

Vybrané problematické aspekty konektivních prostředků v rámci anotace mezivýpovědních významových vztahů v PDT

Pavλίna Jínová

1 Uvedení

V tomto příspěvku je nejprve stručně představen projekt anotace mezivýpovědních významových vztahů a jeho místo v práci na Pražském závislostním korpusu (Prague Dependency Treebank, PDT). Hlavní část příspěvku se zabývá konektivními prostředky, které mezivýpovědní významové vztahy signalizují. Vzhledem k tomu, že práce na anotování mezivýpovědních významových vztahů právě probíhá, má tato část příspěvku spíše formu načrtnutí některých problémů a představení pracovních řešení, než že by chtěla představit řešení již přijatá, nebo dokonce poskytnout sumativní přehled.

2 Projekt anotace mezivýpovědních významových vztahů a jeho místo v PDT

Pražský závislostní korpus představuje lingvisticky bohatě anotovanou databázi textů v rozsahu zhruba 50 000 vět. Texty byly přejeté z ČNK, žánrově se jedná o texty publicistické. Od roku 1995, kdy byla práce na PDT zahájena, proběhla anotace morfologické, analytické a tektogramatické roviny.

Již v rámci anotace tektogramatické roviny – tedy v rámci označování hloubkových syntaktickosémantických vztahů – se pracovalo s některými prostředky, které se kromě výstavby výpovědi podílejí i na výstavbě vyšších celků. Jednalo se především o některé případy koreference,¹ anotaci aktuálního členění a zavedení značky PREC (=reference to preceding text) pro výrazy, které „signalizují návaznost klauze na předcházející kontext.“² Tyto výrazy propojují části textu podobným způsobem jako spojky v koordinačních a některých subordinačních souvětích, ale na rozdíl od nich není jejich dosah omezen na jeden větný celek (celek mezi dvěma koncovými interpunkčními znaménky) – v PDT mají tedy větší dosah než jeden strom. Vzhledem k tomu, že se anotace v PDT až dosud v případě spojovacích prostředků vědomě omezovala na vztahy uvnitř jednoho větného celku (tj. jednoho stromu), vyvstala v další fázi práce na korpusu potřeba teoreticky uchopit a poté anotovat i vztahy mezi více stromy – vztahy textové. Pro lepší představu o problematice je v příkladech (1)–(4) ukázána paralelnost výrazů typu PREC a tradičně uváděných spojovacích výrazů, které jsou v PDT pojímány jako kořeny koordinačních struktur a označeny sémantickou značkou. První z každé dvojice příkladů je součástí PDT, druhý je k němu dotvořen pro ilustraci.

(1) *Původně jsme chtěli za každou cenu využít svého stadionu. **Ale** (PREC) v úterý nás navštívila osmičlenná delegace Chelsea, a když nám promítli na videu chování svých fanoušků, naskočila mi husí kůže.* (PDT, In 94205_125)

(2) *Původně jsme chtěli za každou cenu využít svého stadionu, ale v úterý nás navštívila osmičlenná delegace Chelsea, a když nám promítli na videu chování svých fanoušků, naskočila mi husí kůže.*

(3) *Slovenská elita byla zklamána politickou volbou Slovenska. **Proto** (PREC) většina kvalitních odborníků zůstala v Praze.* (PDT, cmpr9413_012)

¹ Koreferencí se míní vztah výrazů, které poukazují ke stejnému předmětu nebo situaci reálného světa.

V dokončené verzi PDT 2.0 se anotace koreference kvůli komplexnosti tohoto jevu vědomě omezuje pouze na některé případy. Jako případy tzv. gramatické koreference jsou pojímána např. zvrtná zájmena, vztažné prostředky a aktanty u sloves kontroly, tzv. textová koreference byla anotována pouze u jejího nejčastějšího prostředku – u osobních a přivlastňovacích zájmen pro 3. osobu a u zájmen ukazovacích *ten, ta, to* (podrobně viz Mikulová et al.: Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka. UFAL MFF, Praha 2005. s. 935–1050).

² Mikulová, M. et al.: Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka. UFAL MFF, Praha 2005. s. 537.

(4) Protože slovenská elita byla zklamána politickou volbou Slovenska, většina kvalitních odborníků zůstala v Praze.

