

Abstrakt

Tato technická zpráva je dokumentací k elektronickému valenčnímu slovníku českých sloves s tektogramatickou anotací, který vzniká v Centru počítačové lingvistiky od jara 2001. Zpráva obsahuje přehled podobně zaměřených projektů pro ostatní jazyky, shrnuje výchozí situaci pro budování takového slovníku pro češtinu, předkládá přijatý teoretický model valence a jeho použití pro popis valenčních rámců. Dále rozebírá vybrané jazykové jevy související s valenčními rámci, zachycuje pracovní postup při vytváření slovníku a uvádí použité softwarové nástroje. Slovník je zde popsán v současné, stále ještě velmi rozpracované podobě.

Obsah

| | | |
|----------|---|-----------|
| 1 | Úvod | 4 |
| 1.1 | Pojem valence | 4 |
| 1.2 | K čemu vlastně valenční slovník? | 5 |
| 1.3 | K čemu ale další valenční slovník? | 6 |
| 1.4 | Členění zbytku zprávy | 7 |
| 2 | Různé přístupy k popisu valence v ostatních jazycích | 8 |
| 2.1 | Wörterbuch zur Valenz und Distribution deutscher Verben | 8 |
| 2.2 | Słownik syntaktyczno-generatywny czasowników polskich | 9 |
| 2.3 | Valenčný slovník slovenských slovies | 10 |
| 2.4 | FrameNet | 13 |
| 2.5 | Databáze LCS | 14 |
| 2.6 | English Verb Classes and Alternations | 15 |
| 2.7 | PropBank | 18 |
| 2.8 | Japonsko-anglický valenční slovník | 20 |
| 2.9 | “Smolenská databáze slovesných příznaků” | 21 |
| 2.10 | Bulharský valenční slovník | 21 |
| 3 | Výchozí situace pro češtinu | 23 |
| 3.1 | Teorie větných vzorců | 23 |
| 3.2 | Teoretický základ pro náš slovník: Funkční generativní popis | 25 |
| 3.3 | Valenční rámce sloves ve FGD | 26 |
| 3.4 | Existující elektronické valenční slovníky | 31 |
| 3.4.1 | Valenční slovník povrchových realizací ve formátu BRIEF | 31 |
| 3.4.2 | Czech syntactic lexicon | 32 |
| 3.4.3 | Slovesa pro praxi | 34 |
| 3.5 | Další zdroje dat | 35 |
| 3.5.1 | Český národní korpus | 35 |
| 3.5.2 | Pražský závislostní korpus | 36 |
| 3.5.3 | EuroWordNet | 38 |
| 4 | Jádro slovníku – valenční rámce | 41 |
| 4.1 | Funktory a jejich uspořádání v rámci | 42 |
| 4.2 | Morfematické vyjádření | 44 |

| | | |
|----------|--|-----------|
| 4.3 | Typy zachycovaných doplnění | 44 |
| 4.3.1 | ‘Standardní’ valenční rámce | 44 |
| 4.3.2 | Rozšíření FGD o kvazivalenční a typická doplnění | 45 |
| 4.4 | Poznámky k přiřazování funktorů | 48 |
| 4.4.1 | Nově zavedené funktory OBST a RCMP | 48 |
| 4.4.2 | Slovesa emocionální s vazbou <i>někdo/něco někoho/něco něčím</i> | 49 |
| 4.4.3 | Slovesa výměny | 50 |
| 4.4.4 | Slovesa pohybu a funktory PAT a DIR, LOC | 51 |
| 4.4.5 | Slovesa pravení a funktory PAT a EFF | 52 |
| 4.4.6 | Rozlišení ADDR a BEN | 53 |
| 4.4.7 | Rozlišení ADDR a DIR3, DIR1, LOC | 53 |
| 4.4.8 | Rozlišení PAT/ADDR a ACMP | 54 |
| 4.5 | Poznámky k vyčlenění jednotlivých rámců | 54 |
| 5 | Další informace navázané na valenční rámce | 58 |
| 5.1 | Reflexivita | 58 |
| 5.2 | Reciprocita | 62 |
| 5.3 | Kontrola | 65 |
| 5.4 | Vidové dvojice | 67 |
| 5.5 | Diateze | 70 |
| 5.6 | Způsob užití slovesa | 75 |
| 5.7 | Syntakticko-sémantické třídy sloves | 79 |
| 5.8 | Navázání rámců na EuroWordNet | 82 |
| 5.9 | Frekvence rámců | 83 |
| 6 | Budování slovníku – postup a nástroje | 85 |
| 6.1 | Pracovní postup | 85 |
| 6.2 | Prostředí pro ruční anotaci slovníku | 86 |
| 6.3 | Vyhledávací rozhraní pro WWW | 89 |
| 6.4 | XSH | 89 |
| 7 | Shrnutí a práce do budoucna | 95 |
| | Literatura | 97 |

Kapitola 1

Úvod

1.1 Pojem valence

Pojem **valence** je odvozen z latinského **valentia** - síla¹. Zavedl jej v polovině 19. století britský chemik Edward Frankland pro vlastnost atomu, udávající nejvyšší počet chemických vazeb, které může daný atom navázat s jinými atomy. V lingvistice jej poprvé o sto let později použil Tesnière v souvislosti se závislostním rozбором věty ([Tesnière,1959]). Při rozboru vycházel od slovesa, jehož ‘subordonnés immédiats’ rozdělil na ‘actants’ (aktanty - účastníci) a ‘circonstants’ (cirkumstanty - okolnosti). Schopnost slovesa navázat nejvýše určitý počet aktantů (cirkumstantů může vázat libovolné množství) porovnal Tesnière s vlastností atomů a pojmenoval ji valence ([Helbig,Schenkel,1969], str. 10).

A jak pojem valence vysvětlují současné prameny? The Concise Oxford Dictionary of Linguistics:²

The range of syntactic elements either required or specifically permitted by a verb or other lexical unit. Thus the valency of *eat* includes a subject (*I* in *I am eating*) and an object (*cheese* in *I am eating cheese*). An element which is required is an obligatory valent; one which is specifically permitted but is not required is an optional valent. Thus *eat* must be explicitly associated with an object, since there are other verbs with which an object is not possible; but it is an optional valent since, in e.g. *I am eating*, it can be omitted.

Encyklopedie češtiny ([Karlík et al.,2002])

VALENCE. Počet a povaha míst (argumentů), které na sebe sloveso (popř. slovo jiného slovního druhu) váže jako pozice obligatorní n. potenciální. . .

Slovesa pro praxi ([SPP]):

Valencí rozumíme v lingvistice schopnost lexikální jednotky, především slovesa, vázat na sebe jiné výrazy a mj. tak zakládat větné struktury.

Někteří autoři směšují pojem valence s pojmem **subkategorizace**. Pojem subkategorizace zde nebudeme dále používat, přesto uvedeme, jak jej lze odlišit (Geert-Jan Kruijff³):

¹<http://www.m-w.com>

²<http://www.xrefer.com>

³<http://www.coli.uni-sb.de/~gj/Lectures/DG/Slides/subcat+valency.pdf>

Subcategorization describes a head's argument structure at the level of surface structure. Valency describes a head's argument structure at the level of linguistic meaning... However, subcategorization and valency are not isomorphic: e.g. expletive pronouns have no meaningful reflexion.

1.2 K čemu vlastně valenční slovník?

Valenční vlastnosti různých sloves (potažmo jiných slovních druhů) jsou natolik rozmanité, že je nedokážeme předpovědět obecnými pravidly. To je samozřejmě banální pozorování, nicméně potvrzuje, že chceme-li mít k dispozici informaci o valenci českých sloves, nezbyvá než popsat jedno sloveso po druhém, tzn. vytvořit valenční slovník. To je také cílem předkládané práce.

A k čemu vlastně může být informace obsažená v takovém slovníku užitečná? To se v následujícím výčtu pokusíme ilustrovat na několika známých úlohách z oblasti NLP ('Natural Language Processing'):⁴

- **Lemmatizace.** Příklady:

- ▷ *Oba bratři se báli matematiky.*
- ▷ *Oba bratři se stali matematiky.*

Kam patří lemma *matematika* a kam *matematik*, lze v těchto větách rozhodnout jednoznačně právě díky rozdílné valenci.

- **Morfologické značkování (tagging).** Příklad:

- ▷ *Ptala se jeho bratra.*

Forma *jeho bratra* je homonymní (může jít o genitiv nebo akuzativ), *jeho* může vyjadřovat osobní nebo přivlastňovací zájmeno. Jelikož sloveso *ptát se* má vazbu jen s (jedním) genitivem, lze větu disambiguovat jednoznačně.

- **Syntaktická analýza**⁵. Příklady:

- ▷ *Nechala ho spát.*
- ▷ *Začala ho milovat.*

V první větě závisí zájmeno na prvním slovese, protože *spát* nemá valenční člen vyjádřený akuzativem. Ve druhé větě závisí zájmeno na druhém slovese, protože na slovese *začít* nemůže současně záviset infinitiv a akuzativ⁶.

- **Word sense disambiguation.** Příklady:

- ▷ *Nech těch nesmyslů!*
- ▷ *Nech ho spát!*
- ▷ *Nech to ve skříni!*
- ▷ *Nech to na něm!*

Podle tvaru členů závislých na slovese lze v některých případech rozhodnout, o který z významů slovesa jde. Například sloveso *nechat* s genitivní vazbou má význam *skončit s něčím*.

⁴V dobách, kdy ještě nebylo zvykem používat všemocného zaklínadla NLP, dokládali autoři valenčních slovníků jejich potřebnost například na výuce cizích jazyků (např. [Helbig, Schenkel, 1969], str 5). Tento účel je samozřejmě aktuální i dnes, není ale rozhodně jedním z primárních cílů předkládané práce.

⁵Máme na mysli především závislostní syntaktickou analýzu, kterou lze velmi hrubě přiblížit jako hledání toho, které slovo je závislé na kterém (je jeho argumentem nebo jej jinak rozvíjí).

⁶Výjimkou z tohoto pravidla jsou některá časová určení.

- **Strojový překlad.** Každou z výše uvedených úloh je třeba řešit při realizaci strojového překladu.
- Valenční slovník usnadní **budování dalších lingvistických zdrojů**, v našem případě především Pražského závislostního korpusu. Jedním z přínosů práce na slovníku je i to, že v něm některé dosud ne zcela dořešené otázky spojené s anotací textu vyvstávají v daleko hutnější podobě a lze je zkoumat najednou pro celé třídy sloves.

1.3 K čemu ale další valenční slovník?

Prvotní motivace pro vytvoření vlastního valenčního slovníku (jaro 2001) vycházela ze specifických potřeb Pražského závislostního korpusu. Především šlo o to, aby ve slovníku (dále také označovaném jako VALLEX) byly u jednotlivých prvků valenčních rámců obsaženy funktoři (viz kap. 4.1). Výchozí situace se zdála být velmi příznivá, protože už byly k dispozici jiné elektronické valenční slovníky (viz kap. 3.4). Záhy se však ukázalo, že neobsahují natolik spolehlivou informaci (alespoň ne u těch nejčastějších, a tudíž nejsložitějších sloves), aby bylo možné některý z nich vzít a pouze do něj doplnit funktoři.

Ideální valenční slovník by měl mít podle našeho názoru následující vlastnosti:

- **Zachycení hloubkové i povrchové roviny.** Slovník by měl zachycovat vedle popisu realizace jednotlivých valenčních doplnění (předložka, pád, ...) i jejich hloubkovou funkci.
- **Explicitnost.** Měla by existovat jasná, pokud možno lingvisticky podložená kritéria pro vkládání informace do slovníku i interpretaci obsahu slovníku. Slovník by měl být podložen teoretickým modelem.
- **Úplnost.** Slovník by měl popisovat sloveso ve všech jeho významech.
- **Maximální pokrytí.** Slovník by měl obsahovat většinu sloves současného jazyka (přednostně ta častější).
- **Minimální redundance.** Do slovníku by měly být vkládány jen takové údaje, které jsou pro sloveso (nebo skupinu sloves) nějakým způsobem specifické a které nelze z ostatních údajů ve slovníku žádným způsobem vyvodit (například je zbytečné u každého slovesa vyjadřujícího pohyb znovu opakovat, že směrové doplnění *kam* lze vyjádřit předložkovými skupinami *na+4, do+2, k+3, ...*). Jakákoli redundance ve slovníku navíc zvyšuje riziko nekonzistence.
- **Kvalita.** Slovník by neměl obsahovat chyby (ani technické, ani věcné), nebo by měla být deklarována horní mez jejich množství⁷. Vzhledem k povaze oboru bývá ale někdy obtížné definovat, co přesně chyba je. V takovém případě lze přistoupit alespoň k měření 'inter-annotator agreement'.
- **Reálná užitná hodnota.** Mělo by být zřejmé, k jakým aplikacím a jakým způsobem bude možné informaci uloženou ve slovníku využít.
- **Machine readability.** (i) Slovník musí být uložen v dobře strukturovaném elektronickém formátu. (ii) Z hlediska použití v NLP je účelné do slovníku vkládat jen takovou informaci, k jejímuž použití není potřebná například znalost světa, porozumění významu atd.

⁷Přestože se zdá být zbytečné něco takového vůbec vyslovovat, v oblasti lingvistických dat je požadavek deklarované kvality (o bezchybnosti nemluvě) zatím velmi nadstandardní.

- **Human readability.** (i) Slovník musí být pro člověka dobře (rychle a pohodlně) čitelný, orientace v něm musí být snadná. (ii) Slovník musí být pro člověka dobře srozumitelný⁸. Dobrá čitelnost a srozumitelnost jsou nutnou podmínkou například pro efektivní odhalování a opravování chyb.

Jediný z uvedených bodů, na který jsme při budování slovníku od začátku rezignovali, je požadavek velkého pokrytí – v tuto chvíli slovník obsahuje ”jen” přibližně tisíc sloves. Ambicí předkládané práce je spíš vytvořit jakýsi maximálně konzistentní ”zlatý standard”, byť menších rozměrů.

1.4 Členění zbytku zprávy

Druhá kapitola této zprávy obsahuje přehled projektů, v rámci kterých byla popisována valence sloves v jiných jazycích než v češtině.

Třetí kapitola předkládá dosavadní teoretické pohledy na valenci sloves v češtině, shrnuje vlastnosti existujících elektronických valenčních slovníků a dalších zdrojů jazykových dat, které by bylo možné při budování valenčního slovníku použít.

Čtvrtá kapitola uvádí valenční slovník VALLEX a popisuje, jak jsou v něm zachyceny valenční rámce, co je pokládáno za součást rámce a jaké vlastnosti jsou zachyceny u jednotlivých prvků rámce.

V páté kapitole je uveden výčet jazykových jevů, které se snažíme zachytit ve slovníku v attributech valenčního rámce, protože nesou důležitou syntaktickou informaci a náleží k rámci jako celku. Jde například o reflexivitu, kontrolu, syntakticko-sémantické třídy, relativní četnost výskytů a další.

Šestá kapitola se zabývá praktickými otázkami souvisejícími s budováním valenčního slovníku, např. která slovesa zpracovávat, jaké softwarové nástroje lze použít atd.

Sedmá, poslední kapitola shrnuje dosažené výsledky a předkládá úkoly do budoucna.

⁸Problémy se srozumitelností lze očekávat například u slovníků automaticky extrahovaných z korpusu. Pokud se má algoritmus pro extrakci vyhnout nadměrnému zobecnění, může pro časté a komplikované sloveso vychrlit soubor mnoha desítek nebo několika stovek ‘protorámců’. Takový soubor sice asi nějakým způsobem valenci slovesa skutečně odráží, ale kvůli svému objemu bude pro člověka prakticky neuchopitelný.

Kapitola 2

Různé přístupy k popisu valence v ostatních jazycích

V této kapitole je stručně popsáno několik knih a projektů, ve kterých byla zachycována valence sloves pro jiné jazyky než pro češtinu. Jistě nejde o vyčerpávající přehled. Chybí v něm přinejmenším několik dalších projektů, které uvádíme v následujícím seznamu a jejichž dokumentaci se nám nepodařilo opatřit nebo dostatečně prostudovat:

- Valenzwörterbuch deutscher Verben¹. Cílem projektu je detailně popsat 640 německých sloves. Slovník má sloužit především k výukovým účelům.
- Odense valency dictionary². Slovník obsahuje několi tisíc dánských sloves.
- The Proton valency dictionaries³. V rámci projektu Proton byly vytvořeny valenční slovníky pro francouzštinu (3700 sloves) a holandštinu (4200 sloves).
- GREG⁴. Multilinguální slovník GREG má obsahovat pro angličtinu, gruzínštinu, němčinu a ruštinu přibližně tisíc vzájemně ekvivalentních sloves.

2.1 Wörterbuch zur Valenz und Distribution deutscher Verben

Pozoruhodným pokusem (zejména přihlédneme-li k době vzniku) o formální zachycení valence německých sloves je kniha [Helbig, Schenkel, 1969]. Valence každého ze zhruba 350 zpracovaných sloves je v ní popsána na třech stupních:

- I. Na prvním stupni je pro sloveso určen možný počet “spoluhráčů” (‘Anzahl der Mitspieler’, pravděpodobně obdoba aktantů), např. *erwarten*₂ (očekávat), *rauben*₂₍₃₎ (loupit), jde tedy pouze o kvantitativní popis. Počet obligatorních spoluhráčů je bez závorek, počet fakultativních je v závorce.
- II. Na druhém stupni je popsáno “syntaktické okolí” slovesa (formy jednotlivých spoluhráčů). Jsou použity značky jako Sn, As, Sd, Sg (substantivum v nominativu, atd.),

¹<http://www.ids-mannheim.de/gra/valbu.html>

²http://www.id.cbs.dk/forskknin/forskproj/ovd/ovd_en.shtml

³<http://bach.arts.kuleuven.ac.be/PA/proton.html>

⁴<http://www.informatik.uni-stuttgart.de/ifi/is/greg-index.html>

NS ('Nebensatz', vedlejší věta), I (infinitiv bez *zu*), Inf (infinitiv s *zu*) atd. Pokud je spoluhrač fakultativní, je umístěn v závorce. Pokud má několik alternativních forem, jsou odděleny lomítkem.

- III. Na třetím stupni je popsáno "sémantické okolí" slovesa – sémantické rysy jednotlivých spoluhračů: Hum (člověk), +Anim (životný), -Anim (neživotnost), Abstr (abstraktní), Loc (místní), Temp (časový) atd.

Příklad záznamu pro sloveso *hören* je na obrázku 2.1.

| hören | |
|--|---|
| I. | hören ₂₍₃₎ (V1 = wahrnehmen, aufnehmen) |
| II. | hören → Sn, Sa/NS _{das, ob, w} , (I) |
| III. | Sn → + Anim (Das Kind, die Katze hört den Fremden.) |
| | Sa → 1. + Anim (Er hört das Kind, den Hund.) |
| | 2. Abstr (Er hört Musik) |
| | 3. Act (Er hört das Brüllen.) |
| | NS → Act (Er hört, dass er kommt / ob er kommt / wer kommt.) |
| | I → Act (Er hört sie kommen.) |
| I. | hören ₂ (V2 = gehorchen, reagieren) |
| II. | hören → Sn, pS |
| III. | Sn → 1. + Anim (Das Kind, der Hund hört auf ihn.) |
| | 2. Abstr (als Hum) (Die Betriebsleitung hört auf seinen Rat.) |
| | p = auf, |
| | pSa → 1. hum (Die Schüler hören auf den Lehrer.) |
| | 2. Abstr (Die Schüler hören auf seine Worte.) |
| <i>Anmerkungen:</i> | |
| 1. Bei V1 ist I als 3.Mitspieler nur möglich, wenn als 2.Mitspieler Sa (nicht wenn NS) erscheint. I zusammen mit Sa kann als Ersatz für NS angesehen werden ("Ich höre ihn kommen" – "Ich höre, dass er kommt"). | |
| 2. Vereinzelt ist bei V 1 für Sa auch – Anim, möglich, aber nur bei sich bewegenden Objekten ("Er hört das Flugzeug". Aber: "*Er hört den schrank"). | |

Obrázek 2.1: Popis valence slovesa *hören* z knihy [Helbig,Schenkel,1969] (upraveno).

