

# Používání PDT-Vallexu pro analytickou anotaci

Verze 1.0, 20.6.2007, Zdeňka Urešová

## 1. Účel

Účelem této malé „pomocné“ příručky je zvýšit konzistenci anotace na analytické rovině tam, kde tomu může pomoci existence valenčního slovníku sloves PDT-Vallex. Jedná se zejména o určování analytických funkcí (atribut a fun) větných členů závislých na slovese. Především jde o rozlišení funkcí Obj a Adv (případně i Atv, AtvV, Pnom a „dvojitých funkcí s Adv a Obj) a o přiřazení funkcí ke slovu „se“, které (analyticky) závisí na slovese. Při analytické anotaci vám tedy PDT-Vallex příliš nepomůže se stavbou analytického stromu, ale můžete ho používat jako pomůcku při přiřazování analytických funkcí.

Tato příručka nic nemění na pokynech a pravidlech analytického anotačního manuálu – ten je stále v platnosti, pokud to přímo pro projekt ČAK nebylo určeno jinak nebo pokud manuál nebyl doplněn dalšími pokyny. Jedná se tedy opravdu pouze o zpřesňující dodatek. V případě jakýchkoli pochybností při anotaci platí analytický manuál; budeme samozřejmě rádi, když nám ohlásíte jakékoli nesrovnalosti (např. když pokyn zde uvedený je podle vás v rozporu právě s analytickým manuálem).

## 2. PDT-Vallex

Valenční slovník PDT-Vallex byl vytvořen jako pomůcka pro anotování na tektogramatické rovině. Je v něm tudíž zachycena sémantická (hloubková) valence sloves, která se používá na tektogramatické rovině. Zároveň ale valenční rámec sloves obsahuje i povrchovou valenci jednotlivých valenčních členů na analytické (povrchové) rovině.

**Sémantická valence** je zachycena pomocí funktorů. Funktory se podle typu závislosti dělí na aktanty (aktor - ACT, adresát - ADDR, patient - PAT, efekt - EFF a origo - ORIG) a na volná adverbialní doplnění (např. místa - LOC, DIR, času - TWHEN, THO..., způsobu - MANN, ...). Jako aktanty se také chovají CPHR (pro doplnění, které je součástí složeného predikátu) a doplnění pojmenovaná jako DPHR (pro doplnění, které je součástí frazému).

**Povrchová valence** je zachycena pomocí morfematické realizace formy. Ta se zapisuje do závorky za funktorem, ke kterému se příslušná forma vyjádření vztahuje. U aktantů je zápis této formy povinný, u volných doplnění se forma zapisuje jen tehdy, je-li pro daný význam funktoru neobvyklá. To znamená, že většinou je v závorce za funktorem „prázdná“, která indikuje obvyklou povrchovou realizaci pro dané volné doplnění.

Nejběžnější formy jsou: přímý pád: např. ACT(1) – aktor je v nominativu, předložkový pád: např. PAT(na+4) – patient je v akuzativu s předložkou „na“, vedlejší věty spojkové: PAT(↓že) – patientem je vedlejší věta připojená spojkou že, PAT(↓c) – patientem je vedlejší věta obsahová atd.

Pro každý **význam** slovesa obsahuje slovník jiný **valenční rámec**. Vlastní valenční rámec je složen z výše uvedených komponent a obsahuje tedy: pojmenování valenčního(-ch) členu(-ů) (funktor(-y)), jeho(jejich) povrchovou realizaci (nebo několik možných alternativ povrchové realizace) a vyznačení případné fakultativnosti valenčního doplnění na hloubkové sémantické rovině (standardně se každý člen rámce považuje za sémanticky obligatorní).