Teoretické uchopení a anotace vztahů mezi více stromy jsou cílem dvou v současné době probíhajících projektů. Ve starším z nich – projektu anotace rozšířené koreference – jsou označovány vztahy mezi jmennými frázemi, které poukazují ke stejnému předmětu nebo situaci reálného světa. Na rozdíl od předchozích anotací se již neberou v úvahu pouze zájmena, ale také substantiva a adjektiva. Navíc se zavádí pojem tzv. asociační anafory (bridging anaphora), jež umožňuje i zachycování takových vztahů jako např. část – celek (*les – strom*), kontrast (*v Austrálii – v Evropě*) atd., jež nejsou v pravém slova smyslu koreferenční, nicméně na stavbě textu se podílejí nezanedbatelnou měrou.³

V druhém projektu zaměřeném na uchopení a anotaci vztahů mezi více stromy, kterým se budeme zabývat podrobněji – projektu anotace mezivýpovědních významových vztahů – jsou centrem zájmu významové vztahy mezi výpověďmi. Jako výpověď, která může být součástí významového vztahu je přitom v rámci tohoto projektu prozatím chápána pouze výpověď obsahující finitní sloveso. Centrem zájmu jsou syntakticky nezávislé výpovědi (ať už ve smyslu větného celku mezi dvěma koncovými interpunkčními znaménky nebo ve smyslu ucelené části koordinačního souvětí), ale také závislé klauze, které ale netvoří valenční doplnění predikátu své řídicí klauze. Mezivýpovědní významové vztahy zahrnují jednak tradiční syntaktické vztahy,⁴ jednak tzv. vztahy rétorické. Protože syntaktické vztahy výpovědí spojených do jednoho větného celku zachycuje dostačujícím způsobem dosavadní anotace v PDT, všímá si anotace mezivýpovědních významových vztahů především syntaktických vztahů mezi výpověďmi, které do jednoho celku spojeny nejsou (příklady těchto vztahů jsou uvedeny v (1) a (3) výše). Kromě těchto vztahů jsou v textech vyhledávány a poté anotovány i vztahy, kterým Mladová⁵ ponechává v návaznosti na Ashera⁶ pojmenování rétorické. Tyto vztahy patří „svým charakterem spíše do oblasti kompozice textu/diskurzu. (...) [T]ýkají se buď temporality v diskurzu nebo přímo jeho 'rétorické' výstavby.“⁷ Mezi rétorické vztahy patří např. shrnutí, jiná formulace téhož nebo různé druhy rozpracování již řečeného (uvedení příkladu, uvedení detailu). Příklad (5) ukazuje rétorický vztah exemplifikace mezi dvěma syntaktickými celky – druhý celek uvádí příklad toho, jak bude lékařská komora střežit odbornost svých členů.

(5) Komora chce být do budoucna především „strážcem odbornosti“ svých členů, mimo jiné zavedením kreditního systému. Lékař by musel **např.** každých pět let prokazovat, že se zúčastnil nějakého kongresu apod., aby mu byla licence automaticky prodloužena. (PDT, Ind94103_016)

V první fázi anotování mezivýpovědních významových vztahů v textu, která právě probíhá, se vztahy mezi výpověďmi označují pouze na základě přítomnosti konektivního prostředku. Anotace vztahů probíhá podle schématu vypracovaného týmem řešitelů projektu (Mladová, Zikánová, Bedřichová) v návaznosti na anotaci textových vztahů ve filadelfské databázi Penn Discourse TreeBank 2.0⁸ a postupuje od analýzy prostého textu (nalézání a označování možných mezivýpovědních konektivních prostředků v prostém textu) ke značení mezivýpovědních vztahů přímo na tektogramatických stromech. Syntaktické stromové struktury, které daný konektivní prostředek spojuje, jsou propojeny

³ Podrobněji viz Nédolužko, A. 2008. Anotování rozšířené textové koreference a asociační anafory v Pražském závislostním korpusu [online]. UFAL MFF, Praha [citováno 24. 05. 2010]. Dostupné z WWW:

<http://ufal.mff.cuni.cz/~nedoluzko/koref_annot/manual_RK_kratky.pdf>.

⁴ Jedná se např. o slučovací, odporovací, důsledkový, vylučovací nebo gradační vztah. Anotace mezivýpovědních významových vztahů tyto sémantické charakteristiky do značné míry přejímá a dále rozpracovává.

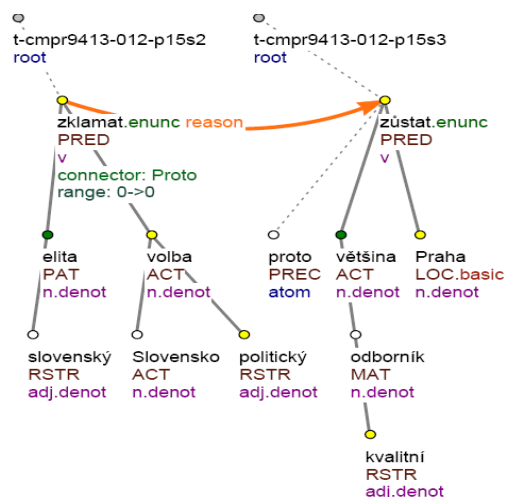
⁵ Mladová, L.: Diskurzivní vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0. UFAL MFF, Praha 2008.