2.2 Słownik syntaktyczno-generatywny czasowników polskich

Pětídílní valenční slovník polských sloves ([Polański,1992]) vycházel v letech 1980-1992; je tedy jedním z prvních slovníků, zpracovávajících valenční vlastnosti sloves ve slovanských jazycích. V názvu slovníku je obsaženo slovo "generativní" – autoři tak odkazují na dva z postulátů generativní lingvistiky: 1) explicitnost popisu jazyka 2) zdůraznění tvůrčího charakteru jazyka (nezávislost na uzavřeném textovém korpusu, využití jazykové intuice).

Heslo obsahuje tyto informace:

- infinitiv heslového slovesa
- označení podhesel (jednotlivých významů)
- větné schéma složené ze symbolů jednotlivých částí větné struktury (např. NP (‘Noun Phrase’) s indexovým označením formy, nebo významu), srov. jeden z význ. sloves *cmokać – cmoknać (mlaskać – mlasknout)*:
 $NP_N - \{(NP_I) + (na \cap NP_{Acc}) + (NP_{Caus})\}$
Obejrzał jeszcze raz rozbity samochód, cmokał przy tym ustami i vzdychał. Woźnica cmoknął na konie i ruszyli w drogę. Dziewczyna cmoknęła z zachwytu.
 Kromě symbolu NP ve větných schématech figurují např. OR (přímá řeč), $K \cap S$ (tázací částice nebo zájmeno a věta) a další.
- sémantická charakteristika nominálních členů (sémantické rysy typu [+/-Anim] [+/-Abstr]; pokud vystupují rysy zároveň, jsou zachyceny pod sebou, pokud se vzájemně vylučují, vedle sebe. Tato část je poměrně podrobně rozpracována, soubor sémantických rysů není v celkové podobě dán na začátku slovníku výčtem, srov. sémantický rys objektu sloves *dobudowywać – dobudować*: [fragment *budowli*].
- informace o možnosti tvoření pasiva (pokud existence nebo neexistence pasiva u daného slovesa není systémová)
- příklady
- frazeologismy

Ukázka slovníkového hesla je na obrázku 2.2.

2.3 Valenčný slovník slovenských slovies

První díl slovníku [Nižníková, Sokolová, 1998] obsahuje gramatickou a sémantickou charakteristiku 625 nejfrekventovanějších slovenských sloves (celkem 2060 lexii). Heslo má tuto strukturu:

- heslové sloveso
- všechny jeho lexie
- každá jednotlivá lexie a její význam
- sémantická struktura
- synonyma
- valenční struktura
- charakteristika participantů
- příklady
- transformované struktury

Oproti některým jiným valenčním slovníkům podává Valenčný slovník slovenských slovies také charakteristiku slovesa (jednotlivých lexii) založenou na čtyřech základních sémantických příznacích: akčnost (+/-A), dynamickost (+/-D), mutačnost (+/-M), rozčlenenost (+/-R). Na základě kombinace těchto příznaků člení autoři slovesa do tří skupin: ”akčné”, ”procesné”, ”statické”.

Ve slovníku je také podrobně rozpracována sémantická charakteristika jednotlivých participantů. Kromě sémantických rysů uváděných v hranatých závorkách, např. [HUM], [QUANT] apod., je totiž u jednotlivých participantů uváděna jejich sémantická role. Velmi podrobná

CZOCHRAĆ SIĘ

I. 'targać sobie włosy, drapać się'

NP_N — $\{(NP_I) + (po \cap NP_L)\}$

NP_N → [+ Anim]

NP_I → $\begin{bmatrix} + \text{Anim} \\ \text{Pars} \end{bmatrix}$

NP_L → $\begin{bmatrix} + \text{Anim} \\ \text{Pars} \end{bmatrix}$

Przykłady:

Chłop czochrał się (rękami) po głowie i milczał długo. – Małpa czochrała się pazurami, po całym ciele. – Pies czochrał się tylną łapą (za uszami), dokuczały mu pchły.

II. 'o zwierzętach: ocierać się o coś twardego'

NP_N — $o \cap NP_{Acc} + (NP_I)$

NP_N → $\begin{bmatrix} + \text{Anim} \\ - \text{Hum} \end{bmatrix}$

NP_{Acc} → $\begin{bmatrix} - \text{Abstr} \\ - \text{Anim} \end{bmatrix}$

NP_I → $\begin{bmatrix} + \text{Anim} \\ - \text{Hum} \\ \text{Pars} \end{bmatrix}$

Przykłady:

Konie sapaly za ścianą, krowy czochrały się o deski. – Pies czochrał się pyskiem i bokami o nogę stołu.

Obrázek 2.2: Vzorek z polského valenčního slovníku [Polański,1992].

| RÁSTĚ ndk | |
|---------------|---|
| rásť 1 | rastom sa zväčšovať, vyvíjať sa |
| rásť 2 | vyrastať, dospievať |
| rásť 3 | vyskytovať sa, rodiť sa niekde |
| rásť 4 | vznikať, vzmáhať sa, zvelaďovať sa |
| rásť 5 | zväčšovať svoj objem, intenzitu, význam |
| rásť 6 | zdokonal'ovať sa vo vývine |
| rásť 7 | vznikať, vytvárať sa, vyvíjať sa |
| rásť 1 | rastom sa zväčšovať, vyvíjať sa SŠ: -A, +D, +R, +M SYN: zväčšovať sa, vyvíjať sa VŠ: /Sn/ - VF Sn: živý organizmus [ANIM/PLANT]: PROCnd <i>Ako tie deti rastú!</i> <i>Pozri, už rastie tráva.</i> <i>To dieťa rastie ako z vody.</i> TRANSF: 0 |
| rásť 2 | vyrastať, dospievať SŠ: -A, +D, +R, +M SYN: vyrastať, dospievať VŠ: /Sn/ - VF - ADVloc Sn: osoba, ktorá vyrastá: PROCnst: ADVloc/mod: miesto, spôsob <i>Rástol som na dedine.</i> <i>Naše deti rástli v hojnosti.</i> TRANSF: 0 |
| rásť 3 | vyskytovať sa, rodiť sa niekde SŠ: -A, -D, +R, -M SYN: vyskytovať sa, rodiť sa niekde VŠ: /Sn/ - VF - ADVloc Sn: ten/to, kto/čo niekde rastie [ANIM/PLANT]: STATnst ADVloc: miesto <i>V lese rastú huby.</i> <i>Kde rastú pekné dievčatá?</i> <i>Niektoré liečivé rastliny rastú vo voľnej prírode.</i> TRANSF: 0 |
| rásť 4 | vznikať, vzmáhať sa, zvelaďovať sa SŠ: -A, +D, +R, +M SYN: rásť 4 vznikáť, vzmáhať sa, zvelaďovať sa VŠ: /Sn/ - VF - (ADVloc/mod) Sn: to, čo vzniká [KONKR/REG]: PROCnst ADVloc/mod: miest, spôsob <i>V meste rastú nové štvrte.</i> <i>Rastú nové firmy a podniky.</i> TRANSF: 0 |
| rásť 5 | zväčšovať svoj objem, intenzitu, význam SŠ: -A, +D, +R, +M SYN: zväčšovať sa, narastať VŠ: /Sn/ - VF Sn: to, čo rastie [QUAL/SIT/MENT]: PROCnst <i>Jeho vplyv stále rástol.</i> <i>Rastú rady nezamestnaných.</i> <i>Výroba rastie iba pomaly.</i> TRANSF: 0 |
| rásť 6 | zdokonal'ovať sa vo vývine SŠ: -A, +D, +R, +M SYN: zdokonal'ovať sa, vyvíjať sa VŠ: /Sn/ - VF - ADVmod/asp Sn: ten, kto rastie [HUM]: PROCnst ADV mod/asp: spôsob, aspekt, oblasť <i>Spisovateľ rástol ľudsky aj umelecky.</i> <i>Človek rastie prekonávaním prekážok.</i> TRANSF: 0 |
| rásť 7 | vznikať, vytvárať sa, vyvíjať sa SŠ: -A, +D, +R, +M SYN: zdokonal'ovať sa, vyvíjať sa VŠ: /Sn/ - VF - z Sg Sn: to, čo vzniká [ANIM/QUAL]; PROCnst z Sg: to, z čoho Sn rastie [ANIM/ORIG]: OBJ <i>Rastlinky rastú zo semienka.</i> <i>Z milého šteňaťa rástol nebezpečný pes.</i> <i>Rastie z neho zlodej.</i> <i>Rastie z teba pekný kvietok!</i> TRANSF: 0 |

Obrázek 2.3: Vzorek ze slovenského valenčního slovníku ([Nižníková, Sokolová, 1998]).

klasifikace sémantických rolí se odvíjí od základní klasifikace sloves, např. levovalenční participant akčního slovesa je vždy agens, procesního slovesa - procesuál, statického - statuál; v rámci skupiny "agens" je vyčleněno 16 sémantických rolí (srov. manipulátor, donor, komunikant a dal.); obdobně podrobná je i klasifikace pravovalenčních participantů.

Ve slovníku je u každého slovesa (v rámci jednotlivých lexíí) obsažena také informace o tom, jaké derivované struktury je možné k danému slovesu tvořit, včetně zápisu povrchové formy odvozené struktury. V úvahu je bráno reflexivní pasívum, konstrukce s neosobním zvrátným tvarem, tzv. "dispozičné konštrukcie" (*Vysoká sa robila Karolovi ľahko.*); dále opisné pasívum, rezultatívni konstrukce (*Otcovstvo je už zistené.*)

Příklad slovníkového hesla pro sloveso *rásť* je na obrázku 2.3.

| Causation | | | |
|-----------------|---------------|-----------------|-----------------|
| Cause | | Affected | Effect |
| <i>The wind</i> | <i>caused</i> | <i>the tree</i> | <i>to sway.</i> |

| Communication | | | | | |
|---------------|--------------------|--------------------|-----------------|----------------------------|-------------------|
| Speaker | | Message | Addressee | Topic | Medium |
| <i>Pat</i> | <i>communicate</i> | | <i>with Kim</i> | <i>about the festival.</i> | |
| <i>Pat</i> | <i>communicate</i> | | <i>with Kim</i> | | <i>by letter.</i> |
| <i>Pat</i> | <i>communicate</i> | <i>the message</i> | <i>to me.</i> | | |

| Reciprocity | | |
|--------------------|----------------|------------------|
| Protagonist-1 | | Protagonist-2 |
| <i>Pat</i> | <i>fought</i> | <i>with Kim.</i> |
| <i>Pat and Kim</i> | <i>fought.</i> | |

Obrázek 2.4: FrameNet: Příklady obecných rámců

2.4 FrameNet

FrameNet⁵ je projekt, jehož cílem je vytvořit pro angličtinu bohatý slovník využitelný k různým úkolům NLP, viz [Fillmore,2002a], [Fillmore,2002b]. Je zaměřen především na zpracování sloves a substantiv s valenčními požadavky ('frame-bearing nouns').

Základní charakteristika

FrameNet sdružuje lexikální jednotky (páry: slovo - jeden z jeho významů) do skupin podle toho, zda mají tzv. paralelní sémantickou charakteristiku (tedy např. slovesa *tell*, *inform* a *notify*, resp. jejich příslušné významy, budou spolu s mnoha dalšími patřit do jedné skupiny).

Pro každou takovou skupinu je určena struktura společného **sémantického rámce** ('semantic frame', viz obrázek 2.4) a jednotlivé komponenty odpovídající **sémantickým rolím** ('semantic roles'); jsou charakterizovány příslušné významy slov. Omezený seznam obecných sémantických rolí (jako Agent, Patient, Goal, Theme, Instrument apod.) je nahrazen sémantickými rolmi specifickými pro jednotlivé rámce (např. 'Speaker', 'Addressee', 'Message' a 'Topic' u rámce 'Speaking')

Pro jednotlivé sémantické rámce jsou určeny možné způsoby, jakými jednotlivá doplnění mohou splňovat sémantické a syntaktické kombinatorické požadavky příslušné lexie (např. slovesa *tell*, *inform* a *notify* v příslušných významech, jejichž rámec obsahuje sémantické role 'Speaker', 'Addressee', 'Message' a 'Topic', mohou vyjadřovat Adresát jako přímý objekt slovesa).

Zpracování jednotlivých slov je **testováno na British National Corpus**⁶ (BNC) a dalších dílčích zdrojích. Jednotlivým významům slov jsou přiřazovány věty z korpusu ilustrující jejich typické užití, tyto věty jsou sémanticky anotovány (tj. jsou určeny složky, které vyjadřují jednotlivé komponenty příslušného rámce s označením jmen sémantických rolí).

Rámce mohou vytvářet **hierarchie**. Specifičtější rámce dědí vlastnosti obecnějších rámců.

⁵<http://icl.pku.edu.cn/doubtfire/semantics/FrameNet/Project/FrameDescs.htm>

⁶<http://www.hcu.ox.ac.uk/BNC/>

Co FrameNet obsahuje

Každé lexikální jednotce (páru slovo - jeden jeho význam), která nese valenční požadavky, jsou přiřazeny následující informace:

- příslušnost k určitému sémantickému rámci; různé lexie jednoho slova (lexému) tvoří lexikální jednotky, které mohou patřit do různých sémantických rámců (např. sloveso *argue* patří jednak do rámce ‘Quarreling’, jednak do rámce ‘Reasoning’);
- sémantické role, tzv. ‘frame elements’ (FE) příslušného sémantického rámce (v případě ‘Quarreling’ např. ‘Interlocutors’, ‘Topic’ atd.);
- možné realizace elementů rámce (na základě vět z BNC);
- ”svázání” (‘binding’) jednotlivých elementů sémantického rámce s obecnými rámci, jejichž vlastnosti dědí (např. FE z rámce ‘Quarreling’ dědí některé vlastnosti obecného rámce ‘Conversation’, jiné vlastnosti se přenášejí z obecného rámce ‘Disagreeing’).

2.5 Databáze LCS

LCS^{7,8} (‘Lexical Conceptual Structure’) byla navržena jako jazykově nezávislá sémantická reprezentace umožňující zachytit význam predikátů a sémantiku propozic ([Jackendoff,1990], [Jackendoff,1983]). Lingvistická perspektiva LCS, její úzký vztah k syntaxi umožňuje vybudovat obsáhlý systém sémantických kompozičních pravidel. Dovoluje klasifikovat slovesa na různých úrovních granularity.

Základní charakteristika

Základními elementy LCS jsou konceptuální kategorie, sémantická pole a primitivní prvky (‘primitives’).

Konceptuální kategorie. Konceptuální kategorie charakterizují jednotlivé lexikální jednotky. Mezi nejdůležitější konceptuální kategorie (také nazývané ‘conceptual POS’) patří např. ‘thing’, ‘event’, ‘state’, ‘place’, ‘purpose’, ‘manner’ či ‘time’. Kategorie jsou přiřazovány na základě kontextu (např. *after [time the meeting]*, ale *[event the meeting] will be held at noon*). Mohou tvořit hierarchie – např. konceptuální kategorie ‘thing’ zahrnuje specifitější ‘human’, ‘animal’ a ‘object’.

Konceptuální primitivní prvky. Velký počet konceptů je pokryt velmi malým počtem konceptuálních primitivních prvků – hlavní jsou ‘BE’ (reprezentuje ‘state’) a ‘GO’ (‘event’), dále např. ‘STAY’ (‘BE’+‘duration’), ‘CAUSE’, ‘EXCH’ (výměna) či ‘ORIENT’ (orientace objektu). Druhý typ konceptuálních primitivních prvků popisuje předložky (‘AT’, ‘ON’, ‘FROM’, ‘UNDER’, ‘VIA’ ...)

Sémantická pole. Hlavními sémantickými poli LCS jsou ‘localization’ (+loc), ‘time’ (+temp), ‘possession’ (+poss) a ‘characteristics of entity, properties’ (+char,+ident). Primitivní prvky mohou vytvářet pole, např. ‘GO’_{+loc} popisuje změnu místa, ‘GO’_{+char,+ident} změnu vlastností (např. váhy, barvy).

⁷http://umiacs.umd.edu/~bonnie/LCS_Data base_Documentation.html

⁸<http://www.ilc.pi.cnr.it/EAGLES96/rep2/node10.html>

Centrálním pojmem LCS je koncept pohybu ‘motion’ – primitivní prvky reprezentující ‘localization’ a ‘movement’ mohou být pomocí analogie a generalizace přeneseny na jiná sémantická pole.

Slovesné třídy založené na LCS vzorcích (‘LCS patterns’)

LCS zachycuje sémantiku sloves jako kombinaci sémantické struktury (orientovaný graf s kořenem a uzly nesoucími primitivní prvky a pole) a sémantického obsahu (specifikovaný konstantami). Sémantická struktura je charakteristická vždy pro všechna slovesa jedné sémantické třídy, sémantickým obsahem se slovesa mohou lišit.

Lexikální jednotka reprezentující sloveso je tedy orientovaný graf s kořenem, který nese informaci o subjektu, argumentech (objektech) a modifikátorech (‘modifiers’) a o jejich obligatornosti. Dále jsou uvedena omezení na konceptuální kategorii každého doplnění a tematická role, kterou zaplňuje.

Příklad: lexikální jednotka pro sloveso *cut down* (=snížit)

```
(act_on loc (* thing 1) (* thing 2)
  ((* [on] 23) loc (*head*) (thing 24))
  (cut+ingly 26)
  (down+/m))
```

Vysvětlivky: ACT_ON primitivní prvek, v poli location;

subjekt – thing, tematická role ”agent” (=1);

jediný argument – thing, tematická role ”theme” (=2)

23=MOD-LOC částice uvádějící místo, zde předl. *on*

24= MOD-LOC částice uvádějící místo, nevyžadována slovesem

poslední dva uzly - způsob ”locational act_on”, tedy *cutting in a downward manner*

Příklad: LCS pro větu *United States cut down (the) quota.*

```
(act_on loc (us+) (quota+)
  ((* [on] 23) loc (*head*) (thing 24))
  (cut+ingly 26)
  (down+/m))
```

LCS rozlišuje logické argumenty (ag, exp, th, src, goal, info, perc, loc, poss, time, prop) a logické modifikátory (mod-poss, ben, instr, purp, mod-loc, manner, mod-prop), které označuje mnemotechnickými značkami (‘labels’).

Příklad: sloveso *cut down*: *_ag_th,mod-loc(on)*

(podtržítko ‘_’ znamená obligatornost, čárka ‘,’ fakultativnost, v závorce je uvedená předložka, kterou doplnění vyžaduje)

2.6 English Verb Classes and Alternations

Levinová se v [Levin,1993] soustředila na vytvoření sémantických tříd anglických sloves. Vychází z hypotézy, že syntaktické chování slovesa je sémanticky determinované, resp. některé

sémantické rysy sloves mají výraznou souvislost s jejich syntaktickým chováním. Tuto hypotézu dokládá na velkém množství anglických sloves (cca 3200), kdy syntaktická kritéria vedla k vytvoření sémanticky homogenních tříd.

Úhelným kamenem při popisu syntaktických vlastností sloves u Levinové je pojem **alternace**, která zachycuje změnu argumentové struktury sloves, její transformaci. Zaujímají ji změny v argumentové struktuře. Definovala celkem 79 alternací v angličtině. Navržený soubor alternací není jazykově univerzální. Každý jazyk má vlastní specifický systém alternací, přesto se v určitých jazycích mohou nacházet alternace podobné. Angličtina je charakterizována vysokým počtem alternací.

Na základě počtu a druhu alternací, do kterých slovesa dané třídy vstupují, Levinová konstituuje sémantické třídy. Slovesa stejné sémantické třídy musí vstupovat do stejného počtu i druhu alternací.

Levinová operuje pouze s pojmy **argumentová struktura** a **subjekt** a **objekt**. Ne-
přisuzuje jim žádné role. V některých zcela zřejmých případech charakterizuje sémantiku argumentů pojmy jako agent, instrument, místo apod. Těchto pojmů neuzívá systematicky, snaží se jimi pouze ozřejmit povahu argumentů vstupujících do určité alternace.

