

K nové úrovni bohemistické práce: Využití anotovaného korpusu*

Část 1.

Eva Hajičová, Jarmila Panevová a Petr Sgall

1. Pražský závislostní korpus a perspektivy bohemistické práce

1.1. Úvodní poznámky:

Zásadní význam Českého národního korpusu (ČNK) pro bohemistiku a pro českou lingvistiku vůbec dnes už naprosto není předmětem pochyb.¹ Jeho část zpracovávaná v podobě Pražského závislostního korpusu (Prague Dependency Treebank, PDT) bude brzy obsahovat desetitisíce a pak i statisíce českých vět označkových na různých mluvnických úrovních - na úrovních morfematiky i tzv. analytické syntaxe a v menší části (zatím tisíce vět) taky významové stavby věty (tektogramatiky).²

Tektogramatické anotování je založeno na teoretickém pojmovém rámci funkčního generativního popisu (FGP), který navazuje na funkčně strukturní syntax Pražské školy a usiluje o potřebnou míru explicitnosti. K jeho výhodám v porovnání s jinými teoretickými přístupy patří poměrná jednoduchost chápání větné stavby jako založené na (hloubkové) valenci (podrobnou charakteristiku základních aspektů FGP podávají Panevová, 1974-75; 1980, Sgall ad. 1986, Hajičová 1993)³ a zahrnující i aktuální členění věty (viz Hajičová 1986; Hajičová ad. 1998; !! SaS; Sgall ad. 1980) a na ně navazující pohled na kohezi promluvy, diskurzu (viz Hajičová ad. 1981; 1985; 1998). Ve prospěch této koncepce mluví i to, že nepředpokládá žádný zvlášť složitý vrozený mechanismus jako nezbytný pro osvojení jazyka ze strany dítěte (Sgall, 2001). Může tedy přispět k zachycení jazykového systému jako soustavy s poměrně jednoduchou strukturací. Zatímco jádro jazyka, soubor jevů prototypických, bezpříznakových, je strukturováno relativně velmi průhledně,

obsahují rozsáhlé soubory periferní jevy příznakové, sekundární, omezené kontextovými podmínkami, a jsou strukturovány daleko méně průhledně, s nejasnými hranicemi a s řadou pravidel více nebo méně specifických, až po jednotlivé výjimky. Proto není divu, že pro řadu oblastí jazykového systému ještě nebyly nalezeny optimální způsoby klasifikace a celkového popisu jevů, popř. je někteří badatelé v té či oné části světa v rámci některé teoretické koncepce našli, ale nepodařilo se ještě získat souhlasná stanoviska jiných, takže i tady jsou nutné další diskuse. Korpusy anotované na několika rovinách, včetně významové stavby věty, poskytují nová, materiálově široce podložená východiska pro takové diskuse.

V dubnu 2002 obsahuje soubor PDT anotovaný z hlediska analytické syntaxe už texty o délce 90 000 vět; tektogramaticky bylo v PDT anotováno 5500 vět v tzv. velkém souboru, ve kterém se zachycují funktoxy (druhy syntaktické závislosti mezi autosémantickými slovy ve větě); pro 2000 z těchto vět už bylo zachyceno i aktuální členění (včetně kontrastivního základu věty a výpovědní dynamičnosti). Menší, tzv. vzorový soubor (s úplnějším tektogramatickým rozbohem, včetně upřesnění gramatémů, tj. hodnot morfologických kategorií) čítá zatím jen 200 vět. Pro anotování nejsou věty vybírány jednotlivě, ale vždycky je z Českého národního korpusu více méně náhodně vybrán souvislý úsek textu čítající 50 vět. Tzv. velký soubor tektogramaticky anotovaných vět tedy obsahuje 110 segmentů textů. Jde o část ČNK s žánrovým složením 40% publicistických textů, 20% ekonomických zpráv a analýz, 20% populárně-vědeckých textů a 20% textů z informačních technologií. Uvedená data jsou přístupná pomocí vyhledávacího programu Netgraph, který je dostupný na adrese ufal.mff.cuni.cz-pdt. Úplná dokumentace je k dispozici na téže adrese. Pokyny pro anotátory jsou obsaženy ve dvou příručkách, viz poznámku 2; tektogramatickou příručku označujeme v dalším jen jako Manuál.

Existence PDT a jeho další rozšiřování a prohlubování je východiskem pro zcela novou úroveň studia a poznání češtiny. Může podstatně přispět:

(a) pro z p ů s o b bohemistického v ý z k u m u mluvnické stavby a slohového

rozvrstvení současné češtiny: každá diplomová práce, disertace, habilitace nebo jiná monografie zabývající se určitou otázkou českého jazyka a českých textů získává daleko výhodnější východisko v tom, že místo dosavadní práce s excerpty a kartotékami může vycházet z automaticky získaných anotovaných souborů vět nebo textových úseků, ve kterých byl daný jev (popř. skupina jevů podobných) zaznamenán; v počátečních stádiích budou v anotacích (zejména v tektogramatických) chyby, ovlivněné zejména nejistými rozhodnutími anotátorů nebo ne zcela precizními pokyny, popř. i mezerami v dosavadním poznání češtiny, ale monografické studie přispějí k jejich budoucímu odstranění a k úpravě odpovídajících procedur, která výskyt chyb sníží;

(b) pro prohloubení poznání češtiny po mnoha stránkách: bohemistice se otvírá možnost v širokém měřítku soustavně studovat nejrůznější jevy nejen na základě jednotlivých vět, ale na základě jejich výskytů v kontextu, studia celých promluv (diskurzů, textů, zatím zejména psaných, ale postupně, doufejme, i mluvených); s tím je spojena možnost systematictějšího poznání slohového bohatství češtiny a jejího funkčního rozvrstvení, a to nejen v rámci spisovné normy, ale i ve vrstvách dalších; díky tomu, že ČNK obsahuje i jevy ze starších stadií vývoje češtiny, bude umožněno i bohatší poznání vývojových kořenů řady jevů češtiny, a to nejen na úrovni slov a jejich tvarů, ale i větných konstrukcí, frazémů, idiomů a jevů povahy promluvové, z oblasti sdělovacího procesu, 'parole';

(c) pro řadu dalších oborů od literární vědy, stylistiky a rétoriky až po vědu právní, psychologii, logiku, kognitivní vědu i historii: všestranné poznání a popis velkého souboru českých textů a postavení jednotlivých jevů v nich poskytne všem těmto i dalším oborům příležitost řešit své jednotlivé problémy ve větší konkrétnosti a úplnosti, než bylo dosud možné;

(d) pro lepší poznání češtiny i z hlediska potřeb daných dnešními a budoucími technickými možnostmi: PDT poskytne daleko lepší východisko pro další počítačové zpracování češtiny pro účely takové, jako je vyhledávání informací v textech, komunikace s

inteligentními roboty a databázemi, budování automatických encyklopedií, strojový a strojem podporovaný překlad ap., tzn. pro nové oblasti, jejichž rychlý a mnohostranný vývoj přináší nový vztah mezi mechanickou prací, kterou dosud musí vykonávat člověk a která může být automatizována, a na druhé straně prostorem pro činnost tvořivou a pro svobodný výběr činností.

Anotování korpusu, zejména na tektogramatické rovině, je záležitostí velmi složitou a dosud zpracované specifikace nezachycují všechny jemné distinkce, které je třeba ve významové struktuře věty odlišovat. Anotátoři stojí před úkolem konsistentní analýzy surových vět z běžných textů včetně všech jejich složitostí a nepravidelností; pro všechny takové jevy je třeba najít prostředky umožňující zachytit potřebné distinkce v závislostním stromě, což není snadné, a snad se to dosud někde adekvátně nedaří.