⁶ Asher, N.: Reference to Abstract Objects in Discourse. Kluwer Academic Publishers, Dordrecht 1993.

⁷ Mladová, L.: Diskurzivní vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0. UFAL MFF, Praha 2008. s. 43.

⁸ Stránky projektu Penn Discourse TreeBank: <http://www.seas.upenn.edu/~pdtb/>.

šipkou, jejich vztah je sémanticky interpretován a daný konektor je připojen k šipce. Obrázek 1 ukazuje anotaci vztahu mezi syntaktickými strukturami uvedenými výše v příkladu (3). Šipka nese sémantickou značku REASON – jedná se o vztah příčiny a následku, je k ní připojen konektivní prostředek *proto*. Zároveň je také označen rozsah jednotlivých částí, které daný vztah zahrnuje (*range: 0 > 0*) – v tomto případě jsou součástí vztahu pouze oba uvedené stromy.



Obrázek 1: Ukázka anotace mezivýpovědního významového vztahu mezi tektogramatickými stromy v PDT

V další fázi se počítá s označením implicitních vztahů (vztahů nesignalizovaných žádným konektorem), podrobnějším zpracováním víceslovných konektivních prostředků a dalšími úkoly.

3 Konektivní prostředky mezivýpovědních významových vztahů

Jak už bylo řečeno, je v rámci právě probíhající fáze anotace mezivýpovědních významových vztahů pro označení vztahu zásadní přítomnost konektivního prostředku. V následujících odstavcích nejprve stručně představíme jednotlivé skupiny výrazů, které mohou být konektivním prostředkem mezivýpovědních významových vztahů tak, jak jsou pojímány v rámci projektu jejich anotace v PDT, a poté zmíníme několik problémů, které se v otázce konektorů během anotace vyskytly.

3.1 Stručný přehled mezivýpovědních konektivních prostředků

Je nutné připomenout, že anotace diskurzu je projekt neuzavřený, a proto ani seznam konektivních prostředků nemůže být v tuto chvíli vyčerpávající. I když je předčasné pokoušet se o úplnou klasifikaci konektivních prostředků, uvedeme zde alespoň zhruba skupiny výrazů, se kterými je nutno počítat, aby se lépe ukázala šíře problematiky. V celém projektu anotace mezivýpovědních významových vztahů se v této otázce vychází zejména ze skupiny výrazů anotovaných jako PREC v PDT, ale inspirací při přípravách anotace byly také prostředky nazývané v MČ II částice strukturující text,⁹ Hrbáčkův přehled konektorů¹⁰ a některé další zdroje.¹¹ Snad nejpřehlednějším, i když ne zcela bezproblémovým třídícím kritériem je v případě zkoumaných výrazů jejich slovnědruhová platnost.¹² Pro tuto chvíli nechme stranou úvahu, zda se zkoumané výrazy svým užitím ve funkci mezivýpovědních konektorů nestávají všechny např. součástí částicových výrazů. Tento aspekt celé problematiky není nyní

⁹ Mluvnice češtiny II. Academia, Praha 1986. s. 237–238.

¹⁰ Hrbáček, J.: Nárys textové syntaxe spisovné češtiny. Trizonia, Praha 1994. s. 56–58.

¹¹ Podrobně viz Mladová, L.: Diskurzí vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0. UFAL MFF, Praha 2008.

¹² Při náčrtu klasifikace vycházíme pro jednoduchost zejména z pojetí slovních druhů v Mluvnici češtiny II. Academia, Praha 1986.

předmětem našeho zájmu. Na slovnědruhovou příslušnost zde pohlížíme spíše z hlediska původu daných výrazů.

Při zkoumání možných mezivýpovědních konektivních prostředků je třeba stále mít na zřeteli, že u většiny z nich je příslušnost k této skupině záležitostí konkrétního kontextu. Během anotací je pro posouzení příslušnosti prostředku k mezivýpovědním konektorům užíváno toto kritérium: Výraz je mezivýpovědním konektorem, jestliže otevírá nejméně dvě pozice, které jsou v daném případě obsazeny nejméně dvěma syntakticky vzájemně valenčně nezávislými výpověďmi obsahujícími verbum finitum. Pro části textu, které konektor spojuje, se užívá označení argumenty. S funkcí konektivního prostředku nutně souvisí i to, že vyjadřuje sémantický vztah argumentů.