V tištěné verzi PDT-Vallexu je heslo (sloveso v daném významu) vždy uvozeno **tučně vysázeným** tzv. *t\_lematem* (infinitivem slovesa, případně doplněným podle významu *\_se* nebo *\_si*) s horním indexem označujícím pořadí významu v rámci jednoho *t\_lematu*. Následuje vlastní valenční rámec vysázený písmem podobným typu Arial (bezpatkové) – tj. posloupnost FUNKTORŮ (velkými písmeny) a za nimi (v závorce) forma vyjádření (nebo několik alternativních vyjádření formy, oddělených středníkem ‘;’), případně ještě indikace fakultativnosti (otazník ‘?’ před jménem funktoru). Na novém řádku v rámci hesla je pak (v závorce) písmem podobným Times New Roman uvedeno synonymum nebo jiné slovo vhodné jako „definice“ daného významu. Příklady jsou pak uvedeny *kurzívou*. Tištěná forma slovníku obsahuje i údaj o frekvenci rámce (tj. daného významu slovesa ve formě <frekv.>x) v Pražském závislostním korpusu, ale ten je opravdu jen pro orientaci a obvykle jde o velmi malé číslo, které nemá průkaznou vypovídací hodnotu.

### 3. Využití PDT-Vallexu pro analytickou anotaci

Na rozdíl od mnoha tektogramatických funktorů se na analytické rovině musíte rozhodnout jen mezi několika možnostmi, a to zda dané valenční doplnění je subjektem - Sb, objektem - Obj, příslovečným určením - Adv, doplňkem - Atv/AtvV nebo je to Pnom.

Přitom je důležité si uvědomit, že v prototypickém případě (aktivní slovesná konstrukce bez diatezí a jiných „transformací“) platí následující vztahy mezi funktoři a analytickými funkcemi:

**ACT** většinou odpovídá analytické funkci subjektu (Sb), ale v neprototypických případech může aktor odpovídat i objektu (Obj)

**ADDR** většinou odpovídá analytické funkci objektu (Obj)

**PAT** většinou odpovídá analytické funkci objektu (Obj), ale v neprototypických případech může patient odpovídat i subjektu (Sb). Pokud je PAT valenčním doplněním slovesa být, může jít o analytický predikát nominální (Pnom).

**ORIG** odpovídá analytickému objektu (Obj)

**EFF** bývá většinou objektem (Obj)

**CPHR** většinou odpovídá analytické funkci objektu (Obj)

**DPHR** může nabývat různých analytických funkcí, je třeba se řídit především povrchovým vyjádřením a ostatními pravidly analytického anotačního manuálu; například slovo v nominativu bude typicky subjekt (Sb), slovo v genitivu, dativu a akuzativu (pokud je samozřejmě přímo nebo přes předložku nebo spojku závislé na slovese) bude objektem (Obj), slovo v instrumentálu nebo předložková fráze bude nejspíše příslovečným určením (Adv). Mohou ale nastat i jiné případy (Pnom u být, Atv/AtvV u doplňků apod.). Jinými slovy, na doplnění slovesa (jedno nebo někdy i více), které jsou pro tektogramatickou rovinu popsány jako frazeologické spojení a uvedeny ve slovníku s funktorem DPHR, pohlízejte tak, jako kdyby šlo o obvyklou kompoziční frázi, a označení „DPHR“ tím pro analytickou anotaci vlastně „ignorujte“.

**Volná doplnění** obligatorní jsou uvedena v rámci, neobligatorní jsou někdy uvedena jako pomůcka v příkladu u valenčního rámce. Všechna volná doplnění jsou analytickým příslovečným určením (Adv) nebo, v případě doplňkového určení – COMPL, doplňkem (Atv/AtvV).

#### 4. Příklady vztahů (aktivní slovesná vazba):

##### ACT může být

Sb:	spát	ACT(1) spím
	brát	ACT(1) PAT(4) brát zaměstnance
Obj:	bušit	ACT(3) PAT(1) buší mi srdce
	zábst	ACT(4) PAT(1) zebou ho nohy

##### ADDR může být

Obj:	učinit	ACT(1) PAT(4) ADDR(3) učinit jí radost
	udělit	ACT(1) PAT(4) ADDR(3) udělil jim medaili

##### PAT může být

Sb:	bolet	ACT(4) PAT(1) bolí ho srdce
Obj:	clít	ACT(1) PAT(4) clít zavazadla
Pnom	být	ACT(1) PAT(...a1;.a7..) být zdravý

