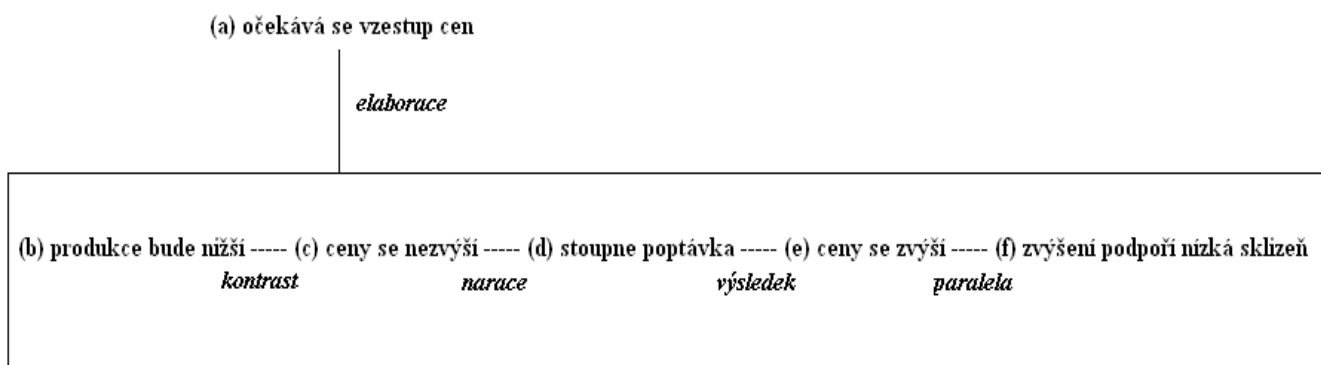
*Discourse Coherence*), (2003: 21), podle něhož je diskurz nejsrozumitelnější (a tedy nejsnáze umožňuje správnou interpretaci) tehdy, když jsou jeho segmenty propojeny více rétorickými vztahy nebo explicitnějšími rétorickými vztahy. Tento princip se užívá zvláště při tzv. bridgingu (tj. u koreferenčních vztahů vyjádřených nepřímo) či při disambiguaci slovního významu v situacích, kdy je aktuální sdělení v rozporu s intuitivně preferovanou interpretací recipientů. Interpretace, která maximalizuje viditelnost rétorických vztahů, je více koherentní (a tedy srozumitelnější) než jiné interpretace. Existence více rétorických vztahů mezi dvěma segmenty tedy dle SDRT zvyšuje možnosti správné interpretace diskurzu.

Jak jsme již zmínili, teorie SDRT nepodává detailní analýzu všech rétorických vztahů, které používá. (Asher a Lascarides 2003: 442). Zároveň se v ní uvádí, že ačkoli je diskurz koherentní „přilepením“ každé propozice k jiné části diskurzu (*glue logic*, 2003: zejména 184-189), není nutné v analýze diskurzu v každém případě plně identifikovat všechny rétorické vztahy, pokud nemají disambiguační funkci (2003: 450). Nevýhodou SDRT vzhledem k našemu účelu je, že velmi málo pracuje s autentickými daty, mnoho příkladů je vytvořeno za účelem názorného vysvětlení popisovaných jevů. Pokud SDRT uvádí autentické příklady diskurzu, jsou to většinou dialogy, na kterých se demonstruje především jejich logická výstavba. SDRT se také téměř nevěnuje problematice syntaktické parataxe a hypotaxe, což je vzhledem ke stanovenému hledisku oprávněné.

Nyní se pokusíme podle vzorové analýzy diskurzu analyzovat dva krátké publicistické texty z Pražského závislostního korpusu. Analýza má poukázat na výhody a problémy, které s sebou tato teorie přináší.

(8) PDT

- (a) *Očekává se vzestup cen cukru.* [= nadpis]
- (b) *Podle odhadu britské obchodní firmy E. D. and F. Man bude objem letošní produkce cukru v zemích EU nižší než loni.*
- (c) *Přesto by se však jeho ceny ještě v prvním čtvrtletí příštího roku neměly zvyšovat.*
- (d) *Teprve potom se projeví stoupající poptávka ze strany Ruska a Číny,*
- (e) *což bude působit na vzestup cen.*
- (f) *Jejich pohyb směrem vzhůru navíc podpoří nízká úroveň sklizně na Kubě i v samotných zemích EU, místo očekávaných 17.55 mil. tun pouze 14.8 milionu.*



**Obr. 9** Analýza rétorické struktury diskurzu dle SDRT I.

V analýze diskurzu (8) se ukázalo jednoduché rozhodnout o struktuře i o sémantické charakteristice rétorického vztahu, protože ve většině případů byl vztah explicitně (lexikálně, gramaticky) naznačen. Věta (a) je nadpis článku, proto zpravidla elaboruje obsah celého následujícího textu. Nejistota může vzniknout tam, kde (i) řešíme velkou míru hypotaxe – zde nám totiž nebude stačit hrubé vymezení rétorických vztahů; (ii) kde je málo signálů o charakteru vztahu, což si ukážeme rozbořem následujícího textu:

(9) PDT

(a) *Narozeniny jako provokace* [= nadpis]

(b) *Jinak než provokací se hra Jaroslava Duška Narozeniny, uvedená v neděli na stanici*

*Vltava v režii Michala Pavlíka, nedá nazvat .*

(c) *Spotřebitelské fráze rodinného života soupeří s hravostí snu a fantazie.*

(d) *Rodiče si pletou, jak vypadá jejich syn, kterého už dávno nevnímají*

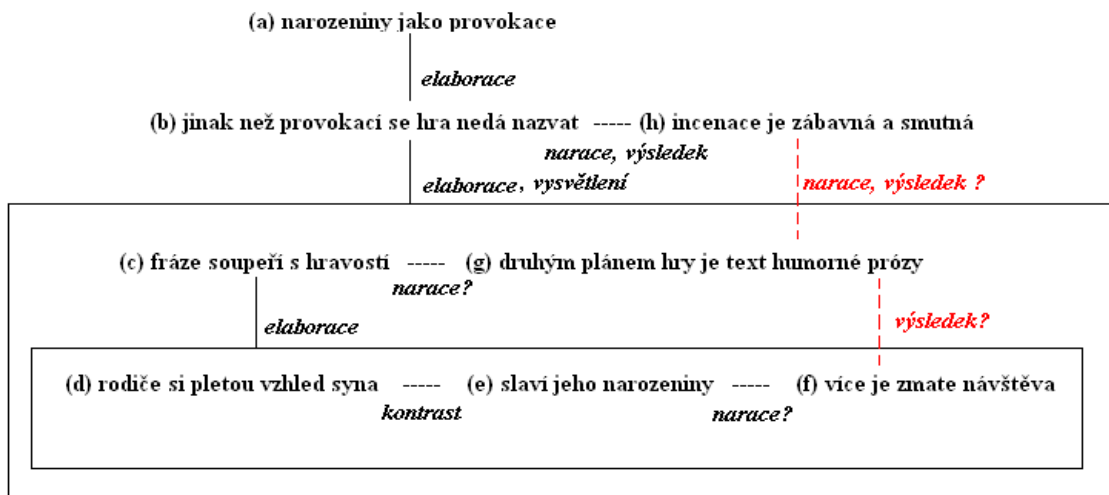
(e) *- přesto slaví jeho narozeniny.*

(f) *Ještě více je zmate návštěva smyšlených bytostí z éteru.*

(g) *Druhým plánem hry je totiž dialogizovaný text humorné prózy Rosti Volného, který*

obrací vše naruby.

(h) *Inscenace je zábavná i smutná právě tím, jak se nám podobá.*



**Obr. 10** Analýza rétorické struktury diskurzu dle SDRT II.

Určení rétorických vztahů v diskurzu (9) je značně komplikované, neboť v něm nalzáme málo prostředků vyjádření koherence, diskurz se jeví jako částečně nekoherentní. Toto zachycení struktury diskurzu (9) je pouze jedním z možných návrhů. Mezi klauzemi (b) a (c) a (c) a (d) lze jen s obtížemi vypořádat diskurzivní vztah jakéhokoli charakteru, nejsou zde žádné anafory, žádná napovídající časová struktura a ani tematicky tyto tři klauze nevykazují žádný společný prvek. Laicky řečeno, kdybychom tyto klauze od sebe oddělili, neměli bychom spolehlivé vodítko, jak je seřadit v původním sledu. Vztah elaborace mezi těmito klauzemi nevyplývá při čtení textu, nýbrž je zřetelný až po rozboru, tedy v rámci celé struktury. Dále není jasné, k čemu se bezprostředně vztahují klauze (g) a (h). Červenou barvou a přerušovanou čarou jsou v diagramu označeny ty možnosti jejich vztahu k ostatním klauzím, které v tomto zachycení nebyly upřednostněny, avšak jsou také oprávněné. Otazníky jsou všude tam, kde nebyla jasná sémantická charakteristika rétorického vztahu, nebo kde je vztah velmi slabý. Například klauze (h), která má podávat určité závěrečné hodnocení divadelní hry, o které článek pojednává, by mohla v textu stát i na třetím místě a uvádět obsah divadelní hry (dle našeho názoru explicitněji, než tak činí klauze (b)). Stejně tak je zcela

kontextově nezapojená klauze (b). Tato volná „posouvateľnost“ některých klauzí v rámci diskurzu svědčí o tom, že z hlediska srozumitelnosti pro čtenáře byl text napsán špatně.<sup>24</sup> °tom jsme se přesvědčili i menším průzkumem: Text jsme rozdělili na jednotlivé klauze (a) až (h), zaměnili jsme jejich pořadí a požádali několik rodilých mluvčích češtiny, aby určili původní pořadí klauzí v textu. Ani jeden z nich nebyl schopen toto pořadí určit napoprvé správně.

Přesto jsme pomocí rétorických vztahů modelu SDRT byli schopni alespoň v hrubých rysech zachytit strukturu málo koherentního diskurzu. Jasně vodítko jsme měli tam, kde jsme rozpoznali charakter rétorického vztahu jako syntaktický, například mezi klauzemi (b) a (c), (d) a (e) v diskurzu (8) a mezi (d) a (e) v diskurzu (9) a kde byl diskurzí vztah explicitně naznačen anaforickými prostředky.

Jak se zdá, některé rétorické vztahy v pojetí SDRT mají také syntaktický charakter a mohli bychom říci, že jsou srovnatelné s některým ze syntaktických vztahů či s nějakou jejich kategorií: rétorický vztah *kontrast* v pojetí SDRT by sémanticky odpovídal adverzativnímu (ADVS), konfrontačnímu (CONFR), kontradikčnímu (CONTRD), gradačnímu (GRAD), srovnávacímu (CPR) či přípustkovému (CNCS) vztahu v tektogramatické reprezentaci v PDT. Takové rétorické vztahy jsou tedy již zachyceny a podrobněji klasifikovány jakožto vztahy syntaktické, v PDT však zatím pouze mezi klauzemi, nikoli mezi větami.

Jiného druhu jsou ty rétorické vztahy, které zachycují především temporalitu (současnost, následnost, předčasnost, děj na pozadí...) a princip výstavby diskurzu (shrnutí, rozvedení, výčet, exemplifikace, jiná formulace...). Avšak i syntakticky pojaté diskurzí vztahy mohou obsahovat a často obsahují informaci o časovém sledu propozic. Podívejme se na následující příklady (přejaté a přeložené z Asher a Lascarides 2003: 6):

(10)

(a) *Petr upadl. Pavel mu pomohl vstát.*

(b) *Petr upadl. Pavel ho strčil.*

Oba diskurzy jsou koherentní, tedy dávají smysl proto, že jsme schopni správně identifikovat

---

24 To dokládají i stylistické chyby: valence slovesa *plést si* vyžaduje doplnění typu „koho/co“ s „kým/čím“. Spojení *druhým plánem hry* evokuje přítomnost plánu prvního. Co ho však dle autora tvoří, se lze z textu jen domnívat. Také není zcela jasné, k čemu z předchozího textu se váže konektor *totiž*.



koreferenční vztah u zájmen a diskurzí vztahy. Oba diskurzy jsou indikativu préterita a ani v jednom nejsou žádnými lexikálními prostředky vyjádřeny časové vztahy. Přesto snadno poznáme, že v (a) se jedná o následnost dějů (narace) a že v (b) děj v druhé klauzi předchází ději v klauzi první (*explanation*, vysvětlení). Jak uvádějí Asher a Lascarides (2003: 273), v tomto případě hraje roli sémantika užitých sloves. V (b) je tak navíc vyjádřen vztah příčiny a následku, který je syntaktické povahy. Jak je vidět, diskurzí vztah v (b) zachycuje temporální výstavbu diskurzu i určitou kauzální charakteristiku, o které se domníváme, že může být jak syntaktické, tak i rétorické povahy. Diskurzí vztah v (a) zachycuje pouze časovou následnost klauzí a syntaktický charakter nemá – pokud nebudeme ve všech takových návaznostech v diskurzu vidět syntaktickou konjunkci.

Domníváme se, že teorie rétorických vztahů SDRT, jak jsme ji zde v hlavních rysech představili, přesahuje modely diskurzu založené na syntaxi a ukazuje způsob, jak nakládat se segmenty, mezi kterými chybí (i) syntaktická relace (ii) *cue phrase* (2003: 451), neboli explicitní lexikální identifikátor diskurzího vztahu. Dále nám tato teorie pomáhá zjistit nejen to, kdy je diskurz nekoherentní (a tedy nesrozumitelný), ale také proč tomu tak je. Ačkoli její zaměření na modelování struktury diskurzu nenabízí vhodné prostředky pro rozbor autentických (nejen) korpusových textů, pokusíme se v následujícím uvažování o analýze diskurzu zachytit a popsat ty vztahy, které nejsou definovány v hloubkově syntaktickém popisu v PDT.

### 3.2 Pojetí rétorických vztahů v české lingvistice

Akademická Mluvnice češtiny uvádí, že významovou strukturu celistvého textu tvoří složitá síť vztahů různého charakteru. Některé typy relací mezi textovými jednotkami se uplatňují už ve struktuře souvětí jako vztahy mezi jednotlivými větami. Ty považujeme za vztahy syntaktické, zachycené v PDT pomocí funktořů (MČ III 1987: 628). Za velmi výraznou je považována síť časových relací, která vytváří svébytnou temporální strukturu textu. Tyto vztahy vyjadřují situování událostí v čase vzhledem k okamžiku promluvy nebo vzájemnou současnost či nesoučasnost událostí. Méně výrazná je síť vztahů prostorových (MČ III 1987: 629). Dalším typem významových textových vztahů jsou ty, které se řadí do oblasti nazývané

kompozice textu: „Textová jednotka přináší ve vztahu k obsahu jednotky předchozí např. exemplifikaci nebo ilustraci (mnohdy enumerativně formulovanou), korekturu nebo rektifikaci, analogii, přirovnání, specifikaci, definici, zobecnění, explikaci, argumentaci (...), interpretaci faktu apod. Jednotka vztažená k řadě předchozích jednotek, ev. k celému textu může obsahovat vyvození výsledků, závěr textu, shrnutí.“ (tamtéž 629).

Josef Hrbáček (1994: 52) popisuje skupinu vztahů, které nazýváme rétorické, jako mezivýpovědní vztahy obsahově sémantické (či vztahy rematické). Podává již určitou klasifikaci těchto vztahů, rozlišuje čtyři základní typy:

- vztahy enumerativní (výčtové)
- vztahy explikativní (vysvětlovací)
- vztahy temporální (časové)
- vztahy kauzální (příčinné)

Dodává, že vnitřně bývají tyto vztahy značně různorodé, a proto nepředkládá jejich další klasifikaci. Tyto vztahy jsou vzápětí demonstrovány na příkladech, příklady vždy dokládají vztah vyjádřený explicitně pomocí určitého lexikálního výrazu (konektoru<sup>25</sup>) a vztah implicitní. Domníváme se, že Hrbáčková klasifikace je problematická. Ukažme si to na jeho příkladech (1994: 53):

Vztah vysvětlovací (explikativní) implicitní:

(11) *Bylo vidět, že na podobné exkurze je dokonale zvyklý. Byly jedním ze zdrojů jeho příjmů.*

Vztah příčinný (kauzální) implicitní:

(12) *Zajdu hned odpoledne k Jolaně. Potřebuju vědět, jak na tom budeme dál.*

Zdá se, že první příklad obsahuje i komponentu příčiny a druhý příklad naopak umožňuje interpretaci explikativní. Máme za to, že kauzální vztah je primárně povahy syntaktické a jako takový je i mezi větami (tj. které jsou oddělené tečkou) zachytitelný dle anotačního schématu PDT. Vztah explikativní hodnotíme jako vztah povahy rétorické s tím, že syntaktický vztah kauzální je jednou z možností jeho konkrétní realizace (tj. vysvětlení udáním příčiny). Explikativní vztah však zahrnuje i obsah jiného než kauzálního charakteru, např. různé způsoby konkretizace.

---

25 Hrbáček ovšem používá pojmu *konektor* pro operátor formálních, nikoli sémantických textových vztahů. Tyto výrazy se ale v praxi ve velké většině překrývají.

V příkladech (11) a (12) se tedy dle našeho názoru jedná pokaždé o přítomnost dvou diskurzních vztahů mezi stejnými dvěma segmenty: jeden vztah je povahy syntaktické, druhý je povahy rétorické – a oba se podílejí na výstavbě sémantické struktury textu.

### **3.3 Příklady rétorických diskurzních vztahů mezi větami z Pražského závislostního korpusu**

V následujících příkladech nevyklučujeme přítomnost ještě jiného diskurzního vztahu než ten, který demonstrujeme mezi dvěma konkrétními segmenty diskurzu. Příklady jsou pouze ilustrační a nevyklučují jinou klasifikaci rétorických vztahů. Zde pouze dokládáme oprávněnost teorií zabývajících se rétorickými vztahy v diskurzu na autentickém jazykovém materiálu.

narace – časová následnost (temporální vztah)

(13) *Instalované měřiče teplé vody registrují všechnu protékající vodu, i když bylo třeba dlouho odpouštět studenou, než přitekla zase teplá. Vzápětí následoval další regulační stupeň a vše se opakovalo.* PDT

(14) *Přišli jsme si pro rozhovor za spisovatelem. Ptali jsme se.* PDT

paralela – současnost (temporální vztah)

(15) *Bylo krásné brzké odpoledne. Lehounce pršelo.* PDT

(16) *Před revolucí bývávaly součástí studia praxe. Studenti si mohli vyzkoušet některé věci, získat rozhled.* PDT

elaborace – rozpracování, rozvedení: výčet (seznam, list, enumerace), uvedení detailu, specifikace, exemplifikace, ilustrace, popř. komentář

(17) *Kupující, který se snaží posoudit cenu, by si měl také položit otázku, v jakém je obraz stavu.*

*Existují 4 možnosti. Ta nejlepší, která také opravňuje k nejvyšší ceně, je nepoškozené dílo v původním stavu. Téměř stejně ceněný je obraz v nepůvodním stavu, s restaurátorskými*

*zásahy, které jsou ovšem kvalitní. Menší hodnotu má samozřejmě dílo v původním stavu, leč poškozené. Vůbec nejhorší možnost nastává, je - li obraz v nepůvodním stavu a špatně zrestaurovaný. PDT*

*(18) Na rok 1994 bylo schváleno celkem 10 programů. Vedle dosavadních programů jsou to nové programy Speciál a Záruka. PDT*

explikace – vysvětlení – také vztah předčasnosti

*(19) Stává se však, že zákazník vyhodí peníze, protože zařízení na doupravu pitné vody je k ničemu. Kapalina, která z něj vystoupí, je stejná jako na začátku procesu. PDT*

*(20) Hlavním důvodem, který vedl Komerční banku (KB) k vypovězení takzvaných termínovaných vkladů s pevnou úrokovou sazbou v rozmezí od zhruba 14 do 18 %, je ten, že se zmíněný vkladový produkt neslučuje s podnikatelskými záměry KB. Vyplynulo to z včerejšího prohlášení náměstka generálního ředitele KB Karla Bednáře. PDT*

restatement – jiná formulace téhož, similarity – podobnost, analogie, přirovnání

*(21) Jedna šestka už Telekomunikacím padla: Technika postoupila dál. PDT*

Tento typ vztahů je v PDT částečně zachycen v rámci apozice.

generalizace – zobecnění: shrnutí – sumarizace, vyvození, závěr

*(22) Podle Naděždy Blažíčkové, která vede oddělení českého malířství 19. století v Národní galerii v Praze, se dá vysvětlit dnešní relativně malá poptávka po obrazech tím, že sběratelé, mezi kterými je nejvíce lékařů, nemají mnoho peněz. Poměrně velká část poptávky odpadla, když k nám ze zahraničí začali jezdit chudší turisté. Proti dřívějšku se však zase objevili noví zájemci o umění z řad podnikatelů, bank, spořitelen a realitních kanceláří. Zkrátka doby na investování do obrazů bývávaly lepší, ale obliba některého umění a vkus lidí jsou věci nevyzpytatelné. PDT*

### 3.4 Shrnutí

Rétorické vztahy v Asherově modelu diskurzu jsou dvojí povahy. Jednak jsou povahy syntaktické – tyto vztahy jsou již zachycené v PDT v rámci tektogramatické reprezentace, ačkoli mají jinou, zpravidla podrobnější sémantickou klasifikaci. PDT se však zatím omezuje na jejich zachycení uvnitř věty, tedy od tečky k tečce. Jednak jsou to vztahy povahy nesyntaktické, patřící svým charakterem spíše do oblasti kompozice textu/diskurzu. Ponecháváme jim označení *vztahy rétorické*. Tyto vztahy tektogramatická reprezentace PDT nezachycuje, týkají se buď temporality v diskurzu nebo přímo jeho „rétorické“ výstavby. Vztahy vyjadřující temporalitu se často objevují zároveň s určitým vztahem syntaktickým, mohou však stát samy o sobě (příklad (10)). Podrobněji se temporálním diskurzním vztahům věnujeme v kapitole o konektorech s časovým významem (5.6).

Vztahům samotné „rétorické výstavby“ diskurzu je třeba věnovat více pozornosti, neboť ještě nebyly podrobně popsány. Jsou to vztahy, které se běžně projevují i při spojování/připojování vyšších textových úseků v rámci celku textu. Mezi většími úseky textu se však mohou uplatňovat i takové významové vztahy, jež jsme dosud označovali jako syntaktické. Například dva po sobě jdoucí odstavce v argumentačním textu spolu mohou být ve vztahu opozice, neboli *kontrastu*. Sémanticky je druh takového spojení stejný, jako příslušný syntaktický vztah koordinační (ADVS, CONFR atd.) v rámci jedné věty nebo mezi dvěma sousedícími větnými celky. Mezi dvě sousedící věty, ze kterých jednoznačně vyplývá druh jejich vzájemného vztahu, je ještě možné doplnit konektor a hodnotit tento celek stejně, jako by se jednalo o souvětí. Avšak zdá se, že čím větší textové jednotky spojujeme, tím více se ztrácí možnost jednoznačné interpretace vztahu mezi nimi a s tím i možnost tak exaktního popisu těchto komplexnějších vztahů v diskurzu. V tomto případě by k lepší orientaci zajisté přispěla i plánovaná anotace tzv. „topics“, tj. obsahových témat v diskurzu (viz. kap. 2.4.1) Tak nově vyvstává problém vzájemného vztahu rétorických a syntaktických vztahů vzhledem k hierarchii vztahů v diskurzu. Zdá se, že roli při tomto vymezení již nehraje pouze sémantický druh určitého vztahu (tj. zda se jedná o kontrast nebo o shrnutí vzhledem k předchozímu kontextu), ale i velikost spojovaných jednotek, neboli hierarchie vztahů v diskurzu.

O tom, jak vztahy rétorické výstavby diskurzu podrobněji charakterizovat a klasifikovat, by tedy měla rozhodnout rozsáhlá a důkladná analýza textů. Jsme si vědomi toho, že jazykový

materiál korpusu, který obsahuje převážně publicistické texty, neposkytuje dostatečně široké pole pro tato zkoumání, nehledě na odlišná kritéria zkoumání, která předpokládáme pro mluvené a zvláště dialogické diskurzy.

Pro účely tohoto výzkumu postačuje, že jsme doložili existenci rétorických vztahů a jejich vztah k diskurzním vztahům syntaktickým na základě revize modelu SDRT, na základě ještě nevydané sémantické klasifikace diskurzních vztahů v Penn Treebanku 2.0, dále dle Hrbáčkova rozdělení obsahově sémantických mezivýpovědních vztahů a též dle příkladů z akademické Mluvnice češtiny. Zároveň se opíráme o sémantickou analýzu diskurzních konektorů zpracovanou v této práci na základě dat z PDT 2.0.

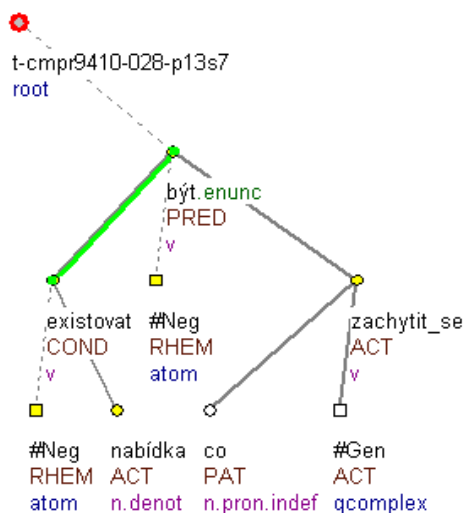
## Kapitola 4. Diskurzí vztahy zachycené v PDT

Jak jsme již zmínili v kapitole 2.4.1, tektogramatická rovina v PDT zachycuje v rámci podkladové syntaktické struktury věty i aktuální členění a navíc vztahy koreferenční. Těmito aspekty diskurzu se v dalším zabývat nebudeme. V této kapitole popíšeme, které vztahy s možným využitím pro anotaci diskurzu jsou na tektogramatické rovině zachyceny právě v anotaci syntaktické struktury věty.