Mezi mezivýpovědní konektivní prostředky byly v průběhu práce na projektu anotace mezivýpovědních významových vztahů zařazeny (vždy s ohledem na kontext) tyto výrazy/prostředky:

- a) spojky souřadící: *a, ale, však, nebo, proto...*
- b) spojky podřadící: *ačkoliv; místo, aby; s tím, že...*
- c) částicové výrazy (včetně rematizačních výrazů): *totiž, ovšem, zkrátka, dokonce, také, například...*
- d) příslovce: *potom, následně, stejně, současně, tak...*
- e) některá užití zájmen: *kromě toho, k tomu, naproti tomu, tím...*
- f) ustálené víceslovné konektivní prostředky vzniklé spojením různých výrazů: *na jedné straně, stručně řečeno, jinými slovy...*
- g) písmenné nebo číselné prvky pro vyjádření vztahu výčtu: *a), b), 1., 2....*
- h) dvě interpunkční znaménka: dvojtečka a pomlčka.

Zatímco některé z těchto výrazů lze do anotací zahrnout bez výhrad a jejich sémantická interpretace je v naprosté většině případů zřejmá, jsou některé tyto výrazy předmětem diskusí, zda vůbec nebo v jakém případě mezi mezivýpovědní konektory patří, případně jak uchopit jejich sémantiku. Na tři takové problémové jevy nyní ukážeme.

3.2 Vybrané problematické aspekty mezivýpovědních konektivních prostředků

3.2.1 Významová diverzita (a nejednoznačnost) spojovacího prostředku *tedy*

Výraz *tedy* v mnoha kontextech nepochybně mezi mezivýpovědní konektivní prostředky patří.¹³ V některých těchto kontextech je ale zajímavou otázkou, jaký sémantický vztah mezi částmi, které spojuje, je jím vyjadřován. Tento výraz je zároveň velmi dobrou ukázkou toho, že některé mezivýpovědní konektivní prostředky mohou nabývat více významů – v případě výrazu *tedy* dosahuje škála možných významů spíše větší šíře, než je obvyklé. Ilustrujme jednotlivé významy, které výraz *tedy* prostředkuje, na konkrétních příkladech z korpusu PDT.¹⁴

¹³ Kromě těchto případů byly během anotace zaznamenány i kontexty, v nichž výraz *tedy* otevírá pouze jednu pozici, a proto ho v těchto případech není možné za mezivýpovědní konektivní prostředek považovat. Taková užití ukazují příklady (6) a (7). Pro lepší představu o kontextu je v nich ponechána i věta předcházející výpovědi, v níž je výrazu *tedy* užito.

(6) *Potom uvidíš cosi, co by mohlo být: telefonní budkou nebo WC nebo kabinkou pro peep show nebo plynovou komoru nebo... Vstoupíš tedy dvnitř - se smíšenými pocity, v nichž se dá rozeznat nervozita z kontaktu s neznámým i zvědavost.* (PDT, mf920922_105)

(7) *Co kdybychom například v tomto týdnu ze solidarity s postiženými všichni zrušili svá konta a své vklady u této instituce. Já to tedy určitě udělám, neboť kdo mi zaručí, že další podobnou akci nepostihne tato banka právě mne?* (PDT, ln94206_103)

¹⁴ V těchto i následujících příkladech jsou argumenty vztahu, který konektivní prostředek signalizuje, značeny takto: První argument je vždy značen tučným písmem, druhý přerušovaným podtržením. Konektor je zvýrazněn šedým podložením.

Velmi často prostředkuje výraz *tedy* vztah příčiny a následku/důsledku. V těchto kontextech je bez problémů nahraditelný jiným konektivním prostředkem – *proto*. Toto užití ukazují příklady (8) a (9).

(8) **Generální ředitel Českých drah Emanuel Šíp zatím neobdržel žádné oznámení o tom, že by byl obžalován v souvislosti se svým jednáním před loni připravovanou stávkou železničářů. Nehodlá se tedy vyjadřovat k informaci, podle níž pražský úřad vyšetřování dokončil v této věci šetření a předal spis městskému státnímu zastupitelství s návrhem na podání obžaloby.** (PDT, ln94204_53)

(9) **To se ovšem nepodařilo, neboť k jednání se dostavili pouze tři z devíti členů rady (ostatní byli v té době na dovolené). Nemohlo se tedy nic vyřešit.** (PDT, ln94200_153)

V jiných kontextech nacházíme *tedy* jako prostředek signalizující přeformulování již řečeného (v terminologii anotace mezivýpovědních významových vztahů se jedná o vztah rektifikace). V příkladu (10) je tento vztah přítomen mezi dvěma větnými celky, v příkladě (11) ho nacházíme mezi částmi jednoho větného celku.