Předkládáme proto v této stati předběžný dílčí přehled otázek, které nám analýza textů v PDT připomněla a které mohou být dořešeny teprve na základě dalšího empirického výzkumu nebo zásadních úvah a popř. i technických zlepšení popisného aparátu, nutných pro adekvátní zachycení obtížnějších úseků jazykového systému. Jsme si vědomi, že z velké části se o těchto otázkách v bohemistice už diskutovalo, a nemůžeme tu dosavadní poznatky a stanoviska k jednotlivým okruhům podrobně probírat. Bereme podle možnosti v úvahu, jak se tyto okruhy probírají v klasických pracích o české syntaxi (u V. Mathesia, Vl. Šmilauera, Fr. Kopečného, I. Poldaufa, M. Dokulila, Fr. Daneše a Zd. Hlavsy, v akademické Mluvnici češtiny i v mluvnicích brněnských), ale nemůžeme vyhledávat a rozbírat řadu dalších, více méně roztroušených příspěvků a využívat závažných poznatků v nich obsažených. Bude ovšem nutné při další práci nezůstávat jen u materiálu z korpusu, ale vyhledat i dosavadní zpracování jednotlivých otázek a opřít se o ně tam, kde je to možné.

Při rozboru a popisu jazyka je třeba, jak jsme už připomněli, pamatovat na to, že vedle jeho poměrně přehledně strukturovaného jádra (vlastnosti slov týkající se závislostních vztahů mezi částmi věty i morfologických významů a těmto souborům odpovídajících

vyjadřovacích prostředků na úrovni morfématiky, tvarosloví) existují rozsáhlé a komplikované oblasti periferní a že jak hranice těchto oblastí, tak i hranice jádra samého i jeho jednotlivých složek nejsou jednoznačně narýsovány, ale jsou často spojeny s nejasností, popř. s jemným odstupňováním (srov. např. Sgall 2001). Musíme tedy počítat s tím, že k zásadním problémům popisu patří právě existence přechodných pásem mezi oblastmi jazykových jevů. To se týká i hranice mezi gramatikou a tvořením slov, srov. různé pohledy na to, jestli mají čistě gramatickou nebo slovotvornou povahu vztahy mezi tvary jako *dělat a mít uděláno, dělany, dělání*, nebo mezi *nový a nově*. Uvnitř mluvnické stavby samé, v jejím jádru, tzn. v oblasti závislostních vztahů, jsou nejasné hranice mezi tzv. valenčními a nevalenčními doplněními: vedle doplnění vnitřních (tj. aktantů) u daného řídicího slova obligatorních jsou i fakultativní jako *číst někomu*, a jsou i obligatorní doplnění „volná“ jako *octnout se někde*. Můžeme za valenční považovat všechny aktanty a u jednotlivých řídicích slov i jejich ostatní obligatorní doplnění; vedle toho lze mluvit o kvazivalenci u vazeb jako *zemřít na zápal plic*, viz v odd. 2.1.1. V soustavě morfologických významů se takové nejasnosti týkají v češtině např. hranice mezi kategoriemi jako čas a vid atd.

Existence takových přechodných pásem vůbec neznamena, že by mezi jednotlivými částmi jazykového systému nebyly podstatné rozdíly. Musíme si ale být vědomi, že velká část pojmů, se kterými jsme zvyklí při popisu jazyků pracovat, jsou (řečeno s VI. Skaličkou) syndromy různých vlastností nebo vztahů (opozic), a ne jakési základní, nerozložitelné jednotky; rysy, ze kterých se skládají, vystupují někdy společně, jindy ne, takže ne vždycky najdeme jejich úplnou shodu. Není možné hned rozčlenit všechny takové jednotky na jejich elementární rysy, ale stupňovitost opozic mezi jednotkami vede k tomu, že někdy musíme vidět hranici mezi nimi jako záležitost relativní, podobně jako mezi nářečími (i v dobách jejich plné, ještě nenarušené existence) existují svazky ne zcela shodně probíhajících izoglos. Tak např. v oblasti vidu není ani opozice „dokonavosti“ gramatikalizována do stejné míry jako třeba kategorie pádu, ale ani u pádu není úplně jasno v třídění jednotlivých hodnot: v češtině snad platí, že opozice jako *ve dne : o dni : o celém včerejším dnu* jsou omezeny

lexikálně (frazologicky) a rozdíl jako v *Nymburce* : o *Nymburku* : v *Hamburku* že jsou dány určitými vývojovými fázemi, popř. místními rozdíly. Ale např. Jakobsonových osm ruských pádů v porovnání s šesti tradičními ukazuje, že obecně je i toto třídění složitější. Nejrůznější otázky tohoto druhu je zřejmě třeba dál promýšlet, což často vyžaduje i zpracování většího počtu dokladů o výskytu jevu v různých kontextech.

Otevřené problémy většinou můžeme v tomto článku jen stručně připomenout; jsme si přitom vědomi neúplnosti jejich výčtu, tj. toho, že tu předkládáme spíš ilustrace k problematice dalšího zpracování češtiny než přehled všech nebo všech hlavních otázek. U některých bodů, jejichž řešení (bez podrobného prohledání dosavadní bohemistické literatury, které ovšem bude nezbytné) dosud neznáme, naznačujeme možné prozatímní odpovědi, ale u jiných se ani nesnažíme návrhy řešení uvádět. To už patří k úkolům budoucích monografických prací, které vyjdou z korpusového materiálu a můžou přispět jak ke zlepšené formulaci teoretického rámce popisu češtiny (popř. jazyka vůbec), tak i k obohacení anotovacích procedur, intelektuálních i automatických.

Chceme tu nejdřív ilustrovat danou problematiku na příkladech anotovaných vět (v odd. 1.2). Pak probíráme především dosud nevyřešené (v dnešní soustavě anotací v PDT jen prozatímně zpracované) body týkající se vztahů na úrovni významové stavby věty, tj. na tektogramatické rovině (viz odd. 2) a dále vztahů mezi touto rovinou a rovinou morfématickou, tj. mezi významem a výrazem (odd. 3; ani v odd. 2 se ovšem vztahům významové stavby k výrazovým prostředkům nemůžeme úplně vyhnout, protože v morfématickém vyjádření je hlavní východisko ke studiu syntaktických a významových vztahů).

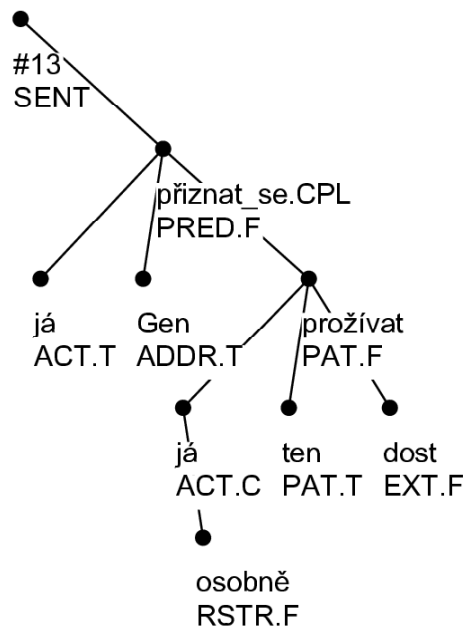
1.2. Ilustrace:

Pro připomenutí zásad FGP a odpovídajících postupů, o kterých se diskutovalo ve výše uvedených publikacích, uvádíme dva příklady tektogramatických reprezentací vět z PDTražského závislostního korpusu. Připomeňme, že tyto zápisy, tektogramatické