Domníváme se, že anotace podkladové syntaktické struktury věty signalizuje diskurzí vztahy konektivní třemi způsoby:

- (i) závislostní hranou – jsou to některé vztahy syntaktické závislosti mezi klauzí řídící a klauzí závislou;
- (ii) uzlem pro kořen souřadné struktury a tzv. „nezávislostními“ hranami k přímým členům souřadné struktury – tedy vztahem syntaktické souřadnosti mezi klauzemi;
- (iii) tektogramatickým funktorem PREC, který u výrazu, kterému je přiřazen, signalizuje funkci připojovací, tedy odkaz k předchozímu kontextu.

### 4.1 Syntaktická závislost



Neexistuje-li nabídka, není se čeho zachytit. PDT

**Obr. 11** Syntaktická závislost jako diskurzí vztah

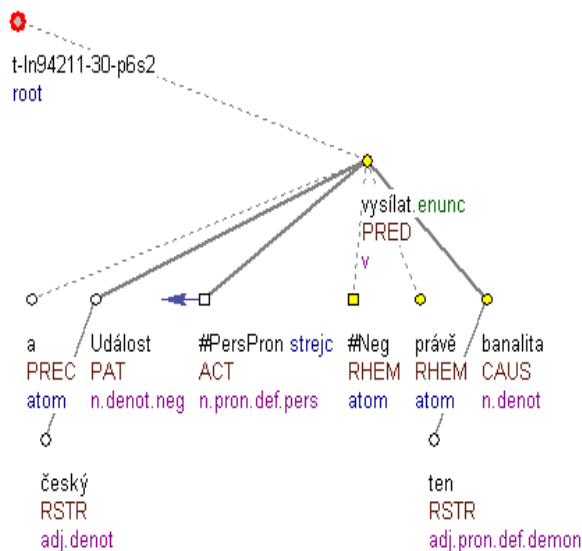
Prvním typem diskurzního vztahu zachyceného v syntaktické struktuře věty je vztah závislosti mezi řídicí a závislou klauzí. Na obrázku 11 je tento vztah znázorněn zeleně zvýrazněnou hranou, která spolu s funktorem COND u závislého uzlu označuje vztah závislosti a jeho druh – v tomto případě jde o podmínku.

Závislostními vztahy jako vztahy diskurzními se zabývá Šárka Zikánová (2007, technická zpráva). Ze závislostních vztahů považujeme za diskurzní pouze vztah řídicí klauze k některým typům závislé klauze adverbialní, za diskurzní naopak nepovažujeme klauze obsahové (které zastupují některé z valenčních doplnění slovesa) a klauze vztahné (které rozvíjejí pouze jmennou skupinu, nikoli primárně děj popisovaný predikací). V rámci sémantiky závislých klauzí adverbialních jsou diskurzními vztahy doplnění kauzálního typu, zůstává však ještě otevřenou otázkou, zda diskurzní vztah vyjadřují různá okolnostní doplnění děje v řídicí klauzi jako vyjádření způsobu a místní a směrová určení. Zdá se, že diskurzní vztahy závislostní svou podstatou odpovídají těm vztahům, které nejsou vztahy mezi slovesem a jeho valenčními doplněními. Dle pojetí slovesné valence, které počítá se směrovými doplněními sloves jako s valenčními (např. Šindlerová 2007: 65), by pak posouzení závislostního vztahu jako vztahu diskurzního vyplývalo z posouzení valence konkrétního slovesa. Této problematice je však třeba se ještě podrobně věnovat.

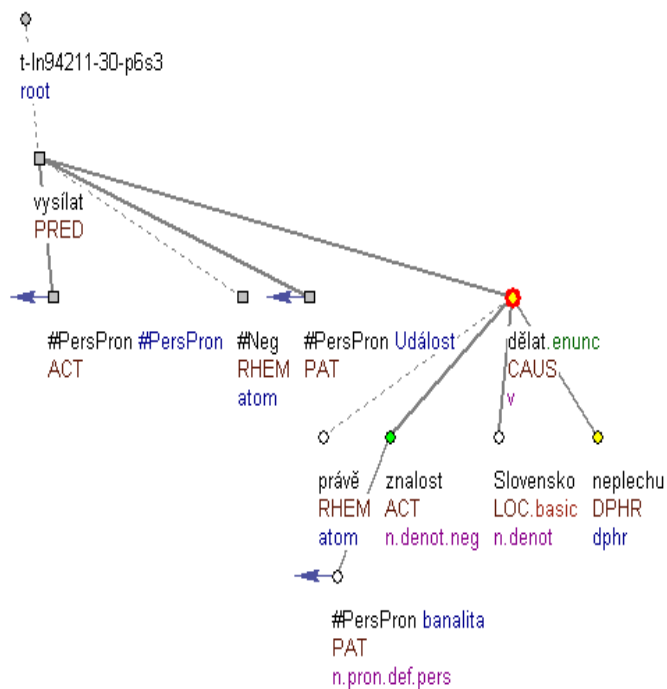
Vztahy závislostní jsou těmi vztahy v diskurzu, které se projevují mezi klauzemi v rámci jedné věty. Pokud je řídicí člen klauze v dané větě elidován (vyskytuje se např. ve větě předcházející), chápe se tento vztah jako parcelace, neboli jeden ze způsobů osamostatnění části výpovědi. V tektogramatické reprezentaci je v takové větě zachycena elipsa řídicí klauze a vztah klauze řídicí a závislé je reprezentován stejným způsobem jako v rámci jediné věty. Na obrázku 12 jsou zachyceny dvě za sebou jdoucí věty, z nichž druhá je parcelátem. Uzlu s t-lematem *dělat* (červeně označen), tedy kořeni závislé příčinné klauze je přiřazen výraz *protože*, který je jakožto podřadící spojka v hloubkové struktuře věty skrytý.



A nevyšílají české Události právě pro ty banality .



Protože právě jejich znalost by mohla na Slovensku dělat neplechu .

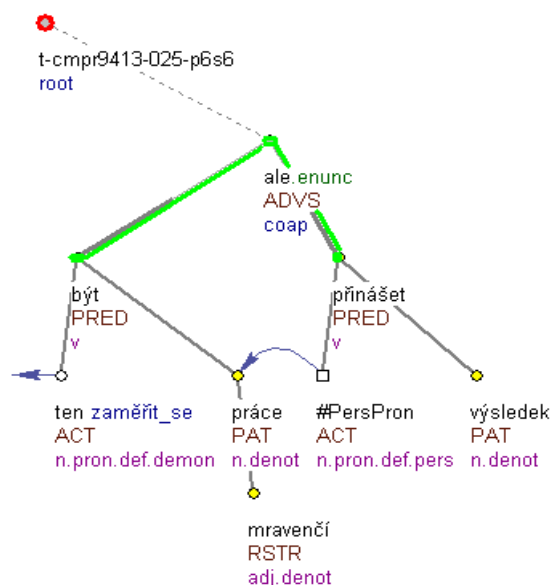


Obr. 12 Parcelace

## 4.2 Souřadnost

Druhým typem diskurzního vztahu zachyceného v syntaktické struktuře věty je souřadné spojení klauzí. Obr. 13 ukazuje (opět zeleně zvýrazněnou) koordinaci klauzí. Diskurzním konektorem se spojovací funkcí je zde spojka *ale*, která je jakožto souřadící výraz reprezentována uzlem s funktorem ADVS. Členy této koordinace jsou tedy spojeny vztahem adverzativním.

Je to mravenčí práce , ale přináší výsledky .



**Obr. 13** Souřadné spojení jako diskurzní vztah

Vzhledem k tomu, že některé závislostní vztahy mezi klauzemi zahrnujeme do vztahů diskurzních, a tudíž i některé podřadící spojovací výrazy chápeme jako konektory (což nemusí být přístup samozřejmý, např. Josef Hrbáček (1994) považuje za textové pouze vztahy koordinační a podřadící spojovací výrazy obecně nemohou fungovat jako textové konektory), nabízí se otázka, zda je na rovině diskurzu třeba rozlišovat mezi hypotaxí a parataxí. Zikánová ve své zprávě o anotaci PDTB ve Filadelfii (2007) podrobuje tradiční české rozdělení druhů syntaktických vztahů na souřadné a podřadné i některé funktoři užívané v PDT 2.0 revizi, zejména poukazuje na problematický vztah adverbativního poměru (ADVS) a přípustky (CNCS), vztahů konfrontačních vyjádřených hypotaktickými či parataktickými prostředky (CONTRA, CONFR, CONTRD) atd. Dále navrhuje pro anotaci diskurzu nezohledňovat při zachycování diskurzních vztahů jejich závislostní či koordinační charakter. Při propojení anotace diskurzu s anotací tektogramatické roviny stejným způsobem, jako je tektogramatická rovina propojena s rovinami nižšími, tato informace zůstane zachována.

### 4.3 Diskurzí vztahy signalizované funktorem PREC

Pokud bychom větu z obrázku 13 rozdělili na jednotlivé klauze, respektive bychom před výraz *ale* vložili koncový signál, nejednalo by se v tomto případě o parcelaci. Obě tyto klauze mohou stát samostatně, není mezi nimi vztah závislosti, pouze druhá klauze je k první připojena. Výraz *ale* uvozující druhou větu by dostal funktor PREC, čímž se signalizuje, že mezi větami existuje určitá spojitost (která však není dána syntaktickou závislostí). Třetím typem diskurzího vztahu zachyceného v syntaktické struktuře věty (ačkoli pouze jediným funktorem) je tedy **navazování či připojování klauzí v různých větách pomocí výrazu s funktorem PREC<sup>26</sup>**.

Tato sémantická kategorie je velmi heterogenní. Vytváří však rozsáhlou základnu pro třídu výrazů, jež nazýváme diskurzí konektory. Diskurzími konektory a výrazům s funktorem PREC v rámci třídy diskurzí konektorů se podrobně věnuje následující část této práce.

Na první pohled by se mohlo zdát, že diskurzí vztahy založené na koordinaci klauzí na jedné straně a diskurzí vztahy vyjádřené výrazem s funktorem PREC na straně druhé jsou stejného charakteru. Jako jediný rozdíl bychom pak chápali přítomnost koncového signálu (tečky, klesavé kadence) mezi dvěma argumenty tohoto vztahu – například jak jsme ukázali u věty z obrázku 13. Takový názor je oprávněný jen částečně. V následujících rozborech chceme ukázat, že třída diskurzí konektorů může v jistém smyslu přesáhnout možnosti a meze koordinace dvou klauzí, a tedy možnosti výrazů s koordinační funkcí ve větě (souřadicích spojek a dalších souřadicích spojovacích výrazů).<sup>27</sup>

### 4.4 Diskurzí vztahy uvnitř věty a přes hranice věty

Jak je vidět, kromě parcelovaných závislých klauzí (viz obr. 12) vztahy syntaktické závislosti i vztahy koordinační reprezentují ten druh diskurzího vztahu, který spojuje dvě jednotky **v rámci jedné věty**. V případě tektogramatického zachycení to znamená v rámci jednoho stromu (dokonce i věta parcelovaná se zachytí jako jeden strom, elidovaná řídicí klauze

26 Výrazy s funkcí PREC obvykle připojují dvě věty k sobě „přes tečku“. Není však výjimkou, že signalizují návaznost dvou klauzí v rámci jedné věty, jak ještě ukážeme dále.

27 Srov. MČ III: 447.

se zkopíruje z předchozího stromu). Tektogramatická reprezentace tedy vyčerpávajícím způsobem zachycuje a popisuje diskurzní vztahy mezi klauzemi v jedné větě (v jednom stromě).

Třetí způsob signalizace diskurzních vztahů v PDT, přiřazení funktoru PREC některým konektorům, je pouze označování operátorů těchto vztahů. Diskurzní vztahy reprezentované výrazem s funktorem PREC mají totiž zpravidla dosah za **hranice jedné věty**, čili jednoho stromu. Anotace tektogramatické struktury nepřekračuje hranice věty a nelze tedy bezprostředně ze stromu zjistit, co s čím daný operátor spojuje, respektive co k čemu připojuje. Také současný stav vyhledávacích nástrojů pro PDT umožňuje jen velmi omezené zkoumání jevů, které přesahují rámec jednoho stromu.

## Kapitola 5. Diskurzni konektory

Fungování textových či diskurzni konektorů jakožto lexikálních prostředků spojování textových jednotek zatím v české lingvistice bylo věnováno jen několik prací. Výrazy plnící konektivní funkci v textu byly detailněji zkoumány v mnoha jiných ohledech (např. spojky a jiné spojovací výrazy z hlediska parataxe a hypotaxe, částice z hlediska problematiky jejich slovnědruhového vymezení a dalšího třídění, další výrazy z hlediska jejich role při vyjadřování aktuálního členění věty apod.). Samotným textovým konektorům se blíže věnovali slovenští lingvisté, např. Mistrík (1975/6) a Bajzíkova (1975), u nás pak mezi jinými Daneš (1985: 171-186), Adamec (1995) a Hoffmannová (1983 a především 1984). Hoffmannová se zprvu soustřeďuje na třídu tzv. textových orientátorů, které především ukotvují text v časových a prostorových souvislostech (např. výrazy *zde, tam, ted', dříve, zatím, výše, níže*). Tuto skupinu v pozdější práci označuje za subkategorii v rámci širší skupiny konektorů, které vymezuje a rozděluje na základě jejich strukturních vztahů na konektory vlastní a implikativní. Konektory vlastní vyjadřují obvykle základní „logické“ vztahy obdobně, jako jsou vyjádřeny ve struktuře souvětí, konektory implikativní se mohou vztahovat i k obsahu v textu nevyjádřenému, mají funkci hierarchizace obsahu, vyjadřují přítomnost pragmatického prvku apod. (1984: 111). Další subkategorizace je dle Hoffmannové složitá, neboť obě skupiny konektorů mohou plnit v různé míře několik funkcí.<sup>28</sup> Jsou to funkce kompoziční (makrostrukturní), funkce orientace v textu (zejména časové a prostorové vztahy) a funkce kontaktní (včetně metatextových komentářů). Studie Jany Hoffmannové jsou zatím zřejmě nejpodrobnějším a nejucelenějším příspěvkem k popisu textových či diskurzni konektorů v češtině. V této práci pracujeme prozatím s poněkud užším pojmem diskurzni konektoru, necháváme stranou rozsáhlejší metatextové komentáře a orientátory vyjadřující prostorové vztahy.<sup>29</sup>

V této kapitole se pokusíme vymezit diskurzni konektory jako jednotnou skupinu na **základě zkoumání korpusového materiálu** (anglického Penn Discourse Treebanku a českých dat

---

28 „Es handelt sich eigentlich um die Differenzierung einer großen offenen Gruppe der Konnektoren, die in sich – wiederum in verschiedenen Proportionen und mit unterschiedlicher funktionaler Dominanz – mehrere Funktionen vereinigen“. (Hoffmannová 1984: 112)

29 Vycházíme ze sémantiky tektogramatického funktoru PREC a z výrazů, jejichž funkce s funkcí PREC hraničí.

z Pražského závislostního korpusu). Chceme tuto skupinu vymezit v základních rysech vůči jiným jevům (vůči větným členům, vůči deiktickým výrazům apod.) a určit její jádro a periferii. Pokusíme se tyto výrazy popsat dle jejich sémantické funkce ve výstavbě diskurzu, zamyslíme se nad tím, zda vždy spojují/připojují právě dvě jednotky diskurzu a pakliže přijmeme předpoklad, že se jedná svým způsobem o anaforické výrazy, pokusíme se zjistit, co a v jakém rozsahu bývá zpravidla jejich antecedentem. Vyjdeme z výrazů s funkcí PREC v tektogramatické rovině PDT a ze seznamu anglických výrazů s funkcí PREC, který byl vytvořen pro účely anotace anglické části paralelního korpusu Prague Czech-English Dependency Treebank (PCEDT) na základě pojetí *discourse connectives* v PDTB. Jsme si vědomi toho, že analýzou prostředků vyjádření určitého diskurzivního vztahu nevěnujeme pozornost vztahům implicitním, lexikálně neznačeným. Tato práce je však zaměřená na jevy patrné z tektogramatické anotace PDT, které mohou mít diskurzivní charakter. Dále jsme si vědomi toho, že do skupiny diskurzivních konektorů je třeba zahrnout i některé spojky a spojovací výrazy podřadící. Přesto je v této studii podrobněji nezkoumáme: Jednak nepřekračují hranice věty nebo tak činí pouze jakožto parceláty (viz kapitola 4.1) a jednak je obtížné je v PDT zkoumat – na rovině hloubkově syntaktického popisu jsou skryté.

## 5.1 Explicitní a implicitní vyjádření diskurzivních vztahů

Mohla by vzniknout námitka, že operátory diskurzivních vztahů jsou pouze povrchové formy, které signalizují určitý sémantický vztah, a nemusí se jim tedy věnovat taková pozornost jako sémantickému vztahu samotnému. Je to názor zajisté oprávněný; i z hlediska ekonomie zachycování diskurzivních vztahů v PDT je dobře možné samotné konektory skrýt. Problém nastane při anotaci. Jestliže připustíme, že v rámci věty jsme i bez přítomnosti explicitního konektoru velmi často schopni jednoznačně odvodit druh sémantického vztahu, není tomu tak i mezi větami. Čím delší a strukturně komplikovanější text máme, tím obtížněji v něm správně interpretujeme vztah mezi těmi úseky, kde není explicitně vyjádřen konektor. Často je více možných interpretací a nelze jednoznačně rekonstruovat autorův záměr. Často jsou diskurzivní vztahy slabé (viz v kapitole 3 o škálovém charakteru některých diskurzivních vztahů), často je text koherentní pomocí jiných druhů diskurzivních vztahů (tematické řetězce, asociace, koreference atd.) a syntaktický diskurzivní vztah mezi některými úseky buďto vůbec neexistuje, anebo ho nejsme schopni zachytit.

## 5.2 Tektogramatický funktor PREC

### 5.2.1 Základní charakteristika

Výrazy s funktorem PREC jsou základem pro vymezení skupiny diskurzních konektorů. Následující oddíl se tedy věnuje charakteristice těchto výrazů, problematice jejich vymezení a třídění a dále zejména roli, kterou hrají pro budoucí anotaci diskurzu. Tektogramatickým funktorem PREC jsou označeny takové jazykové výrazy, které spojují klauzi, ve které se vyskytují, s předchozím kontextem. Definice z manuálu pro anotaci tektogramatické roviny zní takto:

„Definice funktoru PREC: Funktor PREC (reference to preceding text) je funktor pro takový atomický uzel, který reprezentuje výraz signalizující návaznost klauze na předcházející kontext.“ (Mikulová et al. 2005: 537)

Tato definice a následně i vymezení výrazů s funktorem PREC při anotaci jsou spojené určitými problémy:

- (a) Na předchozí kontext může obecně navazovat daleko více jazykových prostředků, než zahrnuje funktor PREC. K tomu blíže v kapitolách 5.3 – 5.8.

Určitým negativním vymezením je zde dle definice atribut „typ uzlu“: výrazy s funktorem PREC jsou uzly atomické. Toto vymezení by pomohlo při rozlišení atomických uzlů od výrazů, které vyjadřují návaznost na předcházející kontext jiným způsobem, například deikticky – takové uzly jsou totiž komplexní. Definice atomického uzlu nám ovšem příliš nepomůže, jelikož v praxi vymezuje atomické uzly od komplexních až právě jejich funktor.<sup>30</sup>

- (b) Jedná se o formálně heterogenní třídu, do které spadají souřadící spojovací výrazy, dále některé adverbialní výrazy a některé částice (viz. kap. 5.2.3.). Tyto výrazy mají ve velké většině případů samy o sobě ještě jiný význam, který je „přenálepковán“ významem PREC. Jinými slovy, výrazy s funktorem PREC se mohou až na několik výjimek roztřídit do svých „původních“ sémantických tříd (tj. mohou dostat jiný funktor). Spojky užitě jako PREC mají stále původní specifickou syntaktickou

---

<sup>30</sup> Definice atomických uzlů je v manuálu vágní, je pouze výčtová. (Mikulová et al. 2005: 9) Atomické výrazy nejsou zasazeny do věty syntaktickou závislostí, ale spíše modifikují význam věty nebo její části „shora“, mimo základní syntaktickou strukturu; komplexní uzly naproti tomu reprezentují výrazy autosémantické – včetně zájmen a zájmených příslovcí – vyjadřující morfologické kategorie a zapojené v závislostní struktuře.

spojovací funkci, o udělení funktoru PREC souřadícímu spojovacímu výrazu rozhoduje pouze nepřítomnost první klauze ve větě, kde se vyskytuje tento výraz. Souřadící spojky v iniciální pozici tak budou mít vždy funktor PREC.<sup>31</sup> Adverbia užitá jako PREC si uchovávají i charakteristiku adverbialní, nejčastěji temporální atd. V některých případech je proto rozlišení mezi PREC a jiným, zejména adverbialním funktorem velmi obtížné.

- (c) Výraz s funktorem PREC a kontext, ke kterému odkazuje, se mohou nalézat v jednom souvětí, ačkoli ve většině případů tyto výrazy odkazují přes hranice věty. PREC tedy kombinuje vztahy uvnitř věty a vztahy mezivětné (tj. „přes tečku“), aniž je to v tektogramatické rovině explicitně popsáno.
- (d) Je-li v iniciální pozici ve větě jednoduché spojka podřadící, funktor PREC nedostane, protože podřadící spojovací výrazy jsou v tektogramatické reprezentaci „skryté“, tj. nemají vlastní tektogramatický uzel (viz kapitola 4.1). Tyto případy dodatečně připojené věty závislé, neboli parcelovaného podřadného souvětí se hodnotí jako elipsy klauze řídící. Vedly ovšem anotátory k chybným úvahám, hodnotili je jako PREC stejně jako výrazy souřadné. Do jisté míry je tato anotace oprávněná, podřadící spojka v takto dodatečně připojené podřadné struktuře také vykazuje rysy PREC, tedy navazování klauze na předchozí kontext. Srov. příklady (23) a (24). Správná anotace vět v příkladu (24) je na obrázku 12.

(23) *Proto je obzvlášť tristní poznání, že vlády na krátící se termín blokace zákona o bankrotu zřejmě jednoduše zapomněly. Nebo mu nevěnovaly dostatečnou pozornost.* PDT

(24) *A nevysílají české Události právě pro ty banality. Protože právě jejich znalost by na Slovensku mohla dělat neplechu.* PDT

---

31 Souřadící spojky mají pevné místo mezi spojovanými větami, na rozdíl od spojek podřadících, které mohou uvozovat celé spojení. Spojka souřadící v iniciální pozici je tedy musí odkazovat za hranice věty, ve které se vyskytuje.



## 5.2.2 Korpusové výzkumy – Analýza výrazů s funktorem PREC v PDT 2.0

Následující analýza vychází ze seznamu anglických výrazů s funkcí PREC, který byl zpracován na základě seznamu anglických explicitních *discourse connectives* v první verzi PDTB (viz Přílohy). Seznam anglických výrazů s funkcí PREC, který je určen pro anotaci anglické části paralelního korpusu PCEDT, byl přeložen do češtiny. V následující analýze byly v Pražském závislostním korpusu 2.0 vyhledány všechny přeložené ekvivalenty, jednak označené (i když chybně) funktorem PREC a jednak ty, které mají v PDT jinou sémantickou charakteristiku. Z analýzy byly vyloučeny pořadivé spojky.

V první části analýzy jsme vytvořili seznam českých výrazů, které mohou (ale nemusí) mít funkci PREC. Dle t-lemat těchto výrazů byla určena četnost jejich výskytu. Téměř žádné t-lemma nemá v korpusu pouze jednu sémantickou charakteristiku, u t-lemat, která se v korpusu vyskytla s funktorem PREC, byly vyhledány i ostatní jim přiřazené funktory. Tak vznikl statistický přehled o tom, jak který výraz sémanticky hraničí s jiným.

V některých případech je anotace velmi nekonzistentní. Příčina je dvojí: buď má jedna forma více různých funkcí (obzvláště širokou škálu funkcí vykazuje výskyt slova *tak*, viz příklady (25) – (35) ) nebo je anotace nesourodá z důvodu nedostatečně specifikovaných anotačních pravidel. Nesourodá anotace je v některých případech právě u funktoru PREC. Výrazy, které jsou v korpusu zjevně užity stejným způsobem a za stejným účelem, tak byly interpretovány různě, jelikož sémanticky umožňovaly přiřazení několika funkcí.

- (25) *Tak je to správně.* MANN
- (26) *Byl tak malý, že nebyl skoro vidět.* EXT
- (27) *Když nepřijdeš, tak půjdeme sami.* PREC
- (28) *Tak nám zabili Ferdinanda!* ATT (?)
- (29) *Přišel jak Jirka, tak Pavel.* CONJ
- (30) *Přišel jak Jirka, tak i Pavel.* GRAD
- (31) *A tak dále.* DPHR
- (32) *Pršelo, a tak jsme zůstali doma.* CM
- (33) *Tak, tak!* PARTL

+ v roli aktantů:

- (34) *Je vaší povinností učinit tak do konce roku.* PAT
- (35) *Přesto se tak dosud děje zcela ojedinele.* ACT

U výrazů, jejichž četnost výskytu jako PREC je větší než četnost výskytů s jiným funktorem (v tabulce 1 ve žlutých řádcích), můžeme tvrdit, že jsou prototypickými PREC, čili již jejich „původní“ či převažující funkce je navazování na předchozí kontext. Jak se ukázalo, nebývají to spojky, (ty mají vysoký absolutní výskyt, a tudíž je jejich vysoký výskyt jako PREC stále malé číslo), ale právě výrazy adverbialní či částice. Výjimkou je polyfunkční *tak*.