(10) **Přijeli podepsat smlouvu o spolupráci a uviděli plakáty s krvelačnou orlicí a ostrými drápy. Tedy, zahlédli je jen koutkem oka, protože radnice téměř všechny mezitím strhla.** (PDT, ln94205_68)

(11) **Ačkoliv neštěstí trvalo 1.7 sekundy, což je při té rychlosti hrozně dlouho, nesnažil se zatáčet - tedy auto nevyvinulo snahu změnit směr.** (PDT, ln94103_028)

Dalším významem, který je v případě výrazu *tedy* nutno brát v úvahu, je ekvivalence – tedy vyjádření téhož jinými slovy. Tento význam *tedy* typicky prostředkuje mezi neslovesnými částmi větných celků,¹⁵ jak ukazuje příklad (12), ale o tomto významu je nutné přemýšlet i v případě, kdy *tedy* propojuje dvě samostatné výpovědi. O tento vztah se podle našeho názoru jedná v případě (13) i (14). V příkladě (14) přistupuje k ekvivalenci i hodnocení autora textu (držet se pořizovací ceny nemovitostí je podle něj stejné jako vycházet z účetní fikce).

(12) **Obecně lze říci, že se kabinet vesměs přiklání k zákonům útlým, tedy ke stručnému deklarování některých obecnějších principů.** (PDT, ln94209_44)

(13) **Tvrzení, že právě zde najdou staré i nové demokracie klidný přístav na rozbouřeném evropském moři, ale dostávají rány z mnoha stran. Od dánského odmítnutí Evropské unie až po rozpad Československa. Evropa jako celek tedy zdánlivě nabízí určité jistoty pro klidný a úspěšný vývoj. Události poslední doby jsou ale úplně jiné než v posledních desetiletích.** (PDT, mf920922_116)

(14) **Je to způsobeno tím, že ve svých úvahách se drží striktně pořizovacích cen nemovitostí, které činí v průměru 1500 Kč/m při průměrném stáří bytového fondu 46 let. Vychází tedy z účetních fikcí, které nemají s dnešní realitou nic společného.** (PDT, ln94103_066)

Už příklad (14) lze považovat za případ, který připouští i jiné interpretace než tu uvedenou. Existují však také kontexty, v nichž význam, který *tedy* prostředkuje, je ještě méně jasný, a nelze ho proto jednoznačně interpretovat – například vztahy v (15) a (16) je možné vidět někde na pomezí ekvivalence (vyjádření téhož) a vyplývání (velmi slabého důsledku nebo i vysvětlení). Příklad (15) je podle našeho názoru blíž ekvivalenci, příklad (16) blíže vysvětlení.

(15) **Jak již bylo řečeno, na rozdíl od dosavadního stavu se v návrhu nového zákona připouští, aby zřizovatelem vysokoškolské instituce byl nejen stát, ale i jiná právnická osoba. Pak se v procesu zřizování rozlišují dva kroky- jednak vlastní založení instituce, jednak její uznání státem.**

¹⁵ Pro zřejmost znovu uvádíme, že tyto případy nejsou do anotace mezivýpovědních významových vztahů zahrnovány.

Je možno tedy hovořit o vysokoškolských institucích státem zřízených a institucích státem uznaných. (PDT, In94210_92)

(16) *Toto testování se týká prakticky všech přípravků, u nichž je předpoklad, že s nimi bude člověk přicházet do přímého kontaktu. Je to tedy i celá tzv. chemie pro domácnost - prací a čisticí prostředky, jsou to ale také i velmi toxické a agresivní látky ze skupiny agrochemikálií, nátěrových hmot apod.* (PDT, In94200_124)

Interpretovat význam prostředkováný výrazem *tedy* je velmi obtížné i v případech, v nichž je tohoto výrazu užito jako konektivního prostředku mezi velkými částmi textu. Zde se vyhraněný význam smývá ještě více než v případě propojení menších argumentů. Zvláště v těchto případech je tedy třeba připustit možnost, že se v anotacích bude vyskytovat více stejně relevantních interpretací.

3.2.2 Některé problémy s řazením rematizátorů mezi mezivýpovědní konektory ilustrované na rematizátoru *také*

Jako rematizátory jsou v PDT označovány výrazy, „jejichž funkcí je signalizovat ve větě kategorie aktuálního členění, a to kategorie sdělně nejzávažnější – ohnisko a kontrastivní základ.“¹⁶ Rematizátory stojí téměř vždy bezprostředně před výrazy, které jsou v ohnisku výpovědi – v ohnisku přitom mohou stát jak nominální skupiny, tak i skupiny verbální. A právě rematizátory, které lze vztáhnout ke skupině obsahující určitý slovesný tvar, je podle našeho názoru (stejně jako podle názoru Mladové¹⁷) vhodné pojímat zároveň jako mezivýpovědní konektivní prostředky, protože jejich funkční paralelnost s některými tradičními spojovacími výrazy (v případě výrazu *také* zejména s konektorem *a*) je z jejich užití zřejmá. Z anotací již vyplynulo, že každý rematizátor, který je potenciálně i mezivýpovědním konektorem, je nutné zkoumat zvlášť a v konkrétních kontextech. Na tomto místě budeme část naznačené problematiky ilustrovat na užitích rematizátoru *také* v PDT.