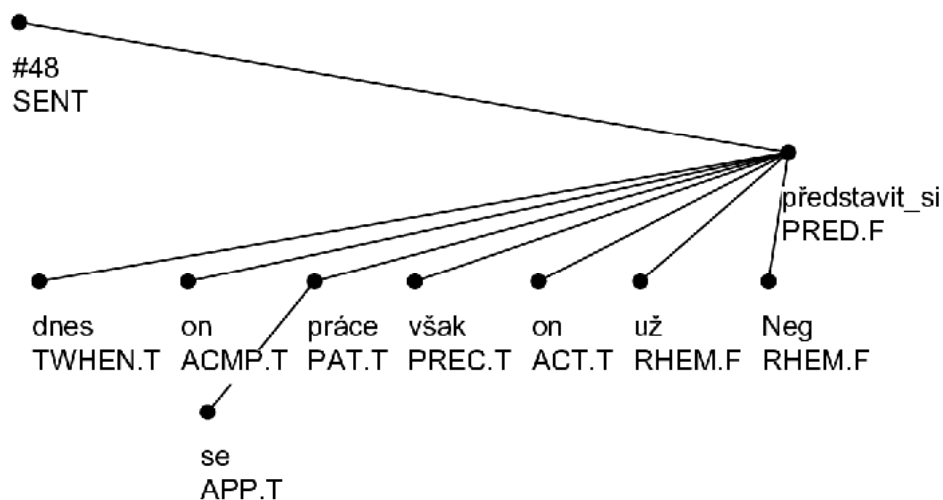
stromové struktury (Tectogrammatical Tree Structures, TGTS) se liší od teoreticky postulovaných tectogramatických reprezentací tím, že z technických důvodů obsahují jednak specifický řídicí uzel pro celou větu (sloužící hlavně pro její identifikaci v korpusu) a jednak u koordinovaných spojení i uzel odpovídající souřadící spojce (k tomuto druhému bodu viz oddíl o koordinaci ve druhé části článku).

Prvním příkladem je věta (a), která kromě analýzy takových jevů, jako je nulová podoba podmětového zájmena, předmětová závislá klauze (ve funkci PAT), všeobecný adresát, restriktivní přívlastek a příslovečné doplnění rozsahu (Extent, EXT) ilustruje i jevy z oblasti aktuálního členění, včetně netypického výskytu kontextově zapojených prvků (kontrastivního *já* a nekontrastivního *to* uvnitř ohniska (jádra, focus, F), viz obr. 1:

(a) Přiznám se, že já osobně to dost prožívám.



Obr. 1.



Obr. 2.

Věta (b) ukazuje, jak chápeme pořadí prvků v *T* (v tom se lišíme např. od postupu A. Svobody, odůvodnění těchto rozdílů podal Sgall (1986): chápeme slovo *dnes* v této větě jako vlastní *T*, nulové zájmeno podmětu (Actor) a spojení bez něho považujeme ze průvodní prvky *T*; specifické pozice slov *už*, *si* a patrně i *však* je možné vysvětlit tím, že jde o příklonky (i když *však* a *už* nemají povahu příklonky ve všech svých výskytech); viz obr. 2.

(b) *Dnes už si však bez něho svoji práci nedovedou představit.*

Hodnoty gramatémů ve stromech neuvádíme, s výjimkou CPL (komplexní, dokonavý vid) v obr. 1. Seznam zkratek je obsažen v Příloze 1.

2. Nejasné hranice mezi jevy významové stavby

Otevřené otázky tektogramatické roviny (TR), vyžadující další monografické zpracování a upřesnění, popř. zjemnění, obohacení dosavadní klasifikace, se týkají zejména valence sloves, tj. třídění závislostních syntaktických vztahů (funktorů) a jejich rozlišování, viz odd.

2.1. Potřeby zjemnit dosavadní předběžné třídění jsme si byli od počátku vědomi, viz zejm. Sgall ad. (1986, s. 161). Z dalších otázek připomínáme v druhé části článku (v odd. 2.2 a dalších) problémy valence substantiv a dalších slovních druhů, dále přímé a nepřímé řeči, aktuálního členění věty a kontrastu, koordinace a reciprocitu, koreference gramatické a textové, zachycení pasiva, i problematiku slovní zásoby a tvoření slov a otázky gramatémů, tj. morfologických kategorií jako vid, modalita, číslo, stupňování.

2.1. Problémy funktorů:

Poznamenejme, že v Manuálu se jako *p r i m á r n í* funkce (se značkou *p*) prostých i předložkových pádů a podřadicích spojek chápou (zatím podle tradice a odhadu) bezpříznakové případy, u kterých nejde o omezení na určitou třídu kontextů; ostatní funkce jsou tam ilustrovány typickými příklady s kontextovými měřítky předběžně odhadnutými. U většiny těchto sekundárních funkcí další výzkum, který je nezbytný, asi potvrdí, že jsou omezeny lexikálně (frazologicky) a že tedy kontexty, které je připouštějí, bude možné vymezit na základě seznamů. Musíme ovšem počítat s tím, že všude nebo téměř všude najdeme zmíněná přechodná pásma, zejm. jevy omezené stylově, generačně, územně, nebo příležitostně individuální odchylky od běžného úzu. Vymezení kontextových měřítek a jejich kontrola jsou věci dalších perspektiv bohemistického výzkumu.

Než přejdeme k otázkám jednotlivých funktorů, poznamenejme, že pracujeme s obecným pojmem *v š e o b e c n é h o* aktantu, tedy vedle všeobecného konatele (typicky vyjádřeného tzv. zvrtným pasívem, např. *O tom se mluví už dlouho*) máme i všeobecný *patiens*, adresát, výsledek a původ, viz Panevová (1992; 1998). Chápeme všeobecný aktant jako specifickou lexikální jednotku (označujeme ji zkratkou *Gen* za angl. *general participant*), která je u aktantů jiných než *ACT* zpravidla vyjádřena nulou, např. *Ještě nemám Gen.PAT uklizeno, Babička ráda vypravuje Gen.ADDR pohádky*. U volných

doplnění předpokládáme možný výskyt jednotky Gen tam, kde je doplnění u daného řídicího slova obligatorní, a tam, kde jde o vztah 'kontroly', viz např. diskusi o Benefaktivu (BEN) v odd. 2.1.2.

Všeobecný aktant asi může být jak tam, kde má sloveso daný aktant ve svém valenčním rámci jako obligatorní, tak tam, kde je fakultativní, jak je tomu např. u fakultativního Adresátu slovesa *prodat* ve větě *Jana prodává Gen.ADDR Gen.PAT u Bati*.

Svou valenci mají i substantiva, a zejména u dějových a konatelských substantiv je snad potřeba s takovým všeobecným aktantem počítat aspoň u obligatorních volných doplnění: chceme-li např. valenci slova *pobyt* vidět jako blízkou té, která je zřejmá u slovesa *pobývat*, pak budeme i u substantiva pracovat s obligatorním doplněním LOC a ve spojeních jako *pobyty dlouhodobé, rekreační* budeme vidět doplnění všeobecné, Gen.LOC.

Vymezení všeobecného doplnění je založeno na tom, že nejde o větný člen referující ke konkrétní obsahové jednotce, ale obecně k jednotkám pro danou valenční pozici typickým (kdo mluví, co se uklízí, komu se vypravuje atd.). Vedle toho jsou ale časté i případy, ve kterých výraz s nulovým vyjádřením (popř. u ACT s obecným vyjádřením 3. os. pl., *oni*) referuje ke konkrétní entitě, kontextem více méně jasně dané, ale slovně ani v něm přímo nevyjádřené. Takovou jednotku označujeme jako Unsp (*unspecified*), např. *U tety Unsp.ACT mu dali najíst*. Diskusi o jednotce Unsp a další ilustrace podávají Řezníčková (2001) a Marková a Panevová (v tisku); její bližší vymezení bude ale možné až po prostudování rozsáhlého materiálu.