Příklady: slovesa *break*, *cut*, *hit*, *touch* patří mezi slovesa tranzitivní, mají dva argumenty - subjekt a objekt:

- ▷ *Margaret cut the bread.*
- ▷ *Janet broke the vase.*
- ▷ *Terry touched the cat.*
- ▷ *Carla hit the door.*

Pouze *cut* a *break* vstupuje do **tzv. střední alternace** (jeden z typů tranzitivní alternace):

- ▷ *The bread cuts easily.*
- ▷ *Vase breaks easily.*
- ▷ **Cats touch easily.*
- ▷ **Door frames hit easily.*

Slovesa *cut* a *hit* se mohou objevit v **konativních alternacích** (opět podtyp tranzitivních alternací), zatímco *break* a *touch* nikoli:

- ▷ *Margaret cut at the bread.*
- ▷ *Carla hit at the door.*
- ▷ **Janet broke at the vase.*
- ▷ **Terry touched at the cat.*

Další alternace, **body-part possessor ascension** jsou schopná všechna slovesa kromě *break*:

- ▷ *Margaret cut Bills arm.* - *Margaret cut Bill on the arm.*
- ▷ *Terry touched Bills shoulder.* - *Terry touched Bill on the shoulder.*
- ▷ *Carla hit Bills back.* - *Carla hit Bill on the back.*
- ▷ *Janet broke Bills finger.* - **Janet broke Bill on the finger.*

Daná slovesa tak vykazují vzhledem k výše zmíněným alternacím jiné chování. Každé z těchto sloves se řadí do jiné sémantické třídy společně se slovesy schopných vytvářet stejné druhy a počty alternací.

Příklad: Sémantickou třídu sloves kontaktu, podtyp dotek, (např. slovesa *caress*, *graze*, *kiss*, *nudge*, *pat*, *pinch*, *prod*, *sting*, *tickle*, *touch*, *stroke*) vymezuje Levinová s použitím možnosti

nebo nemožnosti (*) alternací takto (jako ukázková věta slouží *Carrie touched the cat.*):

1. ***With / Against Alternation:**

- ▷ **Carrie touched the stick against the cat.*
- ▷ *Carrie touched the cat with the stick.*

(Do tohoto typu alternace mohou vstupovat jen slovesa se třemi argumenty, které lze charakterizovat jako 'agent', 'místo', 'nástroj'. Agent je vždy subjekt, ostatní jsou objekty. Agent hýbe nástrojem tak, aby nástroj vešel do kontaktu s určitým místem.)

2. ***Through / With Alternation:**

- ▷ **Carrie touched the stick through / into the cat.*
- ▷ *Carrie touched the cat with the stick.*

(Do této alternace vstupují opět tříargumentová slovesa: 'agent', 'místo', 'nástroj'. Agent pohybuje nástrojem tak, aby nástroj nejen vešel do kontaktu s místem, ale aby do místa vniknul.)

3. ***Conative Alternation:**

- ▷ *Carrie touched the cat.*
- ▷ **Carrie touched at the cat.*

(Konativní alternace je podtyp tranzitivní alternace. Objekt tranzitivní konstrukce se stává v netranzitivní konativní konstrukci objektem uvozeným předložkou *at*. Tato alternace se vyskytuje u sloves pohybu a kontaktu.)

4. **Body-Part Possessor Ascension Alternation:**

- ▷ *Carrie touched him on the shoulder.*
- ▷ *Carrie touched his shoulder.*

(V této alternaci je možné vyjádřit část těla a jeho "vlastníka" dvěma způsoby: "vlastník" je přímým objektem slovesa a jeho část těla je pak uvozena prepozicí s významem místa, anebo je část těla přímým objektem slovesa a "vlastník" v genitivu, popř. je vyjádřen jako přivlastňovací zájmeno)

5. ***Causative Alternation:**

- ▷ *Carrie touched the cat.*
- ▷ **The cat touched.*

(Kauzativní alternace má několik podtypů. Vyskytuje se u sloves, která mohou být užitá tranzitivně a netranzitivně s akcentem na příčinu.)

6. ***Middle Alternation:**

- ▷ *Carrie touched the cat.*
- ▷ **That cat touches easily.*

(Tuto alternaci tvoří tranzitivní a netranzitivní užití slovesa. Netranzitivní varianta "střední konstrukce" je charakteristická nepřítomností agentu, avšak přítomností adverbálního či modálního elementu.)

7. ***Instrument Subject Alternation:**

- ▷ *Carrie touched the fence with a stick.*
- ▷ **The stick touched the fence.*

(Subjekt a jeden z objektů mohou být charakterizovány jako instrumenty. Jedna varianta obsahuje oba "instrumenty", zatímco druhá varianta pouze "bezprostřední" nástroj.)

8. Unintentional interpretation not available:

a. Reflexive Object:

▷ *Carrie touched herself.*

b. Body-Part Object:

▷ *Carol touched her hair.*

(Do této alternace vstupují slovesa se "zvrátným objektem".)

9. *Resultative Phrase:

▷ **Carrie touched the door open.*

10. Zero-related Nominal:

▷ *a touch / *give a touch × a pat / give a pat*

Slovesa patřící do třídy sloves kontaktu dotekem neimplikují náraz. Vyskytuje se u nich menší množství alternací než u sloves kontaktu nárazem, např. není možná konativní alternace.

Levinová slovesa třídí do sémantických tříd na základě druhu a počtu alternací, které jsou pro ně charakteristická. Slovesa v takto vytvořených slovesných třídách sdílejí nejen stejné syntaktické vlastnosti vzhledem k daným alternacím, ale i některé sémantické rysy. Na základě klasifikace sloves dle alternací vznikají sémanticky koherentní slovesné třídy. Levinová jich vyčleňuje zhruba 48; počet však je nepřesný, protože každá třída obvykle obsahuje další podtřídy.

2.7 PropBank

Cílem projektu Proposition Bank⁹ (PB, [Kingsbury et al.,2002],[Kingsbury,Palmer,2002]) je přidat rovinu sémantické anotace do složkově anotovaného korpusu Penn Treebank¹⁰, přesněji řečeno do jeho části, která obsahuje texty z Wall Street Journal o celkové velikosti zhruba milion slov.

V současnosti jsou v rámci PB anotovány pouze slovesné predikáty. Pro daný predikát (sloveso) je nejprve třeba rozlišit jednotlivé významy (senses), ve kterých ho lze použít. Argumenty slovesa v jednotlivých významech jsou očíslovány od Arg0 do Arg5. "Modifikátory" (modifiers) slovesa jsou označovány jako ArgM a dále tříděny na LOC (location), TMP (time), MNR (manner), DIR (direction), CAU (cause), NEG (negation marker), MOD (modal verb), EXT (extent), PRP (purpose), ADV (general-purpose modifier).

Anotace dat z Penn Treebank pak spočívá v tom, že pro každé sloveso v textu je vybrán jeden z jeho významů a jeho argumentům a modifikátorům (terminály složkového stromu) jsou přiřazeny výše uvedené značky. Sloveso samo obdrží značku rel. Zjednodušený příklad anotace:

He was drawing diagrams and sketches for his patron.

Arg0: he
Rel: drawing
Arg1: diagrams and sketches
Arg2-for: his patron

⁹<http://www.cis.upenn.edu/~ace/>

¹⁰<http://www.cis.upenn.edu/~treebank/>

Kromě takto anotovaných textů vznikají souběžně tzv. Frames Files, což je v podstatě obdoba valenčního slovníku. V nich jsou shromážděny všechny významy jednotlivých sloves s popisem a příklady. Argumenty jsou u jednotlivých významů slovesa seříděny podle toho, jak často se vyskytovaly v nějakém vzorku textu nebo jak jsou sémanticky významné. Každý argument dostane navíc mnemonickou značku, která popisuje jeho roli. Číslo argumentu ani jeho značka nemají společný ("globální") význam pro celý slovník, ale jen pro jednotlivá slovesa nebo nejvýše třídy sloves. Arg2 u jednoho slovesa nemusí mít stejnou roli jako Arg2 u jiného, mnemonickou značku *hitter* u Arg0 slovesa *hit* lze zase stěží použít ještě někde jinde.

Takto vypadá v PB záznam pro první význam slovesa *go* (slovesa zpracovaná v PB jsou volně přístupná na Internetu¹¹):

Roleset go.01 "motion":

Roles:

Arg1:entity in motion/goer

Arg2:extent

Arg3:start point

Arg4:end point

ArgM-LOC:medium

ArgM-DIR:direction (usually up or down)

Examples:

start and end points (-)

What flights go from Seattle to Boston via Minneapolis?

Arg1: what flights

REL: go

Arg3-from: Seattle

Arg4-to: Boston

Argm-LOC: via Minneapolis

extent (-)

Imports have gone down 33%

Arg1: Imports

ArgM-DIR: down

Arg2-EXT: 33%

extent and end point (-)

Woolworth went up 1 3/4 to 59 1/2.

Arg1: Woolworth

REL: went

ArgM-DIR: up

Arg2-EXT: 1 3/4

Arg4-to: 59 1/2

with direction (-)

A~lot of people would like TRACE to go back to 1970.

Arg1: TRACE -> a lot of people

¹¹http://www.cis.upenn.edu/~cotton/cgi-bin/pblex_fmt.cgi

| Japonská strana slovníku | role (propojení) | Anglická strana |
|--|------------------|-------------------------------|
| iku₁ | | |
| S1 ga agent, vehicle, animal | N1 | NP Subj |
| S2 ni e made -road, -rail, theatre places, place | N3 | PP <i>to</i> Acc |
| S3 kara yori -road, -rail, places, place | N4 | PP <i>from</i> Acc |
| iku₂ | | |
| S1 ga agent, vehicle, animal | N1 | NP Subj |
| S2 ni e made -road, -rail, theatre places, place | N8 | PP <i>along/around</i> Acc |

Obrázek 2.5: Slovníkový záznam pro japonské sloveso *iku* podle [Bond,Shirai,1997]

REL: go
ArgM-DIR: back
Arg4-to: 1970

2.8 Japonsko-anglický valenční slovník

Pro potřeby projektu ALT-J/E (Automatic Language Translator - Japanese to English) byl vytvořen rozsáhlý anglicko-japonský valenční slovník ([Bond,Shirai,1997]). Slovník obsahuje poloautomaticky vytvořené valenční rámce pro nejméně 16 000 japonských sloves, rámce pro další tisíce sloves byly přidávány automaticky. Slovníkové položky se skládají z predikátu, jednoho nebo více prvků rámce ('case slots') a informace o modalitě. Na japonské straně je pro každý prvek rámce specifikována syntaktická forma a sémantická omezení.

Syntaktická forma je určena jako typ složky: klauze, jmenná složka nebo adverbium. U jmenných složek je uvedena částice, se kterou se složka může vyskytovat ('Particles are a closed class of post-positional case markers that mark Japanese noun phrases').

Záznamy v obou jazycích jsou provázány pomocí pádových rolí ('case-roles', 'deep cases', 'theta roles'). Ve slovníku je použit tento seznam rolí¹²: N1 (Agent), N2 (Object-1), N3 (Object-2), N4 (source), N5 (Goal), N6 (Purpose), N7 (Result), N8 (Locative), N9 (Reciprocal), N10 (Quotative), N11 (Material), N12 (Cause), N13 (Instrument), N14 (Means), QUANT (Quantity), TIME (Time), ADV (Adverb), TN1 (Time-position), TN2 (Time-source), TN3 (Time-goal).

Ukázka slovníkového záznamu pro sloveso *iku* je na obrázku 2.5.

Autoři připouští, že ve slovníku vzniká několik typu nekonzistencí, například:

There are also problems caused by the conflation of the degree of valency (how closely related the case-element is to the verb) and the case-role. The accusative-case ... in a verb such as *tazuneru* "visit" is obligatory, so should be marked with

¹²V projektu ALT-J/E se používá 24 různých rolí, ve slovníku jen 14.

N2, but should be locative, which calls for N8. Due to this conflict in the definition, some verbs of this type are marked as N2, and some N8...

2.9 “Smolenská databáze slovesných příznaků”

V kolektivní práci [Silnickij et al.,1999] je popsána databáze slovesných příznaků, která vznikla v rámci komplexního výzkumu na univerzitě ve Smolensku od roku 1975. Jsou v ní obsažena slovesa deseti jazyků: angličtiny, francouzštiny, němčiny, ruštiny, arménštiny, turečtiny, arabštiny, indonézštiny, čínštiny a japonštiny. Pro každý jazyk je zpracováno 800 až 2250 sloves. U každého slovesa v jednotlivých jazycích jsou zachyceny různé příznaky. Některé příznaky se vzájemně vylučují (jsou v binární opozici). Příznaků je celkem 64 a jsou rozděleny do následujících skupin:

- fonetické příznaky
- morfologické příznaky
- etymologické příznaky
- chronometrické příznaky (v které etapě vývoje jazyka sloveso vzniklo)
- syntaktické příznaky
 - tranzitivita/netranzitivita
 - schopnost vázat předmět v jiném pádě než v akuzativu
 - schopnost vázat obligatorní příslovečné určení
 - schopnost vázat vedlejší větu předmětnou
- sémantické příznaky
 - tématické příznaky - dělí slovesa na tři sémantické makrotřídy: “energetické” (slovesa svázaná s přenosem materiální energie - pohybem, fyzikálními procesy atd.), “informatické” (slovesa odrážející zpracování informace v psychice člověka - vidět, znát, chtít), “ontologické” (slovesa pracující s vyšší mírou abstrakce)
 - chronostrukturní příznaky - odrážejí množství a funkční typ sémantických složek (prvků) slovesných výrazů; rozlišují tři chronostrukturní třídy slovesných výrazů: procesiva, kauzativa, operativa.

Autoři se na základě korelace jednotlivých příznaků zabývají typologií slovesných systémů.

2.10 Bulharský valenční slovník

Valenční slovník [Popova,1987] obsahuje asi 1000 nejčastějších bulharských sloves. U každého slovesa jsou uvedeny následující údaje:

- tranzitivita a vid
- slovní popis významu

ПРИНАДЛЕЖА, -йш, -й, -ят; *несв.; непрех.*

1. Притежание съм в пряк или преносен смисъл на някого или нещо.

I. П+принадлежа+O₂.

II. П=С, М.

O₂=на+C₁, на+Мп, Мкд.

III. С=лице, предмет, идея.

C₁=лице, идея, дейност.

IV. Принадлежеше вече на Мария (Д. Димов). Принадлежат не на мъжа (П). То вече принадлежеше на всички (П). Принадлежеше на тебе (Ц. Цанев). Животът ѝ принадлежи на онова голямо нещо (С). . . . което по право ми принадлежи (Б. Райнов). Аз ти принадлежах дори без честна дума (П).

2. Спадам, числя се, отнасям се към нещо.

I. П+принадлежа+O₂.

II. П=С, М.

O₂=към+C₁, към+Мп.

III. С=лице, предмет, явление и др.

C₁=група, категория, система.

IV. Семейството принадлежеше към висшето дрезденско общество (А. Константинов). От тази точка най-добре можеше да се види към коя система принадлежи Мусала (Ив. Вазов, РСБКЕ). Той не принадлежеше към това племе (Е. Константинов). . . . към който принадлежи (Марксистко-ленинска философия).

Obrázek 2.6: Vzorek z bulharského valenčního slovníku [Popova,1987].

- I. rámeček (argumenty, které sloveso vyžaduje)
- II. morfoložická informace o argumentech
- III. sémantická informace o argumentech
- IV. příklady užití slovesa

Rámeček je popsán jako výraz, kde prvním prvkem je subjekt, pak následuje predikát, přímý a nepřímý objekt, podřadící klauze nebo 'adjunct' predikátu. Na morfoložické rovině je určeno např. to, že subjekt může být podstatné jméno nebo zájmeno, na sémantické rovině může jméno nabývat hodnot typu osoba, zvíře, atd.

Vzorek ze slovníku je na obrázku 2.6.

Slovník byl nedávno převeden do elektronické XML podoby ([Balabanova,Ivanova,2002]).

Kapitola 3

Výchozí situace pro češtinu

V této kapitole budou popsány dva teoretické směry, které byly vyvinuty pro češtinu a v rámci kterých byla formálním způsobem popisována valence českých sloves (teorie větných vzorců a funkční generativní popis). Dále shrneme vlastnosti tří českých elektronických valenčních slovníků, které existovaly v době zahájení prací na slovníku VALLEX (slovník BRIEF, slovník Hany Skoumalové a Slovesa pro praxi¹). V poslední části kapitoly jsou uvedeny další tři zdroje lingvistických dat pro češtinu, které lze potenciálně využít při budování valenčního slovníku (Český národní korpus, Pražský závislostní korpus a EuroWordNet).

3.1 Teorie větných vzorců

Teoretické poznatky z oblasti české syntaxe a sémantiky získané při dlouhodobém výzkumu struktur českého jazyka, který prováděl kolektiv vedený prof. F. Danešem, jsou shrnuty v práci Větné vzorce v češtině ([Daneš,Hlavsa,1987]). Pro krátké shrnutí dané teorie zde proto využijeme tuto monografii, která zároveň představuje teoretickou základnu pro systematický popis syntaktických a sémantických struktur (částečně uskutečněný např. ve valenčním slovníku Slovesa pro praxi, více viz kapitola 3.4.3):

”Větná syntax češtiny se chápe jako systémově uspořádaný soubor větných vzorců a pravidel, tj. obecných a jazykovým společenstvím přijatých organizačních principů, podle nichž mluvíci utváří jedinečné výpovědi”. (s.41) Autoři rozlišují formálně gramatický vzorec (GVV) a sémantický vzorec (SVV), které společně tvoří komplexní větný vzorec (KVV). GVV obráží minimální, avšak úplnou syntaktickou strukturu. Jejím nenaplněním by nevznikla gramaticky správná věta, podmínkou vytvoření gramaticky správné věty je tedy nutné obsazení všech pozic GVV. Gramatický větný vzorec je tvořen predikátorem (sloveso, nebo spojení kategoriálního slovesa s výrazem jmenným) a funkčními pozicemi, které predikátor svou valencí vyžaduje. Souhrn valenčních schopností daného predikátoru se nazývá **valenční potenciál**. Obdobou valence na sémantické rovině je **intence**. Sémantický větný vzorec je pak spojením predikátu (významové složky predikátoru) a participantů, které predikát vyžaduje svou intencí (ne však nutně všech, viz dále u SF). Souhrn participantských pozic daného predi-

¹Slovesa pro praxi sice nejsou elektronickým slovníkem v pravém slova smyslu – jsou vydána pouze knižně –, přesto je zde uvádíme. A to nejen proto, že v elektronické podobě vznikají, ale především proto, že ona elektronická předloha zachycuje v každém hesle explicitní logickou strukturu. Tím se liší například od SSJČ, který dnes sice rovněž existuje v elektronické podobě, ale značky v něm uvedené mají význam pouze pro formátování textu (tučné písmo, kurzíva atd.)

kátu se nazývá **intenční pole**. Úplné intenční pole predikátu lze zachytit tzv. **sémantickou formulí**, formule je shodná pro celé třídy predikátů. Např.

spravovat: ”vyjadřuje přechod nějaké entity ze stavu funkčního nepořádku do stavu funkčního pořádku, přičemž toto převádění uskutečňuje nějaká další entita”

$$(s.53) \quad SF : xA(\overline{yKz})T(yKz)$$

$x \sim$ působitel dějového přechodu, $y \sim$ nositel vlastnosti, $z \sim$ vlastnost, K symbolizuje 'mít vlastnost', T označuje přechod mezi oběma relacemi, A je relátor 'být agentem'.

Sémantickou formuli tedy můžeme interpretovat takto: x způsobuje, že y , které nemělo vlastnost z , přejde do stavu, kdy tuto vlastnost bude mít, tedy u věty: *Karel spravuje auto.*: Karel (x) způsobuje, že auto (y), které nemá vlastnost funkční (z), tuto vlastnost mít bude. Stejnou sémantickou formuli by měly např. predikáty rozšířit, ochočit.

SVV je modifikací, která se uskutečňuje skrze GVV; pouze ty elementy SF, které mají i svou odpovídající pozici v GVV, patří do SVV.

Např. *Karel rozsvítil lampu.*

$SF : xA(\overline{yNd})T(yNd)$; $d \sim$ děj, $N \sim$ být nositelem.