Připomeňme, že hlavním kritériem při rozhodování, zdali nějaký výraz je, či není mezivýpovědním konektorem, je skutečnost, zda daný výraz v daném kontextu otevírá dvě pozice, které jsou naplněné vzájemně valenčně nezávislými výpověďmi obsahujícími verbum finitum. Podle tohoto kritéria není těžké rozhodnout, že v případě (17) výraz *také* mezivýpovědním konektorem v pravém slova smyslu není (k předchozí větě připojuje pouze nominální frázi „*prostory stanice Praha-Smíchov*“, sloveso je v obou větách stejné), zatímco v případě (18) je vhodné *také* jako mezivýpovědní konektor pojímat – připojuje k předchozímu kontextu novou informaci vyjádřenou větně (děti se někdy nechtějí svěřovat rodičům a rodiče na ně také mají méně času).

(17) *Podle Mandíkových slov lze komerčně využít zhruba deset hektarů pozemků v železniční stanici Praha-Žižkov. Využít lze také prostory stanice Praha- Smíchov.* (PDT, In94210_9)

(18) *Děti se s některými záležitostmi nechtějí svěřit rodičům, i když žijí v normálně fungující rodině. Jisté věci se v určitém věku lépe říkají někomu anonymnímu. Dnes mají také mnozí rodiče méně času na své ratolesti než dřív.* (PDT, In94204_107)

Problematickým se toto kritérium stává v kontextech, ve kterých věty propojené výrazem *také* obsahují slovesa, jež nejsou identická, ale jsou zároveň významově velmi blízká. V těchto případech je nutné rozhodnout, zda je jejich význam dostatečně vzdálen na to, abychom jako připojenou k předchozímu kontextu vnímali celou větu, nebo jsou si tak blízká, že připojena je pouze nějaká její nominální část. V případě (19) není podle našeho názoru výraz *také* mezivýpovědním konektorem (firma vyrábí tresty do lihovin atd. a také cukrářské pasty – vynecháním slovesa nezmizí z textu žádná podstatná informace). Podobným případem je i (20) – vláda připravuje snížení všech regulovaných

¹⁶ Mikulová, M. et al.: Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka. UFAL MFF, Praha 2005. s. 1096.

¹⁷ Mladová, L.: K problematice vztahu rematizátorů a textových konektorů. Čeština doma a ve světě, 3 a 4, 2008, s. 126–133.

kapitol státního rozpočtu a změnu zákona o státním rozpočtu, význam sloves v obou relevantních částech (*připravovat* a *navrhnout*) je však již vzdálenější než v případě (19).

(19) *Vyrábějí se zde především resti do lihovin, limonád, sirupů a pečiva. Firma také produkuje cukrářské pasty.* (PDT, ln94205_22)

(20) *Proto, jak informoval ministr Kočárník, vláda ČR připravuje snížení všech regulovaných kapitol státního rozpočtu o zhruba [9,5]9.5 miliardy korun. Kabinet Václava Klause navrhne také změnu zákona o státním rozpočtu, která bude zaměřena na snížení dotací do rozpočtů měst a obcí.* (PDT, mf920925_009)

Naproti tomu v příkladu (21) už jsou významy sloves v obou celcích natolik vzdáleny, že v případě výrazu *také* vystupuje jeho mezivýpovědní konektivní funkce vcelku zřetelně.

(21) *Povinností budoucího nájemce tohoto areálu o rozloze 103 tisíc metrů čtverečních bude mj. péče o všechny nemovitosti včetně jejich údržby a oprav. Nájemce bude také muset vyřešit podmínky parkování pro návštěvníky tržnice a splnit podmínky Pražského ústavu památkové péče při úpravách objektů vzhledem k tomu, že jde o kulturní památku.* (PDT, ln94204_1)

Kromě uvedených kontextů je třeba s výrazem *také* jako mezivýpovědním konektivním prostředkem počítat i v případech jako (22), kde sice stojí před jménem *argument*, ale je zřejmé, že toto jméno vstupuje do vztahu s předchozím větným celkem včetně následující klauze – tedy v podstatě jako celek s verbem finitem.