Deset nejčastějších výrazů s funktorem PREC shrnuje tabulka 1:

výraz	výskyty v PDT	
	PREC	jiný funktor
však	1482	300
a	526	16 777
ale	518	1733
totiž	493	21
tak	350	836
tedy	346	230
proto	334	148
ovšem	274	99
navíc	226	34
přítom	220	33

**Tabulka 1** Nejfrekventovanější výrazy s funkcí PREC v PDT 2.0

Kompletní seznam t-lemat s funktorem PREC v PDT 2.0 je kvůli své rozměrnosti zařazen až na konec dokumentu (Příloha 3). Tato tabulka shrnuje absolutní počet výskytů vybraných t-lemat v PDT 2.0, dále údaje o rozložení anotace funktorů v rámci jednoho t-lematu a případně poznámku k analýze či ke specifickému chování výrazu.

Hlavním účelem vytvoření kompletního seznamu českých výrazů anotovaných v textech PDT 2.0 jako PREC byla možnost určit tak jádro a periferii této sémantické skupiny, pozorovat chování jednotlivých výrazů, především jejich oscilaci mezi několika sémantickými charakteristikami, a v neposlední řadě možnost lepšího teoretického uchopení budoucí skupiny diskurzních konektorů. Analýza výskytů výrazů s funkcí PREC v syntakticky anotovaném korpusu přinesla především tyto informace: údaje o počtu a frekvenci výskytu jednotlivých výrazů této skupiny, přehled sémantických kategorií, s nimiž

tyto výrazy hraničí, jejich slovnědruhovou charakteristiku, údaje o funkční homonymii, o převládajícím připojovacím či spojovacím charakteru těchto výrazů, o okrajových jevech (víceslovné – rozvité či několikanásobné, deiktické, s rematizační funkcí atd.) V průběhu korpusových vyhledávání byly odhaleny i některé nesrovnalosti v anotačním schématu (viz např. výraz *takže*, kap. 5.2.3) a chyby v anotaci.

### **5.2.3 K slovnědruhové charakteristice výrazů s funkcí PREC a diskurzních konektorů obecně**

Většina českých mluvnických příruček se o slovnědruhovém zařazení konektorů zmiňuje pouze obecně. Víceméně shodně uvádějí, že jsou to výrazy formálně totožné se souřadícími spojkami a dále různé druhy příslovcí a částic. V druhém díle *Mluvnice češtiny* (MČ II: 237) se dočteme o částicích s primární funkcí strukturovat text. *Příruční mluvnice* mluví o spojovacích adverbích a vytýkacích a modifikačních částicích ve funkci konektorů, spojkové výrazy na hranici vět (často v iniciální pozici) chápe jako navazovací částice. (PMČ: 293). Hrbáček (1994: 57) dodává, že konektory jsou pouze ta příslovce, která nemají větněčlenskou platnost. U některých temporálních adverbii může být větněčlenská platnost oslabena a výraz tím nabývá spíše konektivní charakter. Dále Hrbáček zmiňuje, že konektivní funkci mohou mít i některé předložkové zájmenné a jiné výrazy (*naproti tomu, kromě toho, v této souvislosti...*), hypotaktická spojka *takže* uvozující nepravou větu vedlejší a některé výrazy vztažné (*přičemž*). Velmi podrobně se problematice slovnědruhové charakteristiky spojovacích výrazů věnuje Hošnová (2005). Ve své monografii, která je zaměřena především na vývoj spojovacích prostředků v novočeské syntaxi, podává v úvodu přehled českých přístupů ke slovnědruhovému vymezení těchto výrazů a zamýšlí se nad kritérii těchto vymezení (zejména 22 - 36). Zabývá se především spojovacími výrazy ve funkci junktorů, tj. „prostředky vyjadřujícími mezivýpovědní vztahy v rámci souvětí“ (Hošnová 2005: 16), a tedy nikoli v první řadě připojovacími konektory. Hošnová však mimo jiné upozorňuje na skutečnost, že „při klasifikaci neohebných slovních druhů formální kritérium [...] uplatnit nelze, a proto je nutné opřít se pouze o hledisko sémantické a o syntaktickou funkci.“ (2005: 23).

Ačkoli je nám jasné, že jednoznačné vymezení slovního druhu je problematické a jasných hranic mezi slovními druhy nelze dosáhnout ani při jednotném kritériu tohoto vymezování, nespokojujeme se s pouhým konstatováním, že skupinu výrazů s konektivní funkcí tvoří

„souřadící spojovací výrazy, dále některé adverbialní výrazy a některé částice“ (Mikulová et al. 2005: 537). V této kapitole ukazujeme, které slovní druhy se v jakém podílu účastní diskurzních vztahů, respektive které výrazy jsou jakožto konektory diskurzních vztahů funkčně homonymní.

Pro explicitní určení slovního druhu výrazů ve funkci konektorů jsme vyšli z morfologického značkování v Pražském závislostním korpusu (Hajič 2004)<sup>32</sup>, jehož systém značek odpovídá morfologickému značení v Českém národním korpusu. Určování slovního druhu v rámci morfologického tagu v PDT proběhlo v první fázi automatickými procedurami, v druhé pak dvojí nezávislou ruční anotací a následně kontrolou. Zkoumali jsme konkrétně, jakých možných morfologických tagů (tedy prvních dvou pozic, které určují a specifikují slovní druh) nabývají výrazy s tektogramatickým funktorem PREC. Omezili jsme se na prvních dvacet nejčastějších výrazů s funkcí PREC v PDT (tabulka 2) a poté na zvláštní případy (tabulky 3,4 a 5).

J<sup>^</sup> = spojka souřadící. Jak je vidět z tabulky 2, ne u všech souřadících spojek převažuje jejich spojující funkce uvnitř věty (CONJ, ADVS apod.). Výrazy, které mají tag J<sup>^</sup> a zároveň jsou ve žlutém řádku, alespoň v textech PDT častěji připojují věty, než spojují klauze (například výrazy *však*, *proto*, *nicméně*, *jenže*). Dále se ukazuje, že ze základních spojek se v připojovací funkci daleko více uplatňuje a , *ale* a *však* než například *nebo*, *anebo*. Toto je pravděpodobně dané i tím, že vylučovací vztah (DISJ) je pro diskurz méně typický než aditivní či adverzativní vztahy. Tyto jsou reprezentovány také řadou dalších konektorů, jak spojkami, tak různými příslovci.

Db = příslovce (bez určení stupně a negace; *pozadu*, *naplocho*). Výrazy v tabulce 2 označené tímto tagem zpravidla primárně plní připojovací funkci, tj. funktor PREC převažuje nad výskytem jiných funktorů (*totiž*, *tedy*, *navíc*, *přitom*, *naopak*, *pak*). Výjimkou jsou zde takové výrazy, které často mají větněčlenskou platnost a hraničí nejčastěji s temporálními funktory (*zároveň*, *potom*), téměř shodný počet výskytů jako PREC a jako temporální doplnění má adverbium *zase*.

---

32 Kromě monografie Jana Hajiče viz též k morfologickému značkování v PDT 2.0: *Manual for morphological annotation* - <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/m-layer/html/index.html> a <http://ufal.mff.cuni.cz/pdt2.0/doc/pdt-guide/cz/html/ch02.html#t-layers-morpho>

výraz	slovnědruhov charakteristika (dle tagů m- roviny PDT) <sup>33</sup>	výskyty v PDT			
		PREC	jiný funktor		
			celkem	CM	nejvíce se kryje s funktoem
však	J^	1482	300	0	ADVS (296)
a	J^	526	16 777	2	CONJ (16 241)
ale	J^	518	1 733	0	ADVS (1 343)
totiž	Db	493	21	1	APPS (14)
tak	J^, Db	350	836	134	MANN (334)
tedy	Db	346	230	229	CM (229)
proto	J^	334	148	131	CM (131)
ovšem	J^, TT	274	99	0	ADVS (67)
navíc	Db	226	34	11	CM (11)
přítom	Db	220	33	7	TWHEN (26)
naopak	Db	189	50	13	CM (13)
pak	Db	160	18	18	CM (18)
dále	Db, Dg	109	211	17	THL (111)
přesto	Dg	107	48	22	CNCS (26)
zase	Db	65	60	2	TWHEN (53)
zároveň	Db	61	179	38	TWHEN (105)
jenže	J^	60	25	0	ADVS (25)
nicméně	J^	59	23	0	ADVS (23)
potom	Db	41	493	1	TWHEN (492)
(a)nebo	J^	38	969	0	DISJ (813)

Vysvětlivky k tabulce 2: J^ = spojka souřadící, Db = příslovce bez určení stupně a negace, Dg = příslovce s určením stupně a negace, TT = částice; CM = „conjunction modifier“, tj. modifikátor souřadného spojení

**Tabulka 2** Dvacet nejfrekventovanějších výrazů s funkcí PREC v PDT 2.0, jejich hranice s jinými funktoem a jejich slovnědruhov  
charakteristika

33 Tj. hodnota atributu m-tag v morfologické rovině PDT 2.0 u výrazů ve funkci PREC. První dvě pozice v m-tagu určují a specifikují slovní druh. Vysvětlivky ke zkratkám se nacházejí v kapitole 5.2.3.

Dg = příslovce (s určením stupně a negace). V tabulce patří k tomuto typu pouze výrazy *dále* a *přesto*, v dalším zkoumání pak výrazy *konkrétně*, *současně*, *stejně*, *spíše/nejspíš* s funkcí PREC. Jedná se o okrajové konektory, které primárně plní jiné funkce. Stupňovatelná a negovatelná příslovce se tedy na připojování vět podílí poměrně málo.

TT = částice. Částice se v morfologickém značkování PDT ani ČNK podrobněji nespecifikují. Z výrazů s funkcí PREC je v PDT pouze velmi málo částic, z celkových 21 výskytů částic se jich 13 váže k výrazu *ovšem*. V těchto případech se však může jednat o afirmační částici nebo i o výraz vyjadřující postoj mluvčího. Dále se jako částice popisují ojediněle se vyskytující výrazy: *ba*, *natož*, *nuže*, *jen*, a *to*. V takto řídkých případech je vždy nutné pečlivě prozkoumat správnost anotace. Například kombinace výskytu morfologického tagu TT a výskytu funktoru PREC u výrazu *jen* (pouze 2 výskyty v celém korpusu) již napovídá, že se o konektor nejedná, srov. věty:

(36) *Jen jed'te!* (anotováno jako PREC – podle nás nesprávně)

(37) *Jen ať si sní...* (anotováno jako ATT)

C= = číslo psané číslicemi (značkováno jako slovní druh číslovka). Číslovky mají v PDT funkci PREC tehdy, pokud uvozují položky nějakého seznamu či výčtu. Tyto výrazy stojí na začátku vět či klauzí, v jiných případech se totiž nejedná o konektory, ale o uvození vícenásobných větných členů. Číslovky tedy nejsou diskurzivními konektory v tomto případě:

(38) *Musíme koupit: 1. zeleninu, 2. pečivo a 3. mléko.*

NN = substantivum, obyčejné. Blíže specifikují morfologické kategorie daného substantiva další pozice v morfologické značce. Ze substantiv se ve funkci konektorů v PDT objevují zejména dvě skupiny výrazů: jednak jsou takto označena písmena abecedy, která stejně jako číslice uvozují položky určitého seznamu, jednak je to skupina nominálních víceslovných výrazů, zejména předložkových, jejichž řídicí uzel je jméno s funktorem PREC. Jsou to vesměs ustálené obraty navazovacího charakteru typu: *jinými slovy*, *na jedné/druhé straně*, *v neposlední řadě* atd. Seznamům a nominálním skupinám jakožto konektorům se věnují kapitoly 5.4.2. a 5.8 o okrajových typech konektorů.

PD = zájmeno ukazovací („ten“, „onen“, ...). Funktorem PREC jsou také označené některé ustálené pronominální výrazy, především různé tvary zájmena *ten*, například obraty typu *k tomu*, *a to*, *a tím*, *místo toho*, *naproti tomu* atp. O těchto výrazech pojednáváme rovněž v kapitole 5.8.

J, = spojka podřadicí (vč. „aby“ a „kdyby“ ve všech tvarech). Spojky podřadicí jsou

v anotaci tektogramatické roviny skryté, tj. nejsou viditelné v tektogramatickém stromě. Tektogramatické uzly by tedy neměly mít tuto morfologickou značku nikdy. Přesto se takové uzly vyskytují, a to i ve funkci konektoru. Jedná se buď o chybnou anotaci – například u podřadicí spojky *protože*, nebo o specifické případy u výrazů jako například *takže*, *příčemž*, *aniž*. Hodnocení těchto výrazů jako hypotaktických prostředků je zpravidla relativizováno, mimo jiné proto, že uvedené výrazy vykazují určité specifické chování. Rozdělení hypotaktických a parataktických prostředků je stále předmětem lingvistických debat, také vzhledem k tomu, že kritéria pro rozlišení těchto skupin se v užívání jazyka stírají.<sup>34</sup> Držíme se zde anotačních zásad pro tektogramatickou rovinu a hodnotíme výrazy *takže*, *příčemž* a *aniž* jako souřadící (Mikulová et al. 2005: 238, 244).

Z morfologické anotace v PDT jsme zjistili, že z hlediska slovních druhů jsou výrazy s funkcí PREC rozmanitější, než jak určuje definice funktoru PREC. Konektivní funkce je v PDT připsána i některým, zejména ustáleným, nominálním a pronominálním spojením a některým číslovkám. Domníváme se, že pokud rozšíříme skupinu těchto výrazů s o některé podřadicí spojky, dostaneme co se týče slovního druhu úplnou charakteristiku diskurzních konektorů. Dále jsme zjistili, že některé tradičně pojaté souřadící spojky častěji připojují, než spojují, tedy mají alespoň v textech PDT častěji funkci připojovacího konektoru (ve smyslu připojování vět přes tečku) než junktoru (zejména výrazy *však*, *proto*, *jenže*, *nicméně*).

#### 5.2.4 Binarita

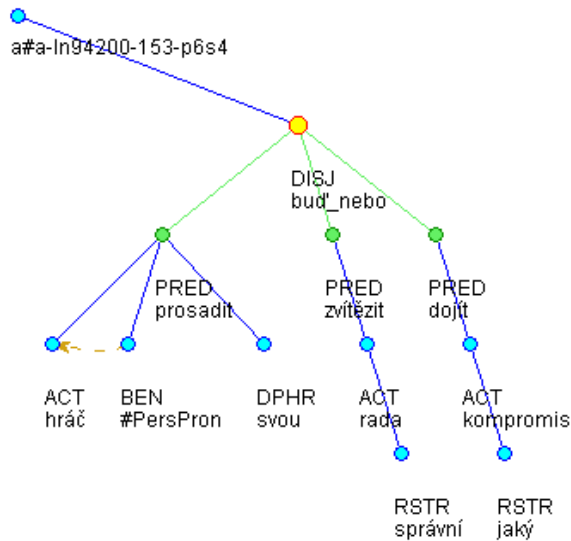
Penn Discourse Treebank hodnotí diskurzní vztahy jako vztahy výlučně binární (Prasad et al. 2006), tj. každý konektor je predikátem binárního vztahu a spojuje tedy právě dva argumenty. Diskurzní částice (*discourse markers*), které se vztahují pouze k jednomu abstraktnímu objektu, nejsou podle schématu PDTB diskurzními konektory, např. *well*, *strangely*, *in all likelihood*, *frankly* (2006: 6).

Pro budoucí anotaci diskurzu v rámci PDT však nepočítáme s výhradně binárními diskurzními vztahy. Binaritu podle našeho názoru porušují dva typy diskurzních vztahů. Prvním jsou vícenásobné koordinace ve větách typu:

---

34 Například postavení souřadících a podřadících spojek ve větě a užívání spojek *protože* a *neboť*.

(39) *Bud' si hráči prosadí svou, nebo zvítězí správní rada, nebo dojde k nějakému kompromisu.* PDT



**Obr. 14** Příklad nebinárního diskurzního vztahu

Druhým typem jsou seznamové vztahy, ve kterých jsou jednotlivé segmenty uvozovány číslicemi, písmeny abecedy nebo výrazy *zprvė, zadruhé, zatřetí* apod. Tento typ vztahů hodnotíme tak, že každá položka seznamu má vztah k položce předchozí a zároveň k hypertématu, neboli ke klauzi či větě, která zpravidla celý seznam uvozuje. Srov. např.:

(40) PDT

(a) *K tomu, aby zaměstnavatel pracovníkovi za škodu opravdu odpovídal, musí být splněny tyto podmínky:*

(b) *1. Zaměstnanci musí vzniknout škoda, tj. musí dojít k určitému snížení hodnoty jeho majetku (v některých případech mu vzniká i právo na náhradu ušlého zisku).*

(c) *2. Zaměstnavatel nebo jiná fyzická či právnická osoba, která jedná jeho jménem, musí porušit své právní povinnosti.*

Uzel s t-lematem 2, který je označen funktoem PREC, uvozuje větu (c), připojuje ji k větě (b) a zároveň má vztah ke větě úvodní, která oznamuje uvedení výčtu. Tato úvodní věta,



respektive dvojtečka upozorňující na následný výčet, se naopak vztahuje ke všem členům tohoto výčtu. Konektivní funkce dvojtečky v takovýchto případech zatím popsána nebyla, neboť dvojtečka je zde koncovým signálem. Uvnitř věty by však byla zachycena jako kořen souřadné struktury s funktorem pro apozici. Z hlediska kompozičních vztahů v diskurzu by reprezentovala relátor ve vztahu elaboračním výčtovém (srov. kap. 3). Otevřené zůstávají dvě otázky: (i) zda je tato dvojtečka identifikátorem stejných vztahů stejných segmentů jako číslice uvozující jednotlivé položky výčtu a nebo se vztahuje k položkám seznamu jako k celku (ii) zda například pátá položka seznamu má vztah ke všem čtyřem předchozím a hypertématu, nebo pouze k předchozí položce a k hypertématu. U první otázky se kloníme k první možnosti vzhledem k zachycení několikanásobné apozice jakožto souřadné struktury způsobem, který je analogický k příkladu souřadné struktury s výrazem s *bud' nebo*. Takové dvojí značení stejného diskurzního vztahu je jistě redundantní, přínos by mělo pouze kvůli informaci, že číslice uvozující položky seznamu odkazují anaforicky, zatímco dvojtečka odkazuje kataforicky. U otázky druhé se kloníme též k první možnosti, jelikož např. výraz *zapáté* jasně implikuje existenci předchozích čtyř (respektive pěti) argumentů výčtového vztahu.

### 5.2.5 „Antecedenty“ výrazů s funkcí PREC

Po objasnění problematiky počtu argumentů diskurzního vztahu je třeba také popsat, jak argument diskurzního vztahu chápeme a jak vymezujeme jeho rozsah v našem pojetí diskurzu.

Asher ve svém pojetí argumentu diskurzního vztahu, které využívá Penn Discourse Treebank, označuje za argumenty diskurzního vztahu *abstraktní objekty* (AO, viz kap. 2.1.3), tj. zejména *events, states and propositions* – děje, stavy a propozice. Dle českých mluvnických příruček jsou textové vztahy vztahy mezivýpovědními či mezipropozičními. (MČ III aj.).

Již jsme naznačili, že základním vyjádřením propozice, neboli základním textovým segmentem, je pro nás klauze. Dodejme, že prozatím chápeme jako diskurzní jednotky pouze klauze slovesné, tedy nikoli klauze neslovesné (nominativní – s kořenem s funktorem DENOM, vokativní – VOCAT a citoslovečné PARTL).<sup>35</sup> U slovesných klauzí může být

---

35 V PDT 2.0 není dostatečný počet dokladů o připojování neslovesných klauzí pomocí výrazů s funktorem

predikát (PRED) kromě určitého tvaru slovesa (VF) vyjádřen infinitivem, participiem, citoslovcem, interpunkčním znaménkem či může být elidován (srov. Mikulová et al. 2005: 182-186). Výrazy s funkcí PREC se vyskytují ve větěné struktuře nejčastěji jako přímý potomek predikátu (ovšem s nezávislostní hranou), mohou se však také vyskytovat v zanořené struktuře (kde rozvíjí řídicí člen závislé klauze). V takových případech „závisí“ vždy na slovesném tvaru, který je řídicím členem této zanořené struktury,<sup>36</sup> a připojují stejně jako v nezanořené struktuře nejčastěji klauzi, v níž stojí, ke klauzi v textu předcházející, ať už v rámci stejné nebo předchozí věty.

Pokud bychom chápali, podobně jako např. Asher a Webber, výrazy s funkcí PREC jako výrazy anaforické, je nutné určit jejich antecedenty.<sup>37</sup> Jak se domníváme, v pozici antecedentu diskurzního konektoru (dle Ashera *cue phrase*) však nebude stát jednotlivý výraz či spojení, jak je tomu ve velké většině případů například u pronominálního odkazování, nýbrž větší segment – klauze, věta.

Opřeme se opět o výzkum korpusových dat. Je však nutné předeslat, že vyhledávací nástroje pro PDT 2.0 nejsou primárně zaměřeny na vyhledávání jevů širších, než dokáže zachytit jeden strom. Je tedy možné, že některé řídké či ojedinělé případy diskurzního připojování v našem výzkumu nezachytíme. Podíváme se tedy na „antecedenty“ diskurzních vztahů a pokusíme se zjistit, zda jsou zpravidla diskurzními argumenty jednotlivé klauze nebo i segmenty menší či větší, dále jaké slovesné klauze (co se týče jejich závislosti a nezávislosti v syntaktickém popisu) jsou připojovány výrazem s funktorem PREC a ve kterých případech funktor PREC vůbec nesignalizuje diskurzní vztah.

Data z Pražského závislostního korpusu potvrzují, že nejčastěji výrazy s funkcí PREC navazující na slovesné klauze s predikátem vyjádřeným určitým tvarem slovesa. V rámci tohoto navazování mohou nastat následující možnosti (úsek, ke kterému se vztahuje konektor, je v následujících příkladech vždy označen tučným písmem, konektor je podtržen. Všechny příklady pochází z PDT 2.0):

---

PREC.

36 Oproti způsobům vyjádření predikátu různými tvary slovesa se může jakožto kořen v závislé konstrukci objevit i přechodník. V korpusu však nebyly nalezeny doklady, kdy by byl výraz s funkcí PREC rozvitím přechodníku.

37 „In more recent work, Webber et al. (in press) argue that one can get very simple discourse structure by distinguishing the building of these structures from the resolution of anaphoric cue phrases. While SDRT has also maintained that cue phrases are anaphoric (...), part of the task is to figure out the set of possible antecedents to cue phrases like *then* and *too*, and this can't be determined (as Webber et al. themselves must admit, by the tree structures they build for discourse.)” (Asher a Lascarides 2003: 104)

## 1. Návaznost výrazů s funkcí PREC na jednu klauzi<sup>38</sup>

### 1.1 Výraz s funkcí PREC navazuje na jednu nezávislou klauzi

A. v rámci stejné věty

(41) *O změnách mechanismu úvěrů se podle něho bude jednat již dnes v rámci první diskuse o státním rozpočtu, ovšem uvažované podmínky hypoték zatím Dyba nezná.*<sup>39</sup> PDT

B. na jednoduchou větu (klauzi)

(42) *Srbské zdroje nehlásí ze Sarajeva rozsáhlejší ozbrojené střety. Bosenský a chorvatský rozhlas však mezitím informovaly o pokračujícím ostřelování metropole srbskými děly.* PDT

C. v souřadném souvětí

(43) *Podle Pavla Zacha z ČNZP mají prevenci ve svém programu všechny pojišťovny, ale **málokdo v posledních letech absolvoval skutečně důkladné preventivní vyšetření.** Přitom například většinu zhoubných nádorů lze při včasném podchycení léčit.* PDT

D. Těžko rozlišitelné jsou případy, u nichž je nutné rozhodnout, zda se výraz s funkcí PREC vztahuje pouze na samostatnou klauzi řídicí bez klauze závislé či na celé podřadné souvětí:

(44) *O **studium svého syna jste se začal zajímat pozdě,** protože oficiální termín přihlášek na střední školy a učňovská zařízení vypršel s koncem února. Přesto ještě máte šanci.* PDT

Domníváme se, že pokud nebude nalezen argument proti zahrnutí závislé klauze do dosahu následujícího konektoru (zde výraz *přesto*), je výhodnější tyto závislé klauze do jejich dosahu zahrnovat. Nestane se tak, že při zachycování diskurzních vztahů budeme muset vyřadit z jednoho argumentu tohoto vztahu celý podstrom. Vznikne tak však zanořená struktura, neboť uvnitř první věty již existuje příčinný diskurzní vztah a závislá klauze se tak účastní dvou diskurzních vztahů najednou.

---

<sup>38</sup> Tj. na větu jednoduchou, viz kap. 2.4.3.

<sup>39</sup> Zde je spojovacím výrazem čárka (ADVS), výraz *ovšem* má funktor PREC a visí na slovese *znát*.

## 1.2 Výraz s funkcí PREC navazuje na jednu klauzi závislou:

A. v rámci stejné věty

(45) ***Protože máme vypěstovaný určitý obranný reflex proti superlativům a sebechvále, tak jsme proti takovému obsahu částečně imunní.*** PDT

Zde se jedná především o výrazy *tak, pak a potom*, které v podřadném souvětí uvozují postponovanou řídicí klauzi a se často kombinují s podřadícími spojovacími výrazy kauzálními a podmínkovými: *protože... tak, (i) kdyby ... tak, (i) když ... tak, když ... potom, pokud ... pak, jakmile ... tak* apod. Tyto výrazy ovšem jen modifikují význam podřadící spojky (která je sama již konektorem) a zdá se, že samostatně se žádného jiného vztahu neúčastní. Nejsou nijak obligatorní pro určení druhu závislostního vztahu a v psaném projevu se často vynechávají.