GVV: (S_{nom}) - VF - S_{acc}

$x \sim S_{nom}$, $y \sim S_{acc}$ (pro y , které se ve SF objevuje dvakrát, je však v GVV jen jedna pozice)

Korelací sémantického větného vzorce s odpovídajícím gramatickým větným vzorcem vzniká tzv. **komplexní větný vzorec** (KVV). Výše uvedený příklad *Karel rozsvítil lampu.* má následující KVV:

$$\frac{\text{GVV: } S_{nom} \quad - \quad \text{VF} \quad \quad - \quad S_{acc}}{\text{SVV: } \text{agens} \quad - \quad \text{act. caus.} \quad - \quad \text{patiens}}$$

Analýza pozic GVV je založena na protikladech **konstitutivní** \times **nekonstitutivní** a **obligatorní** \times **neobligatorní**. Obligatorní pozice (taková, jejímž neobsazením by vznikla gramaticky nesprávná věta) je vždy konstitutivní; neobligatorní pozice konstitutivní je **potenciální**. Nekonstitutivní členy (nelze již mluvit o pozicích) nepatří ke GVV a jsou označeny jako **fakultativní**. Ve větných vzorcích jsou tak zachyceny obligatorní pozice a pozice potenciální (v závorkách).

Příklady:

- $trpět_1$ GVV: (S_{nom}) - VF - (S_{acc})
 Raději trpět nouzi než posluhovat pánům.
- $trpět_2$ GVV: (S_{nom}) - VF - S_{instr} / na S_{acc}
 Babička velmi trpěla revmatismem.
- $trpět_3$ GVV: (S_{nom}) - VF - S_{instr}
 Doly trpí nedostatkem pracovníků.

- *trpět*₄ GVV: (S_{nom}) – VF – S_{dat} – abySENT / S_{acc} ;
 S_{acc} – nevětné vyjádření informace
Trpěl zaměstnancům, aby chodili pozdě do práce / jim absence.
- *trpět*₅ GVV: (S_{nom}) – VF – na S_{acc}
Bedřich trpí na děvčata.

Významnou část práce Větné vzorce v češtině představuje sémantická klasifikace predikátů založená na protikladu "relativně stálý stav" - **predikáty statické, neděje** (*mít, obsahovat*) × "pohyb, změna" - 1) **procesy** (*skákat, litovat*) 2) **události**; jde o přechod od nějaké situace výchozí k situaci závěrové (*rozsvítit, vysvětlit*). Podrobné analýze jsou pak podrobeny tyto skupiny sloves:

- **slovesa pohybu** např.:
 - *jet* GVV: (S_{nom}) – VF – (S_{instr}) – (praep S)
 - Jan jel autem. Jel do Plzně. Jel rychle.*
- **slovesa manipulování** např.:
 - *dovést* GVV: (S_{nom}) – VF – S_{acc} – S_{dat} / praep S / ADV
 - Matka dovedla děti do školky.*
 - *dostat* GVV: (S_{nom}) – VF – S_{acc} – (od S_{gen})
 - Dcera dostala hezký dárek.*
- **slovesa mluvení, myšlení a vnímání** např.:
 - *říci* GVV: (S_{nom}) – VF – o S_{loc} – (S_{dat}) – S_{acc} / že SENT / rel SENT
 - Tomáš (mu) o něm řekl, že je spolehlivý / že je to lump / že se za války choval dobře / jak se choval za války / kde se skrývá.*
- **slovesa s obecným významem změny** např.:
 - *změnit* GVV: (S_{nom}) – VF – S_{acc} – v / na S_{acc}
 - Válka změnila zdravé lidi v mrzáky.*
- **slovesa obecně pojatých a elementárních dějů** např.:
 - *dělat* GVV: (S_{nom}) – VF – S_{acc}
 - V neděli nic nedělám.*

3.2 Teoretický základ pro náš slovník: Funkční generativní popis

Teoretický základ pro vytváření valenčního slovníku je dán Funkčním generativním popisem češtiny (FGD, Functional Generative Description). FGD, který navazuje na tradice pražské lingvistické školy, se rozvíjí od 60. let 20. století (viz zejména [Sgall,1967], [Sgall et al.,1986], [Sgall,1992]). Jde o závislostní typ formalismu, který byl navržen pro účely teoretického popisu struktury českých vět pomocí generativní procedury. Velmi dobře se osvědčuje i jako

podklad pro analytické procedury - FGD slouží jako teoretický základ při budování Pražského závislostního korpusu (PDT, Prague Dependency Treebank), tedy pro analýzu vět na dvou rovinách větné stavby.

Základní charakteristikou FGD je **stratifikační přístup** k popisu jazyka - popis jazyka je rozčleněn do několika rovin. Každá z rovin je množinou zápisů vět, každá rovina má svou syntax (různé elementární jednotky se skládají v jednotky komplexní - jsou ve **vztahu kompozičním** (vztah C)). Nejvyšší rovina odpovídá významovému plánu jazyka, nejnižší plánu výrazovému. Jednotky jednotlivých sousedních rovin jsou ve vzájemném **vztahu formy a funkce - vztahu reprezentace** (vztah R; jednotka vyšší roviny je funkcí jednotky nižší roviny, jednotka nižší roviny je její formou).

Valenční struktura jednotlivých lexikálních jednotek je zachycena na **rovině hloubkové struktury**, která odpovídá významovému plánu jazyka a zachycuje významové jednotky, jejich lexikální a morfologické vlastnosti a vzájemné vztahy. Zápis věty na TR je zbaven homonymie (víceznačnosti). Z nižší **roviny morfematické** se do valenčních rámců promítá možné morfematické vyjádření jednotlivých doplňků (jejich morfematická forma).

FGD využívá **závislostního formalismu** - věta je zachycena jako závislostní strom, což je spojitý acyklický orientovaný graf s kořenem, kde do každého uzlu kromě kořene vede právě jedna hrana. Hrany stromu reprezentují vztah (syntaktické) závislosti (podřízenosti) - za závislý (podřízený) člen se považuje ten, který lze vynechat, aniž by věta ztratila (syntaktickou) správnost². Uzly stromu jsou lineárně uspořádány, jejich pořadí odpovídá (hloubkovému) slovosledu.

Teorie valence byla v rámci FGD rozpracována nejprve pro slovesa jako pro centrální jednotky věty (viz zejména [Panevová,1974], [Panevová,1980] a [Panevová,1992]), posléze byla studována i valence dalších slovních druhů - substantiv a adjektiv. Zde se budeme věnovat pouze valenčním rámcům sloves.

3.3 Valenční rámce sloves ve FGD

Sloveso je ve FGD, podobně jako v jiných popisech, chápáno jako centrum věty - proto je potřeba studovat v první řadě syntaktické vlastnosti sloves. Přitom je kladen důraz na formulaci ověřitelných kritérií pro určování jednotlivých doplňků, které tvoří valenční rámec, i pro určování jejich vlastností. Tento požadavek je zvláště podstatný, chceme-li tento popis využít pro automatické zpracování.

Jako **slovesná doplnění** jsou označovány výrazy (jednoduché i komplexní), které mohou rozvíjet nějaké sloveso (v závislostní reprezentaci věty tvoří dceřinné uzly slovesa; seznam typů doplňků viz kap. 4.1).

Slovesný valenční rámec (v širokém smyslu) je tvořen všemi doplňkami, které mohou dané sloveso v daném významu rozvíjet. Dále budeme termínu valenční rámec užívat v jeho užším smyslu - **slovesný valenční rámec** je tvořen **aktanty** a **obligatorními volnými doplňkami** slovesa, tzv. valenčními doplňkami slovesa či jeho valenčními členy (jednotlivé typy doplňků jsou popsány v následujících odstavcích). Přitom každé sloveso má alespoň jeden valenční rámec, často ale má rámců více. Má-li sloveso n valenčních rámců, má alespoň n významů.

²Ve dvojicích, kde nelze vynechat žádný ze členů, rozhoduje analogie: např. protože existují slovesa, která nevyžadují aktor (*prší*), je vhodné považovat aktor vždy za větný člen závislý na slovese

Charakteristika valenčních rámců spolu s možnými povrchovými formami aktantů (jejich možným morfematickým vyjádřením) musí být obsažena v slovníkovém hesle daného slovesa.

Slovesná doplnění se dělí podle následujících kritérií na aktanty a na volná doplnění.

Aktanty jsou taková doplnění slovesa (tzv. vnitřní doplnění, argumenty, participanty), která jsou charakterizována dvěma podmínkami:

1. nemohou se vyskytovat více než jedenkrát (bez koordinace nebo apozice) jako rozvití konkrétního slovesa;
2. jejich kombinace je charakteristická pro jednotlivá slovesa (musí být specifikována ve valenčním slovníku).

FGD na základě empirických zjištění rozlišuje na tektogramatické rovině pět aktantů. Jsou to aktor (konatel, Agens, Actor, dále ACT), patient (zasažený objekt, PAT), adresát (Addressee, ADDR), původ (ORIG) a výsledek (EFF). Všechny pět aktantů rozvíjí např. sloveso *předělat*. Naopak, existují i slovesa, která ve svém valenčním rámci nemají žádný aktant, např. sloveso *pršet*.

- ▷ *Matka*.ACT *předělala* *dětem*.ADDR *loutku*.PAT
z Kašpárka.ORIG *na čerta*.EFF.
- ▷ *Venku* *prší*.

Naproti tomu **volná doplnění** (adverbiální doplnění, adjunktivy - odpovídají příslovečným určení) - např. místa, času, způsobu, prostředku, podmínky atd., seznam viz [Hajičová et al.,2001] - se mohou objevovat u téhož slovesa vícekrát a mohou rozvíjet jakékoliv sloveso (případná omezení jsou obsahové, nikoli gramatické povahy):

- ▷ *V Praze se sejdeme na Hlavním nádraží u pokladen.*
- ▷ *Včera přišel večer domů pozdě.*
- ▷ *Kvůli dešti musel čekat pod střechou, protože neměl deštník.*

Aktanty i volná doplnění mohou být u jednotlivých sloves **obligatorní** (tj. pro dané sloveso povinně přítomny v hloubkové struktuře věty) nebo **fakultativní**.³ Kritériem pro obligatornost aktantů a volných doplnění je tzv. **dialogový test** (viz [Panevová,1974]). Tento test slouží jako měřítko pro obligatornost doplnění, je-li zkoumaný člen vypuštěn na povrchové rovině:

A: *Přátelé už přišli.*

B: *Kam?*

A: **Nevím.*

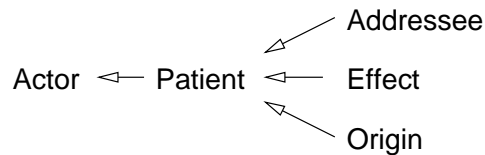
Už otázka mluvčího B signalizuje neúplnost výpovědi mluvčího A, jeho odpověď pak činí dialog zcela deviantní (A musí vědět, o jakém místě mluví) - sloveso *přijít* má obligatorní doplnění směru "kam". Jiná situace nastává u doplnění směru "odkud":

A: *Přátelé už přišli.*

B: *Odkud?*

A: *Nevím.*

³Je ovšem potřeba rozlišovat mezi obligatorností a nutností (povrchového) vyjádření konkrétního doplnění. Některá obligatorní doplnění nemusí být vždy realizována - říkáme jim (obligatorní) valenční členy vypustitelné (např. aktor je v české větě vypustitelný vždy).



Obrázek 3.1: Posouvání aktantů ('shifting').

Zde je odpověď mluvčího A zcela v pořádku - sloveso *přijít* nemá obligatorní doplnění směru "odkud". Dialogový test pomáhá určit všechna doplnění konkrétního slovesa, která musí být přítomna ve větě, aby (vytržená z kontextu) byla úplná, a tím stanovit jeho valenční rámec. Fakultativní aktanty je třeba odlišovat od **všeobecného aktantu** (General Participant) - všeobecný aktant je jen zvláštním případem lexikálního obsazení obligatorního aktantu ([Panevová,1992], [Panevová,Řezníčková,2001]). Příkladem všeobecného aktantu je např. pacient slovesa *prodávat* v následující větě:

A: *Soňa prodává u Bati.*

B: *Co?*

A: *Nevím, boty. To, co se u Bati prodává.*

Mluvčí A sice neví, co konkrétně je pacientem, je ale schopný ho specifikovat (v tomto případě jako určitou skupinu výrobků).

Pro určení aktoru a pacientu FGD užívá **syntaktická kritéria** (v duchu přístupu L. Tesnière [Tesnière,1959]), pro určení dalších aktantů užívá **kritéria sémantická** (tento přístup využívá zejména Fillmore ([Fillmore,1968], [Fillmore,1977]), u nás např. Daneš a Hlavsa ([Daneš,Hlavsa,1987])). FGD vychází z toho, že se realizuje tzv. posouvání aktantů - prvním aktantem je vždy aktor, druhým pacient. U sloves s více aktanty přistupují ohledy sémantické (viz obrázek 3.1).

Má-li tedy sloveso jediný aktant, jde o aktor, slovesa se dvěma aktanty mají aktor a pacient; teprve u sloves se třemi a více aktanty se uplatňují sémantická kritéria.

Příklady: V následujících větách se pacient posouvá do pozice aktoru (tj. aktoru odpovídá doplnění, které by primárně bylo vyjádřeno pacientem):

- ▷ *Kniha.ACT vyšla.*
- ▷ *Jan.ACT spadl se skály.*

V dalších větách je pozice aktoru zaplněna, do pozice pacientu se posouvá adresát, efekt, nebo původ:

- ▷ *Jan.ACT oslovil Marii.PAT (na ulici).*
- ▷ *Učitel.ACT vyvolal žáka.PAT.*
- ▷ *Petr.ACT vykopal jámu.PAT.*
- ▷ *Z banálního nachlazení.PAT se vyvinulo závažné onemocnění.ACT.*
- ▷ *Jan.ACT (za tu dobu) vospěl z jinocha.ORIG v muže.PAT.*

Následují rámce několika častých českých sloves (nejsou zde uvedeny rámce odpovídající některým okrajovým významům;⁴ fakultativní členy rámce jsou v závorce). V příkladových

⁴Indexy u sloves rozlišují jednotlivé rámce sloves, a tedy i jejich významy.

věťách označujeme i fakultativní volná doplnění:

- chodit*₁ ACT
Petr.ACT chodí rychle.MANN.
Chod' pomalu.MANN! (ty=ACT)

- chodit*₂ ACT INTT
Marie.ACT chodí na borůvky/na nákup/nakupovat.INTT.

- chodit*₃ ACT PAT
Petr.ACT chodí s Marií.PAT. (idiom)

- čekat*₁ ACT PAT (ORIG)
Petr.ACT čekal od Jirky.ORIG omluvu/že přijde.PAT.

- čekat*₂ ACT PAT
Rodiče.ACT čekají na dítě.PAT před školou.LOC.
Kdo.ACT čeká na dítě.PAT před školou.LOC?
Čekej na mě.PAT před školou.LOC. (ty=ACT)

- čekat*₃ ACT PAT
Věřitel.ACT jim.BEN čeká s dluhem.PAT. (idiom)

- čekat*₄ ACT PAT
Marie.ACT čeká s Petrem.ACMP dítě.PAT.(idiom)

- hovořit*₁ ACT (ADDR) PAT
Petr.ACT o tom.PAT hovořil s přítelem.ADDR.

- hovořit*₂ ACT
Petr.ACT hovoří plynule/anglicky.MANN.

- hovořit*₃ ACT (ADDR) (PAT)
*Petr.ACT o tom/na toto téma.PAT hovořil k publiku/na pří-
tele.ADDR (a ten ho nevnímal).*

- informovat* ACT ADDR PAT
Petr.ACT informoval rodiče.ADDR o svém návratu.PAT.

- informovat se* ACT PAT
Petr.ACT se na ně/o nich.PAT informoval.

- jednat*₁ ACT
Petr.ACT začal jednat (=konat).

- jednat*₂ ACT ADDR PAT
*Petr.ACT s nimi.ADDR jedná (=vyjednává) o investi-
cích.PAT.*

- *jednat*₃ ACT PAT MANN
 Učitel.ACT jedná (=zachází) se žáky.PAT špatně.MANN.
- *jednat se* ACT PAT
 Jedná se mi.ACT o princip.PAT.
- *nabízet* ACT ADDR PAT (EFF)
 *Petr.ACT jim.ADDR nabízí služby.PAT za úplatu/jako úplatu/
 úplatou.EFF. (ale: nabízí mu auto za pakatel.EXT)*
- *odpovídat*₁ ACT ADDR (PAT) EFF
 *Petr.ACT jim.ADDR na dotaz.PAT odpovídal, žeEFF
 (ale:odpovídal mu na jeho dotaz činy.MEANS)
 Petr.ACT mu.ADDR odpovídal pravdu/že ...EFF*
- *odpovídat*₂ ACT PAT
 Řešení.ACT odpovídá (=je ve shodě) požadavkům.PAT.
- *odpovídat*₃ ACT (ADDR) PAT
 *Rodiče.ACT odpovídají (=mají odpovědnost) za své děti.PAT.
 (ale: odpovídá za ztrátu svým majetkem.MEANS)*
- *říkat*₁ ACT (ADDR) (PAT) EFF
 Petr.ACT mu.ADDR říkal o Marii.PAT, že je chytrá.EFF.
- *říkat*₂ ACT ADDR PAT
 Petr.ACT mu.ADDR říkal o katastrofě.PAT
- *říkat*₃ ACT (ADDR) PAT
 Petr.ACT (si) mu.ADDR marně říkal (=žádal ho) o vodu.PAT
- *vyhrát*₁ ACT PAT (ORIG)
 Petr.ACT na něm.ORIG vyhrál v kartách.REG pět korun.PAT.
- *vyhrát*₂ ACT (ADDR) (PAT)
 *Petr.ACT s ním/proti němu/nad ním.ADDR vyhrál zá-
 pas.PAT. (ale: Buffalo s Haškem.ACMP v bráně vyhráli s Pen-
 guins 2:0)*
- *zahájit* ACT PAT
 Petr.ACT zahájil schůzi.PAT projevem.MEANS.
- *žít*₁ ACT
 Petr.ACT žil v Praze.LOC.
- *žít*₂ ACT PAT
 Petr.ACT žal palouk.PAT kosou.MEANS.

V případě *žit₁*, *žit₂* nejde pochopitelně o dva významy téhož slovesa, ale o homonymii dvou rozdílných sloves.

3.4 Existující elektronické valenční slovníky

3.4.1 Valenční slovník povrchových realizací ve formátu BRIEF

Slovník valenčních rámců ve formátu BRIEF ([Pala,Ševeček,1997]) byl připraven na Masarykově univerzitě kompilací několika tištěných slovníků, především SSJČ. Obsahuje asi 15 000 sloves. Formát BRIEF zde popíšeme jen stručně, podrobný popis lze nalézt v [Horák,1998]. Lemma slovesa a jeho rámce jsou na jednom řádku a jsou odděleny <v>. Rámce jsou odděleny čárkou. Rámec je posloupností prvků rámce, které jsou oddělené pomlčkou. Prvek rámce může mít několik atributů, resp. uspořádaných dvojic jméno atributu - hodnota atributu. Jméno atributu je vždy malé písmeno, hodnota atributu je buď velké písmeno, nebo jiný řetězec uvnitř složených závorek. Atributy jsou tyto: h (sémantický rys), c (morfologický pád), r (předložka), s (klauze nebo infinitiv), e (negace klauze), i (idiom), v (jiná omezení), z (komentář).

Vzorek valenčních rámců ve formát BRIEF (slovník nemá rozhraní, které by "zpříjemňovalo" čtení pro člověka:

```
běžet <v>hTc2r{z},hTc2r{do},hTc4r{na},hPTc3r{ke},hPTc4r{pro},hPTc7r{za},
hTc2r{kolem},hPc3r{proti},hPc3r{proti},hTc4r{o},hTc4z{motor pravidelně}
bičovat <v>hPTc4
bídačit <v>hPc4,hPc4-hTc7
bídačit se <v>
biflovat <v>hTc4,hTc2r{z},hTc4r{na},hTc4-hTc2r{z},hTc4-hTc4r{na}
biflovat se <v>hTc4,hTc2r{z},hTc4r{na},hTc4-hTc2r{z},hTc4-hTc4r{na}
bilancovat <v>hTc4
bílit <v>hTc4,hTc4-hTc7
biřmovat <v>hPc4
bít <v>hPTc4,hPTc4-hTc7,hPc6r{po},hTc4r{na},hTc7-hTc4r{o},hTc7-hTc2r{do}
bít se <v>hPTc7r{s},hPTc7r{s}-hTc3r{kvůli},hPc7r{s},hTc4r{o},
hPc7r{s}-hTc4r{o},hPTc4r{za}
blábolit <v>hTc4,hTc6r{o},hTc4-hTc6r{o}
```

Slovník BRIEF byl (alespoň podle informací autorů) prvním dostupným slovníkem, který zachycoval syntaktické chování českých sloves, byl svým formátem vhodný pro počítačové využití a měl dostatečné pokrytí, takže umožňoval masivní automatickou analýzu českých vět.