(22) *Mě osobně mrzí, že hráči sáhli ke stávce, aniž by nám dali příležitost s nimi jednat. Nelíbí se mi také argument, že všude se vyplácí prémie za zápasy, tak proč by to u nás mělo být jinak.* (PDT, ln94200_153)

Vedle všech naznačených problémů bude třeba vyřešit i jevy ilustrované příklady (23) a (24). V prvním z nich výraz *také* připojuje z jednoho úhlu pohledu pouze jmennou skupinu („v Hostivaři a v dalších pražských lesoparcích“), z druhého úhlu pohledu ale funguje v iniciální pozici jako spojka *a*. Bude nutné rozhodnout, který úhel pohledu má pro naši anotaci větší váhu. I v druhém případě (příklad (24)) připojuje výraz *také* v přísném slova smyslu pouze nominální skupinu (americká strana splatí jednak výstavbu týkající se Svobodné Evropy, jednak výstavbu rodinných domků atd.), nedá se však přehlédnout ani fakt, že tento výraz zřetelně ke stavbě textu přispívá – uvádí nové téma, které jinak přímo nenavazuje na předchozí informace v textu.

(23) *V okolí Brna a Kyjova se hojně vyskytují muchomůrky zelené. Také v Hostivaři a v dalších pražských lesoparcích byl nyní výskyt této houby zaznamenán.* (PDT, ln94211_44)

(24) *Americká strana má po dokončení projektu splatit výstavbu týkající se Svobodné Evropy. Pro první skupiny zaměstnanců Svobodné Evropy se budou byty kupovat prostřednictvím realitních kanceláří za běžné ceny. Nebudou jim podle primátorova vyjádření pronajímány obecní byty, jichž má Praha málo a jsou určeny pro jiné účely. Americká strana bude také hradit výstavbu rodinných domků a školy pro americkou komunitu v Praze- Nebušicích.* (PDT, ln94210_2)

Ze všech uvedených jevů je zřejmé, že rematizátory představují pro další práci s textovými vztahy široké pole zatím do značné míry neuchopených problémů. Vzhledem ke komplexnosti celé problematiky je zřejmé, že stanovisko, které bude k této otázce nakonec zaujato, musí vyplynout ze zkušenosti s fungováním jednotlivých rematizátorů ve velkém množství přirozených textů. Během anotace, která právě probíhá v PDT, jsou proto rematizátory, které podle daného anotátora plní v konkrétním kontextu i funkci mezivýpovědního konektoru, předběžně značkovány, aby na základě těchto dokladů mohlo být posléze zaujato jednotné stanovisko k těmto výrazům. Vzhledem k technickým možnostem není problém poté všechny anotace s tímto stanoviskem zpětně sladit.

3.2.3 Koreferenční a „mezivýpovědně-významová“ interpretace výrazu *tak*

Výraz *tak* může ve větě fungovat v mnoha různých rolích (*Tak se na to podíváme!*, *Byla tak hezká.*, *Tak se to neříká.* atd.). Většina z nich zůstává zcela mimo oblast našeho zájmu, ale existují nejméně dva typy užití, které se výrazně podílí na vyšší stavbě textu. V prvním z nich se výrazu *tak* užívá jako zájmeného příslovce pro vyjádření způsobu nebo prostředku a může odkazovat mimo jiné i k celé větě. Propojení obou vět zůstává zřetelně mimo oblast mezivýpovědních významových vztahů, patří do oblasti koreference. Toto užití ukazuje příklad (25).

(25) *Především by se mělo sloučit sociální a zdravotní pojištění do jedné položky. Zjednodušila by se tak administrativa s tím spojená.* (PDT, cmpr9413_002)

Výraz *tak* se ale na stavbě textu často podílí i jinak – toto užití ukazuje příklad (26). Sledovaný výraz zde funguje jako synonymum výrazů *proto*, *tedy*, uvádí části textu do vztahu příčina – důsledek.

(26) *Obecně lze říci, že se kabinet vesměs přiklání k zákonům útlým, tedy ke stručnému deklarování některých obecnějších principů. Vše ostatní by mělo být shrnuto do podzákonných norem, tedy různých prováděcích vyhlášek a nařízení. Extenzivních zákonů se spoustou technických detailů by se tak náš právní řád měl co nejrychleji zbavit.* (PDT, ln94209_44)

Z uvedených příkladů se zdá, že oba typy užití jsou snadno rozlišitelné – a opravdu tomu tak v mnoha případech je. Kromě nich se ale vyskytují kontexty, v nichž se nabízí obě interpretace – *tak* zde může fungovat jednak jako příslovečné určení způsobu, jednak jako konektivní prostředek pro příčinný vztah mezi větnými celky. Je proto vždy na zvážení, která interpretace je v daném kontextu více na místě. V příkladu (27) je podle našeho názoru příhodnější interpretace způsobová (Jak se stalo, že byl Radovan Novák svědkem tragedie? Byl komisařem při závodu Formule 1.), v příkladu (28) se jako lepší interpretace jeví značení vztahu příčina – důsledek (zemědělství se nedotkla kolektivizace, a proto zde zůstal tradiční venkovský styl života). V příkladu (29) jsou pro nás obě interpretace stejně validní (protože je však třeba se v anotaci vždy na některou stranu přiklonit, byl v tomto případě značen vztah příčina – důsledek). I v obou předchozích příkladech je ostatně třeba připustit možnost interpretovat je opačně.