Většinou jen stručně charakterizujeme jednotlivé dosud otevřené otázky, aniž bychom uváděli argumenty pro to či ono z možných řešení. Ty se zčásti najdou v dřív publikovaných pracích. V Příloze 2 (u druhé části článku) jsou uvedeny postupy zatím užívané v jednotlivých otázkách, charakterizované v Manuálu.

2.1.1. Hranice patientu:

Dobře se k ilustraci otevřených problémů hodí otázka, jak dalece máme vazebná, r e k č n í, doplnění (tj. ty závislé členy, jejichž tvar je podmíněn slovem řídícím) považovat za PAT (popř. snad za EFF nebo jiný aktant) např. ve spojeních jako *vztahovat se k čemu*, *loučit se s čím*, *pomstít se na kom* (viz už Šmilauer 1947, odd. 62-65 aj.).⁴ Jde o vazby, u kterých v sémantice vztahu závislého slova k řídícímu snad by bylo možné vidět (aspoň na první pohled) překážky pro to, abychom jeho závislostní platnost zařadili jako přímý předmět (dějem zasažený, vytvořený ap.), ale právě vazebná povaha vztahu dobře neumožňuje vidět tu jen příslovečné určení (volné doplnění).

Tak např. rozdíl mezi *píchnout se nůžkami* a *píchnout se o nůžky* můžeme snad zachytit jako rozdíl dvou významů slovesa (dvou Filipcových lexíí), z nichž každému odpovídá jak jiný valenční rámec, tak jiná obsahová (sémanticko-pragmatická, kognitivní) interpretace, i různé způsoby morfématického vyjádření. Jiná možnost je vidět tu rozdíl dvou různých funktorů, z nichž pouze instrumentál odpovídá klasickému volnému doplnění prostředku/nástroje (MEANS). Toto chápání umožňuje pro vazebný typ *píchnout se o něco*, *uhodit se o něco*, *zakopnout o něco* uvažovat o novém typu doplnění (s významem Překážky), který má rovněž kvazivalenční povahu (není volně spojitelný s jakýmkoli slovesem, ale není obligatorní). Konstrukcím s instrumentálem a s předložkovou vazbou odpovídá totiž jiná obsahová (sémanticko-pragmatická, kognitivní) interpretace, nejen různé způsoby morfématického vyjádření. Zřejmý významový rozdíl obou vazeb zahrnuje i možnou různost pravdivostních podmínek: jestli se někdo úmyslně píchne nůžkami, neodpovídá skutečnosti, že by se píchl o nůžky; význam druhé vazby zahrnuje neúmyslnost. U slovesa *zavadit (o něco)* jde ovšem o člen obligatorní (který můžeme chápat jako PAT).

Podobná úvaha se pak bude týkat celé řady dalších příkladů, ve kterých snad nemusíme pracovat s několika lexíemi (s několika významy slova), jako *brousit nůž (PAT) o kámen (MEANS)*, *zavadit lžičkou (MEANS) o sklenici (PAT)*, nebo (spíš s PAT než s CAUS) *umřít na zápal plic, onemocnět chřipkou* atd. Rozdíl mezi *brousit o kámen* a *brousit*

brouskem, i mezi spojeními *psát perem*, *psát na stroji* a *hrát na housle* může být zachycen jako lexikálně (frazologicky) vymezené podmínky pro morfématický způsob vyjádření funktoru MEANS.

Je třeba hledat ověřitelná a obecně platná kritéria pro rozlišení mezi příklady právě uvedenými (a podobnými), ve kterých snad je možné pracovat s rámcem obsahujícím PAT, a jinými, ve kterých předložkový pád (nebo jiný způsob vyjádření) je sice vazebně určen, ale jeho jedinou funkcí je sémanticky zřetelné volné doplnění (má nanejvýš synonymní varianty), jak je tomu u *střílet z revolveru* (MEANS), *z děla* (vazbu *střílet revolverem*, *puškou* snad můžeme ve smyslu právě zmíněné lexikální podmíněnosti výrazu považovat za synonymní se *střílet z*).

Teprve rozbor většího množství příkladů z korpusu umožní lepší orientaci v těchto problémech. Zatím jsou zřetelné jen některé podmínky nutné (obtížné by bylo mluvit o podmínkách dostatečných):

(i) O PAT jde jen tam, kde je způsob vyjádření vymezen rekcí, ať už je to reke akuzativní, nebo jiného pádu, bez předložky (genitiv, dativ, instrumentál) nebo s ní.

(ii) Příslopečné doplnění můžeme vidět jen tam, kde je přítomná jeho typická sémantika; proto je zřejmé, že zejména u předložky *o* s lokálem často jde o PAT: nedovedli bychom říct, jaký druh příslopečného vztahu (volného doplnění) je přítomen u spojení jako *mluvit o něčem*, *přemýšlet o něčem*.

Samo rozlišení, jestli v daném případě jde o rekcii nebo ne, není však vždycky snadné, a zrovna tak není vždycky jasné, můžeme-li mluvit o specifické sémantice některého adverbialního vztahu. V některých případech může pomoci zřetel k tomu, je-li přirozenou dopňovací otázkou, na kterou by daná věta odpovídala, otázka pádová, tedy rekční (např. *Na co se díval?* - *Díval se na tu knihu.*) nebo otázka příslopečná (např. *Kam se díval?* - *Díval se do té knihy.*)

Široké přechodné pásmo mezi patientem a volnými doplněními jistě nemůže být

adekvátně popsáno tím, že budeme počítat s třetí možností (např. "mezi" PAT a MEANS nebo PAT a CAUS), protože hranice ani tak nebude jasná a budeme tu mít jen další přechodná pásma. Jak říká Šmilauer (1947, s. 225), je možné podat "jen výběr z velké rozmanitosti případů; o leckterý z nich by bylo možno – ne však užitečno – se přít." Pro praktické účely týkající se anotování korpusu jsme se však přece jen rozhodli pracovat s pojmem *k v a z i v a l e n c e*, který umožní velkou část takových přechodných jevů předběžně zachycovat jako specifickou oblast a usnadnit tak jejich budoucí studium.⁵ Až bude shromážděno velké množství jejich výskytů v rozsáhlém korpusu a budou se systematicky porovnávat jejich kontexty, objeví se i nové možnosti, jak chování takových vazeb blíže charakterizovat a popř. i klasifikovat.

Jiným tématem pro další výzkum je, zda patří pod kvazivalenci (snad jako funktor jiný než PAT) taky spojení jako např. *zpívat co na povel, otevřít na žádost, na zaklepání*, u kterých je nutné se rozhodnout, máme-li je vidět jako aktant, nebo jako volné doplnění příčiny, CAUS (i když vyjádřené vazebně). Dáváme zatím přednost jejich chápání jako volných, protože patrně nejsou omezena na určitou skupinu sloves.

Bude třeba dál probírat i chápání PAT nebo jiného aktantu ve spojeních jako *stvoření světa; pomník T. G. Masaryka*, tj. širší nebo užší pojetí deverbálních substantiv (ke kterým by popř. byly zařazeny i názvy artefaktů, viz Panevová, 2000).

Ke spojení s objektovou vedlejší klauzí poznamenejme jen docela stručně, že u vět jako *bál se toho, že zaprší* vidíme *zaprší* jako PAT k *bát_se*, odkazovací slovo *toho* nemá v tektogramatickém stromu odpovídající uzel. U *bát se, aby nezapršelo* bude PAT bez negace, tj. zatím jako synonymní s právě uvedeným spojením s *...že zaprší*.