Slovník vychází z morfematické realizace slovesných doplnění, a touto koncepcí je silně omezen:

- Pro každé sloveso slovník poskytuje pouze informaci o možných kombinacích forem slovesných doplnění.
- Jednotlivá doplnění sice nesou informaci o sémantických rysech (např. osoba, věc, vlastnost, množství, místo, čas), není jim však přiřazena informace o hloubkové roli ('semantic role').
- Morfematické varianty jednotlivých doplnění se zachycují jako různé rámce⁵.

⁵Výjimkou jsou jen varianty idiomů a frazémů, které jsou odděleny symbolem |.

- Také zachycení fakultativnosti může být předmětem výhrad - obsahuje-li sloveso ve svém rámci fakultativní doplnění, ve slovníku BRIEF se tento rámec rozpadne do dvou rámců, s příslušným doplněním a bez něj. Podobně i pro všeobecné aktanty. To ovšem znamená, že rámce sloves nekorespondují s jednotlivými významy polysémních sloves.

Příklad: zpracování slovesa *bránit* ve slovníku BRIEF:

`bránit<v>hTc3, sI, hPc3-sUeN, hPc3-hTc6r{v}, hPTc4, hPTc4-hPTc3r{proti}, hPTc4-hPTc7r{před}`

Po přepsání do srozumitelnější podoby by soubor rámců vypadal takto:

- (1) *bránit něčemu*
- (2) *bránit inf.*
- (3) *bránit někomu, aby...*
- (4) *bránit někomu v něčem*
- (5) *bránit někoho(4)/něco*
- (6) *bránit někoho(4)/něco proti někomu/něčemu*
- (7) *bránit někoho(4)/něco před někým/něčím*

Pro srovnání uvádíme, jak je sloveso *bránit* zpracováno v SSJČ a ve VALLEXu:

SSJČ:

bránit₁ koho, co; co, koho proti komu, čemu, před kým, čím

bránit₂ komu, čemu; komu v čem, inf.

VALLEX:

- ACT(1) PAT(4) EFF(před+7,proti+3)
Brání město před Švédy, proti útokům.
- ACT(1) ADDR(3) PAT(v+6,Inf,aby)
Bránil mu v tom všemi silami. Brání mu přijít.
- ACT(1) PAT(3)
Brání jejich štěstí.

3.4.2 Czech syntactic lexicon

Český syntaktický lexikon ([Skoumalová,2001], [Skoumalová,2002]) obsahuje zpracování asi 15 000 českých sloves. K takovému zpracování vytvořila Hana Skoumalová širokou teoretickou základnu v rámci své disertační práce, a vymezila tak podrobná pravidla pro zápis vlastností sloves zejména v oblasti reflexivity, diateze a kontroly. Autorka vychází ze slovníku BRIEF, pro přiřazení funktořů jednotlivým členům rámce vyvinula algoritmus; výstup posloužil pro zkoumání možností automatického zpracování materiálu. V zápisu slovesného rámce figurují tyto informace: rod, reflexivita, subjekt (pokud ho sloveso vyžaduje), funktořy, gramatémy, diateze.

Příklad: záznam pro sloveso *brzdit*:

`brzdit R--s[i1]1(hPTc1)2[hPTc4]%`

Vysvětlivky:

R aktivní rod
 -- nezaplňená pozice pro refl. částice
 s[i1] inherentní subjekt v pozici označ. 1
 1, 2 aktanty (1 – Actor, 2 – Patiens)
 h sém. rysy daného doplnění: P – osoba, T – věc, zvíře
 c pádový tvar
 % možnost vytvoření opisného pasiva
 \$ možnost vytvoření reflexivního pasiva

Autorka slovníku se vyrovnává s omezeními zdrojového slovníku BRIEF, který dále automaticky zpracovává. Procedury, které používá, jsou vytvořeny na základě lingvistických pozorování o prototypických formách jednotlivých doplnění a jejich možných kombinacích:

- Ve výsledném slovníku jsou sloučeny rámce, které byly označeny jako varianty jediného rámce (varianty morfemického vyjádření i rámce, které jsou ve zdrojovém slovníku zdvojeny kvůli fakultativnosti některého doplnění, příp. kvůli všeobecnému doplnění).
- Pro část slovesných doplnění byly přidány odhady jejich hloubkových rolí.

Výhodou tohoto přístupu je konzistentní zpracování velkého množství sloves, což je nutná podmínka pro další využití slovníku v NLP, a poměrně spolehlivé výsledky pro nepřiliš složitá slovesa (slovesa s jedním nebo několika málo významy). Nutno říci, že v absolutních číslech (tj. celkový počet sloves v češtině) taková slovesa tvoří většinu.

Zůstává otevřenou otázkou, zda vůbec lze dostatečně spolehlivě automaticky zpracovat i slovesa s mnoha významy, která pokrývají velké procento sloves v běžně zpracovávaných textech.

Hlavní problémy, které se vyskytují v současné verzi slovníku:⁶

- Algoritmus pro vyčleňování jednotlivých valenčních rámců podstatným způsobem přispívá k správnému odhadu jednotlivých rámců slovesa, není však vždy zcela spolehlivý.

Příklad: *bránit*

bránit <v>[hTc3|sI] (pozn: *bránit něčemu* — *bránit dělat*)

bránit <v>[hPc3]-[sUen|hTc6r{v}] (pozn: *bránit někomu v něčem* — *bránit někomu, aby ...*)

bránit <v>[hPTc4]-[hPTc3r{proti}|hPTc7r{před}] (pozn: *bránit někoho/něco proti někomu/něčemu* — *bránit někoho/něco před někým/něčím*)

Doplnění v infinitivu přiřazené do prvního rámce patří do druhého rámce.

- Přístup založený na možných kombinacích morfemických forem nedovoluje vyčlenit významy s totožnou valencí (to ale neumožňuje ani zdrojový slovník).

Příklad: sloveso *absolvovat*

BRIEF:

absolvovat <v>hTc4 (*absolvovat něco*)

Skoumalová:

absolvovat R--s[i1]1(hPTc1)2[hTc4]%(někdo/něco absolvoval něco)

SSJČ:

⁶Slovník se dále vyvíjí, komentujeme verzi popsanou v [Skoumalová,2001].

*absolvovat*₁ ...co *absolvovat školu* (= dokončit školu)
*absolvovat*₂ ...co *absolvovat schůzi* (= zúčastnit se schůze)

VALLEX:

- ACT(1) PAT(4) *absolvovat studium* (= zakončit)
- ACT(1) PAT(4) *absolvovat operaci* (= zažít)

- Na základě vyvinutého algoritmu lze určit hloubkovou strukturu pouze u některých rámců, příp. jen u části doplnění, a to ne vždy zcela spolehlivě.⁷

Příklad:

působit R--1[hPTc1]2CI[hTc7]2A[hPTc4r{na}]&
 (někdo/něco.ACT působí na někoho/něco.??? něčím.???)

Neodhaleno, že předložková skupina na+4 vyjadřuje patient, instrumentál volné doplnění nástroje (MEANS).

působit R--1[hPTc1]2[hTc4]3[hPc3]&
 (někdo/něco.ACT působí někomu.ADDR něco.PAT)

Neúplně vyčlenění rámce, potřeba doplnit formu pro patient (klauze uvozená že, aby).

působit R--1[hPTc1]2P[sU]2JR[hTc4r{jako}]&
 (někdo/něco.ACT působí jako něco.???, aby.???)

Rámec je potřeba změnit na Actor (Nom) Patient (jako+Nom); doplnění vyjádřené klauzí patří k druhému rámci.

3.4.3 Slovesa pro praxi

Valenční slovník Slovesa pro praxi [SPP] obsahuje detailní analýzu 767 nejčastějších českých sloves. Obsah slovníku čtenáři přiblížíme příkladem zpracovaného hesla (obrázek 3.2) a krátkým úryvkem z předmluvy slovníku:

Valenční potenciál každého větného predikátu má v našem slovníku tři (vzájemně související) roviny prezentace: obecný vzorec, explicitní vzorec a morfologicko-syntaktickou analýzu obsazení jednotlivých valenčních pozic, k níž se přidává ještě složka sémantická: informace o selektivních sémantických rysech či tendencích, které omezují výběr konkrétních výrazů určených obsadit valenční pozice. Všechny údaje jsou pak doloženy v příkladové části.

Autoři rozlišují asi 50 selektivních sémantických rysů (přípustné jsou i některé jejich kombinace). Pro ilustraci jich zde několik uvedeme:

- anim – živé bytosti
- circ – cirkumstantnost, okolnostní určení
- coll – kolektivnost
- fin – finalita, účelovost, cílenost, cílovost
- med – médium, prostřednictví, prostředek (nemateriální)

⁷Z celkového množství rámců má přibližně jedna třetina jednoznačně přiřazeny všechny hloubkové role.

ŘÍDIT SE ned.

I. „dodržovat, zachovávat“

Val 1 – VF – Val 2 – Val 3

**někdo – se řídí – něčím/podle něčeho/podle někoho
– při něčem/v něčem**

Val 1: S nom [hum > coll, instit]

Val 2: S instr / podle S gen [opus, inform, sit U med]
// podle S gen [hum] // tím, co SENT

Val 3: při, v S loc [sit]

Iluminátoři kodesů se řídili při pořizování barev pevnými předpisy. – Kolumbus se při své plavbě řídil podle mapy Pavla Toscaneliho. – Sibelius se ve své tvorbě řídil jen vnitřním hlasem. – Člověk se v životě neřídí jen vůlí nebo rozumem. – V takových situacích se řídím především vlastním instinktem. == Řídím se v manželství tím, co mi radíval kdysi můj otec. – Tvůj kamarád není směrodatný, podle něj se v životě rozhodně neřídí. – Městský úřad se při vydávání stavebního povolení řídí příslušnou vyhláškou.

Též: Val 1 – VF – Val 2 Bylo třeba, abž se všichni čs. občané našimi zákony řídili a je respektovali. – Tím návodem se neřídí, je k ničemu. – Domovní samospráva se řídí pokyny představenstva družstva.

II. „být usměrňován“

Val 1 – VF – Val 2

něco – se řídí – podle něčeho/něčím

Val 1: S nom [sit] v [qual]

Val 2: podle S gen / S instr [inform, sit U med]

Systematickými pokusy byla nalezena některá pravidla, podle nichž se řídí vznik spekter. – Lidské jednání se řídí představami. – Výroba se bohužel ještě všude neřídí poptávkou. – Formy boje se řídily potřebami účinného vedení války proti Hitlerovi.

Obrázek 3.2: Vzorek ze slovníku Slovesa pro praxi

- mod – modus, způsob
- opus – výtvor, výsledek lidské činnosti, tvorby
- orig – původ
- plant – rostlina
- reciproc – reciprocita
- signum – znak, znakovost, znamení
- totum – totálnost, celkovost, celek

3.5 Další zdroje dat

3.5.1 Český národní korpus

Český národní korpus⁸ (ČNK) je rozsáhlý soubor počítačově uložených textů, který slouží k jazykovému výzkumu. Pro naše účely používáme jeho část SYN2000, která obsahuje současné

⁸<http://ucnk.ff.cuni.cz>

české texty z různých oblastí češtiny o celkové velikosti 100 milionů slov. Korpus SYN2000 je lemmatizovaný a morfologicky označovaný.

Při budování valenčního slovníku lze využít korpus k různým účelům, například k určení frekvence jednotlivých sloves, k vyhledávání autentických dokladů některých konstrukcí, dále jako zdroj “inspiračního” materiálu při navrhování valenčních rámců nebo jako zdroj testovacího materiálu při jejich ověřování. V současnosti také probíhají experimenty zaměřené na automatické extrahování valenčních rámců z korpusu.

Při užívání korpusu ovšem brzy narazíme na překážky způsobené tím, že věta v korpusu je jen lineární posloupností slov a značek a není nad ní postavena žádná struktura. V ČNK nelze například spolehlivě formulovat ani tak jednoduchý dotaz jako “najdi všechny věty, ve kterých je sloveso *chovat se* (ale ne *chovat*)”. V podobných případech by byl nadějnějším zdrojem především Pražský závislostní korpus (viz další kapitola).

3.5.2 Pražský závislostní korpus

Pražský závislostní korpus⁹ (Prague Dependency Treebank, PDT) je dlouhodobý výzkumný projekt, jehož cílem je komplexní anotace části ČNK. PDT využívá jako teoretický základ Funkčního generativního popisu (viz kapitola 3.2).

Anotační schéma je rozděleno na tři úrovně:

1. morfologická anotace - ke každému slovu je přiřazeno lemma a morfologická značka, obsahující hodnoty morfologických kategorií,
2. analytická anotace - pro každou větu je vytvořen tzv. **analytický strom** odpovídající závislostně popsané povrchové syntaxi věty; každému uzlu je přiřazena analytická funkce,
3. tektogramatická anotace - pro každou větu je vytvořen tzv. **tektogramatický strom** odpovídající hloubkové struktuře věty; každému uzlu je také přiřazen funktor.

Příklad analytického a tektogramatického stromu je na obrázku 3.3.

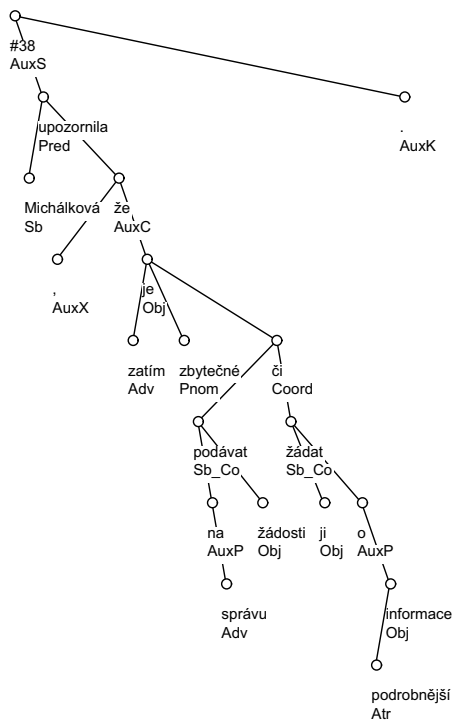
Data z PDT lze z hlediska zkoumání valence použít podobně jako data z ČNK (materiál pro inspiraci, testování, extrakci ...), anotace na analytické a zejména potom na tektogramatické rovině má však řadu předností.

Analytická rovina PDT.

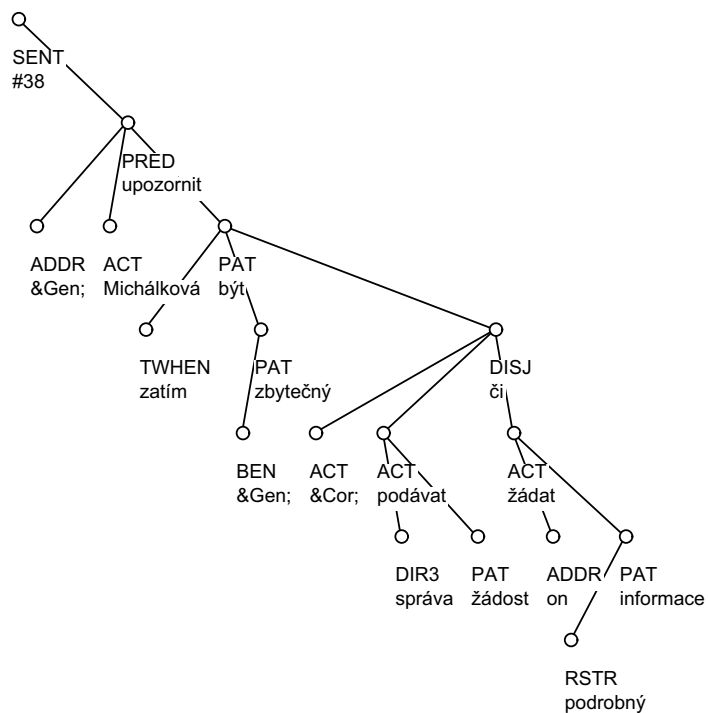
Výhody analytického značkování PDT:

- Odpadá problém s určováním celého lemmatu slovesa u vět s reflexivní slovesnou formou. Je-li *se* ve větě součástí reflexiva tantum, na analytické rovině visí pod slovesem uzel s formou *se* a analytickou funkcí AuxT.
- Kandidáty na valenční doplnění daného uzlu stačí hledat mezi přímými potomky uzlu a potomky, které od rodiče dělí jen koordinační a apoziční uzly (zde je ovšem třeba rozlišit argument koordinace od společného rozvíjení koordinace).
- Skupinu kandidátů na valenční doplnění lze třídit podle hodnoty analytické funkce - uzel s analytickou funkcí Sb prototypicky odpovídá aktoru, uzly s analytickou funkcí Obj by měly odpovídat aktantům a uzly označené Adv by měly reprezentovat volná doplnění.

⁹<http://ckl.mff.cuni.cz>



(a)



(b)

Obrázek 3.3: Příklad analytického (a) a tektogramatického (b) stromu pro větu: *Michálková upozornila, že zatím je zbytečné podávat na správu žádosti či žádat ji o podrobnější informace.*

Při pokusech o automatickou extrakci valenčních rámců z PDT bude jedním z hlavních problémů jeho malá velikost¹⁰. V současné době ovšem existuje několik parserů, které nad morfologicky analyzovanou (případně i disambiguovanou) českou větou postaví analytický strom. Množství správně “zavěšených” uzlů se pohybuje sice jen mírně nad 80 %, lze ovšem očekávat, že při aplikaci parseru pouze na neproblémové věty (např. věty bez koordinací, přímých řečí, zkratk, cizích slov, dvojteček atd.) by dosažená úspěšnost byla daleko vyšší. Pokud by se podařilo vyvinout vhodný algoritmus na získávání valenčních rámců z analytické roviny PDT, bylo by možné jej aplikovat i na analytické stromy vytvořené parserem z těchto neproblémových vět, čímž by množství vstupních dat mnohonásobně vzrostlo.

Tektogramatická rovina PDT.

Tektogramatická rovina (TR), na jejímž vytváření se v současné době pracuje,¹¹ představuje významný zdroj informací o valenci sloves - pro významy sloves, které se objevily v anotovaném textu, tektogramatický strom obsahuje kompletní soubor jejich obligatorních doplnění (v případě, že některá obligatorní doplnění nebyla na povrchu vyjádřena, jsou na TR doplňovány umělé uzly s příslušnými funktoři).

Tektogramatické anotování PDT a vytváření valenčního slovníku (oba projekty vycházejí z FGD) jsou ovšem duální problémy - anotátoři TR ke své práci potřebují valenční slovník a naopak valenční slovník může významně čerpat z PDT.

Valenční slovník anotátorů PDT.

Anotátoři TR potřebují k vytváření tektogramatických stromů valenční slovníky sloves, substantiv a adjektiv. Protože takové slovníky nebyly na začátku projektu (jaro 2001) k dispozici, začali vytvářet vlastní seznamy valenčních rámců. Ve stejné době (a se stejnou motivací) začal pro slovesa vznikat také slovník VALLEX. V současnosti se tyto zdroje vyvíjejí paralelně:

- Valenční slovník anotátorů PDT zpracovává jen ty významy sloves, které se vyskytnou v anotovaném materiálu. Uvádí pouze příslušný valenční rámec a příklady jeho užití. Lze tedy říci, že slovník se vyvíjí ‘do šířky’. V současné době obsahuje 4281 sloves s celkem 6697 rámci.
- VALLEX se soustřeďuje na komplexní analýzu sloves ve všech jejich významech. Zachycuje obohacený valenční rámec a další syntakticky relevantní vlastnosti sloves (reciprocita, kontrola, diateze a další, viz dále). Slovník VALLEX je tedy rozvíjen ‘do hloubky’.

Nezávisle vznikající slovníky byly sjednoceny v lednu 2002, od té doby se rozvíjejí nezávisle. Předpokládáme, že v roce 2003 dojde opět k jejich sjednocení.