##### ORIG může být

Obj:	brát	ACT(1) PAT(4) ?ORIG(od+2) brát od nich dary
Adv:	být	ACT(1) dárek je od přítele

##### EFF může být

Obj:	členit	ACT(1) PAT(4) ?EFF(na+4;do+2;v+4) členit svět na dva tábory
------	--------	---

##### CPHR může být

Obj:	poskytovat	ACT(1) CPHR(dotace, informace,...4) ADDR(3)
	poskytovat jim informace	

##### DPHR může být

Čímkoliv, např.

Sb: předcházet DPHR(pýcha.S1,pád.S4) pýcha předchází pád  
 Obj: balit ACT(1) DPHR(fidlátka.P4) balím tu fidlátka  
 Adv: běžet ACT(1) DPHR(jako[na-1[drátek.P6]];jako[po-1[drátek.P6]]) program běží jako po drátkách

### COMPL může být

Atv/AtvV: cítit se ACT (1) PAT(7;být[.1];být[.7];.a1; .a7; .f)  
 jako žena.COMPL se cítila diskriminována

### LOC je vždy:

Adv: dát ACT(1) PAT(4) LOC() dát přiznání na úřadě

### DIR3 je vždy:

Adv: dát ACT(1) PAT(4) DIR3() dát květiny do vázy

### TWHEN je vždy:

Adv: uskutečnit se ACT(1) akce se uskutečnila minulý týden

Valenční rámce zachycují prototypické formy valenčních doplnění, tj. takové formy, kterých doplnění nabývají, pokud doplňují sloveso v aktivu. To znamená, že valenční rámce nezachycují změny v povrchové formě valenčních členů, pokud k této změně dojde na základě nějakého obecně platného gramatického pravidla. V souvislosti s tím je třeba mít na paměti, že na tektogramatické rovině mají valenční doplnění stále stejnou funkci. Výraz, který je aktorem v aktivu, je aktorem i v pasívu. Výraz, který je patientem v aktivu, je patientem i v pasívu. Na analytické rovině je to ale jinak: výraz, který je subjektem v aktivu, se v pasívu stává objektem, a výraz, který je v aktivu objektem, se v pasívu stává subjektem.

## 5. Určení analytické funkce u „se/si“ jako doplnění slovesa

PDT-Vallex může sloužit také při ohodnocování tvarů „se“ a „si“ u sloves - pomůže vám rozhodnout, zda „se“ dostane analytickou funkci (afun) AuxT, AuxR nebo Obj.

**Objektem (Obj) je „se/si“** tehdy, pokud není součástí slovesa (hesla, neboli t\_lematu) a je ohodnoceno některým aktantovým funktorem. Např. gratuloval si.ADDR/Obj k vítězství, viděl se.PAT/Obj v zrcadle.

**AuxT dostane „se/si“** tehdy, je-li součástí slovesa (t\_lematu). Např. vést si, osamostatnit se.

**AuxR dostane „se/si“** tehdy, není-li součástí slovesa a není-li ohodnoceno žádným aktantovým funktorem.

## 6. Závěrečné pokyny

Ve všech případech, kdy máte pocit, že

- v PDT-Vallexu je něco špatně (význam slovesa, které máte v datech, v PDT-Vallexu není, nebo tam není ani celé sloveso),
- při aplikaci výše uvedených pravidel a pokynů dojde k rozporu mezi analytickým manuálem a touto příručkou (i kdyby to bylo třeba jen v jednotlivém příkladu),
- PDT-Vallex vám poskytuje informaci, kterou neumíte využít (např. protože tato příručka o tom nic neříká),

nám dejte vědět (mailem, poznámkou v „comments“ přímo v datech) – pomůžete nám zkvalitnit manuál, tuto příručku nebo vůbec dopomoci ke kvalitnímu korpusu. Při anotaci v takových případech postupujte podle analytického anotačního manuálu, který je (až na námi přímo schválené/dodané úpravy a změny) „nejvyšší autoritou“.