2.1.2. Hranice mezi jinými funktoři:

Obecně je nejasná hranice mezi aktantem ADDR a volným doplněním BEN, a obtížné je

rozhodování zejména tam, kde musí být BEN chápán jako obligatorní; prozatím počítáme s tím, že je třeba doplňovat Gen.BEN u verbonominálního přísudku tam, kde jde o vztah kontroly (tj. kde nevyjádřený subjekt infinitivu, kterému v TGTS odpovídá specifický symbol pro koreferenci, Cor, je referenčně shodný s některým doplněním řídicího slova, v tomto případě s jeho benefaktivem), např. *Je možné Gen.BEN Cor.ACT to udělat, Je náročné (pro Janu) přinést článek už zítra...* (viz Panevová, 1996).

Podobně není úplně jasné, jak analyzovat spojení se slovem *lze/nelze*; např. u *nelze odejít* snad můžeme infinitiv chápat jako vyjádření ACT a vidět tu i Gen.ADDR (nebo Gen.BEN); jiné možné chápání, s infinitivem jako PAT a s Gen.ACT, má tu nevýhodu, že není snadné najít tu konkrétní lexikální obsazení pozice ACT (snad např. *Nebylo mu lze odejít*). U příkladů jako *ta otázka se dá řešit* můžeme infinitiv chápat jako PAT, substantivum v nominativu jako ACT. Jiné ovšem je spojení *tu otázku se dá řešit*, kde lze substantivum v akuzativu považovat za PAT k infinitivu.

Nejasné, v Manuálu zatím jen prozatímní, je i rozlišení mezi rematizátory (např. *taky, i, jen, negace*) a doplněním postoje (Attitude, ATTIT, např. *naštěstí, bohužel*); k upřesnění pohledu tu snad pomůže ta okolnost, že rematizátor může ve větě být v pozici primární, ve které uvádí réma, ohnisko věty (viz odd. o aktuálním členění), nebo sekundární (zejména uvnitř základu věty), kdežto pozice ATTIT je určena jinými hledisky. Ani v dalších ohledech však dosud není dořešena otázka, které výrazy se (typicky? vždy?) chovají jako rematizátory; je např. *a také* jen spojka (když *také* chápeme jako rematizátor)?

K dalším opozicím, které ještě vyžadují podrobnější výzkum, patří rozdíl mezi restriktivním přívlastkem (RSTR) a doplněním identity (IDENT); mezi příklady jako *pan Novák, poslanec Svoboda* na jedné straně a *termín sloveso, pojem subjekt* na straně druhé můžeme snad rozlišovat tak, že při dvou sousedních (shodných, resp. nesklonných) substantivech bude IDENT (závislé zprava) výraz (i) nesklonovaný (např. *parník Hradčany*) a (ii) který může být synonymně zaměněn genitivem, jako *pojem subjekt/u*; u ostatních dvojic shodných substantiv (tj. snad u všech, kde je shoda i mimo nominativ a genitiv) má

funktor RSTR slovo závislé primárně zleva. I tak ovšem zůstává nejistota u dvojice jméno - příjmení: *Lojzík Vomáčka* má jistě blízko k *Lojzík Vomáčků* (se závislým slovem vpravo, u kterého bychom mohli vidět i vztah posesivity v širším slova smyslu, tedy přináležitosti, Appurtenance, APP), ale naproti tomu ve spojení *Alois Vomáčka* vidíme (aspoň prozatím) *Alois* jako závislé, totiž RSTR (nemluvíme tu o apozici, protože tu chápeme v Mathesiově smyslu jako specifický syntaktický vztah, přítomný jen u volnějších spojení, tedy u typu *Václav, král český*, ne u typu *král Václav*).

Dosavadní uplatnění funktoru COMPL (Complement, doplněk) je třeba chápat jen jako předběžné řešení. Hlubší zpracování doplnku si vyžádá i analýzu uplatňující vztah 'kontroly', a to jak tam, kde je takový koreferenční vztah vyjádřen shodou (např. *Vrátil se poraněn*), tak i tam, kde takové vyjádření není (*Našel ho s pistolí v ruce*, *Uvidíš ho osobně* bude asi analyzováno jako dvojznačné, totiž podobně jako *Našel ho jsa s pistolí v ruce* a *Našel ho jsoucího s pistolí v ruce* apod.). K oblasti doplnku patří i jiná otázka vyžadující další zkoumání a diskuse, totiž Fr. Kopečným iniciovaná analýza stupňovitěho vztahu mezi přívlastkem a doplňkem; srov. i možné postavení adverbia ve funkci doplnku: *děti mají vstup zdarma* (stejně jako *laciný*), *Šel naboso* (stejně jako *bos*).

Jiné náměty, směřující hlavně k dalším diskusím mezi bohemisty a k vyšší míře sjednocení stanovisek i terminologie, jsou:

- (a) způsob (Manner, MANN, např. *jde rychle*) a postoj (Attitude, ATT, např. *bohužel nepřišel*),
- (b) přívlastek nerestriktivní (deskriptivní, DES, např. *stověžatá Praha*), restriktivní (RSTR, např. *bohaté město*), omezení (RESTR, např. *přišli všichni až na něho*) a doplnění přípustkové (Concession, CNCS, např. *přišel, byť pozdě*),
- (c) přirovnání (Comparison, CPR, např. *vysoký jako strom, vyšší než strom*) a rozdíl (Difference, DIFF, např. *o metr vyšší*),

- (d) doplnění účinku (Result, RESL, např. *bylo tam plno, až omdlěli*) a souřadné spojení důsledkové (Consequence, CSQ).

Řada nepravých předložek a složených spojovacích výrazů potřebuje podrobnější třídění svých tektogramatických protějšků (významů), než jaké bylo zatím možné. Budou tu asi nutné i nové funktoři, popř. syntaktické gramatémy; srov. např. otázky, zda je možné spojení jako *pro případ potřeby* zařadit pod doplnění účelu (AIM), nebo jak hodnotit spojení jako *Šel v čele průvodu, ...na základě čeho, ... s ohledem na, ... se zřetelem k* aj.

Bude nutné soustavněji, než to bylo dosud možné, zpracovat i intenzifikátory (INTF), o kterých se zatím v Manuálu pojednává jen docela předběžně.

K Poldaufovu doplnění záměru (Intent, INTT) asi patří (vedle typických příkladů jako *šel nakupovat, půjde na jahody*) i spojení *byl na jahodách* (v jednom ze svých významů), *na kurzu, na prázdninách*. Doplnění INTT a LOC mohou být ve větě vedle sebe, nedají se koordinovat (pokud se zdánlivě dají, např. *stávalo se mu to na jahodách a na plovárně*, dáme asi přednost chápání se dvěma koordinovanými LOC); není ale jasné, jak zachytit vztah mezi INTT a směrovým doplněním typu *kam* (DIR3).

V oblasti časových doplnění jsme rozlišovali už ve starší podobě FGP (viz Panevová, Benešová, Sgall, 1971) mezi těmito funktoři:

TWHEN – kdy (s gramatémy NIL, AFT, BEF, JBEF), TSIN – odkdy, TTILL – dokdy, THL – jak dlouho, TFHL – na jak dlouho, TPAR – čas během.

Během práce na syntaktickém anotování PDT jsme dodali funktoři THO (Temporal – How Often), 'jak často', a na základě argumentace Grepla a Karlíka (1998) i dva další funktoři, totiž TFRWH (T. From When) – 'ze kdy' a TOWH – (T. On When) - 'na kdy'.