3.5.3 EuroWordNet

EuroWordNet¹² je multilinguální lexikální databáze obsahující tzv. wordnety pro několik evropských jazyků: češtinu, estonštinu, francouzštinu, holandštinu, italštinu, němčinu a španělštinu. Wordnety těchto jazyků jsou strukturovány podle vzoru Princeton WordNet¹³ (americký wordnet pro angličtinu). Jejich základní stavební jednotkou je **synset**. - synset obsahuje jedno nebo více slov stejného slovního druhu, která jsou v daném významu zaměnitelná (syn-

¹⁰Největším problémem vůbec bude samozřejmě nalezení vhodného zobecňovacího algoritmu.

¹¹V současné době je tektogramaticky anotováno zhruba 20 000 vět.

¹²<http://www.hum.uva.nl/~ewn>

¹³<http://www.cogsci.princeton.edu/cgi-bin/webwn>

| český synset | anglický synset | anglická glosa |
|--|--|-------------------------------------|
| postavit se, povstat, zdvihnout se, zvednout se, postavit se | arise, rise, get up, stand up, get to one's feet | "The audience got up and applauded" |
| postavit se | react, react against, oppose, act in opposition to | |
| postavit se | uphold, fight for, stand up for, stick up for | of causes, principles, or ideals |
| postavit se | oppose, express opposition to, be against | We oppose the ban on abortion. |

Tabulka 3.1: České synsety se slovesem *postavit se* a odpovídající anglické synsety s glosou (navázáno přes Inter-Lingual Index).

set - set of synonyms, jde ovšem o velmi široké chápání synonymie)¹⁴. Například české sloveso *postavit se* se objevilo ve čtyřech synsetech, viz tabulka 3.1.

Jednotlivé synsety mohou vstupovat do **vzájemných vztahů**, především hyperonymie/hyponymie (vztah obecnějšího a specifitějšího pojmu). Tvoří tak sémantickou hierarchii. Na obrázku 3.4 je jako příklad vzorek hyperonymické hierarchie pro sloveso *jump* v jednom z jeho významů.

Několik "nejobecnějších" pater stromu (patra nejbližze kořenu) této sémantické hierarchie se nazývá **Top ontology**. Rozdělení na první úrovni Top ontology vypadá takto:

- 1stOrderEntity (skutečné předměty a látky reálného světa)
- 2ndOrderEntity (stavy, situace, události)
- 3rdOrderEntity (abstraktní entity jako myšlenky, koncepty, znalost)

Wordnety jednotlivých jazyků jsou propojeny pomocí **Inter-Lingual Index (ILI)**. ILI určuje, které synsety z jednotlivých jazyků jsou vzájemně ekvivalentní.

Příklad vzájemně ekvivalentních synsetů:

| | |
|------------------|---|
| český synset: | <i>prohodit, poznamenat, připomenout</i> |
| anglický synset: | <i>note, observe, make a remark, remark</i> |
| německý synset: | <i>anmerken, bemerken</i> |
| ... | ... |

EuroWordNet sice neobsahuje žádné informace o valenci, ale pro naše účely jej lze použít

¹⁴WordNet i EuroWordNet jsou častým terčem kritiky právě pro příliš subjektivní rozlišování jednotlivých významů slova.

• **jump, jumping** – (the act of jumping; propelling yourself off the ground; "he advanced in a series of jumps"; "the jumping was unexpected")

⇒ **propulsion, actuation** – (the act of propelling)

⇒ **deed, feat, effort, exploit** – (a notable achievement: "the book was her finest effort")

⇒ **accomplishment, achievement** – (the act of accomplishing something)

⇒ **action** – (something done (usually as opposed to something said); "there were stories of murders and other unnatural actions")

⇒ **human action, human activity** – (something that people do or cause to happen)

Obrázek 3.4: Vzorek hyperonymické hierarchie pro slovese *jump* v jednom z jeho významů.

jednak jako pomůcku pro rozlišování jednotlivých významů slovesa, jednak jako prostředek pro navázání na ostatní jazyky.

Kapitola 4

Jádro slovníku – valenční rámce

Heslo ve slovníku VALLEX, předkládaném v této zprávě, je posloupností valenčních rámců daného slovesa (jeden valenční rámec obvykle odpovídá jednomu významu slovesa). Ke každému valenčnímu rámci jsou v podobě atributů přidruženy další informace. Atributy jednotlivých rámců (např. reciprocita, kontrola, vidový protějšek atd.) jsou popsány v další kapitole, tato kapitola se věnuje samotným valenčním rámcům. Výjimkou je pouze atribut *example*, který obsahuje příklady použití rámce (příkladovou větu nebo její fragment) a jehož hodnotu uvádíme pro větší přehlednost pod rámci i v této kapitole.

Valenční rámec definujeme jako posloupnost prvků rámce (doplnění). U každého prvku jsou zachyceny tři jeho vlastnosti:

- funktor (kapitola 4.1),
- morfematické (povrchové) vyjádření (kapitola 4.2)),
- typ¹ doplnění (kapitola 4.3).

Při zápisu valenčního rámce používáme následující pravidla:

- Valenční rámec je zapsán jako posloupnost zápisů valenčních doplnění oddělených mezerou.
- Valenční doplnění je zapsáno jako funktor, za kterým v kulatých závorkách následuje morfematické vyjádření (nebo posloupnost alternativních morfematických vyjádření oddělených čárkou, nebo prázdný řetězec, lze-li použít jakékoli potenciální vyjádření daného funktoru) a typ doplnění oddělené středníkem.
- Morfematické vyjádření může mít následující hodnoty:
 - číslo pádu (jde-li o přímý pád)
 - (nevokalizovaná) předložka a číslo pádu oddělené znakem ”+”.
 - inf (infinitiv)
 - podřadící spojka

Příklad zápisu valenčního rámce v této zprávě:

¹K zápisu typu valenčního doplnění jiní autoři tradičně používají závorky, někdy i několik typů závorek. Tento způsob ale shledáváme málo přehledným v případě, kdy je rozlišování typů bohatší než jen binární opozice obligatorní/fakultativní.

□ *vyprávět*₂ ACT(1;obl) ADDR(3;opt) PAT(o+6;obl)
 vyprávěl přátelům.ADDR o dovolené.PAT

Číselné indexy někdy uváděné za lemmatem slovesa zde používáme pouze příležitostně a jen pro přehlednost textu; nejsou součástí slovníkových dat.

4.1 Funktory a jejich uspořádání v rámci

Při budování valenčního slovníku je bezesbytku převzata koncepce FGD, podle které má každé sloveso alespoň jeden valenční rámec, který se skládá z valenčních doplnění, v 'klasickém' pojetí FGD z aktantů a obligatorních volných doplnění (3.2). Předpokládáme, že existuje omezený seznam tzv. funktorů, které vyjadřují hodnotu daného doplnění na tektogramatické rovině, tj. jméno aktantu nebo volného doplnění (někdy mluvíme nepřesně o funktorech místo o slovesných doplněních). Seznam funktorů byl stanoven pro potřeby anotace na tektogramatické rovině PDT.

Seznam všech doplnění sloves (podle Manuálu pro tektogramatické značkování ([Hajičová et al.,2001]) je v tabulkách 4.1 a 4.2.

| | |
|---------------------|---------------------------|
| ACT (actor, bearer) | agens, hloubkový subjekt |
| ADDR (addressee) | adresát, komu |
| EFF (effect) | výsledek (zvolí kým, zač) |
| ORIG (origin) | původ, z čeho (ne odkud) |
| PAT (patient) | patiens, hloubkový objekt |

Tabulka 4.1: Aktanty (argumenty, participanty, vnitřní doplnění)

Uspořádání funktorů v rámci.

Slovesná doplnění v jednotlivých rámcich jsou uváděna v uspořádání, které vychází z bezpříznakového systémového uspořádání (SO, 'system ordering') ([Sgall et al.,1986]). Upřednostňujeme však dělení podle typu doplnění – napřed jsou uváděny všechny aktanty, následují kvazivalenční doplnění a jednotlivé typy volných doplnění.

Aktanty jsou uváděny v bezpříznakovém pořadí: ACT, ADDR, PAT, ORIG, EFF.

Volná doplnění třídíme podle jednotlivých typů takto – doplnění času, příčiny a důsledku, způsobu, místa, ostatní. Jednotlivé typy jsou uspořádány v tomto pořadí, které v zásadě odpovídá systémovému uspořádání:

- Časová určení uvádíme v následujícím pořadí: TSIN, TWHEN, THL, TTILL, TFHL, THO, TPAR, TFRWH, TOWH.
- Jednotlivá určení příčiny a důsledku jsou uspořádána následovně: COND, CAUS, AIM, CONFR, CNCS, CTERF, INTT.
- Určení způsobu jsou zachycena následovně: REG, MANN, EXT, NORM, CRIT, SUBS, ACMP, MEANS, DIFF, BEN, CPR, ATT, HER, RESL, RESTR.
- Následují volná doplnění místa: LOC, DIR1, DIR2, DIR3.
- Na závěr uvádíme ostatní doplnění: COMPL, DPHR, ETHD.

| | | |
|--|--|--|
| čas TWHEN (temp. when) TSIN (temp. since when) TTILL (temp. till) THL (temp. how long) TFHL (temp. for how long) THO (temp. how often) TPAR (temp. parallel) TFRWH (temp. from when) TOWH (temp. to when) | kdy? odkdy? dokdy? jak dlouho? na jak dlouho? jak často? ze kdy? na kdy? | <i>loni, napřesrok</i> <i>od té doby co...</i> <i>až do, dokud ne, než</i> <i>četl půl hodiny</i> <i>na dva dny</i> <i>často, mnohokrát</i> <i>během, zatímco</i> <i>Z dětství si nepamatuji nic.</i> <i>Přelož výuku na pátek.</i> |
| místo DIR1 (direction-from) DIR2 (dir.-which way) DIR3 (dir.-wh.to) LOC (locative) | odkud? kudy? kam? kde | <i>prošli lesem</i> <i>jednání uvnitř koalice</i> |
| příčina, důsledek AIM (aim) CAUS (cause) COND (condition) CONFR (confrontation) CNCS (concession) CTERF (counterfactual) INTT (intent) | účel příčina podmínka reálná konfrontace připustka irrealná podmínka záměr | <i>aby, pro něco</i> <i>(jest)-li, jestliže, když, až</i> <i>zatímco, kdežto</i> <i>ačkoli</i> <i>kdyby</i> <i>šel se koupat</i> |
| způsob ACMP (accompaniment) ATT (attitude) BEN (benefactive, -tory) CPR (comparison) CRIT (criterion) EXT (extent) DIFF (difference) HER (heritage) MANN (manner) MEANS (means) NORM (norm) REG (regard) RESL (result) RESTR (restriction) SUBS (substitution) | doprovod postojové benefaktiv porovnání kritérium, měřítko míra rozdíl dědictví způsob prostředek norma zřetel účinek omezení zastoupení | <i>s, bez</i> <i>s radostí, vhodně, právem</i> <i>pro koho, proti komu</i> <i>než, jako, stejně jako</i> <i>podle něj, ve smyslu</i> <i>velmi, trochu</i> <i>oč</i> <i>po otci</i> <i>ústně; psát česky</i> <i>psát rukou, na stroji, tužkou</i> <i>ve shodě s, podle (pravidla)</i> <i>vzhledem k, bez ohledu na</i> <i>opálen do hněda</i> <i>kromě, mimo</i> <i>místo koho-čeho</i> |
| další COMPL (complement) DPHR (dep. part of phraseme) ETHD (ethical dative) FPHR (foreign phrase) INTF (intensification) MOD (adverbial of modality) PREC (ref.to preceding text) RHEM (rhematizer) | doplněk část frazému volný dativ cizí fráze navazovací element | <i>Děti nám nechodí včas.</i> <i>To Karel ještě nepřišel?</i> <i>asi, možná</i> <i>tedy, naopak, tudíž, totiž</i> <i>í, také, jenom, nejen, vůbec</i> |

Tabulka 4.2: Volná doplnění (adjuncts, free modifications)

4.2 Morfematické vyjádření

Koncepce valenčních rámců se týká především tektogramatické roviny, rámce zachycují podkladovou strukturu. Pro potřeby automatické syntaktické analýzy je nutné zabývat se také morfematickou realizací jednotlivých valenčních doplnění – také tuto informaci je potřeba zachycovat ve valenčním slovníku.

Slovesná doplnění mohou být vyjádřena buď jednotlivými slovy – zejména podstatnými jmény a zájmeny v určitém pádu, ale i např. přídavnými jmény, příslovci. Dále skupinami slov – rozvitými větnými členy, jmennými či předložkovými skupinami, a koordinovanými větnými členy. Další možností jsou slovesa v infinitivu a vedlejší věty (uvozené podřadícími spojkami, tázacími zájmeny a příslovci).

Aktanty mají pro každé sloveso pevně danou morfematickou formu (zpravidla jedinou, někdy jich může být více) uchovávanou ve slovníku, volná doplnění (obligatorní i fakultativní) jsou určena sémanticky (typicky mají možnost různého morfematického vyjádření). Jednotlivé aktanty mohou být vyjádřeny u daného slovesa prototypicky (v aktivní větě aktor v nominativu, akuzativ pro patient, dativ pro adresáta, např. *Petr.ACT poslal Martinovi.ADDR dopis.PAT*), u jiného neprototypicky (např. adresát v akuzativu u slovesa *informovat*).

Vraťme se k nutnosti (povrchového) vyjádření konkrétního doplnění. Zdá se totiž, že neexistuje člen valenčního rámce, který by byl (absolutně) nevypustitelný (alespoň ve specifických kontextech, jako je např. odpověď na otázku). Na druhou stranu zřejmě pro vypouštění platí jistá omezení – pro konkrétní sloveso realizace některého aktantu na povrchové rovině vyžaduje vyjádření jiného aktantu (alespoň formou zájmena), především patientu.

Např. valenční rámec slovesa *předělat* obsahuje obligatorní aktor, patient, původ a výsledek a fakultativní adresát. Varianty bez vyjádřeného patientu nejsou zřejmě správnými větami:

- ▷ *Matka.ACT předělala dětem.ADDR loutku.PAT z kašpárka.ORIG na čerta.EFF.*
- ▷ **Matka.ACT předělala dětem.ADDR na čerta.EFF.*
- ▷ **Matka.ACT předělala dětem.ADDR.*
- ▷ *Matka.ACT předělala loutku.PAT.*

4.3 Typy zachycovaných doplnění

‘Standardní’ valenční rámce podle FGD zachycují aktanty a obligatorní volná doplnění, nezachycují už fakultativní volná doplnění. Pro NLP je však zřejmě výhodné využívat co nejbohatší slovníkové informace - ve slovníku jsou kromě těchto standardních valenčních doplnění zachycovány další dva typy doplnění, tzv. **kvazivalenční doplnění**, která stojí na hranici vnitřních doplnění (aktantů) a volných doplnění (např. *zakopnout o schod*), a **typická doplnění**, což jsou doplnění pro slovesa charakteristická (např. volné doplnění směru u slovesa *jet* (kam?) nebo *přijít* (odkud)). Oba nové typy doplnění jsou charakterizovány níže.

4.3.1 ‘Standardní’ valenční rámce

Standardní valenční rámce tvoří aktanty (obligatorní i fakultativní) a obligatorní volná doplnění (obr. 4.1).

| | obligatorní | fakultativní |
|---------|-------------|--------------|
| aktanty | + | + |
| volná | + | - |

Obrázek 4.1: Doplnění, která tvoří standardní valenční rámec, jsou označena ”+”

Pro stanovení valenčních rámců využíváme všechna kritéria navržená pro valenční doplnění (zejména princip shiftingu a dialogový test pro obligatornost doplnění ([Panevová,1974])).

Aktanty a volná doplnění lze rozlišit podle následující kritérií.

Vlastnosti aktantů:

- jen jedenkrát jako rozvití konkrétního slovesa (bez apozice, koordinace)
- omezeny na skupiny sloves (jejich kombinace je charakteristická pro jednotlivá slovesa, je specifikována ve slovníku)
- aktanty podléhají posouvání (shifting)
- pomocné kritérium: obvykle rekční (forma daná řídicím slovesem)

Vlastnosti volných doplnění:

- mohou se opakovat
- bez syntaktických omezení, pouze obsahová omezení (vágní vymezení)
- volná doplnění nepodléhají posouvání
- pomocné kritérium: forma daná sémantikou doplnění

Kritérium pro rozlišení obligatornosti a fakultativnosti – dialogový test (viz výše).

Zachycování typu doplnění ve slovníku. Množina aktantů a volných doplnění je disjunktní, proto by explicitní rozlišení ve slovníku nebylo účelné. U aktantů proto zaznamenáváme jen obligatornost / fakultativnost (typ obl, resp. opt); do standardního valenčního rámce patří jen obligatorní volná doplnění (typ obl).

4.3.2 Rozšíření FGD o kvazivalenční a typická doplnění

Kvazivalenční doplnění

Jako kvazivalenční doplnění označujeme doplnění se sémantikou volného doplnění, která ale mají vlastnosti, kterými se blíží aktantům - jsou rekční, vazebné (tj. takové závislé členy, jejichž forma je dána slovem řídicím), pojí se vždy jen s určitou skupinou sloves a nedají se zřejmě opakovat. Patří tedy do přechodného pásma mezi aktanty a volnými doplněními (viz i [Hajičová et al.,2002]). Jde o doplnění převážně fakultativní, nicméně často se vyskytující u konkrétního slovesa. Lze je charakterizovat jako ”obvyklá” doplnění (pro daný význam).

Kritéria pro určení kvazivalenčního doplnění:

1. rekčnost
2. omezení na skupinu sloves
3. nemožnost opakování jako doplnění konkrétního slovesa

| | obligatorní | fakultativní |
|------------------------|-------------|---------------|
| aktanty | + | + |
| kvazivalenční doplnění | + | + |
| volná doplnění | + | pouze typická |

Obrázek 4.2: Doplnění, která tvoří obohacený valenční rámec, jsou označena ”+”

4. typická sémantika volného doplnění

5. nepodléhají posouvání

Na základě předběžných zkoumání se zdá vhodné **omezit kvazivalenci na volná doplnění**, a to právě kvůli jejich typické sémantice. Nicméně jejich vlastnosti, kterými se blíží aktantům (rekčnost, omezení na skupiny sloves, nemožnost opakování), nás vedou k jejich vyčlenění do zvláštní skupiny.

Empiricky byla zatím jako kvazivalenční doplnění vyčleněna doplnění účelu (AIM), nové doplnění překážky (OBST, viz níže), doplnění rozdílu (DIFF) a zřejmě i doplnění účinku (RESL).

‘Standardní’ valenční rámce rozšířené o kvazivalenční a typická doplnění, označujeme jako **obohacené valenční rámce** (enriched valency frames).

Rozlišení kvazivalenčních doplnění ve slovníku. Množina kvaziaktantů je dána výčtem, není tedy třeba explicitně ve slovníku uvádět, že jde o kvazivalenční doplnění. Znamenáváme proto jen obligatornost, resp. fakultativnost (typ obl, resp. opt).

Příklady:

Funktor AIM (účel):

- slovesa výměny:
 - ▷ *nabídnout něco na prodej, k prodeji*
- dvojí závislost:
 - ▷ *poskytnout peníze dětem na knížky.AIM*
 - ▷ *potřebovat peníze na splnění úkolu.AIM*
- další:
 - ▷ *brát_i léky.PAT na chřipku.AIM*
 - ▷ *bere na posílení imunity.AIM (proti chřipce.AIM) léky*
 - ▷ *brát_j lidi na práci.AIM*

Ad kritérium 1: forma doplnění - typicky na+4, k+3

Ad kritérium 2: možné pouze u některých skupin sloves

Ad kritérium 3: lze opakovat! - *nabídl něco na prodej.AIM, aby se toho zbavil.AIM*

Další slovesa: *doporučit pacienta na vyšetření, k vyšetření, hlásit se na práci, hodit se na učitele, na práci, předkládat k uvážení, na uvážení, postavit vodu na čaj, pozvat na návštěvu, připravovat studenty na zkoušku, sejít se na skleničku, svolat zástupce na poradě, k poradě, učit se na ševce, ukládat děti k spánku, určit hráče na penaltu, k vystřídání, užíval vlivu k*

získání peněz, volat lékaře k náhlému případu.