(27) *Poprvé v historii byl Čech sportovním komisařem při závodu Formule 1. Radovan Novák se tak stal v Imole svědkem tragedie Rakušana Ratzembergera a Brazilce Senny.* (PDT, ln94103_028)

(28) *Ani zemědělství se nedotkla kolektivizace. Zůstal tu tak tradiční venkovský styl života včetně namáhavé manuální práce, který už z původní vlasti vymizel.* (PDT, ln94203_35)

(29) *Na tiskové konferenci po zápase, která se konala v kuželně vedle libereckého zimního stadionu, charakterizoval trenér George Kingston druhou část hry následovně: **Naše mužstvo nedokázalo přesně rozehrát. Dostalo se tak do role kuželek, na které se valily koule českých útoků.*** (PDT, ln94202_95)

Viděli jsme, že u výrazu *tak* je stejně jako v případě konektivního prostředku *tedy* v některých kontextech zřejmě třeba připustit několikero možnou interpretaci. Zatímco však u výrazu *tedy* je mezivýpovědní významový vztah mezi argumenty založen téměř vždy a jde jen o jeho sémantickou interpretaci, je v případě výrazu *tak* navíc otázkou, zda je jím mezivýpovědní významový vztah vůbec založen.

4 Závěr

V tomto příspěvku byl stručně představen právě probíhající projekt anotování mezivýpovědních významových vztahů v PDT. Vzhledem k tomu, že označení nějakého významového vztahu mezi částmi textu bylo v první fázi omezeno pouze na případy, kdy je v textu přítomen konektivní prostředek, věnovala se převážná část příspěvku právě jim. I když je ještě předčasné pokoušet se o jakékoliv definitivní seznamy mezivýpovědních konektivních prostředků, byly uvedeny skupiny

výrazů, se kterými je při práci s mezivýpovědními významovými vztahy třeba počítat, aby se lépe ukázala šíře celé problematiky. Jako ilustrace problémů, s nimiž se během anotací setkáváme, byly v poslední části příspěvku představeny tři jevy – významová diverzita (a nejednoznačnost) výrazu *tedy*, některé problematické stránky pojmání rematizátorů jako mezivýpovědních konektivních prostředků ilustrované na rematizátoru *také* a různé interpretace výrazu *tak* jako prostředku textové koherence. Jednotná stanoviska k těmto problémům teprve budou zaujata, cílem tohoto příspěvku bylo upozornit na některé problémové oblasti a ukázat možný směr jejich řešení.

Poděkování

Výzkum představený v tomto článku vzniká za podpory grantu MŠMT ČR LC536 „Centrum počítační lingvistiky“ a ME 10018 „Kontakt - K počítačové analýze struktury textu.“

Na výzkumu představeném v tomto článku se podílejí: prof. PhDr. Eva Hajičová, DrSc.; PhDr. Šárka Zikánová, Ph.D.; Mgr. Lucie Mladová, Mgr. Zuzanna Bedřichová, RNDr. Jiří Mírovský, Ph.D. a Mgr. Pavlína Jínová.

Literatura

- Asher, N.: Reference to Abstract Objects in Discourse. Kluwer Academic Publishers, Dordrecht 1993.
- Hrbáček, J.: Nárys textové syntaxe spisovné češtiny. Trizonia, Praha 1994.
- Mikulová, M. et al.: Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka. UFAL MFF, Praha 2005.
- Mladová, L.: Diskurzí vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0. UFAL MFF, Praha 2008.
- Mladová, L.: K problematice vztahu rematizátorů a textových konektorů. Čeština doma a ve světě, 3 a 4, 2008, s. 126–133.
- Mluvnice češtiny II. Academia, Praha 1986.
- Nědolužko, A. 2008. Anotování rozšířené textové koreference a asociační anafory v Pražském závislostním korpusu [online]. UFAL MFF, Praha [citováno 24. 05. 2010]. Dostupné z WWW: <http://ufal.mff.cuni.cz/~nedolužko/koref_annot/manual_RK_kratky.pdf>.
- Zikánová, Š. Possibilities of Discourse Annotation in Prague Dependency Treebank (Based on the Penn Discourse Treebank Annotation). UFAL MFF, Praha 2007.