Jen stručnou poznámku tu můžeme věnovat vedlejším klauzím časovým a lokálním (k těm srov. Panevová, ??19). Analyzujeme je jako spojení s odkazovacími slovy, "korelativy" (*kde, kdy, odkdy, poté, co* apod., popř. vyjádřenými substantivními spojeními se slovy *místo, doba*); druh závislosti vedlejší klauze je zachycen jako funktoři jejího slovesa,

např.: *Přijeli odtamtud, kde.LOC se stalo.DIR1 to neštěstí, Bydleli tam do té doby, odkdy.TSIN začaly.TTILL ty problémy s jejich sousedy.* Otevřená je ovšem otázka, jak určit hranici mezi takovými korelativy a substantivní platností spojení jako *od té doby, co.*

2.1. 3. Vztah dvou lokálních nebo časových příslovečných doplňků

Problém pro adekvátní zachycení v závislostní syntaxi představují konstrukce typu *vzadu u vchodu, kilometr od řeky, letos v únoru, týden po Vánocích.* Šmilauer (1947) část z nich považuje za apozici (specializující, srov. 1947, s. 363), část pak spadá pod modifikaci místa a času (tzv. místní a analogicky by se jistě dalo mluvit i o časové míře, srov. Šmilauer, 1966, s. 259). Protože apozici ve FGP (a tedy u v PDT) vymezujeme mathesiovsky, tedy jako skrytou predikaci, nemůžeme spojení typu *letos v únoru* chápat apozičně. Konstatování, že jde o časovou, resp. lokální míru, je značně neurčité a nelze je přímo aplikovat při závislostní analýze. Teoreticky nelze souhlasit ani s Hrbáčkovou (1995, s. 76) analýzou, že ve větě *Vlevo u vchodu (stála skříň)* jde u neuzávorkované části o “souřadnost determináční”, jednak proto, že sám termín představuje z našeho hlediska kontradikci, jednak proto, že stejně není ve všech případech jasné, co je uvnitř této konstrukce determinováno čím. Štěpán (1989) navrhuje “složená příslovečná určení času a místa” chápat jako “slabou řetězcovou závislost na slovese”; tyto konstrukce pak podle Štěpána (op.c., s. 14) vytvářejí „syntaktickou jednotku vyšší, než jsou skladební dvojice“. Pojem „řetězcové závislosti“ nepokládáme za natolik průkazný nebo systémový jev, aby stálo za to kvůli němu upravovat formální rámec závislostního popisu. Z názorů předložených v citovaných pracích plyne jednoznačně skutečnost, že tato „časová a místní syntagmata“ mají svou specifickou povahu a že se je nedaří jednoduše popsat klasickými prostředky závislostní syntaxe.

Pokusili jsme se proto aplikovat na uvedené konstrukce kritérium vypustitelnosti, srov. metodu redukční analýzy, formálně zpracovanou M. Plátkem v Jančar ad. (1999) a

zjišťovat tímto způsobem, které z těchto konstrukcí se chovají endocentricky (jsou zastupitelné jednou ze svých složek a zachová se přitom gramatičnost) a které mají povahu exocentrickou (distribuce celku se nekryje s distribucí žádné jeho součásti, nelze v nich tedy vypustit nic, chceme-li gramatičnost zachovat), popř. kdy lze vypustit kteroukoli složku výrazu. Budeme analyzovat tímto způsobem zvlášť syntagmata č a s o v á (A) a zvlášť lokální (B).⁶

A. (i) (1) Stalo se to/Udalo se to/Bylo to (dva měsíce) před Vánoci

(2) Přijďte (půl hodiny) před začátkem představení

(3) Odvezli ji do nemocnice (měsíc) před porodem.

(4) Tu budovu postavili (dva roky) po válce.

(ii) (5) Zůstaňte v nemocnici dva měsíce (před porodem).

(6) Strávila tam dvě hodiny (po zákroku).

(7) Budova tam stála už tři roky (před válkou).

Konstrukce v obou skupinách (i) a (ii) v odd. A se chovají endocentricky (vypustitelnost je označena závorkami), avšak ve skupině (i), kde jde o určení času TWHEN (kdy), lze vypustit “časovou míru”, zatímco ve skupině (ii) obsahující určení THL (jak dlouho) je vypustitelný předložkový časový výraz. Jako řídicí se tedy na základě této metody jeví v (i) vlastní časové určení a na něm je závislý jeho modifikátor, “časová míra”. Ve skupině (ii) je tomu právě naopak: gramatičnost věty se zachovává, když vypustíme vlastní časové určení, jako řídicí se tu tedy jeví určení “časové míry”.

B. U m í s t n í c h syntagmat je situace složitější, protože jde často buď o konstrukce exocentrické, nebo o konstrukce, v nichž lze beze ztráty gramatičnosti vypustit kterýkoli z členů syntagmatu. Zjišťujeme dále, že hranice nevede mezi určeními lokálními (LOC) a

směrovými (DIR), což by se analogií podle A dalo očekávat.

(i) (8) Leží to/Nachází se to/Je to (pět metrů) pod povrchem.

(9) Přesunul lavičku (kousek) za zahrádku.

(10) Postavil altán (těsně) u stanice.

(ii)(11) Přihodilo se to daleko (od civilizace).

(12) Vlevo (od vchodu) stála skříň.

(13) Bydlí blízko (od stanice metra Háje).

(iii) (14) Leží to/Nachází se to/Je to (hluboko) (pod povrchem).

(15) Přestěhoval se (daleko) (za město)

(16) Dítě se posunulo (blíže) (k matce).

(iv) (17) Sejdeme se (na Hlavním nádraží) (v hale).

(18) Lopata leží (na zahradě) (u plotu).

(19) Našel to (pod stolem) (na zemi)

Výrazná je odlišnost chování příkladů ze skupině B (i) a (ii); v (i) se jeví jako řídicí vlastní lokální určení a závislé (vypustitelné) je určení jeho míry. Skupina (ii) se chová právě opačně. Toto chování se však zdá být specifické pro předložkový výraz s *od+Gen*, kdy věta zůstává gramatická pouze při zachování příslovce; Hrbáček (1995) tu mluví o „příslopečných určeních ve vztahu závislosti“ na rozdíl od *vlevo u vchodu* (kde, jak uvádíme výše, mluví o „determinační souřadnosti“. Specifické chování předložky *od* po příslovcích, s nimiž se pojí a jichž není mnoho (*vlevo, vpravo, daleko, blízko, jižně, západně,...*, popř. i *na jih, na západ,...*) připomíná valenci. Ve skupině (iii) se konstrukce sice jeví endocentricky; přímý návod, který člen je řídicí a který závislý, však redukční analýza neposkytuje. Budeme tu

postupovat analogicky jako v B (i), protože tyto dvě skupiny se od sebe liší nepodstatně (stačí vyměnit adverbium v (iii) za měrový akuzativ a získáme konstrukci typu (i)). Budeme tedy i ve skupině (iii) pokládat za řídicí lokální/směrové určení a příslovce za jeho modifikátor. Rovněž ve skupině (iv) nám redukční analýza nedává jednoznačné výsledky, zde se však setkávají dva předložkové výtazy místní, kde jejich syntagmatická sounáležitost je (v porovnání s (i)-(iii)) značně oslabena, a analyzujeme je proto jako dvě (sesterská) na sobě nezávislá lokální určení. O směrových by platilo analogické tvrzení (srov. (20)):

(20) Položil to na postel pod polštář.

Řešení navrhovaná v tomto oddílu bude třeba v průběhu anotování ověřit na rozsáhlejším materiálu, neděláme si zde nároky na vyčerpávající přehled těchto typů konstrukcí.

2.1.4. Další otázky valence slovesa

Ze zbývajících témat můžeme ještě připomenout potřebu jemnějšího členění funktorů a popř. gramatémů. Bude zřejmě třeba rozdělit některé dosavadní funktoři, zejména tam, kde můžeme rozlišit vztah sémanticky volnějšího a těsnějšího určení; srov. už diskusi o tom ve Sgall ad. (1986, s. 161) a zde v odd. 2.1.1 o možnosti uplatnění opozice mezi doplněním Prostředku a Překážky.