Funktor OBST (překážka):

viz zavedení tohoto funktoru(4.4)

Funktor DIFF (rozdíl)

- slovesa vyjadřující změnu stavu, místa, hodnoty apod.:
klesat, přesáhnout, přibývat, stoupnout; prodloužit, rozšiřovat, snižovat, zlepšit, zvednout, zvýšit; (o ceně) stoupat, padat, posílit; posunout
- další slovesa: *prohrát, předat se, přežít, vzít někoho (=předhonit), zvítězit*
Ad kritérium 1: rekční (daná forma o+4)
Ad kritérium 2: není volně spojitelný s jakýmkoliv slovesem
Ad kritérium 3: nelze opakovat

Funktor RESL (účinek) (předběžné rozhodnutí, nejistota v anotaci)

- ▷ *argumentovat proti něčemu*
- ▷ *chránit, bránit, uchovat, zachovat něco před zničením*
- ▷ *zachránit před zničením, od zničení*
- ▷ *zabezpečit, zajistit děti před chudobou, střechu proti dešti*
- ▷ *spojit pod nadvládu*
- ▷ *hájit draky proti špatnému jménu*
- ▷ *pojistit se proti nebezpečí, před nebezpečím, zajistit skříň proti posunutí*

Ad kritérium 1: rekční (forma daná slovesem: proti+3, před+7, od+2)

Ad kritérium 2: není volně spojitelný s jakýmkoliv slovesem

Ad kritérium 3: nelze opakovat

Typická doplnění

Ve spojitosti s volnými doplněními se v teoretických člancích obvykle mluví o "obsahových vztazích", které připouštějí přítomnost jednotlivých volných doplnění ([Panevová,1980]): '[volná doplnění] se spojují se všemi slovesy, pokud to **obsahové vztahy** připouštějí'. Termín "obsahové vztahy" je však příliš vágní, a přitom nejsou k dispozici obecná pravidla pro určení volných doplnění, která se (často) vyskytují u jednotlivých sloves, ani pro doplnění, která se s daným slovesem vyskytnout nemohou. Informace o typických doplnění umožňuje zachovat informace, které jsou uváděny v tištěných slovnících, ale nepatří do 'standardního' valenčního rámce. Typická doplnění obvykle nespecifikují význam slovesa, ale rozvíjejí skupiny sloves. Pro některá doplnění mají prototypickou formu (např. instrumentál pro MEANS - *psát perem, jet vlakem*, či předložková skupina pro+4 pro BEN), jindy je jejich forma daná sémantikou doplnění (např. předložkové skupiny pro určení času, místa). Navíc umožňují zachycení dalších syntaktických informací, jako je reciprocita a kontrola.

Negativní informace, tj. nemožnost určitého doplnění, ve slovníku uváděna není.

Rozlišení typických doplnění ve slovníku. Ve slovníku jsou typická doplnění označována zkratkou typ v závorce za funktořem (středníkem odděleno od možných forem).

4.4 Poznámky k přiřazování funktorů

Pro rozlišování aktantů a volných doplnění využíváme "definiční" kritéria (pro volná doplnění možnost opakovaného výskytu a možnost rozvíjet všechna slovesa, pokud to sémantické vztahy dovolují; pro aktanty jedinečnost daného doplnění u konkrétního slovesa a omezený výčet sloves, které mají příslušné doplnění). Přebíráme též pravidla pro přiřazování funktorů z Manuálu pro tektogramatické značkování ([Hajičová et al.,2001]).

Přesto se vyskytují případy, kdy hranice mezi aktantem a volným doplněním není zcela jasná. Protože lingvistické "cítění" jednotlivých anotátorů se může v nejasných případech lišit, snažíme se stanovit kritéria, která by vedla ke konzistentnímu rozhodování. Jako velmi výhodné se ukazuje zkoumání celých skupin sloves s podobným významem (právě tato skutečnost byla hlavním důvodem pro zavádění syntakticko-sémantických tříd).

Dále se zaměřujeme na formulaci pravidel pro přiřazování jednotlivých funktorů.

4.4.1 Nově zavedené funktoři OBST a RCMP

Masivní zpracovávání sloves při vytváření slovníku a při tektogramatické anotaci PDT vede k úvahám o rozšíření seznamu doplnění. Týká se to následujících případů:

- u slovesa se opakují doplnění s blízkým významem (např. vyjádření prostředku), kvůli významovému rozdílu je však potřeba je rozlišit (např. vydělení doplnění překážky, OBST);
- u skupiny sloves se vyskytuje doplnění, často s danou formou, které stojí na rozhraní několika volných doplnění a v konkrétních větách je velmi obtížné přiřadit mu některý ze stávajících funktorů; kvůli konzistenci přiřazování se zdá vhodné zavést jediný hraniční funktor (např. zavedení funktoru odplaty, RCMP).

Při úvahách o nových funktorech se ovšem držíme principu ekonomičnosti popisu, před zavedením nového funktoru zvažujeme možnost jiného zachycení potřebných rozdílů (např. rozštěpením významu, pomocí gramatémů).

Doplnění překážky, funktor OBST (obstacle)

U sloves jako *klepnout (rukou o stůl)*, *omotat se (vodítkem o vývrat)*, *švihnout (ocasem o podlahu)*, *udeřit (hlavou o chodník)* se vyskytují dvě doplnění s blízkým významem. Obě doplnění, vyjádřená Instr a Pg o+4, mají nástrojovou povahu. Zatímco instrumentálem je vyjádřen nástroj, který "aktor drží v ruce", Pg o+4 vyjadřuje pevnou překážku. Tato skutečnost vedla k návrhu nového doplnění překážky (obstacle, OBST). OBST je rekční (daná forma o+4), spojuje se se slovesy jedné sémantické skupiny (slovesa "negativního kontaktu") a nelze jej opakovat, klasifikujeme ho jako kvazivalenční doplnění.

Další slovesa s doplněním OBST:

- ▷ *píchnout se (o nůžky.OBST) × píchnout se/někoho nůžkami.MEANS*
- ▷ *poranit se (o hřebík.OBST), uhodit se (o něco.OBST), zakopnout o schod.OBST), zavadit (nohou.MEANS (o stůl.OBST)*

Doplnění překážky je obvykle fakultativní, ale např. sloveso *zavadit* má obligatorní OBST.

Doplnění odplaty, funktor RCMP (recompense)

Při anotování široké třídy sloves výměny dochází k velkým problémům při přiřazování funktoru doplnění s formou za+4. Touto Pg je vyjadřováno doplnění stojící na hranici mezi AIM, CAUS, COND a MEANS. Navrhujeme proto zavést nový hraniční funktor odplaty (recompense, RCMP). Forma tohoto doplnění je daná sémantikou Pg a může se vyskytovat u velkého množství sloves, proto je považujeme za volné doplnění.

- ▷ věnovat filmové hvězdě 100 000 za úsměv.RCMP
- ▷ dát někomu provizi, peníze za služby / za ochotu / za výpomoc.RCMP
- ▷ dát někomu auto za peníze.RCMP
- ▷ darovat někomu něco za dobrý výkon.RCMP
- ▷ prodal dům prvnímu zájemci za jeho rychlost.RCMP (=protože byl rychlý)
- ▷ bere za hodinu angličtiny.RCMP 200 Kč.PAT od žáků.ORIG
- ▷ vzal si.BEN od nás.ORIG za svou výpomoc.RCMP malou pozornost.PAT
- ▷ dostat od matky.ORIG něco.PAT za peníze.RCMP
- ▷ dostal od školníka.ORIG důtku.PAT za zlobení.RCMP
- ▷ získat něco za peníze.RCMP

Kromě sloves výměny se RCMP vyskytuje i u dalších sloves: *pracovat za směšnou odměnu, děti přestaly zlobit za zmrzlinu, zůstal doma za to, že dostal dárek.*

4.4.2 Slovesa emocionální s vazbou někdo/něco někoho/něco něčím

Slovesa emocionální mají podle J. Panevové rámeček složený z aktoru v nominativu, adresátu v akuzativu a fakultativního pacientu v instrumentálu:

- ACT(1;obl) ADDR(4;obl) PAT(7;opt)

Tento jediný rámeček má dvojí obsahovou interpretaci:

1. ve větě

- ▷ Muž.ACT mě.ADDR vyděsil (tónem svého hlasu.PAT)
jde o Agens, Experient a Stimul;

2. ve větě

- ▷ Tón mužova hlasu.ACT mě.ADDR vyděsil (svou hroubostí.PAT)
jde o Stimul₁, Experient a Stimul₂.

V interpretaci 2 jde významově o "lexikální pasiva", která se však sama mohou pasivizovat (*Byl jsem vyděšen tónem mužova hlasu je pasivem k větě Tón mužova hlasu mě vyděsil*).

Podobně např. slovesa *bavit, nadchnout, naštvat, nervovat, obtěžovat, ohrozit, ovládnout, ovlivnit, polekat, překvapit, rušit, šokovat, ujistit, unavit, zabezpečit, vyděsit, zklamat, zachránit*

ALE: Toto řešení platí pro emocionální slovesa, naproti tomu např. ve spojení *zranil se nožem*.MEANS je nůž typický nástroj.

4.4.3 Slovesa výměny

Slovesa výměny navrhuje dělit do čtyř skupin:

- ACT (příp. ADDR v dativu) pozbývá PAT a nabývá protihodnotou EFF
- ACT pozbývá PAT
- ACT získává PAT od ORIG
- ACT odnímá PAT (někomu, příp. odněkud)

ACT (příp. ADDR v dativu) pozbývá PAT a nabývá protihodnotou EFF

U prototypických sloves výměny (ACT, příp. ADDR výměnou za PAT získává EFF) je postavení PAT a EFF symetrické, obě doplnění jsou obligatorní.

- *vyměnit*₁ ACT(1;obl) ADDR(3;opt) PAT(4;obl) EFF(za+4;obl)
ruské zajatce.PAT vyměnili za naše lidi.EFF
vyměnil mu.ADDR marky.PAT za dolary.EFF
- *vyměnit*₂ ACT(1;obl) ADDR(s+7;obl) PAT(4;obl) EFF(za+4;obl)
vyměnil s ním.ADDR ranní službu.PAT za odpolední.EFF
(ACT získá odpolední)
- *vyměnit*₃ ACT(1;obl) ADDR(3;opt) PAT(za+4;obl) EFF(4;obl)
vyměňte mi rohlíky.EFF za knedlíky.PAT (chce guláš s rohlíky) (ČNK)
- *vyměnit*₄ ACT(1;obl) ADDR(s+7;obl) PAT(za+4;obl) EFF(4;obl)
vyměnil s ním.ADDR vstupenky.EFF za letenky.PAT (ACT získá vstupenky)

Stejně sloveso *měnit* a dvojice *prohodit, prohodit si*.

ACT pozbývá PAT

- *dát*₁ ACT(1;obl) ADDR(3;obl) PAT(4;obl) RCMP(za+4;typ)
dal filmové hvězdě.ADDR 100 000.PAT za úsměv.RCMP

Stejně pro slovesa *darovat, půjčit, věnovat*. Slovesa *prodat, poskytnout, nabízet, nechat, platit (penězi), přenechat, vydávat* se liší pouze fakultativností ADDR (u nabízet může být PAT vyjádřen též klauzí).

- ▷ *prodal čest.PAT za mísu šošovice.EFF*
- ▷ *věnovat filmové hvězdě 100 000 za úsměv.RCMP*
- ▷ *dát někomu provizi, peníze za služby / za ochotu / za výpomoc.RCMP*
- ▷ *dát někomu auto za peníze.RCMP*
- ▷ *darovat někomu něco za dobrý výkon.RCMP*
- ▷ *prodal dům prvnímu zájemci za jeho rychlost.RCMP (=protože byl rychlý)*

ACT získává PAT od ORIG

- *dostávat* ACT(1;obl) PAT(4;obl) ORIG(od+2;opt) RCMP(za+4;typ)
dostával od matky.ORIG drahé dárky.PAT

Stejně slovesa *brát/vzít* (=přijímat), *kupovat, získat, půjčit si, nabývat* (sloveso *kupovat* má navíc fakultativní adresát v dativu, patient u slovesa *nabývat* může být vyjádřen též genitivem).

- ▷ *bere za hodinu angličtiny.RCMP 200 Kč.PAT od žáků.ORIG*
- ▷ *vzal si.BEN od nás.ORIG za svou výpomoc.RCMP malou pozornost.PAT*
- ▷ *dostat od matky.ORIG něco.PAT za peníze.RCMP*
- ▷ *dostal od školníka.ORIG důtku.PAT za zlobení.RCMP získat něco za peníze.RCMP*

ACT odnímá PAT (někomu, příp. odněkud)

- *odebrat₁* ACT(1;obl) ADDR(3;obl) PAT(4;obl)
odebrali mu řidičský průkaz

Stejně pro slovesa *brát/vzít* (=odnímat) a *zabavit*. Slovesa *okrást, obrat* mají adresát vyjádřen akuzativem a patient předložkovou skupinou o+4.

Zvlášť stojí sloveso *platit₂* (přenesený význam):

- *platit₂* ACT(1;obl) ADDR(3;opt) PAT(4,za+4;obl) EFF(7;opt)
platil za její lásku.PAT ctí.EFF
platil dobrodružství.PAT poškozeným zdravím.EFF

Pozor: Předložková skupina za+4 může být u všech skupin volné doplnění (CAUS, COND, MEANS,SUBS, ?EXT, časové, ?COMPL):

- ▷ *zaplatil večeři za Marii.SUBS*
- ▷ *vyměnil jim marky za 18 Kč.EXT (= po 18 Kč)*
- ▷ *za dva roky.TWHEN vymění dva byty za jeden větší.EFF*

4.4.4 Slovesa pohybu a funktoxy PAT a DIR, LOC

Slovesa pohybu s předponou, která "ztranzitivněla", mají dva rámce:

- s obligatorním patientem:
 - ACT (1;obl) PAT(4;obl)
 - ▷ *přejít most, vyjít kopec, přeskochit příkop*Tento rámec vyjadřuje "celistvost", "definitivnost", "zasaženost".
- s obligatorním doplněním směru (odkud, kudy, kam)
 - ▷ *přejít přes most, projít okolo, přeskochit přes příkop.DIR2, vyjít na kopec.DIR3*Tento rámec vyjadřuje "průběhovost".

4.4.5 Slovesa pravení a funktoři PAT a EFF

Slovesa pravení (verba dicendi) rozdělujeme do čtyř tříd ([Hajičová et al., 2002]), podle vztahu **tématu** (= to, o čem se mluví) a **dikta** (= co se o tom říká). Kritériem pro "roztržení" tématu a dikta je možnost vyjádření EFF akuzativem (alespoň v některých spojeních, jako *mluvit pravdu*). Zavedené třídy:

- Slovesa se dvěma valenčními rámci (s tématem i diktem)
- Slovesa s jediným valenčním rámcem (s tématem i diktem)
- Ostatní slovesa pravení
- Další slovesa, okrajově patřící ke slovesům pravení

Slovesa se dvěma valenčními rámci (s tématem i diktem)

Ve valenčním poli je jak téma (vyjádřené patientem), tak diktum (vyjadřuje je efekt). U významu 2 se může realizovat jen téma.

- vyprávět*₁ ACT(1;obl) ADDR(3;opt) PAT(o+6;opt) EFF(4;obl)
o té příhodě.PAT vyprávěl přátelům.ADDR hrozné věci.EFF
- vyprávět*₂ ACT(1;obl) ADDR(3;opt) PAT(o+6;obl)
vyprávěl přátelům.ADDR o dovolené.PAT

Podobně další slovesa – *číst, dozvědět se, odpovídat, zpívat, povídat, psát, říci, ...*

Slovesa s jediným valenčním rámcem (s tématem i diktem)

Slovesa mají jeden význam, ve valenčním poli mají jak téma (PAT), zpravidla fakultativní, tak diktum (EFF).

- dotat* ACT(1;obl) PAT(k+3;opt) EFF(4;obl)
- prohlašovat* ACT(1;obl) PAT(o+6;opt) EFF(4;obl)
- tvrdit* ACT(1;obl) ADDR(3;opt) PAT(o+6;opt) EFF(4;obl)
- ukázat* ACT(1;obl) ADDR(3;opt) PAT(o+6;opt) EFF(4;obl)

Podobně pro řadu dalších sloves – *domnívat se, oznámit, poznamenat (si), předpokládat, vyjadřovat, zapamatovat si, ...*

Diktem je obsah výpovědi nebo její pojmenování (v akuzativu). Vedlejší věta se netransformuje v nominalizaci vůbec (**prohlašoval svůj příjezd*), nebo se transformuje v akuzativ (*sdělil nám (o svém pobytu), co plánuje = své plány*).

Ostatní slovesa pravení

Tato slovesa mají pouze patient. Pokud se u nich objeví vedlejší věta (jako umělé "roztržení" tématu a dikta, např. *spekuloval o výsledku sporu, že byl nespravedlivý*), je to věta doplňková s funktořem COMPL.

- domlouvat se* ACT(1;obl) PAT(o+6/na+6;obl) ADDR(s+7;obl)
- hovořit* ACT(1;obl) PAT(o+6;obl) ADDR(s+7;opt)
- promýšlet* ACT(1;obl) PAT(4;obl)
- ptát se* ACT(1;obl) PAT(na+4;obl) ADDR(4;obl)

Podobně u mnoha dalších – *informovat, hlasovat, jednat, mluvit, pochybovat, prosit, připustit, uvědomit, žertovat, ...*

Zde existuje hypotéza, že pokud koexistují vedlejší věta a o+6 (jako PAT), jsou synonymní (na rozdíl od první skupiny). Obě tyto formy jsou kognitivně diktum (nikoli téma).

Další slovesa okrajově patřící ke slovesům pravení

U těchto sloves už ani "rozštěp" tématu a dikta nepadá v úvahu.

| | |
|---|---------------------------------------|
| <input type="checkbox"/> <i>blahopřát</i> | ACT(1;obl) PAT(k+3;obl) ADDR(3;obl) |
| <input type="checkbox"/> <i>děkovat</i> | ACT(1;obl) PAT(za+4;obl) ADDR(3;obl) |
| <input type="checkbox"/> <i>doporučit</i> | ACT(1;obl) PAT(4,Inf;obl) ADDR(3;obl) |
| <input type="checkbox"/> <i>mlčet</i> | ACT(1;obl) PAT(o+6) |
| <input type="checkbox"/> <i>slíbit</i> | ACT(1;obl) PAT(4,Inf;obl) ADDR(3;obl) |
| <input type="checkbox"/> <i>vyřídít</i> | ACT(1;obl) PAT(4;obl) ADDR(3;obl) |
| <input type="checkbox"/> <i>zpovídat</i> | ACT(1;obl) PAT(4;obl) |

4.4.6 Rozlišení ADDR a BEN

Pro rozlišení mezi ADDR a BEN používáme následující kritéria:

1. Pokud význam dativu odpovídá významu předložkové skupiny pro+4, považujeme slovesné doplnění v dativu za benefaktor
 - ▷ *zaplatil jí.BEN hodiny = zaplatil pro ni hodiny klavíru*
2. Pokud lze slovesné doplnění v dativu zaměnit přivlastňovacím zájmenem, považujeme jej za benefaktor
 - ▷ *hladil jí.BEN ruku = hladil její ruku*
 - ▷ *zaplatil jí hodiny = zaplatil její hodiny*
 - ▷ *ale zaplatil učitelce/škole.ADDR její hodiny, * zaplatil jí učitelce/škole hodiny*

Příklady s BEN:

- ▷ *hučelo mu v uchu = hučelo v jeho uchu*
- ▷ *chodil babičce na nákup = chodil na nákup pro babičku*
- ▷ *podepsal čtenáři knihu = podepsat knihu pro čtenáře*
- ▷ *poranit někomu játra = poranil jeho játra*
- ▷ *potvrdit někomu domněnku = potvrdil jeho domněnku*
- ▷ *získat pro někoho byt*

ALE: *nesl jí.ADDR květiny* (parafráze *nesl pro ni květiny*) × *nesl jí.BEN tašku* (parafráze *nesl její tašku*)

4.4.7 Rozlišení ADDR a DIR3, DIR1, LOC

Pro rozlišení adresátu a doplnění směru je důležitá (prototypická) životnost aktantu (*platit řemeslníkovi.ADDR opravu pračky, podobně platit bance.ADDR úroky; ale dorazili sem.DIR3 pozdě, podobně dorazili k nim.DIR3 pozdě*).

Pokud se vyskytne nejistota v rozlišování ADDR a směrových doplnění, je potřeba zvážit, zda jde skutečně o jediný význam slovesa, nebo o dvě lexie, kterým potom odpovídají dva valenční rámce.