B. v předchozí větě

(46) ***Víš, on je náš Baryk docela hodnej, ale prosím tě, hlavně si teď sedni a už se nezvedej. Stejně by ti to nedovolil.*** PDT

## 2. Návaznost výrazů s funkcí PREC na větší úsek, než je jedna klauze

A. na koordinaci klauzí

(47) ***Lidé na všech stupních řízení jsou schopní, mají snahu se dále učit. Chybí jim však zkušenosti z dlouhodobého působení.*** PDT

B. na klauzi řídicí spolu s klauzí závislou (viz též v 1. D v příkladech)

(48) ***Podle Majora je v pořádku, když se některé státy nebo skupiny států dohodnou v určitých oblastech na rychlejší integraci než jiné. Všechny však musejí mít právo si vybrat.*** PDT

C. na větší textový úsek

(49) *Jste - li pojištěn na základě zákona o všeobecném zdravotním pojištění, hradí se vám náklady nutného a neodkladného léčení v cizině, a to až do výše nákladů spojených s takovým léčením na území České republiky. Náklady za další léčebné úkony hradit nelze. Může proto dojít k tomu, že náklady vynaložené v cizině budou mnohem vyšší než náklady v České republice.* PDT

(50) *Nejrychlejší cestou by byl překlenovací úvěr, ale banky zpravidla na úhradu dluhů nepůjčují. (Vlastní domek byl zastaven už dříve ke krytí půjčky pro rozjezd kanceláře.) Nenadálou finanční krizi musela podnikatelka řešit jiným způsobem. Především informovala své věřitele (jen hotelům dlužila půl druhého milionu ) o nastalé situaci a požádala je o poshovění. Pokud nechtěli nebo nemohli odklad platby povolit, použila k úhradám zálohy vybrané na připravené zájezdy.*

*Něco málo zaplatila i z osobních půjček poskytnutých příbuznými a přáteli.*

*Hlavní pomoc ale nacházela v dalších obchodních aktivitách.* PDT

Při navazování na delší úsek textu se častěji objevují různé interpretace toho, na jak rozsáhlý textový úsek navazovat. Různé názory jsou v těchto případech oprávněné a částečně se dají sjednotit jasným určením anotačních pravidel. Dosáhnout ve všech takových případech stejného hodnocení rozsahu argumentu je však velmi obtížné, pokud ne nemožné.

Dosud jsme se zabývali případy, kdy výraz s funktoem PREC navazoval na klauze, jejichž řídicí člen byl vyjádřen určitým tvarem slovesa. Jak jsme však již naznačili v úvodu k této kapitole, kromě určitého tvaru slovesa může být řídicí člen slovesné klauze (a tedy diskurzním argument) vyjádřen i infinitivem, participiem, citoslovcem, interpunkčním znaménkem a elipsou. Celkový výskyt takto vyjádřených řídicích členů je na rozdíl od řídicích členů vyjádřených určitým tvarem slovesa nízký a případů, kdy na ně navazuje další klauze pomocí výrazu s funktoem PREC, je ještě méně. Přesto zde uvádíme příklady těchto připojení, pokud jsme je v PDT 2.0 byly schopni najít:

### 3. Návaznost výrazů s funkcí PREC na propozici vyjádřenou jinak než určitým tvarem slovesa

#### 3.1 Návaznost na slovesné klauze s řídicím členem vyjádřeným infinitivem

(51) *Čilý byznys samozřejmě potřebuje neustále přísun zboží. **Kde ho však neustále brát, a přitom nekrást.*** PDT

(52) *Věnovat jim život a veškeré snažení? Aby ostatní viděli, v jakém jedu fáru... tak to tedy ne.* PDT

#### 3.2 Návaznost na slovesné klauze s řídicím členem vyjádřeným participiem trpným

Tyto případy jsou nejčastěji nadpisy novinových článků: *Podnět nezávislých patrně odmítnut; Zvolen patriarcha* apod. Nenalezli jsme však doklad, že by taková klauze byla antecedentem výrazu s funkcí PREC. Uvádíme proto alespoň případ, kdy je participiální klauze v adverzativním vztahu s následující klauzou v rámci jedné věty. Uvedený text je součástí přehledu o jaderných haváriích, predikáty u jednotlivých položek jsou často vyjádřeny participiem.

(53) *1991 – v jaderné elektrárně Kozloduj (Bulharsko) objevena po požáru dvě místa úniku radioaktivity; **zamořeno jen několik čtverečních metrů, ale ADVS úroveň radiace dosáhla dvacetinásobku povolených hodnot.*** PDT

#### 3.3 Návaznost na slovesné klauze s řídicím členem vyjádřeným interpunkčním znaménkem

Případy, kdy výraz s funktorem PREC odkazuje k řídicímu členu vyjádřenému interpunkčním znaménkem, také nemají korpusový doklad. Takové konstrukce však principiálně existovat mohou, viz příklad (54):

(54) ***Doprava: vlastní.** Ale strava bude zajištěna.*

Nyní je třeba prozkoumat domněnku, že některé konektory se mohou vztahovat také k obsahu v textu nevyjádřenému (MČ III: 706). Anotace syntaktické roviny v PDT umožňuje takto zkoumat aktuální a gramatickou elipsu řídicího členu, neumožňuje však zkoumat domnělou návaznost konektoru na presupozici vyplývající z kontextu.

### 3.4 Návaznost na elipsu řídicího slovesa

#### A. návaznost na aktuální elipsu řídicího slovesa

V případě aktuální elipsy se v PDT výraz s funkcí PREC vztahuje k celé slovesné koordinaci, elidované sloveso je tedy v jeho dosahu pouze v rámci podstromu. Případy, kdy by se výraz s funkcí PREC vztahoval přímo na aktuálně elidovaný člen, nemůžeme vyloučit, ale ani jsme pro ně nenalezli v korpusových datech žádný jednoznačný doklad.

(55) *Naše metodika a techniky vyšetřování mohou u jednotlivých vyjmenovaných profesí předpovědět úspěšnost v rozdílné míře pravděpodobnosti. Nejvyšší bude u svářeče a dělníka, nižší u manažera a redaktora. Ale v každém případě podstatně vyšší, než při hodnocení „od boku“.* PDT

#### B. návaznost na gramatickou elipsu řídicího slovesa

Věta s gramatickou elipsou řídicího slovesa stojí na rozdíl od věty s aktuálně elidovaným slovesem zcela samostatně. Mělo by tedy být možné na ni navázat stejně jako na větu s řídicím slovesem vyjádřeným. Takové případy se v PDT vyskytují, i když nepříliš často – příklad (56). Také je snadné takové příklady k větám s gramatickou elipsou domyslet (57), (58).

(56) *Není problém v něčem jiném, třeba v tom, že je hodně prodavaček levých? Levých? **Možná.** Jisté ale je, že je dost levorukých.* PDT

(57) *Nač **ten** spěch? A proč vynakládat tolik zbytečného úsilí?*

(58) *V **Praze**, v **pět hodin**. Nebo také zítra v Brně.*

### 3.5 Návaznost na presupozici

V případě návaznosti na presupozici, čili na obsah, který v textu není nijak zachycen, ale je významem věty či textu nějakým způsobem implikován,<sup>40</sup> zmíníme pouze příklad z literatury.

---

40 Užíváme zde pojmu *presupozice* v poměrně obecném vymezení. Jsme si vědomi toho, že s tímto pojmem odlišně pracují například logické postupy zkoumající pravdivostní podmínky věty a negaci (Hajičová: 1975).

Zdá se, že vzájemným vztahem presupozice a diskurzního konektoru obecně se v teoriích diskurzu zatím nikdo explicitně nezabýval. Komárek (1958) se zabývá tzv. endocentrickými částicemi, tj. částicemi, které implikují určitou presupozici, nezabývá se však jejich potenciální konektivní funkcí. V současné době vzniká na téma částice implikující presupozici diplomová práce Z. Bedřichové. Hrbáček věnuje ve své monografii tomuto tématu krátký odstavec (1994: 57): „Konektory mohou též vyjadřovat vztah k presupozici: *Na vysokou školu se nehlásil. Stejně by se tam nedostal.* Konektor *stejně* odkazuje k obsahu, který není v textu vyjádřen, ale který by se dal vyjádřit v tomto případě přípustkovou větou: *,i kdyby se hlásil, ‘stejně by se tam nedostal.’*“

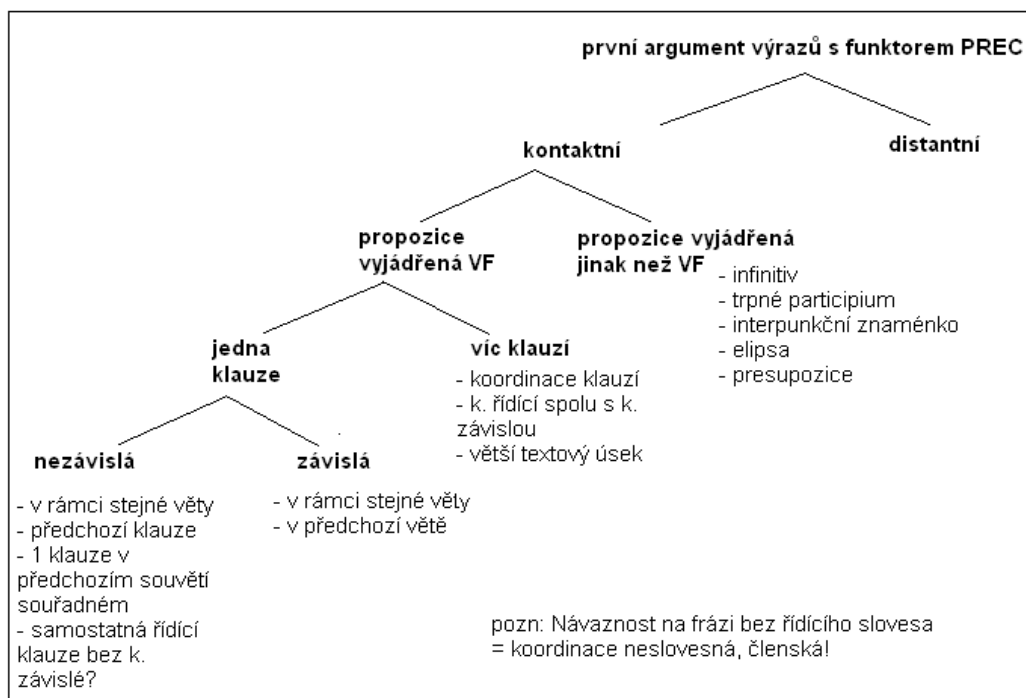
#### 4. Návaznost na distantní úsek textu

Všechny typy navazování výrazů s funktorem PREC na předchozí kontext, jak jsme je dosud popsali, představovaly navazování na bezprostředně předcházející segment textu. Nežřídká se však objevují i případy navazování na distantní úsek textu, viz následující příklady:

(59) *Mít rád umění při investování do obrazů nestačí. Na aukci můžete koupit pod cenou, ale stejně jako u starožitníka vás před padělkem nikdo neochrání. **Vydělal jsem nějaké peníze a protože mám rád umění, chtěl bych si koupit pár hodnotných obrázků.** Nedávno jsem byl ve starožitnictví a některé věci se mi tam líbily. Byly to většinou obrazy z 19. století, moderně moc nefandím. Moje znalosti o obrazech jsou však téměř nulové. Co bych měl koupit, abych neprodělal, když bych to potřeboval zase prodat?* PDT

(60) ***Proto musí odběratel platit tolik, kolik bude stanoveno.** Že se pod tento plášť schová leccos dalšího, netřeba připomínat. Nebo platit nebude, ale musí si zajistit teplo sám.* PDT





**Obr. 15** Povaha „antecedentů“ výrazů s funktorem PREC

### 5.2.6 Povaha „druhého“ argumentu

Uvedenými příklady jsme vyčerpali popis možných antecedentů výrazů s funktorem PREC. Povahu „druhého“ argumentu jsme naznačili již v samotném úvodu k této kapitole: výraz s funktorem PREC se vztahuje prostřednictvím predikace k obsahu celé klauze, v níž se nachází. Zachycen je téměř vždy jako potomek řídícího slovesa, ať již v roli predikátu či v roli řídícího slovesa závislé klauze. Pokud se výraz s funkcí PREC vyskytuje v klauzi závislé, může ji připojovat k předcházející větě nezávislé, jedná se tedy o případ opačný k příkladu pod 2. B., věta (48). Oba tyto doklady ukazují, že konektory mohou spojovat textové segmenty s různou úrovní zanořenosti ve struktuře.

(61) *Zahraniční certifikáty u nás neplatí. Pokud ale žadatel předloží certifikáty CB*

### 5.2.6.1 PREC bez funkce diskurzního konektoru – připojování členské

Různé situace mohou nastat, pokud výraz s funkcí PREC stojí v klauzi s elidovaným řídicím slovesem. Je to signál, že se nemusí jednat o konektor diskurzního vztahu, ale pouze o připojování větných členů. Zde je třeba přesně rozlišovat mezi koordinací členskou a větnou a všimnout si charakteru spojovaného a připojovaného členu. Této problematice se věnuje zejména Šindlerová (2007), která se zabývá především členskou koordinací mezi konjunktivy v rámci věty. Domníváme se, že její poznatky můžeme uplatnit i při rozlišení větné a členské koordinace přes hranice věty. Šindlerová chápe členskou koordinaci jako přímou koordinaci potomků slovesa a slovesnou koordinaci jako koordinaci s elipsou řídicího slovesa, při které je nutné toto sloveso do struktury zkopírovat“ (Šindlerová 2007: 57).<sup>41</sup>

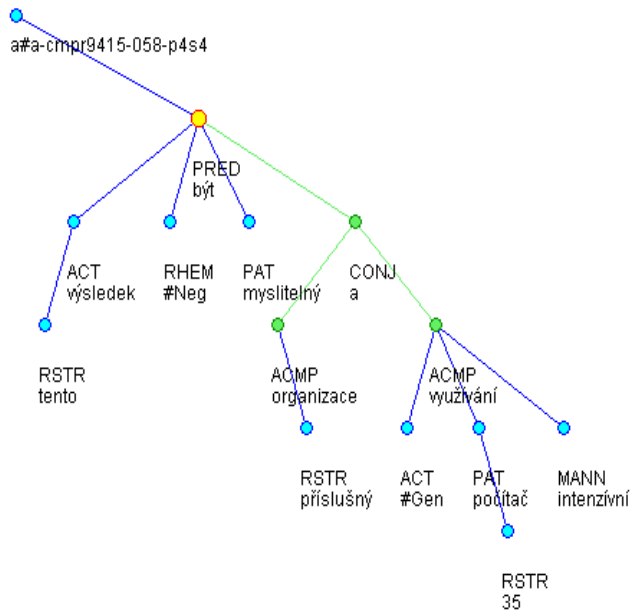
Příklad slovesné koordinace:

(62) *Základní otázky bezpečnosti smlouvají jiní a [smlouvají je] jinde.* PDT, Šindlerová 2007

Chybí-li v povrchové stavbě věty řídicí sloveso, jedná se buď o elipsu aktuální (sloveso lze snadno a jednoznačně doplnit z předchozího kontextu), nebo o elipsu gramatickou (není jasné, které sloveso by mělo být doplněno, nevyskytuje se v předchozím kontextu).

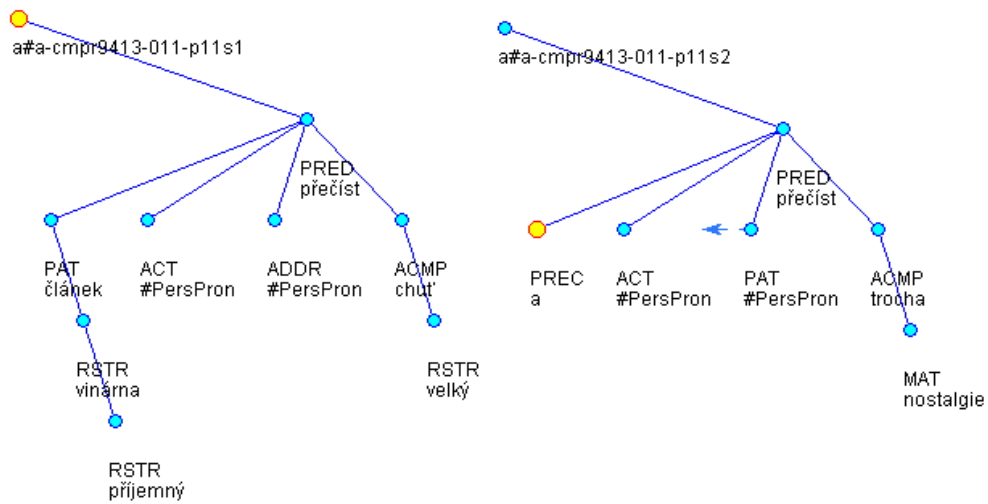
Nejprve se soustředíme na aktuální elipsu řídicího slovesa v klauzi, kde se nachází funktor s významem PREC. V takovýchto případech se na místo chybějícího slovesa zkopíruje odpovídající sloveso z předchozí věty/klauze. To však i v těch případech, kdy se jedná o koordinaci členskou. Roli zde hraje anotační konvence, podle níž musí mít každý strom svůj efektivní kořen s příslušným funktoem (PRED, DENOM, VOCAT, PARTL). Srovnání členské koordinace v rámci jedné věty a v rámci dvou vět ukazují obrázky 16 a 17. Ze zachycení věty, ve které se vyskytuje potenciální diskurzní konektor a zároveň je elidováno řídicí sloveso, se tedy o charakteru koordinace (čili o povaze obou konjunktů) nic nedozvíme.

41 Jiná je definice členské a větné koordinace v Manuálu pro tektogramatickou anotaci, kde je kritérium přítomnost slovesného tvaru v konjunktách větné koordinace (Mikulová et al. 2005: 255). Tedy, jak uvádí Šindlerová (2007: 57, věty typu *Ptám se, zda přijedeš a koho s sebou přivezeš.* jsou přímá koordinace potomků. Tyto případy přímé koordinace slovesných uzlů hodnotíme jako jasně diskurzní a dále se jimi nezabýváme.



Tyto výsledky by nebyly myslitelné bez příslušné organizace a bez intenzivního využívání 35 počítačů.

**Obr. 16** Členská koordinace v rámci jedné věty



Článek o příjemné vinárně jsem si přečetla s velkou chutí. A trochu nostalgie.

**Obr. 17** Členská koordinace (připojení členu) v rámci dvou vět

Pojetí FGP dává při anotaci obecně přednost jednodušší struktuře, tedy i koordinaci členské před větnou. Pro rozlišení koordinace členské (která tedy není diskurzním vztahem) od koordinace slovesné ze zohledňují zejména dvě kritéria:

A. Koordinované členy členské (přímé) koordinace se shodují ve funkci (mají stejný funktor). K tomu však Šindlerová podává přehled případů koordinace volných doplňků (adjunktů), které mohou mít jiný funktor, pokud jsou ale sémanticky kompatibilní: „Dva adjunkty jsou sémanticky kompatibilní, pokud je možné z jejich významů abstrahovat společný rys a na jeho základě propojit jejich sémantické interpretace.“ (2007: 94). Sémantickou kompatibilitu takovýchto doplňků pak testuje jednak možností nalezení společné otázky, pomocí které se lze zeptat na obě doplňky a jednak možností jejich spojování v rámci apozice. Pokud mají tedy členy potenciální členské koordinace různý funktor a jsou sémanticky nekompatibilní, jedná se o souřadnost větnou.

B. Pokud je některý člen koordinace rozvit rematizátorem nebo je rozvit nepřítomný predikát, nejedná se o členskou koordinaci (a tedy se jedná o diskurzní vztah) (Mikulová et al. 2005: 256). Analogicky tedy při připojování ve větách

(63) *Koupím chleba. A mléko.*

půjde o koordinaci členskou, čili „nediskurzni“, zatímco ve větách

(64) *Koupím chleba. A asi i mléko.*

se bude jednat o koordinaci se zkopírovaným predikátem<sup>42</sup>, čili slovesnou a tedy diskurzni. Problematické jsou situace, kdy mezi připojovací výraz a druhý člen potenciálně členské koordinace vstoupí negace či jiný výraz, který je v PDT zachycen jako CM – modifikátor souřadného spojení. Například:

(65) *Koupím chleba. A PREC ne RHEM mléko.*

(66) *Koupím chleba. A PREC také RHEM/PREC<sup>43</sup> mléko.*

(67) *Koupím chleba a CONJ také CM mléko.*

---

42 Tj. predikát by byl zkopírován, pokud by se jednalo o koordinaci v rámci jedné věty. V připojené větě bude predikát zkopírován v každém případě, jak jsme ukázali výše.

43 Zde je anotace nejednoznačná, blíže k této problematice kapitola 5.3.

Při navazování přes hranice věty se z výrazu CM stane buď další PREC nebo rematizátor RHEM, což samo o sobě opravňuje k tomu, považovat druhý člen za člen v jiné pozici či s jinou funkcí. V PDT jsou však koordinace tohoto typu modifikované výrazem s funktorem CM hodnoceny jako členské. Zde se však přikláníme k názoru, že alespoň v případě modifikace koordinace negací jde o koordinaci slovesnou, čili věty (65) parafrázujeme takto: *Udělám dvě věci – koupím chleba a nekoupím mléko*; a nikoli: *Udělám jednu věc – koupím chleba a „nemléko“*. Pokud se v připojované větě s aktuální elipsou řídicího slovesa vyskytne negace, která v předchozí větě není, hodnotíme to jako připojování slovesné, tj. diskurzí.

K rozhodování o připojování členském, neboli „nediskurzí“, a větném lze tedy použít kritérium stejné funkce (nebo sémantické kompatibility) a „nemodifikovanosti“.

Jako slovesné koordinace (tj. diskurzí) pak Šindlerová hodnotí všechny mnohačetné koordinované otázky a mnohačetné koordinované odpovědi, např.:

(68) *Kam, proč a s jakým záměrem chce vláda vysílat na rozhlasových vlnách do zahraničí?*  
PDT, Šindlerová 2007

Pod čísla (69) – (73) uvádíme pro porovnání doklady na členské a větné připojování pomocí výrazu s funkcí PREC z Pražského závislostního korpusu:

PREC signalizuje členské připojování

(69) *Je ředitelem mocné vojenské komise při ÚV KS Číny – ale **armádní špičky** si prý od něj udržují odstup. Stejně tak pekingští byrokraté.* PDT

(70) *Dnes nakupujeme maso **zprostředkovaně**.MANN Čili za vyšší ceny MEANS, než si představují.* PDT

(71) *Některé ekonomické vazby jsem podcenil. Třeba {jsem podcenil} **cenovou politiku** při nákupu surovin, kdy je nutné hodně předvídat. Nebo způsob, jak se uplatnit na trhu, který je hodně konzervativní.* PDT

PREC signalizuje slovesné připojování

(72) *Její **příznivce** určitě **zahřálo u srdce**. Méně však již poraženou opozici.* PDT

(73) *Může se objevit **nesnášenlivost mezi lidmi, nedůvěra, podezírání, malichernost a málo velkorysosti a nadhledu**. A někdy také příliš úzké osobní zájmy.* PDT

Kombinace obou možností: v následujících větách jde nejprve o koordinaci členskou (přímou), poté o koordinaci slovesnou:

(74) ***Pro její hráče i pro dirigenta** to byl jistě výjimečný večer. Stejně tak pro někdejší československé posluchače, jimž je Dvorský vskutku národním umělcem. A možná MOD [to byl výjimečný večer] i RHEM *pro velmi skromně vystupujícího tenoristu...* PDT*

### 5.2.6.2. Gramatická elipsa řídicího slovesa v klauzi s výrazem s funktorem PREC

Stojí-li výraz s funkcí PREC ve větě s gramatickou elipsou řídicího slovesa, je tomuto predikátu doplněno zástupné t-lema #EmpVerb. Gramatická elipsa nijak nenaznačuje možnost členské koordinace závislých členů elidovaného predikátu se členy v předchozí větě, predikáty obou vět jsou rozdílné. Případy gramatické elipsy řídicího slovesa ve větách s funktorem PREC tedy vždy hodnotíme jako koordinaci slovesnou, tj. jako vztah diskurzivní.

(75) *Třeba jsou excelentními matematiky, ale nijak nevynikají v jazycích. Nebo naopak.* PDT

(76) *Základní filozofií pola je „druhý vůz do rodiny“. Čili auto využívané především v městském provozu, při cestách na kratší vzdálenosti nebo za nákupy.* PDT

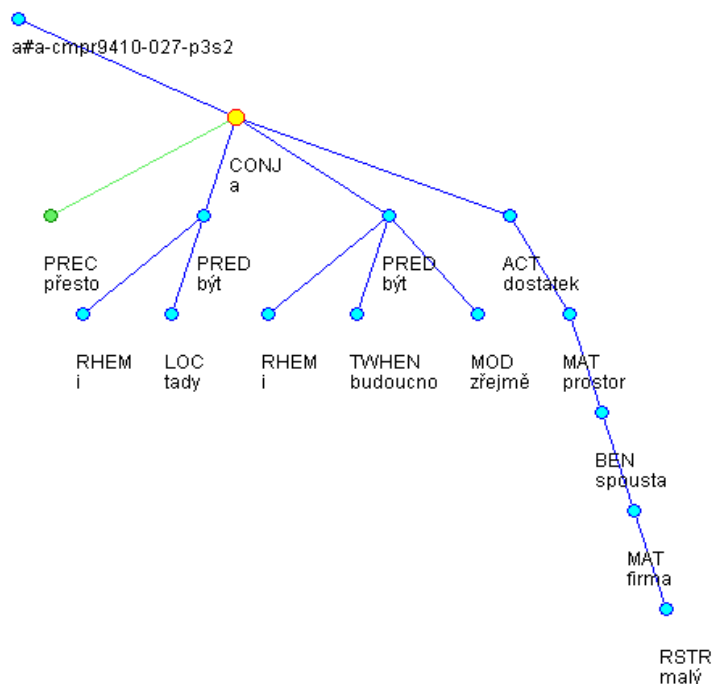
Častá je gramatická elipsa v rozhovorech na začátku otázky, která nebyla položena první. Zde se zpravidla konektor vztahuje k poslední položené otázce.