Diskutovalo se už o možnosti pracovat se dvěma druhy místního doplnění (LOC); pro to mluví jednak příklady, ve kterých se těsné doplnění blíží Způsobu (*zranit se na ruce, bydlet ve stanu*) a jednak možnost společného výskytu těsného a volného LOC v téže klauzi, bez možnosti koordinovat oba LOC: *V koupelně se zranila na ruce, V Bulharsku jsme bydleli ve stanu*. Podobný dvojí výskyt LOC je však možný i bez rozdílu v těsnosti, jak to ostatně

odpovídá obecné charakteristice volných doplnění v FGP, srov. např. *V Bulharsku jsme bydleli u moře*, kde vidíme dvě lokální doplnění stejného druhu, podobně jako u *Na ruce se zranil na dlani* (srov. příklad (20) v odd. 2.1.3). Další otázka je, jestli by se pak příklady jako *jednání uvnitř koalice* měly řadit jako volný nebo těsný LOC.

Vztah mezi těsnějším a volnějším doplněním by se mohl rozlišovat i u funktoru Substituce (SUBST), viz např. *Zaplatila za dceru za/místo jejího otce taneční*. Není zatím jasné, jestli je vhodné takové případy chápat prostě jako dvojí výskyt téhož volného doplnění, s tím, že vztah větší a menší těsnosti je gramaticky strukturován jen na úrovni aktuálního členění (výpovědní dynamičnosti).

Podobně obtížné je i rozlišení mezi *pojmenovali ho Jiří* (Způsob nebo EFF) a *pojmenovali ho po otci* (snad s funktořem pro dědictví, Heritage, HER).

Připomeňme, že opozice funktořů CRIT (kritérium, např. *podle jeho slov*) a NORM (*podle pravidla palce pravé ruky*), se kterou při anotacích pracujeme, už podobné rozlišení odráží.

K tématům pro další analýzu patří i problémy dané tím, že vztah mezi stavem (zejm. výsledným) a dějem mění některé rysy valence u tvarů téhož slovesa, viz např. různé preference u spojení jako:

- (a) *ukládá peníze do banky* (i v *bance*)
- (b) *peníze jsou/má uloženy v bance* (i *do banky*)
- (c) *započítat něco do ceny*
- (d) *něco je započteno v ceně*

Podobně je různost valence někdy spojena s opozicí vidu, jak je už dlkouho známo: u *přečíst* je PAT nevypustitelný, kdežto u *číst* může mít nulovou podobu (zejména při vyjádření všeobecného PAT).

Zmiňovali jsme se už o tom, že nepravé předložky si časem patrně vynutí rozlišení dalších funktořů a syntaktických gramatémů.

Jak už řečeno, množství otázek dosud otevřených – ať už skutečně nevyřešených, nebo takových, ve kterých zatím nebylo dosaženo shody mezi bohemisty, ukazuje na užitečnost další práce s textovými korpusy vůbec, a zejména s korpusy anotovanými, tedy u nás především s PDT.

Oddíly 2.2 a další, stejně jako část 3 a závěrečné úvahy v odd. 4 i Příloha 2 budou otištěny v příštím čísle SaS.

Příloha 1: Seznam zkratk ...

Poznámky:

* Základem tohoto článku je práce na teoretickém popisu češtiny v rámci projektu MŠMT ČR LN00A063. Za cenné příspěvky k tomuto projektu, o které se tu opíráme, děkujeme Alě Bémové, Evě Buráňové, Janu Hajičovi, Veronice Řezníčkové, Zdeně Uřešové a dalším účastníkům.

1 Informace o ČNK přináší ve stručné podobě Čermák (1995), podrobněji viz zejm. Kocek ad. (2001). O významu korpusů pro lingvistiku obecně a zvláště pro posuzování otázek “gramatičnosti” píše Štícha (2001), který ukazuje, jak uplatnění ČNK pomáhá bohemistický výzkum prohlubovat. Připomeňme však, že není třeba zacházet tak daleko, abychom např. po zjištění “nulového výskytu” kataforického (odkazovacího) *to* u některých sloves řídicích “vedlejší větu nebo infinitiv” jednoznačně věty s nimi “vzhledem k porušení pevné, reálně existující a zjistitelné normy označili za negramatické” (s. 171). Lingvisté uplatňující korpusová zjištění přece jen nejsou v pozici archeologů, nejsou odkázáni jen na korpus a mají možnost např. (nejen podle vlastního povědomí, ale i na základě testů se skupinami mluvčích) zjišťovat, zda je spojení s infinitivem jako **Odmítl to, odpovědět jim...* z tohoto

hlediska na stejné úrovni jako *Odmítl to, že jim odpoví...*

2 O analytické rovině, která je technickou, pomocnou složkou anotační soustavy, informuje Hajič (1998); podrobně jsou její jednotky specifikovány v příručce (Bémová ad., 19??, která je přístupná na webové stránce

http://ufal.mff.cuni.cz/pdt/Corpora/PDT_1.0/Doc/amaual/index.html). O tektogramatickém anotování viz zejm. ... SaS, sb.JP, TSD... a podrobnou příručku pro převod z analytické syntaxe na tektogramatickou rovinu Hajičová ad., 2001, na stránce

http://shadow.mff.cuni.cz/pdt/Corpora/PDT_1.0/Doc/tmaual/tmancz.pdf, popř. ...rtf). V teoretickém popisu jazyka je možné počítat s jedinou rovinou větné stavby (s rovinou stavby významové, hloubkové) a vyjadřovací prostředky jejích jednotek vidět na rovině morfématické (Sgall, 1992).

3 K námitkám, se kterými se valenční přístup setkává (viz zejm. Karlík, 2000) můžeme poznamenat, že při chápání valence jako jevu významové (hloubkové) větné stavby velká část problémů přestává být akutní. I my jsme si ovšem vědomi toho, že valence je v základě záležitost slovníku (především jeho tektogramatické složky) a v syntaxi se projevuje kombinatorickými vlastnostmi slov ve větě.

4 Šmilauer uvádí tyto vazby (a mnohé jiné) jako příklady předmětu, ale užívá tento termín ve velmi širokém významu. Jako předmětový označuje např. i dativ prospěchový (odd. 644, např. *V Litomyšli napsal Jirásek spolkům nejeden proslov*), který chápeme jako příslovečné doplnění, BEN.

5 K pojmu kvazivalence viz Straňáková-Lopatková a Žabokrtský (v tisku); autoři pro praktické účely anotování pracují (nejen v uvedené stati, ale zejména ve svém podrobném, zatím nepublikovaném rozboru syntaktických vlastností českých sloves) i s pojmem

'typického' (tj. nezcela obligatorního, ale pro syntaktické vlastnosti řídicího slova charakteristického) doplnění (např. *jít někam, slyšet nějak*). K upřesnění snad může sloužit chápání kvazivalence jako oblasti na pomezí aktantů a volných doplnění, zatímco typická doplnění jsou doplnění volná, charakterizovaná svou poměrně vysokou četností u daného řídicího slova v určitém významu.

6 Cennými připomínkami k diskusi o popisu těchto konstrukcí přispěli posluchači gramatického semináře J.Panevové ve školním roce 2001/02, náš vděk patří zejména J.Dotlačilovi a K.Součkové.

LITERATURA:

BÉMOVÁ, A. – BURÁŇOVÁ, E. – HAJIČ, J. – KÁRNÍK, J. –

PAJAS, P.- PANEVOVÁ, J. – ŠTĚPÁNEK, J. – UREŠOVÁ, Z.

(1997): Anotace na analytické rovině: návod pro anotátory.

Technical Report ÚFAL TR-1997-03, Univerzita Karlova Praha.

Buráňová, E., Hajičová E. a P. Sgall (2000). Tagging of very large corpora: Topic-Focus articulation. In: *COLING Proceedings*, 139-144. Saarbrücken, Universität des Saarlandes.