(77) *A váš názor na celkovou situaci v oblasti informací?* PDT

### 5.2.6.3 PREC jako potomek koordinačního uzlu

Kromě příslušnosti uzlu s funktorem PREC k uzlu pro řídicí sloveso jsou v PDT 2.0 také případy, je kdy výraz s funktorem PREC zachycen jako přímý potomek uzlu pro větnou koordinaci, jinými slovy, vztahuje se dle anotace k oběma koordinovaným klauzím. V těchto případech je připojovaným argumentem větší textový úsek než jedna klauze.

(78) *Málokde vládne tak konkurenční prostředí jako v počítačové branži. Přesto i zde je a zřejmě i v budoucnu bude dostatek prostoru pro spoustu malých firem.* PDT



Přesto i zde je a zřejmě i v budoucnu bude dostatek prostoru pro spousta malých firem. PDT

**Obr. 18** PREC jako přímý potomek uzlu pro souřadnost

#### 5.2.6.4 Nepravé závislé klauze vztažné jako reprezentant problémových okruhů

Výraz s funkcí PREC někdy objevuje i jako potomek slovesa závislé klauze vztažné. Vztah řídicí a závislé klauze vztažné jsme již označili za vztah „nediskurzní“, tyto případy v korpusu nás však nutí se zamyslet nad tím, k čemu se přesně konektor v takových konstrukcích vztahuje. Předpokládáme, že podobně nejednoznačných struktur se vyskytuje více.

(79) *Odběratel v době splatnosti faktury za zboží vystavil šek do banky, který ale nebyl krytý hotovostí na jeho účtu.* PDT

V tomto případě se zdá, že výraz *ale*, který má v tektogramatické reprezentaci funktor PREC, připojuje vztažnou klauzi nikoli ke klauzi řídicí, nýbrž k presupozici. Pisatel i čtenáři předpokládají, že vystavený šek bývá obvykle krytý penězi z účtu. Pokud tomu tak není, upozorní se na to částicí vyjadřující určitý rozpor s nepříznakovou situací (s očekáváním čtenáře).



(80) *Například podnikatel by chtěl dosáhnout u zaměstnance v obchodě tržby 60 tisíc korun. Dotyčný zaměstnanec si však myslí, že může dosáhnout nanejvýš 50 tisíc korun. Tak vzniká konflikt řešitelný podle druhu zboží a agresivity nabídky, který však většinou zůstává neřešen.*  
PDT

Zde, jak napovídá přípona slova *řešitelný*, je naznačena existence možnosti řešení konfliktu vyjádřená postponovaným přívlastkem. Jinými slovy, konflikt je *řešitelný ale většinou neřešený*. Nejedná se tedy, stejně jako v předchozím případě, o vztah klauze řídicí a závislé vztažné. Zde se jedná o vztah dvou vztažných doplnění. Narážíme na jiný problém, na problém kondenzace propozic do neslovesných útvarů.

V dalších případech výskytu potenciálního konektoru ve vztažné klauzi narážíme na problematiku toho, že vztažná klauze je prostřednictvím tohoto konektoru vázána ke klauzi předchozí i jiným sémantickým vztahem, než je vztah jmenná skupina – její rozvití, respektive determinace. Často vztažná klauze vyjadřuje následný děj vzhledem k ději v klauzi řídicí. Tyto vztažné konstrukce se pak jeví jako stylisticky méně vhodné, než když mluvčí/pisatel užije souřadného spojení těchto klauzí. Podobně jako lze větu (79) parafrázovat souřadným spojením klauzí (81), lze takto parafrázovat i další podobné konstrukce (82) a (83):

(81) *Odběratel v době splatnosti faktury za zboží vystavil šek do banky, ten ale [jak se později ukázalo] nebyl krytý hotovostí na jeho účtu.*

(82) *Chtěli jsme hrát nátlakový fotbal, který však ztroskotal na kvalitní obraně Benešova.* PDT

(83) *Chtěli jsme hrát nátlakový fotbal, [naše snaha] však ztroskotal[a] na kvalitní obraně Benešova.*

Přítomnost výrazu s funkcí PREC ve vztažné klauzi (a taktéž fakt, že se tento výraz nevztahuje nutně k predikaci věty řídicí, ale například k presupozici, ke jmenné skupině či k modálnímu slovesu) může signalizovat, se jedná o nepravou závislou klauzi restriktivní, dle MČ III (533) o pseudodeterminační nepravou vedlejší větu. Tektogramatická reprezentace v PDT však rozdíl mezi pravými a nepravými závislými klauzemi restriktivními nezachycuje, a je tedy nutné se na tyto struktury zaměřit samostatně. Na konkrétních textech je v těchto případech možné zkoumat možnosti parafráze, a z nich usuzovat na diskurzivní vztah. Jak však

uvádí Mluvnice češtiny, „mezi pravými vedlejšími větami vztažnými determinativními a [...] vztažnými větami pseudodeterminativními“ je „neostrý přechod“.

### 5.2.6.5 Kataforické konektory

Výrazy s funktorem PREC reprezentují jádro skupiny diskurzivních konektorů, svými vlastnostmi však nemusejí vyčerpávajícím způsobem reprezentovat vlastnosti celé této skupiny. Tak proti Asherově pojetí diskurzivních konektorů (*cue phrases*) jako výrazů anaforických musíme v závěru dodat, že některé konektory, které tektogramatická anotace nezahrnuje do skupiny výrazů s funkcí PREC, mohou být nikoli anaforické, nýbrž kataforické povahy: jedná se především o podřadící spojky stojící v preponované závislé klauzi (mají oba argumenty „před sebou“, tj. vpravo od sebe) a dále o některé konektory časové a kompoziční (srov. MČ III, 707): *nejdříve*, *zprvė*, ale také *například* nebo interpunkční znaménko dvojtečka.

### 5.2.6.6 Shrnutí

Probírali jsme způsob vyjádření a rozsah argumentů výrazů anotovaných v PDT 2.0 funktorem PREC. Zjistili jsme, že v převážné většině případů se výraz PREC vztahuje k obsahu jedné předchozí propozice, a to propozice vyjádřené klauzí s určitým tvarem slovesa jakožto řídicím členem. Nalezli a doložili jsme příklady, které se od tohoto centra odchyľují, a zjistili jsme, že zvláštní pozornost při budoucí anotaci diskurzivních vztahů bude třeba věnovat zejména:

- rozlišení slovesného a členského připojování
- možnosti návaznosti diskurzivní částice na presupozici
- připojování klauzí, jejichž řídicí člen není reprezentován verbem finitem a v této souvislosti především všem deverbativním substantivům a jiným kondenzátům
- hierarchické struktury diskurzivních vztahů
- rozsahu argumentu vzhledem k zahrnutí zanořené klauze do argumentu (viz příklady 1.1 D, 3.B)
- problematickým syntaktickým strukturám – nepravé vedlejší věty, vsuvky apod.

## 5.2.7 Hraniční a problémové případy

Další část korpusového průzkumu se soustřeďuje na ty výrazy ze seznamu přeložených anglických *discourse connectives* dle PDTB 1.0 (viz kap. 5.2.2), které se v Pražském závislostním korpusu nikdy neobjevují s funktorem PREC. Převažují u nich temporální a jiné adverbialní funkory nebo funkory pro atomické uzly. U výrazů v tabulce 3 tedy bude na dalším uvážení, zda a podle jakých kritérií je klasifikovat jako diskurzni konektory. Tabulka 3 popisuje rozložení funktorů u těchto výrazů.

výraz s 0 výskyty jako PREC v PDT 2.0 <sup>44</sup>	nejčastější funkory
jen	2 PREC, 1102 RHEM, DPHR, EXT
rovněž	1 PREC, 273 RHEM, CM
například	1 PREC, 561 RHEM, 147 CM
právě	RHEM, CM, TWHEN
především	467 RHEM, CM
nejprve	62 TWHEN
nejdříve	60 TWHEN
zprvé	2 CM, REG
zadruhé	0 výskytů
poté	TWHEN, nebo skrytý uzel
předtím	TWHEN
mezitím	TPAR
zatím	TWHEN
nakonec	161 TWHEN, 10 ATT, 1 CM
stručně (t-lema stručný)	MANN
krátce	MANN, TWHEN, EXT
zkrátka	ATT
konečně	9 ATT, 1 CM

**Tabulka 3** České výrazy dle seznamu *discourse connectives*, které nemají v PDT 2.0 funkci PREC

<sup>44</sup> nebo jde o výskyty anotované jako PREC chybně

Zamyšlení nad výrazy s tektogramatickým funktorem PREC a analýza shrnutá v tabulkách 1 - 3 daly zejména podnět k otázkám, které je třeba zodpovědět, než budeme definovat skupinu diskurzních konektorů pro PDT.

Diskurzní konektory (vycházející z charakteristiky výrazů s funktorem PREC) je třeba dále popsat ve vztahu vůči:

1. jiným funkcím týchž výrazů:

- významy modifikující souřadné spojení – popsané funktorem CM
- významy rematizující – popsané funktorem RHEM
- významy vyjadřující postoj mluvčího k obsahu výpovědi – popsané funktorem ATT
- adverbialní významy, především temporální

2. všem ostatním jazykovým prostředkům, které naznačují textové vztahy, zvláště vůči koreferenčním vztahům (na základě anotace výrazů s t-lematem *ten* jako PREC).

Ad 1. Tomuto okruhu věnujeme kapitoly 5.3 až 5.6, každá z těchto kapitol se vždy zabývá jednou ze zmíněných sémantických kategorií v PDT.

Ad 2. Z tohoto okruhu se budeme věnovat jen okrajovým jevům zachyceným v tektogramatické rovině funktorem PREC, jsou to ustálené víceslovné nominální a pronominální spojení a dále interpunkci. Věnujeme jim kapitoly 5.7 a 5.8.

### 5.3 Modifikátory souřadných spojení a jejich vztah k diskurzním konektorům

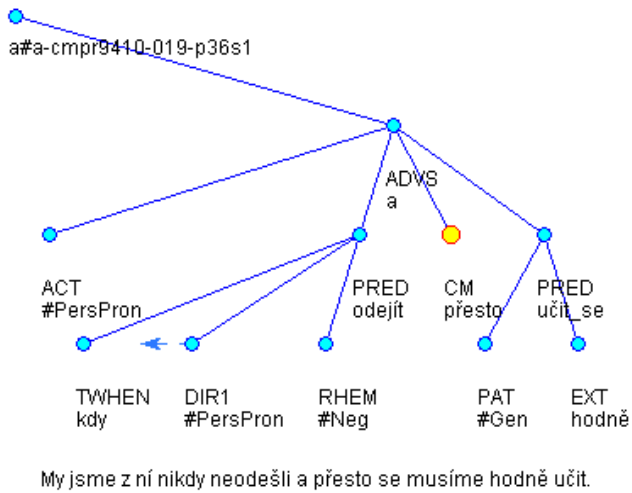
Sémantická kategorie CM (conjunction modifier) je v anotačním schématu tektogramatické roviny v PDT 2.0 přiřazena výrazům modifikujícím souřadící spojovací výrazy. Ve stromovém grafu jsou atomické uzly reprezentující tyto výrazy zachyceny vždy nezávislostní hranou jako přímí potomci souřadícího spojovacího výrazu. Formálně se „jako součást souřadících spojovacích výrazů může vyskytovat většina částic a příslovce s primární funkcí rematizátorů, různé strukturní částice, některá příslovce s primární funkcí příslovečného určení a další výrazy“ (Mikulová et al. 2005: 594). S funktoem CM se neobjevují spojková slova, jak můžeme vyčíst i z tabulky 2. Výrazy, které mají ve sloupci m-tag vyplněnou zkratku pro souřadící spojky (J<sup>^</sup>), se zároveň nikdy nevyskytují s funktoem CM.

Jak však ukazuje tato tabulka, sémantická kategorie CM se lexikálně překrývá nejen s třídou rematizátorů a adverbialních doplnění, ale také s některými výrazy, které se vyskytují zejména ve funkci PREC.<sup>45</sup> O zařazení všech takovýchto výrazů do kategorie CM rozhoduje, zda jsou součástí víceslovného spojovacího výrazu či zda nějaký spojovací výraz v daném kontextu modifikují. Mají i specifické postavení ve větě – stojí zpravidla mezi dvěma koordinovanými uzly, a to za první částí spojovacího výrazu vyjádřeného koordinačním uzlem (jako koordinační uzly vystupují především základní spojky nebo interpunkční znaménko, nejčastěji čárka (t-lemma = #Comma))<sup>46</sup>.

---

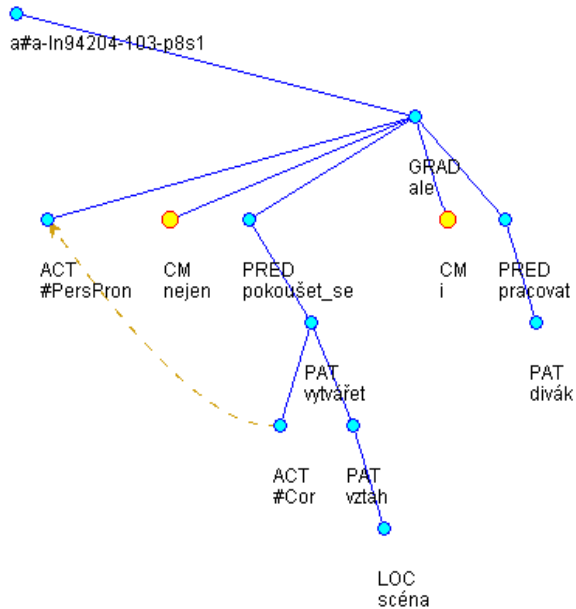
45 Viz sloupec s názvem CM, který udává u dvaceti nejběžnějších výrazů s funkcí PREC počty jejich výskytů s funkcí CM v PDT 2.0.

46 Máme na mysli postavení v hloubkovém slovosledu.



**Obr. 19** Zachycení uzlu s funktorem CM I.

Některé modifikátory souřadných spojení se vyskytují v obou klauzích, které spojují. „Hlavní“ výraz takového víceslovného spojovacího výrazu se zachycuje jako koordinační uzel, ostatní jako CM:



**Obr. 20** Zachycení uzlu s funktorem CM II.

Jak je vidět z druhého příkladu, výrazů, které modifikují souřadné spojení, může být v jedné větě nebo klauzi několik za sebou a různě se mezi sebou kombinují. Typicky se objevují např:

*nejen CM ale GRAD třeba CM i M,*  
#Comma CONJ *případně CM i CM*

Určité kombinace spojovacího výrazu a jeho modifikátoru se vyskytují často, např.:

*sice ale, a proto, a tedy, a tak, a dále, a případně i, a především, nebo zase, nebo i, nebo naopak, nebo alespoň, #Comma případně, #Comma především, ale rovněž, ale také, přesto CM však ADVS, přesto ale, zároveň však atd.*

V počátcích anotace druhé verze PDT byl sestaven seznam výrazů, které se s funktorem CM mohou objevit pod uzlem pro souřadnost.<sup>47</sup> tyto výrazy byly dále subkategorizovány na restriktory (např. *alespoň, jen, hlavně, výhradně, třeba, většinou*, kontextualizátory ( např. *dále, i, ještě, pak, rovněž, také, zároveň*), negační výrazy (*ani, ne, nikoli, nikoliv*, negační morfém *ne-* u sloves<sup>48</sup>) a ostatní modifikátory, přičemž u ostatních bylo popsáno, v jakém koordinačním vztahu se typicky vyskytují. I v těchto subkategoriích se však mísí výrazy fungující v jiném kontextu jako rematizátory a jako konektory.

Zaměříme-li se na vztah výrazů s funktorem CM k výrazům s funktorem PREC, shledáme, že určitá podskupina výrazů označených jako CM jsou vlastně navazovací výrazy, které pouze stojí uvnitř jedné věty mezi klauzemi, zpravidla za čárkou nebo za spojkou. Anotaci interpretujeme tak, že tyto výrazy svoji připojovací funkci ztrácí tím, že ji přejímá uzel pro souřadné spojení. Zda tomu tak opravdu je, nebo zda se jedná o pouhou konvenci, je oprávněná otázka především v případech, kdy je uzel pro souřadné spojení reprezentován čárkou, srov:

(84) *Snažím se projet co nejméně.* [#Comma] CSQ *proto* CM *jezdím na žižkovské nákladové nádraží.* PDT

(85) *Snažím se projet co nejméně.* *Proto* PREC *jezdím na žižkovské nákladové nádraží.*

47 Tento seznam vypracovala pro účely anotace Marie Mikulová. Uvádíme ho v Příloze 4.

48 Tento morfém je v tektogramatickém stromě reprezentován uzlem se zástupným t-lematem #Neg.

Čárka, sama o sobě grafický prostředek členění textu, je sémanticky vyprázdněná a spojuje textové segmenty, aniž by explicitně určovala druh vztahu mezi nimi. Proto se přikláníme k tomu, aby v případech, kdy po čárce roli modifikátoru souřadného spojení zaujímá výraz běžně se vyskytující jako PREC, hodnotit tento výraz jako diskurzí konektor. Teprve jeho přítomnost totiž umožňuje jednoznačně přiřadit čárce funktor pro druh souřadného spojení. Pokud by zde výraz s funkcí CM chyběl, rozhodování o druhu souřadného spojení nemusí být jednoznačné. Pokud je modifikována spojka *a*, je situace podobná, neboť tato spojka je také sémanticky vyprázdněná. Není-li modifikována, pouze spojuje základním vztahem junkce.

Domníváme se tedy, že pokud je výraz s funktorem CM funkčně homonymní s výrazem PREC (tzn. může se vyskytnout i jako PREC) a zároveň v souřadné struktuře napomáhá jednoznačnému udělení funktoru uzlu pro souřadné spojení, měl by se uzel s funktorem CM spolu se souřadným uzlem analyzovat jako víceslovný diskurzí konektor (s funkcí spojovací, nikoli připojovací). Typickými zástupci této skupiny výrazů jsou důsledková spojení *a proto*, *a tedy*, *a tudíž*, *a tak* či adverzativní spojení *a přece*, *a přesto*. Spojka *a* může být v těchto případech nahrazena čárkou.

Posouvat význam spojky (tedy přiřadit uzlu pro kořen souřadné struktury jiný funktor, než by měl bez modifikace) napomáhají i jiné výrazy než ty, které se běžně vyskytují s funktorem PREC, například výrazy funkčně homonymní s rematizátory, srov.:

(86) *Tato složka bude kontrolovat stavební činnost na území hlavního města, a GRAD zejména CM dohlížet na to, zda všechny firmy mají stavební povolení.* PDT

Hodnotit v takových případech jako diskurzí konektor pouze spojku by bylo chybné. Pokud má tedy vliv na určení funktoru kořene souřadné struktury jakýkoli výraz s funktorem CM, měl by být pojímán spolu se spojkou jako jeden konektor. Druh souřadného spojení totiž určuje výhradně jako celek.

Některé modifikátory souřadných spojení, zejména kontextualizátory, význam základní spojky neposouvají, ale pouze zpřesňují či zdůrazňují:

(87) *Prvně jmenovaný import ze Švédska se blýskl nedávno vítězstvím ve Vídni a CONJ také CM v kvalifikaci na derby měl dobrý čas.* PDT



V těchto případech hodnotíme jako diskurzní konektor pouze kořen souřadné struktury. Uzly s funktorem CM v těchto strukturách jsou jakousi volnou modifikací souřadného spojení.

Při třídění výrazů s funktorem CM na ty, které fungují jako součást diskurzního konektoru a na ty, které tuto funkci neplní, je nutné vzít v první řadě v potaz, že velká část souřadných struktur s modifikátorem CM jsou koordinace členské:

*(88) V nákladním autě byly nalezeny náboje, benzín, zásobníky s plynem a také mléko. PDT*

Tyto koordinace nehodnotíme jako diskurzní s dvěma výjimkami: (i) koordinované neslovesné členy jsou kondenzáty slovesných klauzí; (ii) uzel s funktorem CM je záporná částice. K obojímu blíže viz kapitola 5.2.7., respektive podkapitola 5.2.7.1.

## 5.4 Rematizátor ve funkci diskurzního konektoru

Rematizátory (fokalizátory, aktualizační částice) jsou takové výrazy, jejichž funkcí je signalizovat ve větě kategorie aktuálního členění. Zpravidla signalizují informaci novou, kontextově nezapojenou (réma). Ve větě stojí téměř vždy před členem, ke kterému se vztahují (který rematizují). Prototypickými rematizátory, jak se jimi zabývají práce o aktuálním členění, jsou v angličtině výrazy *also*, *only* a *even* (zejména Sgall, Hajičová a Buráňová 1980; Hajičová, Partee a Sgall 1998; Firbas 1992, Hajičová 1995, 1999 a další). Primární funkcí takovýchto výrazů je tedy být ve větě rematizátorem, obvykle vyjadřují zdůraznění či vytýkání a mohou mít užší nebo širší dosah (tedy rematizovat jeden či více členů věty).

Pro účely anotace tektogramatické roviny v PDT byl vytvořen seznam českých výrazů, které mohou mít rematizační funkci. Jedná se ovšem nezdědka o výrazy, které mohou plnit i jiné funkce. Většina rematizátorů je tedy funkčně homonymní. Manuál pro tektogramatickou anotaci se věnuje rozlišení této funkční homonymie u výrazů, které mohou být (i) rematizátorem či adverbialním doplněním (viz tabulka 9.3 u Mikulové et al. 2005: 1099); (ii) rematizátorem či výrazem s funkcí modalitní (2005: 1100); (iii) rematizátorem či výrazem modifikujícím souřadící spojovací výraz (tamtéž).

Existuje však skupina výrazů, která je v tektogramatické anotaci PDT označena někdy jako rematizátor a jindy jako PREC, tedy výraz s připojovací funkcí. Dle korpusových výzkumů se to týká zejména výrazů *také*, *těž*, *i*, *rovněž*, *zároveň*, *spíše*, *nejspíš*, *zase*, *jen* a *naopak*. Toto překrývání v sémantickém označení můžeme posuzovat různě: buďto je v těchto případech anotace nesprávná, a tudíž každý z těchto výrazů plní ve všech svých výskytech pouze jednu funkci (tedy pokud je rematizátorem, nemůže být nikdy konektorem a naopak), nebo se přikloníme k názoru, že anotace v těchto případech neproběhla mechanicky podle většinového přístupu a tedy (i) výraz může mít v určitém kontextu funkci rematizátoru, zatímco v jiném funkci konektoru (ii) výraz může plnit obě tyto funkce najednou ve stejném kontextu.

Následující příklady (89) – (92) ukazují, že existuje určitá skupina sporných případů, kterou je třeba podrobit detailnějšímu zkoumání. Není účelem této studie najít a stanovit přesné kritérium pro určení funkce konektivní a funkce rematizační u potenciálně funkčně homonymních výrazů, pokud vůbec takové kritérium stanovit lze. V této kapitole chceme pouze poukázat na tyto sporné případy a na jejich základě otevřít otázky, které by blíže

osvětlily tuto problematiku. Následující věty se týkají fungování výrazů *také*, *naopak* a *jen* ve specifickém kontextu.

(89) *Petr vyluxoval celý byt. Navíc PREC vytřel podlahy.*

(90) *Petr vyluxoval celý byt a GRAD navíc CM vytřel podlahy.*

(91) *Petr vyluxoval celý byt. Také RHEM (PREC?) vytřel podlahy.*

(92) *Petr vyluxoval celý byt a CONJ také CM vytřel podlahy.*

Ve větě (89) má výraz *navíc* konektivní funkci. Ve větě (91) může mít výraz *také* ve stejné pozici dle tektogramatické anotace PDT jak funktor RHEM, tak i PREC. Pokud se ovšem oba tyto výrazy objeví v rámci jedné věty (90), (92), konektivní funkce je značena u přítomné spojky (nebo čárky) a výrazy za touto spojkou se bez rozdílu hodnotí jako CM – „conjunction modifier“, tedy modifikátor souřadícího spojovacího výrazu. Zdá se, že ve všech čtyřech větách je však výrazy *také* a *navíc* řečeno shodně to, že Petr kromě vyluxování udělal ještě něco dalšího. Slovo *navíc* sice ještě evokuje určitou gradaci dějů, ale základní význam nějaké další Petrovy činnosti vyjadřují oba tyto výrazy. V textech PDT se velmi často objevují věty se slovesem mluvení typu: *Dále/také/rovněž/těž řekl/vedl/zminil/navrhl/sdělil/dodal...*, v nichž jsou tyto výrazy hodnoceny různě.

Složitá je situace také u výrazu *naopak*:

(93) *Jirka naopak vstal.*

Tato věta je v Manuálu uvedena jako vzorová věta pro rematizátor, ačkoli podle anotace je tento výraz (pokud se nejedná o adverbialní doplnění způsobu) daleko častěji konektorem (189 výskytů) než rematizátorem (5 výskytů).

Dalším sporným případem je chování částice *jen* (prototypického rematizátoru) v následujících případech:

(94) *V předpokládané sestavě jeho jméno neuvedl. Neznamená to, že postaví dvanáct hráčů, CONJ jen RHEM v útoky vybere dva ze jmenované trojice. PDT*

(95) *Láska, rodina a mravní kodexy se nemění*, CONJ jen RHEM *technika pokročila*. PDT

(96) *Přijdu rád. Jen ještě nevím, kdy budu moci.*

V těchto případech bychom měli zvážit, zda výraz *jen* neplní adverbativní spojovací funkci s dalším významem omezení. Je nahraditelný výrazy *ale* nebo *však*.

V hodnocení rematizátorů má zásadní význam jejich postavení ve větě a větná intonace. Obecně platí, že rematizátor stojí bezprostředně před výrazem, který rematizuje a tento výraz nese i větný přízvuk. Pokud je však větný přízvuk na rematizátoru samotném, zpravidla se vztahuje k výrazu vlevo od rematizátoru.

Domníváme se, že pokud bychom uvažovali o konektivní funkci některých výrazů, které mají zároveň (ve stejném kontextu) nebo v různých výskytech (v jiném kontextu) funkci konektivní, jedná se o takové rematizátory, které se vztahují k predikaci, nikoli tedy bezprostředně k ostatním členům věty, srov. následující příklady. Tento požadavek vychází ze základního vymezení diskurzních konektorů: spojují či připojují děje, vyjádřené především finitním slovesem ve funkci predikátu.

(97) Petra si umyla také boty.

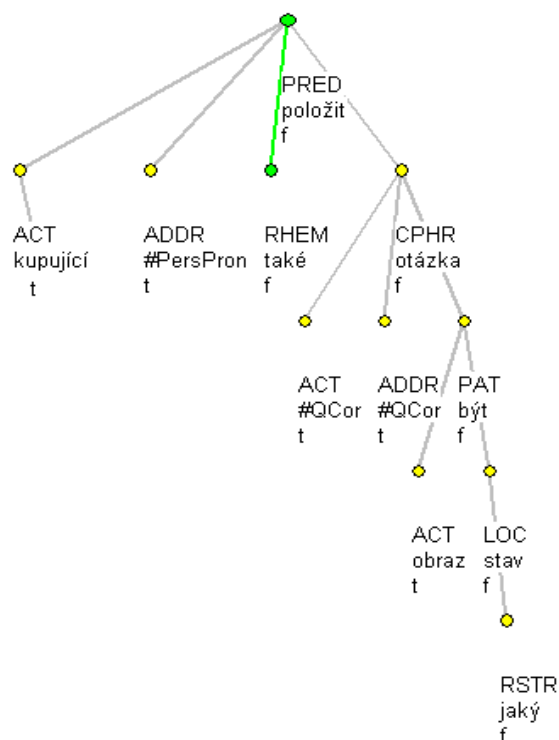
(98) Petra si umyla také červené boty.

(99) Také Petra si umyla boty.

(100) Petra si umyla boty také.

(101) Petra si také umyla boty.

Rematizované členy jsou označeny kurzívou. Větný přízvuk předpokládáme na těchto členech (kromě věty (100), kde může být přízvuk na rematizátoru). Za potenciální konektor považujeme výraz *také* pouze v posledním příkladě, kde má v dosahu sloveso, neboli kde se dozvídáme o další Petřině činnosti. Rematizátor v této pozici ve větě se v tektogramatickém stromě zachycuje jako nejpravější levý potomek predikátu, viz obr. 21.



Kupující by si měl také položit otázku, v jakém je obraz stavu. PDT

### Obr. 21 Rematizovaná predikace

Ze všech takto zachycených rematizátorů PDT 2.0 je nejčastějším rematizátorem slovesa záporná částice *ne*, které ovšem jistě nepřísluší vyskytovat se ve funkci konektivní. Situaci komplikuje mnoho dalších faktorů, jako je subjektivní pořadí členů věty a příznakové umístění intonačního centra, věty s vnořenými klauzemi atd.

Je tedy možné vyslovit domněnku, že některé z rematizátorů, které mají v dosahu predikaci, ale rozhodně ne všechny takové, by mohly současně sloužit jako konektory diskurzu. Je ovšem nutné podívat se na každý z těchto výrazů samostatně. Jak uvádí již Daneš ve svém zkoumání vlivu větné intonace na smysl sdělení (1957: 82) a také Hajičová (1995, 1999), jednotlivé rematizátory vykazují různé chování, např. jiné vlastnosti má částice *jen* (*jenom*) a jiné opět *také* (*též*). Zároveň nelze zkoumat oddělené příkladové věty, „větě dodává jednoznačného významu nakonec jen souvislost, kontext, situace“ (Daneš 1957: 84). Abychom se tedy přiblížili nalezení určitého řešení otázky plnění konektivní funkce rematizátory, je třeba zabývat se konkrétními rematizujícími výrazy v konkrétních textech, a na tomto základě pak podrobit zevrubnému sémantickému třídění nejen výrazy anotované v PDT jako PREC, ale i celou třídu rematizátorů.

## 5.5 Výrazy vyjadřující postoj mluvčího ve funkci diskurzního konektoru

Výrazy vyjadřující postoj (hodnotící či emocionální stanovisko) mluvčího k obsahu výpovědi nebo její části jsou v tektogramatické reprezentaci označeny funktorem ATT (attitude). „Narozdíl od jiných doplnění charakterizují doplnění s funktorem ATT komunikační situaci, signalizují komunikační funkce výpovědi a vnášejí do výpovědi pragmatické významy. Vyjadřují takové postoje mluvčího, jako je podiv, překvapení, uspokojení, obavy, lhostejnost, politování aj.“ (Mikulová et al. 2005: 531). Uzly s funktorem ATT však nevyjadřují modální charakteristiku výpovědi – takové výrazy mají v PDT tektogramatický funktor MOD:

(102) *Je to náhodou ATT hezké.*

(103) *Jenom ATT se opovaž!*

(104) *Pravděpodobně MOD přijdeme.*

(105) *Vím jistě MOD, že Praha mě poznamená.*

Výrazy s funktorem ATT jsou především adverbialní a částicové výrazy, které mají často i jinou funkci:

(106) *Setkali se náhodou CAUS.*

Doplnění s funktorem ATT a MOD vykazují podobné chování jako rematizátory, stupněm pravděpodobnosti obsahu nebo vyjádřením postoje zasahují vždy určitou část výpovědi (určitý podstrom) či výpověď celou. (srov. Mikulová et al. 2005: 376). Pro účely anotace tektogramatické roviny v PDT byl sestaven seznam adverbii spolu s jejich tektogramatickými funktoři.<sup>49</sup> Z tohoto seznamu je patrné, že funktor ATT, stejně jako funktor PREC, se částečně lexikálně překrývá s rematizátory a s adverbialními doplněními především způsobu a času.

Otázka, kterou si zde pokládáme, zní, zda v některých případech nepřevažuje u výrazů s funktorem ATT nad jejich vyjádřením postoje mluvčího k obsahu výpovědi funkce konektivní.

---

49 sestavila E. Buráňová (2001), část seznamu přejímáme jako Přílohu 5

K této otázce nás vede nejen výše zmíněná homonymie, a tudíž překrývání funktorů ATT a PREC u některých výrazů v anotaci (např. *vždyť, stejně, ovšem*), ale také to, že gramatické příručky uvádějí výrazy, které mají v PDT funktor ATT, jako příklady v kapitolách o textových konektorech. V *Mluvnici češtiny* (MČ III: 706, 712) jsou výrazy *ostatně* (v PDT ATT), *stejně* (v PDT PREC či MANN) a *přece* (v PDT ATT, CNCS, CM) hodnoceny jako konektory připojující dodatečné vysvětlení, explikaci s dalšími významovými rozdíly zdůraznění, utvrzení či přípustky, viz příklad z MČ III:

(107) *Sám patos je třeba chápat v různých polohách... Ostatně i naše devatenácté století diskutovalo o těchto polohách.* MČ III

*Příruční mluvnice* zde mluví o různých konektivních funkcích modifikačních částic, konkrétně o vytvoření komunikačního souladu odvoláním na sdělené nebo jinak sdílené informace, hodnoty apod. (PMČ: 694). Viz příklad:

(108) *Já nevím, proč se pořád bojíte. Kyselina se přece neutralizuje louhem.* PMČ

Podobné případy, kdy věta uvozená výrazem s funktoem ATT podává dodatečné vysvětlení či utvrzení, se vyskytují i v PDT 2.0:

(109) *Naší oporou by mělo být i fantastické domácí publikum. Vždyť ATT máme kapacitu stadionu 5000 míst a dva týdny před ligou už jsme prodali 3000 permanentek.* PDT

(110) *Nechci podezřívát Kalvodu z toho, že chtěl lacino přijít k možnosti skryté předvolební kampaně. Ostatně ATT typ pořadu, o který šlo, sleduje naprosté minimum lidí.* PDT

(111) *Finanční či investiční poradce totiž nesmí být závislý na nějakém fondu nebo bance. Musí svým klientům doporučit tu investici, která pro ně bude optimální, a ne tu, za kterou dostane největší provizi. Za to je přece ATT klienty placen, a ne zrovna málo...* PDT

V uvedených případech z PDT lze výraz s funktoem ATT nahradit výrazem *totiž* s tím, druh vztahu k předchozí větě zůstane stejný, ale příznak hodnocení se vytratí.

Mezi částice strukturující text patří dle MČ II (238) i například výrazy *krátce, prostě, vlastně, zkratka*,<sup>50</sup> které se v PDT anotují jako ATT nebo MANN, tj. nikdy jako PREC.

(112) *Spadl z hradby a srazil si vaz nebo co, zkratka už nevstal.* MČ II

Případy, kdy mají tyto výrazy větněčlenskou platnost, lze poměrně snadno rozpoznat od jejich funkcí modifikátorů výpovědi nebo její části. Otázka však zní stejně jako u předchozí skupiny výrazů: zda mimo výskyty s větněčlenskou platností nemají tyto výrazy v různých kontextech různé funkce, případně, zda je nemůžeme hodnotit jako připojovací konektory s vedlejším významem hodnocení mluvčího.

Při řešení této otázky navrhuje zohlednit následující aspekty:

- Jádru výrazů ATT je poměrně stabilní – vyjadřuje postoj mluvčího „silně“, a nemívá jinou funkci (např.: *bohužel, jaksi, naštěstí, pochopitelně, evidentně*).
- Diskurzivní konektory samy o sobě jasně signalizují přítomnost „prvního“ argumentu, přijímají tedy do diskurzivního vztahu dva nebo více argumentů. Tento požadavek jádro výrazů s funktorem ATT v našem pojetí nespĺňuje, zatímco výše zmíněné příklady z PDT ano.
- Funkce částic (nebo adverbii bez větněčlenské platnosti) může být několikerá, dokládá to jak anotace křížením funktorů, tak popis těchto výrazů v gramatických příručkách. Hodnocení funkce částic je však v gramatických příručkách velice nesourodé. Není snadné najít jednotící hledisko pro posouzení určité funkce té které částice jako převládající.

Zdá se, že sémantická kategorie ATT, jak je vymezena v anotačním schématu PDT, obsahuje výrazy, jejichž sémantický příznak vyjádření postoje (hodnocení) mluvčího je různě silný – čím je slabší, tím spíše mohou „vystupovat“ i jiné funkce těchto výrazů: například schopnost reatizace či schopnost vázat se určitým způsobem k předchozímu kontextu. V příkladech (107) – (112) uvedené výrazy pro nás konektivní povahu mají, neboť vyjadřují určitý vztah (vysvětlení, utvrzení, shrnutí atd.) k obsahu předchozí věty. Zároveň vyjadřují i určité hodnocení děje mluvčím.

---

<sup>50</sup> MČ II hodnotí tyto výrazy jako částice s tím, že kritérium rozlišení částice a příslovce je větněčlenská platnost. (Proto zde můžeme porovnávat částice strukturující text dle MČ a seznam příslovci fungujících v PDT jako ATT – větněčlenskou platnost nemají ani v jednom případě.)



## 5.6 Temporální doplnění slovesa jako diskurzní konektory – několik poznámek k temporálním vztahům v diskurzu

Text, diskurz či promluva je jev zakotvený v časových i prostorových souvislostech. Užití temporálních výrazů v diskurzu obecně souvisí s temporální sémantikou diskurzu. Hoffmannová (1986) rozlišuje čas reálného dění (externí) od času textového (interního). Čas textový „je organizován mluvčím [...] na základě jeho mentálního zpracování vyjádření objektivní reality [...], je dán lineární sukcesivitou textu, ale především jeho kompozičním uspořádáním, které se projevuje i v grafickém členění.“ (Hoffmannová 1986: 160)

Vyjádření temporality v diskurzu je zřetelnější než vyjádření prostorových charakteristik – temporalita je signalizována už jen gramatickou kategorií slovesného času. Dále je zde však celá škála slovesných doplnění s určitým časovým významem. Hoffmannová mluví o tzv. časových textových orientátorech, u nichž se „v různých proporcích uplatňuje (a) konkrétní reference, poukaz ke konkrétnímu času (místu) děje, (b) funkce kohezní, odkaz k časovému zařazení (resp. místnímu určení) v jiné, někdy i vzdálené textové jednotce.“ (1986: 161) U některých adverbii se pak podle Hoffmannové konkrétní, mimotextová časová reference úplně ztrácí – pak se jedná o orientátory vnitrotextové, interně časové.

Anglické výrazy *after, before, by then, finally, in the meantime, meanwhile, thereafter* jsou dle anotačního schématu PDTB diskurzními konektory. Jejich ekvivalenty v češtině v PDT však téměř nikdy nemají jiné funkory než časové. Nulový výskyt funktoru PREC je v PDT 2.0 u *zatím, předtím, mezitím, nakonec*; u *potom/poté* je zaznamenáno 41 výskytů PREC z celkových 534 výskytů a u výrazu *pak* je to 160 z 178. Zaujmout jednoznačné stanovisko, zda se temporální výraz spíše vztahuje k vyjádřenému ději nebo zda spíše strukturuje text, nemusí být jednoduché. Ačkoli lze rozlišit strukturaci událostí v čase, reprezentovanou zejména výrazy *nejdříve – potom – nakonec*, od strukturace obsahu, typickými výrazy jsou *především – dále – konečně*, tyto dvě skupiny prostředků se prolínají. To je zřetelné i z různé anotace výrazů *poté, pak, potom, dále, následně, zároveň, současně*.

(113) *Bývalý hlavní vojenský prokurátor Miroslav Kříženecký se stal volebním lídrem ČSSD v komunálních volbách a zároveň TWHEN kandiduje na primátora Českých Budějovic.* PDT

(114) *Město postihla krize a nezaměstnanost. Zároveň* TWHEN začala nová éra svobodných celních zón. PDT

(115) *P. Dvorský zahájí program áriemi od B. Smetany a A. Dvořáka. K této literatuře se hlásím jako k vlastní, řekl Dvorský. Zároveň* PREC připomněl, že v Čechách se mu vždy dostávalo velké pozornosti. PDT

V angličtině se zpravidla staví výše zmíněné temporální výrazy na první pozici ve větě a oddělují se čárkou, což je částečně odděluje od zbývajících obsahu věty. Český slovosled naopak umožňuje různé postavení temporálních výrazů v rámci věty. Anotace českého korpusu ukázaly, že tendence přiřadit funktor PREC je vyšší u výrazů v iniciální pozici. Není to však pravidlem, viz příklad (114). Zejména u výrazu *zároveň* obtížné určit funktor, neboť tzv. prototypicky adverbialní doplnění (viz věta (116)) se v textech PDT příliš nevyskytují.

(116) *Vrátili se zároveň.*

V případě výrazu *zároveň* lze ještě rozlišit, zda vyjadřuje opravdu časovou souslednost dvou dějů (113) či zda jde spíše o následnost. O následnost dějů se jedná zejména u vět obsahujících sloveso mluvení, viz příklad (115).

Konkrétní význam temporálního výrazu ve větě, jeho postavení ve větě ani anotace těchto výrazů v PDT však nejsou jasnými kritérii pro vymezení určitého typu<sup>51</sup> temporálních doplnění jako vnitřně textových, tj. diskurzivních konektorů. Řešením je opět prozkoumat každý výraz této skupiny samostatně, a to v různých kontextech. Lze také zkoumat subkorpus textů s vyznačenou temporální strukturou: z tektogramatických uzlů by zkoumaly pouze uzly se všemi t-lematy, která se mohou vyskytovat jako doplnění s funktorem času a hledalo by se jejich propojení typu (*nejprve – potom – dále – nakonec*). Tímto způsobem by bylo možné pozorovat, jak se potenciální konektory s časovým významem na této struktuře podílejí, případně porovnávat zachycení externí (dějové) temporální struktury se zachycením diskurzivních vztahů ve stejných textech. Domníváme se, že funkci strukturace textu mají jen některé z těchto temporálních výrazů a pouze v některých kontextech, avšak mohou to být

---

51 Nezabýváme se temporálními doplněními typu *včera, příští týden, denně, do pondělka* apod. Tyto výrazy považujeme jednoznačně za adverbialní doplnění slovesa bez možnosti spojování či připojování textových úseků. Jsou však součástí temporálních vztahů v diskurzu.

i takové kontexty, kdy časové adverbium hodnotíme jako větný člen, tj. je v tektogramatické struktuře PDT anotován funktorem s významem času. Vnitřně textovou temporální strukturu pak navrhuje dále zkoumat a v anotaci diskurzu zachycovat odděleně od ostatních typů textových vztahů.

## 5.7 Interpunkční znaménka ve funkci diskurzních konektorů

Zástupná t-lemata se v tektogramatické anotaci PDT 2.0 přiřazují jednak nově generovaným uzlům, a jednak uzlům s povrchově přítomnými výrazy, kterým dle anotační konvence nepřípadně přímo lema (tj. základní forma slova) výrazu z textu. Zmíněný druhý typ tak tvoří jednotná lematizace osobních zájmen zástupným t-lematem #PersPron a dále zástupná t-lemata pro interpunkční znaménka, nejčastěji #Comma pro tečku, #Colon pro dvojtečku, #Bracket pro závorky (jeden uzel pro obě závorky), #Dash pro pomlčku a spojovník, #Semicolon pro středník.

Hovořili jsme o tom, že do skupiny diskurzních konektorů zahrnujeme i výrazy reprezentující kořeny souřadných struktur, tj. spojky souřadící. Z hlediska formy však může jako kořen souřadné struktury vystupovat nejen slovní tvar, ale i interpunkční znaménko. Interpunkční znaménka zpravidla v tektogramatické anotaci nejsou nijak reprezentována. Vlastní uzel s příslušným zástupným t-lematem však mají v několika případech. Jedním z nich je právě, pokud se vyskytují ve funkci spojovacího výrazu nebo operátoru. Těmto uzlům je přiřazen funktor pro druh souřadného vztahu. Ve funkci spojovacího výrazu se mohou vyskytnout všechna interpunkční znaménka kromě tečky. (srov. Mikulová et al. 2005: 925).

V rámci jedné věty (jednoho stromu) jsou interpunkční znaménka ve funkci konektorů v PDT zachycena ve všech případech. Pokud však opustíme hranice věty, mohou nastat případy, kdy interpunkční znaménko spojuje několik větných celků. Tyto případy tektogramatická rovina nezachycuje. Typickým příkladem jsou mezivětná spojení a výčty uvozené dvojtečkou:

(117) *Malá privatizace končí; Jako každý pacient měla své neduhy, ze kterých se dostala. 31. prosinec 1993 se jí však stal osudným.* PDT

(118) *Neschválené hračky mívají ze zdravotního hlediska řadu závad; Často ohrožují děti tím, že pouštějí barvu, z jiných se odpařují nebo dotykem do dětského organismu přenášejí různé škodlivé látky, například těžké kovy. Děti může ohrožovat i snadná hořlavost některých látek, nezapuštěný spojovací materiál u dřevěných i kovových výrobků a podobně.* PDT

Takové případy spojování vět pomocí interpunkce, budeme-li hodnotit jako diskurzní vztahy, bude potřeba vyhledat a ohodnotit na základě samotných textů či pracovat s analytickou rovinou popisu v PDT, kde je zachycena formou analytických uzlů veškerá interpunkce.

## 5.8 Další výrazy

Jak ukazuje korpusový výzkum, v PDT 2.0 byly funktorem PREC označeny i některé výrazy, respektive spojení formálně netypická pro tuto kategorii. První takovou skupinu tvoří předložkové fráze, jejichž řídicí substantivum bylo označeno funktorem PREC. Zde je označujeme jako „nominální“ z hlediska jejich řídicího uzlu a v porovnání se skupinou druhou. Skupinu druhou tvoří pronominální spojení, převážně předložková, jejichž řídicí člen s funktorem PREC tvoří t-lemma *ten*. Obě tyto skupiny mají společný rys jisté ustálenosti v užívání.

První skupinu tvoří spojení, jejichž funkce strukturovat text je nesporná: spojení typu *na jedné straně/na druhé straně* jsou spojeními uvozující určité argumentování, jde zpravidla o vztah kontrastu. Spojení *jinými slovy* je spojením apozičním, v popisu SDRT tento vztah byl nazván vztahem *restatement* či *similarity*. Spojení *v první/naposlední řadě* a *k dovršení* kompozičními prostředky ve výstavbě textu, signalizují začátek či konec a mají také funkci zdůrazňovací. Je patrné, že tato spojení nereprezentují konektivní vztahy primárně syntaktického charakteru, nýbrž že se jedná o prostředky vyjádření vztahů kompozičních, čili rétorických, jak je popisujeme v kapitole 3. Tato spojení ovšem v PDT nejsou zachycena jednotně, často jim je přisuzován význam modifikace výpovědi hodnocením či postojem mluvčího, někdy se hodnotí jako adverbialní doplnění. Také je jisté, že podobně fungujících obrátů lze nalézt více. Ustálenými spojeními, která především plní funkci strukturace textu, tak mohou být i rozsáhlejší obraty jako například: *z toho plyne, přesněji řečeno, smím-li to tak říci, jak se říká, pokud se týká* apod. Těmito obraty se skupina prostředků strukturace textu značně rozšiřuje, jsou také funkčně různorodé. Jde již spíše o metatextové komentáře (srov. Hoffmannová 1984:131) a mají tedy charakter parentezí. Těmito prostředkům vyjádření textové koherence se však v této práci blíže nevěnujeme.

spojení	výskyty v PDT 2.0			anglický ekvivalent dle PDTB	poznámka
	celkem	PREC	jiné		
k dovršení	2	1	COND 1		
jinými slovy	4	1	#EmpVerb +COND, MANN, MEANS	<i>in other words</i>	pokaždé jiná interpretace!
na jedné straně/na jednu stranu	21	1	REG, ATT, LOC	<i>one one hand</i>	
na druhé straně/na druhou stranu	81	15	LOC, REG, ATT	<i>on the other hand</i>	stále stejný význam, ale rozdíly v anotaci – nejspíše chybně
v první/prvé řadě	13	1	REG 5, ATT, LOC, TWHEN, CM, RHEM		velká nekonzistence anotace
v poslední řadě	19	2	TWHEN, LOC, CM, COND, REG, ATT		velká nekonzistence anotace

**Tabulka 4** Nominální víceslovné PREC (se závislým uzlem)

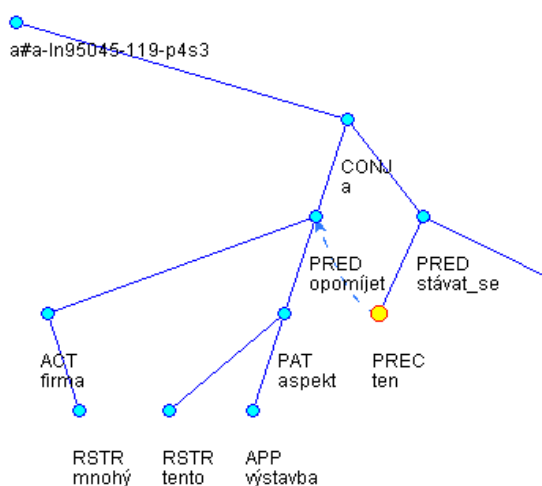
Druhou skupinu tvoří spojení s uzlem s t-lematem *ten* označená funktorem PREC. Ukazovací zájmena obecně přispívají k textové koherenci jako prostředek vyjádření vztahů textové koreference, neboli jejich základní funkcí v diskurzu je odkazovat k plnovýznamovým jednotkám.

V PDT je uzlů s t-lematem *ten* a funktorem PREC 33 (viz tabulka 5). Nejčastěji se zájmeno *ten* s funktorem PREC vyskytuje v konstrukcích a *to na* začátku věty, tyto případy jsou v anotaci hodnoceny jako dva samostatné uzly s dvěma funktoři PREC. U zájmeně *to* s funktorem PREC v těchto spojeních není značena textová koreference. Zpravidla se jedná o věty s dodatečně připojeným doplněním, tedy s aktuálně elidovaným řídicím slovesem. Pokud tato konstrukce spojuje dvě klauze v rámci jedné věty, jedná se o jeden uzel s t-lematem *a to*, který je zpravidla kořenem souřadné struktury. Obě tyto konstrukce zachycují apoziční vztah mezi spojovanými/připojovanými klauzemi.

(119) *Zveřejnění v kratší lhůtě je možné ve výjimečných případech pouze po předchozí domluvě, a to APPS s přírážkou 50 %.* PDT

(120) *Jsou bohatí, ale tisíce Brazilců musí hladovět. A PREC to PREC v zemi, která má spoustu přírodního bohatství.* PDT

Variantou t-lematu *a\_to* je forma *a tím*. Toto spojení má pak jakožto kořen souřadné struktury v anotaci funktor pro důsledkový vztah CSQ. Několikrát je však stejný případ zachycen pomocí dvou uzlů, kdy *a* je kořen souřadné struktury s funktořem CONJ a *tím* má funktor PREC a visí o dvě úrovně níže, u slovesného potomku kořene souřadné struktury.



Mnohé firmy tento aspekt výstavby opomíjejí a tím se stává, že... PDT

### Obr. 22 Zachycení spojení *a tím* dvěma uzly

V souladu s postupem lematizace zmíněných forem jednotnými t-lematy pro kořeny souřadných struktur navrhujeme hodnotit i spojení *a to*, *a tím*, která připojují přes hranice věty, jako diskurzni konektory s jedním uzlem a lematem *a\_to*. Tento postup je ostatně navržen i pro další víceslovné spojovací výrazy typu *a proto*, *a tedy*, *a tak*. To se jeví účelné i v případně připojovacích konektorů.

Spojení *podle toho* a *místo toho*, která jsou českými překlady anglických diskurzni konektorů *accordingly* a *instead*, se v PDT s funktořem PREC nevyskytují – mají vždy funktor dle významu: CRIT (měřítka) a SUBS (substituce). Zájmeno *ten* v těchto případech zpravidla odkazuje (vztahem textové koreference) ke konkrétnímu výrazu (uzlu) v předchozí větě. Adverbiální doplnění vyjadřující měřítka a substituci podle nás neimplikují diskurzni vztah, a tak je textové navazování u těchto výrazů pouze koreferenční.