Čermák, Fr.: Jazykový korpus: Prostředek a zdroj poznání. *SaS* 56, 1995, s. 119-140.

GREPL, M. – KARLÍK, P. (1998), *Skladba češtiny*, Votobia, Olomouc.

HAJIČ, J. (1998): Building a Syntactically Annotated Corpus: The Prague Dependency Treebank. In: *Issues of Valency and Meaning*. Karolinum, Praha, 106-133.

HAIČ, J. – HAIČOVÁ, E. – PANEVOVÁ, J. – SGALL, P. (1998): Syntax v českém národním korpusu. *SaS* 59, 168-177.

Hajičová, E.: Presupposition and allegation revisited. *Journal of Pragmatics* 8, 1984, s. 155-167.

Hajičová, E.: Postavení rematizátorů v aktuálním členění věty, *SaS* 56, 1995, s. 241-251.

Hajičová, E., Oliva K. a P. Sgall: Odkazování v gramatice a v textu. *SaS* 48, 1987, s. 199-212.

HAIČOVÁ, E., PANEVOVÁ, J., SGALL, P. (2001): Manuál pro tektogramatické značkování. Verze IV. Pracovní materiály ÚFAL a CKL MFF UK (viz i anglické znění předchozí verze: A Manual for Tectogrammatic Tagging of the Prague Dependency Treebank. Technical Report TR 2000-09, ÚFAL/CKL MFF UK).

HRBÁČEK, J. (1995): Skladba. *In: Hrbáček J. ad.* (1995), 59-114.

HRBÁČEK, J., HRDLIČKOVÁ, H., MAREŠ, P., SERVÍTOVÁ, J. (1995): Mluvnické rozborů a cvičení v češtině. Katedra českého jazyka FF UK Praha.

JANČAR, P. – MRÁZ, F. – PLÁTEK, M. – VOGEL, J. (1999): On Monotonic Automata with the Restart Operation. *Journal of Automata, Languages and Combiatorics, Vol. 4, No. 4, 287-311.*

KARLÍK, P. (2000): Hypotéza modifikované valenční teorie. *SaS* 61, 170-189.

MARKOVÁ K. - PANEVOVÁ, J. (v tisku): Eshche raz po povodu nulevyx elementov v strukture predlozhenija. *In: Festschrift fuer V.S.Xrakovskij zum 70. Geburgstag. Sankt-Peterburg.*

PANEVOVÁ, J. (1974-75): On Verbal Frames in Functional Generative Description. Part I. PBML 22, 3-40, Part II, PBML 23, 17-52.

PANEVOVÁ, J. (1975): Tzv. vedlejší věty místní a jejich významová stavba. SaS 37, 284-290.

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

PANEVOVÁ, J. (1992): *O nekotoryx tipax obobschennyx aktantov*. In: Festschrift für V.Ju. Rozenčevjg zum 80.Geburtstag. Wiener Slawistischer Almanach 33 (ed. T. Reuther). Wien, 73-78.

PANEVOVÁ, J. (1996): More Remarks on Control. In: Prague Linguistic Circle Papers, Vol. 2 (red. E.Hajičová, O.Leška, P.Sgall, Z.Skoumalová). John Benjamins, Amstredam-Philadelphia, 101-120.

PANEVOVÁ, J. (1998): Ellipsis and zero elements in the Structure of Sentence. In: Tipologija, grammatika, semantika . K 65-letiju V.S. Xrakovskogo (eds. N.A.Kozinceva, A.K.Ogloblin). Nauka, Sankt-Peterburg, 67-76.

PANEVOVÁ, J. (2000): Poznámky k valenci podstatných jmen. In: Čeština – univerzália a specifika 2 (red. Z.Hladká, P.Karlík). Masarykova univerzita Brno, 173-180.

PANEVOVÁ, J. – BENEŠOVÁ, E. – SGALL, P. (1971): Čas a modalita v češtině. AUC – Philologica Monographia 34.

ŘEZNÍČKOVÁ, V. (2001): PDT – Two Steps in Tectogrammatical Annotation. Předneseno

na výročním zasedání SLE, Lovaň.

Straňáková-Lopatková, M. a Žabokrtský Zd.: Valency dictionary of Czech verbs: Complex tectogrammatical annotation. V tisku (vyjde ve sb. ???)

ŠMILAUER, V. (1947): Novočeská skladba. Ing. Mikuta, Praha (citováno podle 2. vydání, SPN Praha, 1966)

ŠTĚPÁN, J. (1989): K složeným příslovečným určením prostoru a času. SaS 50, 10-14.

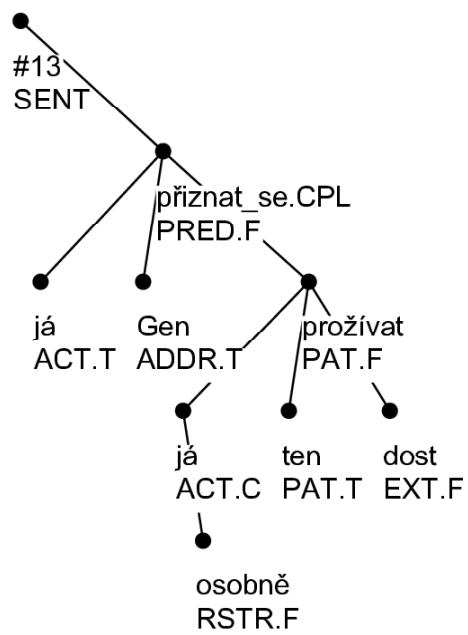
ŠTÍCHA, FR. (2001) Kritéria gramatičnosti (Korpus jako argument a inspirace). SaS 62, 161-175.

Resumé

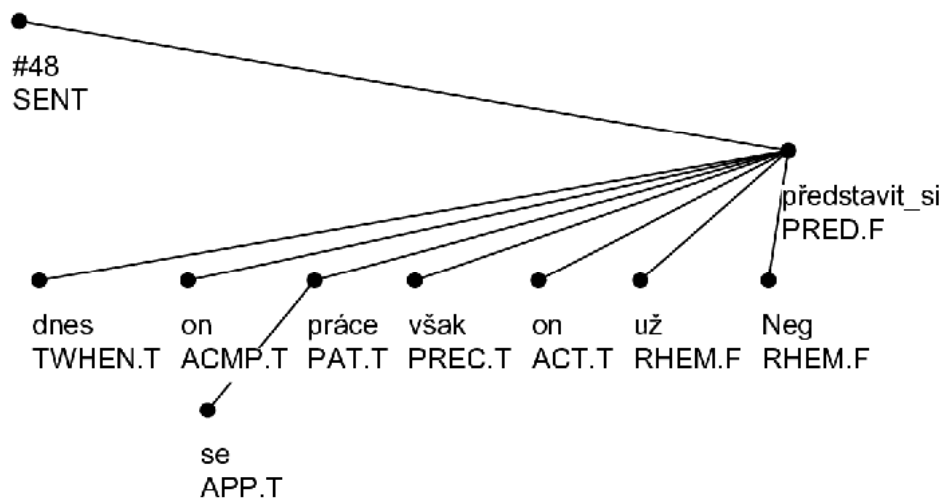
Towards a new level of work in the study of Czech: Working with an Annotated Corpus.

Part 1.

In the Prague Dependency Treebank, a part of the texts from the Czech National Corpus is being annotated on several layers, including the underlying (tectogrammatical) representations. The usefulness of such a treebank is briefly characterized and a large set of topics is discussed for which further monographical research appears to be necessary. The future discussion and elaboration of these topics can be carried out much more effectively with the use of the annotated corpus, and the results thus gained may then serve to an enrichment of the descriptive framework and of the annotation procedure.



Obr. 1.



Obr. 2.