Výrazy *k tomu* a *naproti tomu* však mají sémantické rysy gradace a kontrastu. Odpovídají tak například konektorům *navíc* a *na druhou stranu, avšak, však*. Ze sémantického hlediska tedy tato spojení mají určitou konektivní funkci, otázkou zůstává, zda hodnotit přítomnost vztahu textové koreference a tedy deiktickou povahu těchto výrazů jako dvojí roli jednoho prostředku při vyjadřování koherence textu: jednak vztah textové koherence a jednak vztah konektivní. V zásadě nevidíme důvod, proč by takováto deiktická spojení, pokud jsou jasně nahraditelná nějakým základním konektorem, nemohla tuto dvojí funkci plnit. Srov. příklady (121) - (123):

výraz/spojení	jako PREC v PDT	anglický ekvivalent	značení textové koreference	poznámka
podle toho	0	<i>accordingly</i>		angl. výraz je PREC
místo toho	0	<i>instead</i>		angl. výraz je PREC
a to	15	<i>namely</i>	-	<i>a</i> PREC <i>to</i> PREC v iniciální pozici, jinak zpravidla <i>a_to</i> APPS
k tomu	4	<i>moreover</i>	++- -	<i>ještě</i> RHEM <i>k tomu</i> PREC
naproti tomu	2	<i>by comparison</i> <i>by contrast</i> <i>on the other</i> <i>hand</i>	+ -	
tím	6	<i>thereby, so,</i> <i>(hereby)</i>	+	ve smyslu <i>tak</i> , hraničí s adverbialními funktory 106x MEANS
a tím	5	<i>thereby, so,</i> <i>(hereby)</i>	-	a CONJ <i>tím</i> PREC
a to tím	1			špatná anotace – chybí 1 uzel: a PREC <i>ten</i> PREC

**Tabulka 5** Pronominální víceslovné PREC

(121) Každému sděluji, že člen dnešní vlády má pouze svůj plat, k tomu k dispozici auto a telefon. PDT

(122) Jeho to uklidňuje. Mě to naproti tomu rozčiluje.

(123) Banka ve zvýšené míře uspokojovala poptávku svých klientů po úvěrech ve volně směnitelných měnách, zejména v amerických dolarech, naproti tomu zaznamenala menší zájem o účelové zboží linky pro dovoz ze zahraničí. PDT

## 5.9 Shrnutí

Provedená analýza výrazů s funktorem PREC a výrazů s nimi hraničících v Pražském závislostním korpusu vedla k následujícím zjištěním o kategorii diskurzních konektorů jakožto relátorů vztahů mezi jednotkami v diskurzu:

Diskurzní konektory jsou výrazy, které:

- 1) **spojují** propozice vyjádřené slovesnými klauzemi v **rámci jedné věty**; jde především o spojky nebo víceslovné spojovací výrazy, a to jak souřadící, tak podřadící, řídčeji pouhé výrazy adverbialní. Víceslovné spojovací výrazy jsou zpravidla modifikací základní spojky a posouvají tak její sémantiku, např. *a proto, a tedy, nejen ale i, a přesto* (viz kap. 5.3);

nebo

- 2) **připojují** propozice vyjádřené slovesnými klauzemi **přes hranice věty**; jde především o adverbialní a výrazy. Nejčastějšími z nich jsou však spojky *a, ale* a *však*, u spojky *však* byl zjištěn výrazně vyšší výskyt ve funkci připojovacího konektoru než jako junktoru (viz tabulka 1). I připojovací konektory se často kombinují v jedné větě<sup>52</sup>.

Sémantická kategorie PREC reprezentuje jádro druhého typu diskurzních konektorů. Navazovací/připojovací konektory se od spojovacích liší kromě slovnědruhové reprezentace ještě v několika ohledech:

- Co se týče směru navazování, jsou na rozdíl od spojovacích konektorů připojovací konektory až na výjimky anaforické, tj. odkazují v textu směrem k již řečenému. (viz kap. 5.2.6.4).
- na rozdíl od spojovacích konektorů nemohou připojovací explicitně vyjadřovat závislostní vztahy, jinými slovy, závislé klauze se nespojují se svými řídícími klauzemi přes hranice věty. Pokud se tak stane, je závislá klauze parcelátem, stojí nesamostatně a použitý konektor patří do první skupiny diskurzních konektorů – spojovacích (viz 4.1).
- Připojovací konektory se vztahují k předchozímu kontextu vztahy analogickými k syntaktickým vztahům uvnitř věty, jako je například slučování, gradace,

---

<sup>52</sup> Kombinacemi konektorů připojovacích typu *Ale přece také proto...; Tak tedy přece jen...* jsme se ovšem v této práci podrobněji nezabývali a odkazujeme pouze k pracím Hoffmannové (1983) a Adamce (1995).

adverzativní vztah atd. V rámci připojovacích konektorů však můžeme rozlišit několik okrajových skupin, jež vykazují **kromě syntaktického druhu připojení určitý vedlejší význam**: postoj mluvčího (*vždyť, přece*), či zdůraznění či omezení (*také, jen*). U některých z okrajových typů jsme vzhledem k tomuto druhému významu nebyli schopni na základě vykonaného výzkumu vždy jednoznačně určit, zda se jedná o konektory či nikoli.

- Připojovací konektory mají na rozdíl od spojovacích konektorů (vázaných na jednu větu) možnost vztahovat se jak k bezprostředně předchozímu, tak k distantnímu úseku textu, a co se týče rozsahu, mohou se vztahovat jak ke klauzi, tak k větě, či i k několika větám (viz kap. 5.2.5)

Připojovací konektory se vztahují k předchozímu kontextu především vztahy analogickými k syntaktických vztahům uvnitř věty, jako je například slučování, gradace, adverzativní vztah atd. V rámci připojovacích konektorů však můžeme zřetelněji rozlišit několik okrajových skupin, jež vykazují **kromě syntaktického druhu připojení určitý vedlejší význam**: postoj mluvčího (*vždyť, přece*), či zdůraznění či omezení (*také, jen*). U některých z okrajových typů jsme vzhledem k tomuto druhému významu nebyli schopni na základě naší analýzy vždy jednoznačně určit, zda se jedná o konektory či nikoli.

Dále jsme zjistili, že některé konektory **nevyjadřují žádný ze syntaktických vztahů popsaných na tektogramatické rovině PDT**, ale spíše **vyjadřují vztah klauze/věty**, ve které se vyskytují, k **celkové kompozici textu**. Jsou to zpravidla víceslovné výrazy ustáleného charakteru, které se podílejí na výstavbě textu nejrůznějšími elaboračními vztahy, uvedením do protikladu, jinou formulací téhož, členěním textu na začátek, prostředek a konec apod. (*na jedné/druhé straně, v neposlední řadě, jinými slovy, k dovršení*). (viz kap. 3 a 5.8). Některé z těchto výrazů mají časový příznak a vyjadřují tak vnitřnětextovou temporální charakteristiku (*nejprve, pak, nakonec*) (kap. 5.6). Jako diskurzívní konektory spojovací či připojovací hodnotíme v některých případech i interpunkční znaménka (viz kap. 5.7) a jisté ustálené deiktické výrazy (viz kap. 5.8).

## Kapitola 6. Závěr

Tektogramatické struktury v Pražském závislostním korpusu vycházejí z propracované teorie závislostního hloubkově syntaktického popisu jazyka ve Funkčním generativním popisu (FGP). Teoretický rámec FGP se také věnuje zkoumání různých aspektů diskurzu, z nichž některé jsou zahrnuty do tektogramatické struktury. Konektivním vztahům v diskurzu však bylo v FGP a v české lingvistice vůbec zatím věnováno málo pozornosti. Také proto nový filadelfský projekt diskurzního korpusu vybízí jak k teoretickému zamyšlení nad diskurzem, tak ke zkoumání empirickému.<sup>53</sup>

Tato studie přispívá ke zkoumání diskurzu dvěma způsoby. Nejprve jsme v ní ujasnili pojetí diskurzu a diskurzních vztahů v návaznosti na již dosažené cíle FGP. Diskurzem pro budoucí práci na tomto poli rozumíme užívání jazyka jako systému v procesu komunikace, diskurz je provázanou sítí syntaktických, sémantických a pragmatických vztahů.

Poté jsme na základě filadelfského diskurzního korpusu a zejména anotace anglických *discourse connectives* zjišťovali, které z konektivních vztahů v diskurzu a v jaké míře zachycují i tektogramatické struktury v Pražském závislostním korpusu. Jak shrnujeme v kapitole 4, jsou to některé vztahy syntaktické závislosti, vztahy syntaktické souřadnosti kromě souřadného spojování členského a vztahy připojovací vyjádřené v této struktuře zejména (ale nejen) výrazy s funktorem PREC.

Hlavní část této studie jsme zaměřili zachycení posledně jmenovaných konektivních prostředků připojovacích v tektogramatické rovině PDT. Jejich analýzou jsme získali několik nových poznatků o povaze a fungování diskurzních konektorů v češtině. Zjistili jsme, že se diskurzní konektory připojovací (tj. převážně s funktorem PREC, navazující na obsah v předchozí větě) od konektorů spojovacích (v rámci věty) v několika ohledech liší, např. slovnědruhovou reprezentací, rozsahem a velikostí textových úseků, které spojují či připojují, schopností navazovat na předchozí kontext prostřednictvím závislé klauze. Dále jsme zjistili, že konektivní funkci plní i výrazy, které nevyjadřují žádný ze syntaktických vztahů popsaných funktory v PDT, tyto vztahy se týkají kompozice textu a nazýváme je rétorické. Některé výrazy vyjadřující v textu určité zdůraznění, postoj mluvčího k obsahu propozice apod., a jejich význam v textu tedy primárně není konektivní, dle našeho názoru konektivní

---

53 Účelem této práce ovšem nebylo porovnávat konkrétní způsoby zachycení vztahů v diskurzu ani zkoumat komparativně pojetí kategorie diskurzních konektorů v češtině a v angličtině.

funkci plnit mohou v závislosti na jejich konkrétním užití v daném kontextu.

Přestože výzkum provedený v této studii otevírá mnoho dalších otázek ke vztahu syntaxe a diskurzu, věříme, že se nám k této problematice podařilo přispět alespoň dílčími odpověďmi.

## Kapitola 7. Otázky, problémy, budoucí kroky

Do budoucna je plánováno v rámci mezinárodního projektu, který se týká popisu diskurzu v typově odlišných jazycích a do kterého je zapojena také Karlova univerzita a University of Pennsylvania, využít tektogramatické anotace češtiny v PDT 2.0 k návrhu nové, diskurzívní „úrovně“ popisu. Ačkoli jsme ukázali, že tektogramatická rovina již některé diskurzívní vztahy zachycuje, přináší s sebou vznik nového anotačního schématu mnoho otázek a problémů, z nichž některé naznačila i tato práce.

Mezi tyto otázky patří například vymezení modifikovaných diskurzívních konektorů: z důvodu jednotné lematizace spojkových i připojovacích víceslovných výrazů je třeba posoudit, kdy se jedná o jeden modifikovaný výraz a kdy o dva samostatné výrazy stojící v povrchové stavbě věty vedle sebe. (*a proto x dále pak*) V prvním případě by měla být provedena jednotná lematizace těchto výrazů pod jedním uzlem. Seznam modifikovaných spojovacích výrazů v PDT je připojen jako příloha k manuálu pro tektogramatickou anotaci (Mikulová et al. 2005: 1153). Obsahuje však pouze souřadné spojovací výrazy modifikované uzly s funktoem CM, modifikované připojovací konektory se zachycují jako uzly s funktoem PREC, takže ještě zpracovány a připraveny pro jednotnou lematizaci nejsou.

Dále je problematické vymezení diskurzívních a nediskurzívních vztahů, zejména hodnocení obsahových klauzí a nepravých závislých klauzí jako diskurzívních. Pokud bychom totiž nezachycovali velmi časté konstrukce se slovesem mluvení typu *Oznámil, že...* jako diskurzívní vztahy, vznikla by poměrně velká skupina nespojitých zanořených diskurzívních vztahů v rámci souřadně (i jinak, než pouhým slučováním) spojených závislých klauzí obsahových. Druhým problémem vymezení argumentů je jejich rozsah. Jak jsme již zmínili v kapitole 5.2.5, problém činí mnoho druhů kondenzace propozice. Nejedná se pouze o hodnocení deverbativních substantiv na *-ní, -tí*, ale prakticky o všechny typy nominalizací (Po příchodu domů si dal horký čaj pro zahřátí.) a dále o hodnocení souřadnosti smíšené a spojování neslovesných klauzí.

Jak jsme již naznačili v teoretické části studie, v publicisticky zaměřeném PDT se setkáváme pouze s texty psanými, v drtivé většině monologickými (repertoár navazovacích prostředků se může v monologickém a dialogickém kontextu výrazně lišit (srov. např. Adamec 1995: 60), s texty přizpůsobenými specifickým pravidlům publicistického stylu (např. enormní výskyt trpných participií je dán oblibou je používat v nadpisech) v neposlední

řadě se specifickou slovní zásobou. Pro budoucí práci na poli diskurzu bude podnětné rozšířit zkoumaný jazykový materiál o další stylistické úrovně, a tím otevřít prostor pro zkoumání užívání jazyka v jeho stylové rozmanitosti.



## Summary

The present thesis is a contribution to the widely discussed issue of how the syntactic structure of a sentence and the structure of discourse (text) are related. The syntactic sentence structure along with other language phenomena participates in building a coherent, comprehensible discourse. The author calls the syntactically motivated relations in discourse connective relations. These relations include coordinating relations and some of the subordinating relations within a sentence and, secondly, adjoining of discourse units across the sentence boundary. The explicit means of expressing connective relations are called discourse connectives. It is a group of language expressions that connect or adjoin discourse units while indicating the type of semantic relation between them, i. e. conjunctions, some subjunctions, particles and adverbials, and marginally also some other parts-of-speech. The present thesis describes the semantic category of discourse connectives in Czech on the basis of language data and their syntactic annotation in the Prague Dependency Treebank, and thus aims to contribute to the design of a language corpus annotation scenario capturing the discourse relations in Czech.

## Přílohy

### Příloha 1: Seznam explicitních *discourse connectives* v PDTB 1.0

Explicit Connective	No.	Explicit Connective	No.
accordingly	6	conversely	2
additionally	7	earlier	15
after	577	either..or	4
afterward	11	else	1
also	1748	except	10
alternatively	6	finally	32
although	328	for	3
and	3003	for example	197
as	743	for instance	98
as a result	78	further	9
as an alternative	2	furthermore	11
as if	16	hence	4
as long as	24	however	487
as soon as	20	if	1234
as though	6	if and when	3
as well	6	if..then	38
because	860	in addition	165
before	326	in contrast	12
before and after	1	in fact	83
besides	19	in other words	17
but	3312	in particular	15
by comparison	11	in short	4
by contrast	27	in sum	2
by then	8	in the end	10
consequently	10	in turn	30

Explicit Connexive	No.
indeed	104
insofar as	2
instead	112
later	92
lest	2
likewise	8
meantime	15
meanwhile	193
moreover	101
much as	9
neither..nor	3
nevertheless	44
next	7
nonetheless	27
nor	31
now that	22
on the contrary	4
on the one hand..on the other hand	1
on the other hand	37
once	85
or	98
otherwise	24
overall	12
plus	1
previously	49

Explicit Connective	No.
rather	17
regardless	2
separately	74
similarly	18
simultaneously	6
since	184
so	263
so that	31
specifically	10
still	197
then	340
thereafter	11
thereby	12
therefore	26
though	320
thus	113
till	3
ultimately	18
unless	95
until	162
when	990
when and if	1
whereas	5
while	782
yet	101

## Příloha 2 : A. Seznam anglických výrazů s funktorem PREC pro anotaci PEDT

Tabulka platných PREC přítomných v textech Wall Street Journalu (anotace PEDT)	
accordingly	
additionally	
afterward(s)	
also	PREC x RHEM
alternatively	<i>We could drive all the way. <u>Alternatively</u>, we could fly.</i>
and	
as a result	
as an alternative	
as well	ale nikoliv <i>as well as!</i>
besides	
but	
by comparison	
by contrast	
by then	PREC x TTILL <i><u>By then</u> [1991] TTILL, it also hopes to begin producing...</i>
consequently	
conversely	
finally	PREC x TWHEN
further	
furthermore	
hence	
however	
in addition	
in contrast	
in other words	
in short	(= briefly, in brief) PREC x MANN
in sum	(= briefly, in brief) PREC x MANN
in the end	
in turn	
instead	<i>The committee has rejected our proposal. <u>Instead</u>.PREC, they have brought forward an alternative plan.</i> - ale nikoli <i>instead of!</i>

later	PREC x TWHEN
likewise	
(in the) meantime	PREC x TPAR
meanwhile	PREC x TPAR
moreover	
next	
nonetheless	
nor	<i><u>Nor</u> is the group likely to come forward with a revised offer...</i>
on the contrary	PREC x REG
on the one hand... on the other hand	PREC x REG
on the other hand	PREC x REG
or	
otherwise	
overall	
plus	pouze 1 výskyt jako PREC: <i><u>Plus</u>, one Exxon official said...</i>
previously	častěji TWHEN
rather	PREC x ATT? <i>His purpose was not so much to attack his rivals. <u>Rather</u>, it was to defend his own position.</i>
regardless	Pouze 2 výskyty: PREC x REG? <i>"I think the Soviet peoples are more introverted than the Chinese." <u>Regardless</u>, he said, he still believes the Soviet Union, while falling far short of the efficiency of a Western economy, may well manage to improve considerably. After four years of no progress in this direction, it is doubtful any viable collaboration with Ms. Bartlett will suddenly now be possible. (Mr. Quennell has said he plans to go with the grid, <u>regardless</u>.)</i>
separately	
similarly	PREC x MANN
simultaneously	
so	
specifically	PREC x MANN
still	
then	
thereby	
therefore	
though	
thus	

ultimately	PREC x TWHEN, RHEM, ATT
yet	ve smyslu <i>přesto</i>

## B. *Discourse connectives* z PDTB bez funkce PREC v anotaci PEDT

after TWHEN či jiné časové funktořy  
 although CNCS  
 as TWHEN  
 as if  
 as long as COND, THL, TTILL  
 as soon as TWHEN  
 as though CPR, MANN  
 after and before  
 at last ATT  
 at least RHEM  
 because  
 before (adverbium i předložka) TWHEN či jiné časové funktořy  
 earlier  
 either...or CM + CONJ  
 else (pouze 1 výskyt, funkce souřadicí spojky – CONJ)  
 for  
 for example RHEM  
 for instance RHEM  
 if  
 if and when  
 if...then – pouze *then* má funktoř PREC  
 in fact ATT  
 in particular RHEM  
 indeed ATT  
 insofar as EXT  
 lest CAUS (pouze 2 výskytů)  
 much as  
 neither ...nor CM + CONJ  
 now that  
 of course ATT  
 once  
 since (adverbium i předložka TFRWH)  
 so that RESL  
 thereafter  
 till TTILL  
 until TTILL  
 when  
 when and if  
 whereas  
 while  
 unless

### Příloha 3: Seznam českých výrazů s funktorem PREC v PDT 2.0

Základní informace:

6127 výskytů v 5830 stromech

několikanásobné (typu „dále pak“, „ale přesto“): 382 ve 189 stromech

rozvité (mají závislý uzel): 21

výraz	výskyt v PDT 2.0				
	celkem	PRE C	jiné		
a	17 308	531 (526+ 5)	16 777	<i>and</i>	homonymie t-lematu se „seznamovými“ a)
ale	2251	518	1733	<i>but</i>	
aniž	58	1	CONJ, ADVS, GRAD		
avšak	76	25	ADVS 49, GRAD 2	<i>but, yet</i>	
ba	15	2	GRAD 12, ATT 1		PREC x ATT v iniciální pozici
bud'	2	1	MANN	<i>either</i>	specifické případy
čili	25	5	APPS 19, CSQ 1	<i>or, thus</i>	
dále	320	109	THL 111, CM 17, DIR3 10, TWHEN 7, ACT 2, MANN 2, TFHL 1	<i>further (hereafter)</i>	ACT: a (CONJ) <i>tak</i> (DPHR) <i>dále</i> (ACT)
i	4023	11 (10+1 )	RHEM 2653, CONJ 951, CM 396	<i>also, as well, too</i>	primárně RHEM, homonymie t-lematu se „seznamovými“ i)
jenomže	4	4		<i>but</i>	
jenže	85	60	ADVS 25	<i>but, only, except?</i>	
jinak	152	14	MANN 122 16: REG COND DISJ, ATT, PAT	<i>otherwise, or (else)</i>	
konkrétně	8	3	CM 5	<i>specifically</i>	
leč	15	3	ADVS 12	<i>though?</i>	převážně koordinace NP, jako PREC iniciální pozice
nadto	3	3			

naopak	239	189	MANN 30, CM 13, RHEM 5, DIR3 2	<i>in contrast, on the contrary</i>	
natož	14	1	GRAD 13		
navíc	260	226	CM 11, RSTR 9, MANN 4, ATT 2	<i>furthermore, besides, moreover, in addition</i>	
(a)nebo	1004	35	DISJ 813, CONJ 114, GRAD 37, CONFR 3	<i>or</i>	<i>nebo</i> naopak CONFR
nebot'	225	3	REAS 222	<i>for, because</i>	anglické ekvivalenty jsou podřadící spojky!
nejenže	11	1	CM 10	<i>not only</i>	
nicméně	82	59	ADVS 23	<i>however, nevertheless</i>	
nuže	4	1	PARTL 2, ATT 1	<i>so?</i>	
ovšem	373	274	ADVS 67, ATT 30, CONJ 1, RHEM 1	<i>of course (!), nevertheless</i>	jako afirmační částice – ATT
pak	178	160	CM 18	<i>then, thereafter, afterward(s)</i>	pokud..., pak
podobně	5	3	CM 2	<i>similarly</i>	
potom	534	41	TWHEN 492, CM 1	<i>then, so, thus, hence</i>	jestliže..., potom; kdyby..., potom
proto	482	334	CM 131, CAUS 17	<i>therefore, accordingly</i>	neostrá hranice s příčinou
přesto	155	107	CNCS 26, CM 22	<i>still, yet, regardless</i>	neostrá hranice s přípustkou
přičemž	92	5	87 CONJ		dle PDT manuálu souřadící výraz
přitom	253	220	TWHEN 26, CM 7		
sice	281	9	CM 272	<i>or else</i>	
současně	25	16	CM 9	<i>simultaneously</i>	
spíše/ nejspíš	311	3	RHEM, EXT, MANN		
stejně	69	29	CM 35, ATT 5	<i>(accordingly), anyway</i>	
tak	1186	350	MANN 334, EXT	<i>so, thus</i>	- neostrá hranice



			267, CM 134, PAT 33, ACT 29, RHEM 1, MEANS 1		s doplněním způsobu - když..., tak; protože..., tak
také/taky	1063	20	878 RHEM, 165 CM	<i>also, as well, too</i>	
takže	157	31	CSQ 126	<i>so</i>	
teda	1	1			
tedy	576	346	CM 229, CAUS 1	<i>so, then, hence...</i>	
těž	48	1	RHEM 37, CM 10	<i>also, as well, too</i>	
tj	1	1			nesprávná lematizace, správně: <i>to jest</i> (APPS)
totiž	514	493	APPS 14, REAS 6, CM 1		
tudíž	44	14	CM 30	<i>then, so, hence...</i>	
však	1782	1482	ADVS 296, GRAD 4	<i>however, nevertheless</i>	
vždyť	54	32	REAS 15, ATT 7	<i>after all?</i>	
zároveň	240	61	TWHEN 105, CM 38, RHEM 34, TPAR 1, MANN 1	<i>simultaneously</i>	
zas	6	2	TWHEN 2, CM 2		
zase	125	65	TWHEN 53, RHEM 5, CM 2	<i>to the contrary?</i>	
zato	46	15	CM 31	<i>but, yet</i>	
<b>Seznamy</b>	<b>celkem</b>	<b>PRE C</b>	<b>jiné</b>		
čísla 1 - 13		89			
písmena a - j		32			
za první	6	3	CM 3		
za druhé	10	5	CM 5		
za třetí	1	0	CM 1		
za čtvrté	1	1			

Pozn.: V tabulce jsou u funktořů pro souřadná spojení zahrnuta i spojení nevětná!

## Příloha 4: Seznam českých výrazů s funktorem CM v PDT 2.0

(Marie Mikulová 2005)

### **restriktory**

*alespoň*  
*aspoň*  
*hlavně*  
*jen*  
*jenom*  
*maximálně*  
*nejméně*  
*nejen*  
*nejenom*  
*pouze*  
*právě*  
*především*  
*převážně*  
*přímo*  
*přinejmenším*  
*spíše*  
*takřka*  
*třeba*  
*třebas*  
*většinou*  
*vlastně*  
*vůbec*  
*výhradně*  
*vylučně*  
*vysloveně*  
*zejména*  
*zvlášť*  
*zvláště*

### **OSTATNÍ**

#### **typické CONJ**

*podobně*  
*stejně*

#### **typické CSQ**

*potažmo*  
*proto*  
*tudíž*  
*tedy*  
*tak*

### **kontextualizátory**

*dále*  
*i*  
*ještě*  
*k tomu*  
*pak*  
*potom*  
*rovněž*  
*s tím*  
*také*  
*taktéž*  
*těž*  
*zároveň*

### **negační**

*ani*  
*ne*  
*nikoli*  
*nikoliv*

#Neg;

**typické ADVS***přece**přesto**sice**zato*

+negace

+restriktory

**typické GRAD***dokonce**nadto**navíc**neřku-li*

+negace

+restriktory

**typické DISJ***eventuálně**popřípadě**případně**už*

+negace

+restriktory

**typické CONFR***naopak*

+negace

+restriktory

**typické APPS***jmenovitě**konkrétně**například**přesněji**zkrátka*

+restriktory

**ostatní***jednak**respektive**za\_druhé*

## Příloha 5: Seznam příslovčí s funktoem ATT

(dle seznamu příslovčí v PDT, Eva Buráňová 2001)

beztak (b. se tam nevejdeme)	
bohužel (je to b. tak)	
díkybohu (dopadlo to d. dobře)	
doufejme (dopadne to d. dobře)	(bez čárek, jinak PAR)
doslova (bylo to d. trapně)	(také: MANN – přeložit d.)
evidentně (vliv byl e. podceněn)	
fakticky (je to f. zlé)	(také: MANN – zjišťovat fakticky)
hypoteticky (h. je to pravda)	
jaksi (j. to neberou na vědomí)	
jasně (j., že ano – hovor.)	(také: MANN – měsíc svítí j.)
jednoduše (j. nepůjdu)	(také: MANN – ztvárnit j.)
jednoznačně (je to j. pravda)	(také: MANN- vyjádřete se j.)
jen, jenom (j. se opovaž, j. se neboj)	(také: RHEM – myslí j. na sebe)
klidně (já tam k. půjdu s vámi)	(také: MANN – kůň stál k.)
koneckonců (je to k. jedno)	
konečně (=koneckonců- je to k. jedno)	(také: TWHEN – k. přišel)
mimočodem (je mi to m. jedno; obyčejně PAR)	(také: MANN – m. se zeptat).
náhodou (je to n. hezké)	(také: CAUS – setkat se n.)
nakonec (je to n. jedno)	(také: TWHEN – dostavili se až n.)
naštěstí (n. se to podařilo)	
neoddiskutovatelně (je to n. pravda)	
ostatně (do toho se o. nebudu plést)	
ovšem (o. že přijdu)	(také: PREC – vzpomínám si o.,...)
pochopitelně (p. přijdu)	
pomalů (myslím, že p. půjdu)	(také: MANN – chodit p.)
prakticky (p. jsme prohráli)	(také: MANN - p. vyzkoušet, zařít p.)
právě (to se mi p. nepovedlo)	(také: PREC – p. se chystám ven; RHEM – proč to dal p. jí?)
proboha (nelekej se p.; obyčejně PAR.)	
prostě (nemohu to p. pochopit)	(také: MANN- oblékal se p.)
přece (na tom p. nezáleží)	(také: spojka - a_přece)
přirozeně (to sem p. nepatří)	(také: MANN – chovat se p.)
raději (jména r. neuvádíme)	
samozřejmě (to sem s. nepatří)	(také: MANN – chovat se s.)
takzvaně (je to t. nemorální)	
vážně (v. se to stalo)	(také: MANN- přístupuje ke všemu v.)
vlastně (co se v. stalo?)	
vůbec (jak si to v. představujete?)	(také: RHEM – nechce v. jít)
vyloženě (to mě v. rozčiluje)	(také: RHEM – je to v. nesmysl)
vysloveně (to mi v. jde na nervy)	(také: RHEM – byl to v. uličník)
vždyť (v. je to ostuda)	(také: spojka REAS)
ztraceně (obvykle PAR – z., nevyšlo to)	(také: EXT – z. hezký)
zkrátka (je to z. problém)	(také: MANN – projednáme to z.)
zpropadeně viz ztraceně	
zrovna (to se mi z. nepovedlo)	(také: PREC – z. se chystám ven; RHEM – proč to dal z. jí?)

## Seznam užívaných funktorů

### Funktory pro aktanty

ACT – aktor, původce děje  
ADDR – adresát, příjemce  
EFF – cíl, efekt  
ORIG – zdroj, původ  
PAT – patiens, předmět zasažený dějem

### Funktory pro koordinaci

ADVS – odporovací (adverzativní) poměr  
CONFR – konfrontační poměr  
CONJ – slučování  
CONTRA – kontrast  
CSQ – důsledkový poměr  
DISJ – vylučovací poměr  
GRAD – stupňování  
REAS – důvod, příčina  
APPS - apozice

### Funktory pro efektivní kořeny nezávislých klauzí

PRED – predikát, kořen slovesné klauze  
DENOM – kořen nominativní klauze  
VOCAT – kořen vokativní klauze  
PARTL – kořen citoslovečné klauze

### Funktory pro rematizátory, větné, navazovací, modální adverbialní výrazy a výrazy modifikující význam souřadících spojovacích výrazů

ATT – postoj mluvčího k obsahu výpovědi (viz kap. 5.5)  
CM – modifikátor souřadícího spojovacího výrazu (viz kap. 5.3)  
MOD – modální charakteristika obsahu výpovědi  
**PREC – navazování na předchozí kontext**  
RHEM – rematizátor (viz kap. 5.4)

### Časové funktory

TFHL – „na jak dlouho“  
THL – „jak dlouho, za jak dlouho“  
THO – „jak často“

TFRWH – „ze kdy“  
TOWH – „na kdy“  
TPAR – „během jaké doby, souběžně s čím“  
TSIN – „odkdy“  
TTILL – „dokdy“  
TWHEN – „kdy“

#### Funktory místa a směru

DIR1 – „odkud“  
DIR2 – „kudy“  
DIR3 – „kam“  
LOC – „kde“

#### Funktory pro implikační (kauzální) vztahy

AIM – účel  
CAUS – příčina, důvod  
CNCS – přípustka  
COND – podmínka

#### Funktory pro vyjádření způsobu a jeho specifických variant

ACMP – doprovod, průvodní okolnost  
CPR – srovnání  
CRIT – norma, měřítko  
EXT – míra  
MANN – způsob „jak“  
MEANS – prostředek  
REG – zřetel  
RESL – účinek  
RESTR – omezení nebo výjimečné slučování

#### Další specifické funktoary

COMPL – doplněk  
CONTRD – volné doplnění s konfrontačním významem  
SUBS – substituce  
DPHR – závislá část frazému

## Seznam zkratk

AO – abstract object, abstraktní objekt (Asher 1993)

ČNK – Český národní korpus

FGP – Funkční generativní popis

MČ – Mluvnice češtiny

MDC – Maximizing Discourse Coherence

PCEDT – Pražský česko-anglický závislostní korpus (Prague Czech-English Dependency Treebank)

PDT – Pražský závislostní korpus (Prague Dependency Treebank)

PDTB – Penn Discourse TreeBank

PEDT – Pražský česko-anglický závislostní korpus (Prague Czech-English Dependency Treebank), anglický subkorpus

PMČ – Příruční mluvnice češtiny

PTB – Penn TreeBank

RST – Rhetorical Structure Theory

SDRT – Segmented Discourse Representaion Theory

TR – tektogramatická reprezentace

TrEd – Tree Editor, vyhledávací a editační nástroj pro PDT 2.0

VF – verbum finitum, určitý tvar slovesa

WSJ – Wall Street Journal

## Literatura a prameny

- Kolektiv autorů (1986): *Mluvnice češtiny 2*, Praha: Academia
- Kolektiv autorů (1987): *Mluvnice češtiny 3*, Praha: Academia
- Kolektiv autorů (2000): *Příruční mluvnice češtiny*, Praha: Nakladatelství Lidové noviny
- Kolektiv autorů (2006): *Macmillan English Dictionary for Advanced Learners. International Student Edition*, Oxford: Macmillan Publishers Limited
- Adamec, P. (1995): Konektivní částice a jiné textově propojovací výrazy v současné češtině, In *Přednášky z 37. a 38. běhu LŠSS*, Praha: Univerzita Karlova, s. 59-64
- Asher, N. (1993): *Reference to Abstract Objects in Discourse*, Kluwer Academic Publishers
- Asher, N. – Lascarides, A. (2003): *Logics of Conversation*, Cambridge University Press
- Bajzíkova, E.: O textových konektorech, *Zborník Pedagogickej fakulty v Nitre. 23. Jazyk a literatúra*. Bratislava: Slovenské pedagogické nakladateľstvo, s. 45-55
- Bedřichová, Z. (v přípravě): *Částice implikující presupozici v češtině*, diplomová práce, Praha: Univerzita Karlova
- Bloomfield, L. (1933): *Language*, New York: Holt
- Buráňová, E. (2001): Příslovce – klasifikace podle funktorů. Pracovní materiál pro PDT 2.0
- Celle, A. – Huart, R. (ed.) (2007): *Connectives as Discourse Landmarks*, Amsterdam/Philadelphia: John Benjamins Publishing Company
- Černý, J. (1998): *Úvod do studia jazyka*, Olomouc: Rubico
- Černý, J. (2005): *Malé dějiny lingvistiky*, Praha: Portál
- Daneš, F. (1957): *Intonace a věta ve spisovné češtině*, Praha: ČSAV
- Daneš, F. (1985): *Věta a text*, Praha: Academia.
- Daneš, F. – Hlavsa, Z. (1981): *Větné vzorce v češtině*, Praha: Academia
- Dressler, W. (1972): *Einführung in die Textlinguistik*, Tübingen: Niemeyer



- Dressler, W. – de Beaugrande, R.-A. (1981): *Introduction to Text Linguistics*, London and New York: Longman,
- dostupné z [http://www.beaugrande.com/introduction\\_to\\_text\\_linguistics.htm](http://www.beaugrande.com/introduction_to_text_linguistics.htm)
- Firbas, J. (1992): *Functional sentence perspective in written and spoken communication*, Cambridge: Cambridge University Press
- Grepl, M. – Karlík, P. (1998): *Skladba češtiny*, Olomouc: Votobia
- Hajič, J. (2004): *Disambiguation of Rich Inflection. (Computational Morphology of Czech)*, Praha: Karolinum
- Hajič, J. – Cuřín, J. – Čmejrek, M. – Havelka, J. – Kuboň V. – Žabokrtský, Z. (2004): *Prague Czech-English Dependency Treebank 1.0*, CD ROM, Linguistic Data Consortium, CAT: LDC2004T25, 1-58563-321-6
- Hajič, J. – Hajičová, E. – Panevová, J. – Sgall, P. – Štěpánek J. – Havelka, J. – Mikulová, M. (2006): *Prague Dependency Treebank 2.0*, CD ROM, Linguistic Data Consortium, CAT: LDC2006T01, 1-58563-370-4
- Hajičová, E. (1975): *Negace a presupozice ve významové stavbě věty*, Praha: Academia
- Hajičová, E. (1993): *Issues of sentence structure and discourse patterns*, Praha: Univerzita Karlova
- Hajičová, E. (1995): Postavení rematizátorů v aktuálním členění věty, *SaS 56*, s. 241-251
- Hajičová, E. (1999): Aktuální členění věty a výstavba promluvy, *Čeština - univerzália a specifika* s. 47-54.
- Hajičová, E. – Vrbová, J. (1982): On the role of the hierarchy of activation in the process of natural language understanding, *Proceedings of the 9th conference on Computational linguistics*, Praha, s.107-113
- Hajičová, E. – Čermák, F. (2003): Introduction: Prague School of Linguistics in its classical

- time and today, In Vachek, J.: *Dictionary of the Prague School of Linguistics*, ed. Dušková, L., Amsterdam/Philadelphia: John Benjamins Publishing Company, s. 1-20
- Hajičová, E. – Partee, B. H. – Sgall, P. (1998): *Topic-Focus Articulation, Tripartite Structures, And Semantic Content*, Kluwer Academic Publishers Group
- Halliday, M. A. K. – Hasan, R. (1976): *Cohesion in English*, London: Longman
- Harris, Z. (1951): *Methods in Structural Linguistics*, Chicago: University of Chicago Press
- Harris, Z. (1952): Discourse Analysis. In *Language* 28:1. s. 1-30.
- Hausenblas, K. (1964): On the characterization and classification of discourses, *Travaux linguistiques de Prague 1.*, Prague
- Hausenblas, K. (1971): *Výstavba jazykových projevů a styl*, Praha: Univerzita Karlova: Mír
- Hoffmannová, J. (1983): *Sémantické a pragmatické aspekty koherence textu*, disertační práce, Linguistica VI, Praha ČSAV
- Hoffmannová, J. (1984): Typen der Konnektoren und deren Anteil an der Organisierung des Textes, In *Text and the Pragmatic Aspects of Language*, Linguistica X, Praha: ČSAV, s. 23-39
- Hoffmannová, J. (1986): Temporální sémantika a text, In *Teoretické otázky jazykovědy*, Linguistica XVI, Praha: ČSAV, s. 160-191.
- Hoffmannová, J. (1997): *Stylistika a ... Současná situace stylistiky*. Praha: Trizonia
- Hošnová, E. (2005): *Studie z novočeské syntaxe (konjunkce, pronominalizace)*. Praha: Karolinum
- Hrbáček, J. (1994): *Nárys textové syntaxe*, Praha: Trizonia
- Joshi, A. et al. (2006): Penn Discourse TreeBank 1.0, pracovní verze korpusu a nástroje pro vyhledávání dostupné z <http://www.seas.upenn.edu/~pdtb/pdtb-corpus-1.0/> [k 15. 9. 2007]
- Komárek, M. (1958): K sémantické a syntaktické charakteristice slov ještě, již (už), In *Studie*

ze slovanské jazykovědy. Praha, s. 147nn

Kořenský, J. (1992): *Komunikace a čeština*, Jinočany: H+H

Lopatková, M. – Žabokrtský, Z. – Benešová, V. (2006): *ValLex – Valency Lexicon of Czech Verbs*, dostupné z <http://ufal.mff.cuni.cz/vallex/2.0/doc/home.html>

Lee, A. et al. (2006): Complexity of Dependencies in Discourse: Are Dependencies in Discourse More Complex Than in Syntax? In *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories, TLT 2006*, Prague

Mann, W. C. – Thompson, S. A. (1986): Rhetorical Structure Theory: Description and Construction of Text Structures, In Kempen, G. (ed.): *Natural Language Generation: New Results in Artificial Intelligence*, s. 279-300

Marcus, M. P. et al. (1995): *Treebank-2*. Linguistic Data Consortium, CAT: LDC95T7, ISBN 1-58563-054-3. (Penn TreeBank 2.0)

Mathesius, V. (1939): O tak zvaném aktuálním členění větném, *SaS* 5, s. 171-174

Mathesius, V. (1943): Jazykozpytné poznámky k řečnické výstavbě souvislého výkladu, *SaS* 7, s. 114-129

Mathesius, V. (1947): *Čeština a obecný jazykozpyt*. Praha: Melantrich

Mikulová, M. (2005): Seznam výrazů, které se s funktorem CM mohou objevit pod uzlem pro souřadnost. Pracovní materiál pro PDT 2.0

Mikulová, M. et al. (2005): *Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka*, dostupné z <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/t-layer/pdf/t-man-cz.pdf>

Mistrík, J. (1975/6): Skladba textu, *Slovenský jazyk a literatúra v škole*, 22, s. 209-214

Panevová, J. (1980): *Formy a funkce ve stavbě české věty*, Praha: Academia

Panevová, J. (1992): Koreference gramatická nebo textová? In *Études de linguistique romane et slave*, Kraków, s. 495-506

The PDTB Research Group (2004): *Penn Discourse Treebank: Annotation Tutorial*, technická zpráva, dostupné z <http://www.seas.upenn.edu/~pdtb/manual/pdtb-tutorial.pdf>

Prasad, R. et al. (2006): *The Penn Discourse TreeBank 1.0 Annotation Manual*, IRCS Technical Report IRCS-06-01, Institute for Research in Cognitive Science, University of Pennsylvania, dostupný z <http://www.seas.upenn.edu/~pdtb/papers/pdtb-1.0-annotation-manual.pdf>

*Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue* (2007) Antwerp

Quirk, R. et al. (2004) *A Comprehensive Grammar of the English Language*, Longman

Sgall, P. (1967): *Generativní popis jazyka a česká deklinace*, Praha: Academia

Sgall, P. et al. (1969): *A Functional Approach to Syntax in Generative Description of Language*, New York: American Elsevier

Sgall, P. – Hajičová, E. – Buráňová, E. (1980): *Aktuální členění věty v češtině*, Praha: Academia

Sgall, P. – Hajičová, E. – Panevová, J. (1986): *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*, Dordrecht: Reidel Publishing Company; Praha: Academia

Schiffrin, D. (1987): *Discourse markers*, New York: Cambridge University Press

Schiffrin, D. (1994): *Approaches to Discourse*, Blackwell Publishing

Skalička V. (1937): Promluva jako linguistický pojem, *SaS*, s. 163-166

Skalička V. (1948): The need for a linguistics of „la parole“. In *Recueil linguistique de Bratislava 1*, s. 21-38

Šindlerová, J. (2007): *Zachycení koordinace v závislostním popisu (se zvláštním zřetelem ke koordinaci nestejných kategorií)*, diplomová práce, Praha: Univerzita Karlova

Šindlerová, J. – Mladová, L. – Toman, J. – Cinková S. (2007): An Application of the PDT-scheme to a Parallel Treebank. In *Proceedings of the Sixth Workshop on Treebanks and*

*Linguistic Theories*, TLT 2007, Bergen

Šmilauer, V. (1947): *Novočeská skladba*, Praha: Nakladatel Ing. Mikuta

Tárnyiková, J. (2002): *From text to texture. An introduction to processing strategies*, Univerzita Palackého v Olomouci

Tárnyiková, J. (2007): *Sentence complexes in text: Processing strategies in English and in Czech*, Univerzita Palackého v Olomouci

van Dijk, T. A. (1980): *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*, London: Longman

van Dijk, T. A. (ed.) (1997): *Discourse studies: Vol.1; Discourse as structure and process*, London: Sage

Webber, B. et al. (2005): A Short Introduction to Penn Discourse TreeBank, In *Copenhagen Working Papers in Language and Speech Processing*,

dostupné z <http://www.seas.upenn.edu/~pdtb/papers/nodalida.pdf>

Zeman, D. et al. (2005): *A Manual for Morphological Annotation*, technická zpráva, ÚFAL MFF UK, dostupné z <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/m-layer/html/index.html> a <http://ufal.mff.cuni.cz/pdt2.0/doc/pdt-guide/cz/html/ch02.html#t-layers-morpho>

Zikánová, Š. (2007): *Possibilities of Discourse Annotation in Prague Dependency Treebank (Based on the Penn Discourse Treebank Annotation)*, technická zpráva, Praha: Univerzita Karlova

## ÚFAL

ÚFAL (Ústav formální a aplikované lingvistiky; <http://ufal.mff.cuni.cz> ) is the Institute of Formal and Applied linguistics, at the Faculty of Mathematics and Physics of Charles University, Prague, Czech Republic. The Institute was established in 1990 after the political changes as a continuation of the research work and teaching carried out by the former Laboratory of Algebraic Linguistics since the early 60s at the Faculty of Philosophy and later the Faculty of Mathematics and Physics. Together with the “sister” Institute of Theoretical and Computational Linguistics (Faculty of Arts) we aim at the development of teaching programs and research in the domain of theoretical and computational linguistics at the respective Faculties, collaborating closely with other departments such as the Institute of the Czech National Corpus at the Faculty of Philosophy and the Department of Computer Science at the Faculty of Mathematics and Physics.

## CKL

As of 1 June 2000 the Center for Computational Linguistics (Centrum počítační lingvistiky; <http://ckl.mff.cuni.cz> ) was established as one of the centers of excellence within the governmental program for support of research in the Czech Republic. The center is attached to the Faculty of Mathematics and Physics of Charles University in Prague.

## TECHNICAL REPORTS

The ÚFAL/CKL technical report series has been established with the aim of disseminate topical results of research currently pursued by members, cooperators, or visitors of the Institute. The technical reports published in this Series are results of the research carried out in the research projects supported by the Grant Agency of the Czech Republic, GAČR 405/96/K214 (“Komplexní program”), GAČR 405/96/0198 (Treebank project), grant of the Ministry of Education of the Czech Republic VS 96151, and project of the Ministry of Education of the Czech Republic LN00A063 (Center for Computational Linguistics). Since November 1996, the following reports have been published.

- ÚFAL TR-1996-01** Eva Hajičová, *The Past and Present of Computational Linguistics at Charles University*  
Jan Hajič and Barbora Hladká, *Probabilistic and Rule-Based Tagging of an Inflective Language – A Comparison*
- ÚFAL TR-1997-02** Vladislav Kuboň, Tomáš Holan and Martin Plátek, *A Grammar-Checker for Czech*
- ÚFAL TR-1997-03** Alla Bémová at al., *Anotace na analytické rovině, Návod pro anotátory (in Czech)*
- ÚFAL TR-1997-04** Jan Hajič and Barbora Hladká, *Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structural Tagset*
- ÚFAL TR-1998-05** Geert-Jan M. Kruijff, *Basic Dependency-Based Logical Grammar*
- ÚFAL TR-1999-06** Vladislav Kuboň, *A Robust Parser for Czech*
- ÚFAL TR-1999-07** Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (in Czech)*
- ÚFAL TR-2000-08** Tomáš Holan, Vladislav Kuboň, Karel Oliva, Martin Plátek, *On Complexity of Word Order*
- ÚFAL/CKL TR-2000-09** Eva Hajičová, Jarmila Panevová and Petr Sgall, *A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank*
- ÚFAL/CKL TR-2001-10** Zdeněk Žabokrtský, *Automatic Functor Assignment in the Prague Dependency Treebank*
- ÚFAL/CKL TR-2001-11** Markéta Straňáková, *Homonymie předložkových skupin v češtině a možnost jejich automatického zpracování*

- ÚFAL/CKL TR-2001-12 Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (III. verze)*
- ÚFAL/CKL TR-2002-13 Pavel Pecina and Martin Holub, *Sémanticky signifikantní kolokace*
- ÚFAL/CKL TR-2002-14 Jiří Hana, Hana Hanová, *Manual for Morphological Annotation*
- ÚFAL/CKL TR-2002-15 Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarská and Vendula Benešová, *Tektogramaticky anotovaný valenční slovník českých sloves*
- ÚFAL/CKL TR-2002-16 Radu Gramatovici and Martin Plátek, *D-trivial Dependency Grammars with Global Word-Order Restrictions*
- ÚFAL/CKL TR-2003-17 Pavel Květoň, *Language for Grammatical Rules*
- ÚFAL/CKL TR-2003-18 Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarska, Václava Benešová, *Valency Lexicon of Czech Verbs VALLEX 1.0*
- ÚFAL/CKL TR-2003-19 Lucie Kučová, Veronika Kolařová, Zdeněk Žabokrtský, Petr Pajas, Oliver Čulo, *Anotování koreference v Pražském závislostním korpusu*
- ÚFAL/CKL TR-2003-20 Kateřina Veselá, Jiří Havelka, *Anotování aktuálního členění věty v Pražském závislostním korpusu*
- ÚFAL/CKL TR-2004-21 Silvie Cinková, *Manuál pro tektogramatickou anotaci angličtiny*
- ÚFAL/CKL TR-2004-22 Daniel Zeman, *Neprojektivity v Pražském závislostním korpusu (PDT)*
- ÚFAL/CKL TR-2004-23 Jan Hajič a kol., *Anotace na analytické rovině, návod pro anotátory*
- ÚFAL/CKL TR-2004-24 Jan Hajič, Zdeňka Urešová, Alevtina Bémová, Marie Kaplanová, *Anotace na tektogramatické rovině (úroveň 3)*
- ÚFAL/CKL TR-2004-25 Jan Hajič, Zdeňka Urešová, Alevtina Bémová, Marie Kaplanová, *The Prague Dependency Treebank, Annotation on tectogrammatical level*
- ÚFAL/CKL TR-2004-26 Martin Holub, Jiří Diviš, Jan Pávek, Pavel Pecina, Jiří Semecký, *Topics of Texts. Annotation, Automatic Searching and Indexing*
- ÚFAL/CKL TR-2005-27 Jiří Hana, Daniel Zeman, *Manual for Morphological Annotation (Revision for PDT 2.0)*
- ÚFAL/CKL TR-2005-28 Marie Mikulová a kol., *Pražský závislostní korpus (The Prague Dependency Treebank) Anotace na tektogramatické rovině (úroveň 3)*
- ÚFAL/CKL TR-2005-29 Petr Pajas, Jan Štěpánek, *A Generic XML-Based Format for Structured Linguistic Annotation and Its application to the Prague Dependency Treebank 2.0*
- ÚFAL/CKL TR-2006-30 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razimová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Annotation manual)*
- ÚFAL/CKL TR-2006-31 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Anotace na tektogramatické rovině Pražského závislostního korpusu (Referenční příručka)*
- ÚFAL/CKL TR-2006-32 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Reference book)*
- ÚFAL/CKL TR-2006-33 Jan Hajič, Marie Mikulová, Martina Otradovcová, Petr Pajas, Petr Podveský, Zdeňka Urešová, *Pražský závislostní korpus mluvené češtiny. Rekonstrukce standardizovaného textu z mluvené řeči*
- ÚFAL/CKL TR-2006-34 Markéta Lopatková, Zdeněk Žabokrtský, Václava Benešová (in cooperation with Karolína Skwarska, Klára Hrstková, Michaela Nová, Eduard Bejček, Miroslav Tichý) *Valency Lexicon of Czech Verbs. VALLEX 2.0*
- ÚFAL/CKL TR-2006-35 Silvie Cinková, Jan Hajič, Marie Mikulová, Lucie Mladová, Anja Nedolužko, Petr Pajas, Jarmila Panevová, Jiří Semecký, Jana Šindlerová, Josef Toman, Zdeňka Urešová, Zdeněk Žabokrtský, *Annotation of English on the tectogrammatical level*

- ÚFAL/CKL TR-2007-36** Magda Ševčíková, Zdeněk Žabokrtský, Oldřich Krůza, *Zpracování pojmenovaných entit v českých textech*
- ÚFAL/CKL TR-2008-37** Silvie Cinková, Marie Mikulová, *Spontaneous speech reconstruction for the syntactic and semantic analysis of the NAP corpus*
- ÚFAL/CKL TR-2008-38** Marie Mikulová, *Rekonstrukce standardizovaného textu z mluvené řeči v Pražském závislostním korpusu mluvené češtiny. Manuál pro anotátory*
- ÚFAL/CKL TR-2008-39** Zdeněk Žabokrtský, Ondřej Bojar, *TectoMT, Developer's Guide*
- ÚFAL/CKL TR-2008-40** Lucie Mladová, *Diskurzivní vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0*