Např. u slovesa *brát*: *bere dítěti hračku* (=odnímat), oproti *bere z krabice zboží* (=odnímat a tím vyprazdňovat) se liší významem, jde o dvě lexie s rozdílnými valenčními rámci.

4.4.8 Rozlišení PAT/ADDR a ACMP

Pro rozlišení PAT, příp. ADDR a ACMP platí obecná kritéria pro rozlišení aktantu a volného doplnění. Pokud přesto vznikají nejasnosti, je zatím nutné spoléhat na citění anotátora.

- ▷ *hrát si s dětmi.ACMP, ale hrát si s míčem.PAT, hrát si s dětmi.ACMP s míčem.PAT*
- ▷ *zajet s autem.ACMP do garáže.DIR3*
- ▷ *jít se zbožím.ACMP na trh.DIR3 × jít s dětmi.ACMP na procházku.DIR3*
- ▷ *přijít s nápadem.PAT (idiom)*
- ▷ *bojovat o něco.PAT s někým/proti někomu.ADDR*
- ▷ *diskutovat s někým.ADDR (o něčem.PAT)*
- ▷ *teroristé se sešli s velitelem.PAT × teroristé se sešli s granátami.ACMP u opasku (reciproční se)*
- ▷ *válčit s někým.PAT*
- ▷ *dohodnout s někým.ADDR něco.PAT*

4.5 Poznámky k vyčlenění jednotlivých rámců

Základní pojmy

Sloveso v češtině může mít jediný význam, často ale má významů několik.² Jednotlivým významům odpovídají lexie³, tzv. monosémické jednotky. Lexie je dána valenčním rámcem a významem - změna kteréhokoliv z těchto údajů znamená změnu lexie. Má-li slovníková jednotka (lexém) více významů, hovoříme o polysémickém lexému.

Vyčleňování rámců se tedy děje na základě rozlišování významů polysémických lexémů (jednotlivým valenčním rámcům odpovídají lexie). Oporou při vyčleňování rámců jsou možné kombinace morfematických forem doplnění (využíváme slovníku morfematických realizací valenčních rámců ve formátu BRIEF, viz kapitola 3.4.1). Dále vycházíme z materiálu poskytnutého SSJČ, SSČ a Slovesy pro praxi. Vzhledem k tomu, že valenční slovník preferuje lexikálně-syntaktický přístup, dochází často k redukci počtu významů, a tedy i rámců oproti SSJČ, na některých místech i oproti SSČ.

Hranice mezi jednotlivými významy polysémického lexému není ostrá, anotátorovi však může pomoci morfematická realizace – změna morfematické realizace často indikuje i významový posun.

²V této kapitole jsme vycházeli z následující literatury: [SSJČ], [SSČ], [SČFI], [SPP], [Čermák,Blatná,1995], [Skoumalová,2001], [Panevová,1999], [Filipec,Čermák,1985], [Filipec,1973], [Macháčková,Svozilová,1990], [Sgall et al.,1986].

³lexie = jednotlivá významová užití lexému. [PMČ], s.90

□ *hlásit se*₁ ACT(1;obl) PAT(o+4;obl)
 hlásit se o svá práva

□ *hlásit se*₂ ACT(1;obl) PAT(k+3;obl)
 hlásit se k příbuzným

Jedná se však pouze o tendenci, nikoli o obecné pravidlo:⁴

□ *učit* ACT(1;obl) ADDR(4;obl) PAT(o+6;opt) EFF(3,4;obl)

Naproti tomu různé lexie mohou mít totožnou valenci (tj. stejný repertoár doplnění a jejich stejnou realizaci).

□ *absolvovat*₁ ACT(1;obl) PAT(4;obl)
 absolvovat školu (= zakončit)

□ *absolvovat*₂ ACT(1;obl) PAT(4;obl)
 absolvovat operaci (= zažít, projít něčím)

□ *seznámit se*₁ ACT(1;obl) PAT(s+7;obl)
 seznámit se s novým kolegou (= udělat si s někým známost)

□ *seznámit se*₂ ACT(1;obl) PAT(s+7;obl)
 seznámit se s rozvrhem (= obeznámit se)

Slovesa, která se nevyskytují bez zvrtné částice – reflexiva tantum jsou ve slovníku zastoupena zvláštním lemmatem (např. *smát se*, *bát se*). Pokud sice existuje sloveso bez zvrtné částice, má ale odlišný význam, zachycujeme sloveso se zvrtným *se/si* jako zvláštní lemma (např. *vracet se* oproti slovesu *vracet*). Pokud je zvrtná částice u lexému (popř. jen u jedné lexie polysémického lexému) pouze fakultativní a pokud je lexém se zvrtnou částicí a lexém bez zvrtné částice synonymní, pak tyto lexémy nejsou reprezentovány jako dvě lemmata, ale pouze jako jedno lemma (např. *koukat (se)*, (více viz kapitola 5.1).

Zachycení ve slovníku

Lexikální jednotka - slovesný lexém je uváděn ve shodě s českou lexikografickou praxí v infinitivu. Heslová stať ve valenčním slovníku typicky odpovídá jednomu lexému. V případě, že se dané sloveso vyskytuje i v lexémech formálně vícečlenných (idiomy a frazémy⁵), jsou všechny tyto lexémy zpracovány uvnitř jediného hesla (lexémy formálně vícečlenné řadíme na konec hesla). Jednotlivá hesla se skládají z valenčních rámců a dalších informací příslušných jednotlivým rámcům.

⁴Někteří autoři ovšem považují změnu morfematické realizace za spolehlivé kritérium pro rozlišení významu, např. [Kingsbury et al.,2002]: "This [word-sense] disambiguation is done largely on the basis of differing argument structures and as such avoids the subjectivity of other word-sense distinctions such as those found in WordNet."

⁵Rozlišení viz kapitola 5.6

Příklady

Valenční slovník tedy obsahuje lexémy monosémické povahy, např.

- ▷ *slíbit někomu.ADDR něco.PAT*

a slovesa polysémické povahy, např.

- ▷ *odstoupit₁ ode dveří.DIR1 (= ustoupit)*
- ▷ *odstoupit₂ z funkce.DIR1/PAT (= resignovat)*
- ▷ *odstoupit₃ od smlouvy.PAT (= upustit od něčeho)*
- ▷ *odstoupit₄ majetek.PAT příbuzným.ADDR (= přenechat)*

Kritéria pro rozhodování

Každá změna valenčního rámce znamená změnu významu, tedy jinou lexii. Pomůckou mohou být změny v morfematické realizaci, popř. způsob užití slovesa; primární či posunuté významy slovesa by měly mít vlastní valenční rámec. V případě nejistoty se řídíme spíše syntaktickými vlastnostmi slovesa než jeho lexikálně-sémantickou povahou.

Otevřené otázky

1. Vyčleňují se rámce pro "abstrahovaná užití" sloves vyjadřující schopnost, míru, příp. nástroj:
 - ▷ *psát perem, hovořit anglicky, psát bez chyb*Tyto případy však dle teorie odpovídají spíše případům, kdy je obligatorní aktant vyjádřen nulou, tedy jedná se o všeobecné aktanty.
2. V případě "široké směrovosti" se kloníme spíše k rozdělení do dvou rámců:
 - ▷ *vycházet z lesa.DIR1 × vycházet z předpokladu.DIR1/PAT???*
 - ▷ *dovést děti do školky.DIR3 × dovést projekt k úspěšnému konci.DIR3/EFF*Morfematická realizace se v těchto případech nemění, avšak dochází ke změně ve způsobu užití - v prvním případě se jedná o primární užití, zatímco ve druhém spíše o posunuté užití.
3. Domníváme se, že je vhodné rozlišovat dva rámce i pro primární a idiomatická užití slovesa, i když formálně se morfematická realizace u těchto lexii shoduje:
 - ▷ *jít do lesa (primární užití)*
 - ▷ *jít do sebe (frazém)*
4. Vzácně se vyskytla slovesa v SSJČ uváděná jako monosémická, pro která však rozlišujeme několik slovesných významů (rámce se liší funktoři), např. *pocházet* (mít původ):
 - ▷ *Jan pochází z vesnice.DIR1*
 - ▷ *Nemoc pochází z viru.PAT*
 - ▷ *Kniha pochází z roku 1945.TFRWH*

Empirická zjištění

Valenční slovník obsahuje celkem 1017 sloves a 2935 valenčních rámců. Z tohoto počtu má 25 sloves alespoň 10 valenčních rámců, u 992 sloves se vyskytuje méně než 10 rámců. Na každé

sloveso připadá v průměru 2,88 rámce. Slovesa s nejvyšším počtem rámců: brát - 22, vzít - 19, přijít - 17, vyjít - 16, dělat - 15, dát - 13, jet, mít - 12.

Kapitola 5

Další informace navázané na valenční rámce

Tato kapitola se zabývá vybranými jazykovými jevy, které ve valenčním slovníku zachycujeme nebo plánujeme zachycovat jako atributy valenčních rámců:

- reflexivita
- reciprocita
- kontrola
- odkaz na vidový protějšek
- možné diateze
- způsob užití slovesa
- syntakticko-sémantická třída slovesa
- odkaz na synset EuroWordNetu
- frekvence rámce.

5.1 Reflexivita

Základní pojmy

Reflexiva tantum¹ - slovesa, která se nevyskytují bez zvrtné částice vůbec (*bát se, smát se, stěžovat si*), nebo jejichž význam je značně vzdálen od významu sloves bez této částice (*odehrát se - odehrát, chovat se - chovat*).

Reflexivní forma slovesná (RFS) - spojení 3. osoby slovesa s morfémem *se*. Mluvnice [Mluvnice II], s. 174 a [Mluvnice III], s. 240 připouštějí užití tohoto tvaru také v 2. os. v administrativním stylu:²

▷ *Vyzýváte se, abyste...*

popř. v reprodukované řeči s ironickým zabarvením i v 1. os.:

¹V této sekci jsme vycházeli z následující literatury: [GrepI,Karlík,1998], [Hajičová et al.,2001], [Komárek,2001], [Králíková,1981], [Mluvnice II] s. 174-179, [Mluvnice III] 1987, [PMČ], [Skoumalová,2001], [Skoumalová,2002], [Šmilauer,1966], [Oliva,2001].

²Podle našeho názoru je však takové užití zmíněného tvaru nepřirozené, neboť evokuje užití objektového *se*. Viz též [Šmilauer,1966], s. 150-151.

▷ *Já se prý vyzývám, abych...*

Podle mluvnice [Grepel, Karlík, 1998], str. 135:³ "RFS lze užít u většiny predikátorů implikujících agens nebo procesor (kromě reflexiv)."

Reflexivní slovesná forma obvykle (kromě případů personifikací) vyžaduje osobní ACT. Pod tuto hodnotu spadá jak možnost tvoření zvrtného pasiva (*Dům se staví.* - od přechodných sloves), tak neosobních vazeb (*Do školy se chodí každý den.* - od předmětových nepřechodných a nepředmětových sloves)^{4,5} (více viz 5.5).

Odvozená reflexiva: V [Skoumalová, 2001] jsou odvozená reflexiva charakterizována jako "slovesa, která se chovají jako reflexiva tantum, pocházejí však od vlastních reflexiv⁶. Jejich lexikální význam se však natolik změnil, že nemohou být dále chápána jako vlastní reflexiva." ([Skoumalová, 2001], s. 28).

Zachycení ve slovníku

Atribut reflexivita (refl.).

Navrhujeme zpracovávat pro každý rámeček, nikoli pro celé lemma.

Hodnoty: AuxT, AuxR, Obj (Objse, Objsi), bezd., odvozdal.

AuxT - reflexiva tantum. Navrhujeme zachycovat tato slovesa vždy do zvláštního lemmatu.

Existuje několik bezobjektových sloves, u nichž je nereflexivní a reflexivní podoba synonymní: *koukat* - *koukat se* (srov. [Hajičová et al., 2001], heslo reflexivní sloveso). U takových případech navrhujeme z důvodů systematickosti vytvořit pro zvrtnou podobu samostatné lemma⁷.

AuxR - možnost vytvoření reflexivní formy slovesné (tj. zvrtného pasiva nebo neosobních vazeb, viz výše).

Objse, Objsi - tvar zájmena *se* (Objse), *si* (Objsi) ve funkci objektu. Navrhujeme zaznamenat do atributu možnost užití zvrtného zájmena, označujícího objekt totožný s ACT, pomocí zkratky Obj (např. *mýt se*, *koupit si* v jednom z rámečků sloves *mýt*, *koupit*). Za vhodné bychom považovali rozlišení dativního a akuzativního tvaru: Objse, Objsi (v zápisu některých rámečků se mohou vyskytovat obě hodnoty).

Druhy odvozených reflexiv.

V rámci velké skupiny odvozených reflexiv navrhujeme rozlišovat dvě podskupiny:

- **bezd.:** bezděčná činnost - do této skupiny patří jednak slovesa se životným ACT typu *zabil se*, *utopil se* vyjadřující nezáměrný děj, jednak děje samovolné u neživotných aktorů *vlny se šíří*.

Typ *zabít se*, *utopit se*: jak je patrné, těmito tvary může být také vyjádřen význam *záměrně zabít sebe*, *utopit sebe*. Pro zpracování do val. slovníku navrhujeme, aby v rámci

³Další podmínky pro tvoření a užití refl. formy jsou shrnuty v části o diatezi.

⁴Srov. také v [Skoumalová, 2001], s. 44. Na rozdíl ve významu upozorňuje např. [PMČ], s. 324, [Mluvnice II], s. 176

⁵Stranou necháváme tzv. podnícené pasivum, u něhož je RFS možná ve všech osobách - srov. *Léčím se u dr. Pažourka už dva roky.* ([PMČ], s. 525).

⁶*True reflexive verbs* je v disertaci H. Skoumalové skupina obdobná, nikoli však totožná s našimi slovesy s *se/si* coby objektem.

⁷H. Skoumalová nazývá tato slovesa "reflexiva s fakultativním *se*, *si*" a uvádí, že fakultativnost zvrtné částice je možná jen pro některé významy, pro jiné je její přítomnost obligatorní, nebo naopak nemožná: *Na co (se) koukáš? Kouká ti podolek.* - **Kouká se ti podolek.* Podobně *myslet* - *myslet si*. ([Skoumalová, 2001], s. 29-31)

s PAT, popř. ADDR byla v atributu refl. označena jednak hodnota Obj (tedy záměrně - sebe, sobě), jednak bezd. (bezděčně).

Typ *větev se zlomila, studna se naplnila vodou, vlny se šíří*. Problémy mohou nastat zejména v homonymii s pasivní konstrukcí. Tuto konstrukci však není možné beze změny významu nahradit opisným pasivem.

Pro slovesa bezděčné činnosti nebudeme vytvářet zvláštní lemma ani zvláštní rámec, omezíme se na zápis hodnoty "bezd." v příslušném rámci.

- **odvozdal.:** Další velká skupina odvozených reflexiv, sloves, která mohou existovat i bez *se, si*, mají však význam poněkud odlišný od slovesa bez refl. částice (ale ne zcela, jako u AuxT; základní společný význam dosud pocítujeme) a *se/si* nelze nahradit pomocí *sebe/sobě*, např. *vrátit se, nachodit se, napracovat se* apod. Zároveň tato slovesa nespádají do bezděčných reflexiv. Navrhujeme vytvořit zvláštní lemma.

Další funkce *se/si*.

Zvratné *se/si* může být použito i pro vyjádření dalších jazykových jevů, jež v tomto atributu nezachycujeme:

- reciprocita - zvratné *se, si* může také vyjadřovat reciprocitu. Ve slovníku je tato možnost řešena v rámci samostatného atributu (viz kapitola 5.2).
- tzv. mediopasivum - užit tvar slovesa se zvratnou částicí *se*, subjekt je vyjádřen tvarem dativu, může být i všeobecný, přítomen je hodnotící výraz typu dobře, špatně, snadno, např. *Večer se mi přemýšlí dobře*. Podle H. Skoumalové ([Skoumalová,2001], s. 47) lze tuto konstrukci vytvořit od jakéhokoli nedokonavého slovesa, u něhož existuje zvratné pasivum. Nepovažujeme tedy za nutné označovat ve slovníku zvláště možnost tvoření mediopasiva.
- typ *psát se s něčím, nosit se s něčím*: Jde o ustálený způsob vyjádření míry příznaku (obtížnost, vysoká intenzita), který není nutné zachycovat.

Závěr: Zvláštní lemma tvoříme pro všechna AuxT a pro podskupinu odvozených reflexiv označených jako odvozdal. Hodnoty AuxR, Objse, Objsi a bezd. budou zaznamenány u příslušných rámců.

Příklady

- | | |
|------------------|--|
| □ <i>smát se</i> | ACT (1;obl) PAT(3;opt) MANN(;typ) -refl.: AuxT |
| □ <i>budovat</i> | ACT(1;obl) PAT(4;obl) BEN(3,pro+4;typ) ORIG(z+2;opt) LOC(;typ) -refl.: AuxR, Objsi |
| □ <i>utopit</i> | ACT(1;obl) PAT(4;obl) LOC(;typ) -refl.: AuxR, bezd., Objse |

Kritéria pro rozhodování

Objekt:

Pokud můžeme nahradit krátký tvar zájmen *se/si* tvarem dlouhým *sebe/sobě*, klasifikujeme se, si jako objekt (v zápisu rámců PAT, ADDR/BEN). Namísto tvaru zvratného zájmena je možno užít jiného vyjádření:

- ▷ *koupit si jízdenku - koupit mu jízdenku, koupit dítěti jízdenku*
(*koupil jízdenku sobě, ne Janovi*).

Přikláníme se tak k "tradičnímu" přístupu, vycházejícímu z [Šmilauer,1966]⁸. Tyto případy budou spadat pod některý z rámců slovesa bez zvratné částice (tedy např. *koupit, mýt*).

Problémy a otevřené otázky

A. Homonymie zvratných *se/si*. Pro zjednodušení budoucí práce s tímto atributem uvádíme možnosti homonymie a možnost souvýskytu některých hodnot atributu.

Možnosti homonymie (podle článku K. Králíkové [Králíková,1981] a s jejími příklady (KK), dále podle práce [Skoumalová,2001] a s jejím příkladem (HS) a podle [Mluvnice II]; upraveno pro naše termíny):

1. AuxR - Obj: *Jejich práce se posuzuje několikrát do roka.* × *Posuzuje se velice přísně.* (*sám sebe*) (KK)
2. AuxR - reciprocita⁹: *Někteří nositelé proudu se neutralizují.* (KK)
3. AuxR - Obj - reciprocita: *Něco takového se vidí zřídka.* *Viděl se v zrcadle.* *Viděli se v divadle.* (KK)
4. AuxR - bezd.: *Napětí se zmenší a přenese se na mřížku.* (KK) *Musíme počkat, až se naplní všechny sklenice.* ([Mluvnice II])
5. AuxR - bezd. - Obj: *něco zaseklého se uvolní pomocí něčeho* (AuxR) - *atomové vazby se uvolní* (bezd.) - *vedoucí oddělení se pro ten den uvolnil, neuvolnil však ostatní pracovníky* (Obj). (vše KK)
Zabila se husa a hodovalo se. - *Zabil se pádem ze střechy.* - *Zabil se z nešťastné lásky.* ([Mluvnice II])
6. Obj - bezd.: *Ve všech lidech se zklamal, trpěl depresemi a nakonec se utopil ve vaně.* × *Dítě spadlo do řeky a utopilo se.* (KK)
7. AuxT - reciprocita - Obj: *Jan a Marie se milují.* (HS)

Nemožnosti souvýskytu:

⁸Argumenty K. Olivy v článku [Oliva,2001] nám připadají pro koncepci valenčního slovníku málo přesvědčivé; uplatnění jeho teorie, že všechna *se/si* jsou pouze částicemi, by znamenalo zaznamenávat bez rozdílu výskyt všech sloves se zvratnou částicí: zvláštními lemmaty by pak musely např. být: *mýt - mýt se - mýt si*. Test zaměnitelnosti *se/si* s *sebe/sobě* K. Oliva napadá zejména v případech koreference v poměrně složitých konstrukcích. Při rozhodování, zda je třeba vytvořit zvláštní lemma, nebo jde pouze o jiné vyjádření aktantu/volného doplnění v rámci jednoho rámce, tento test plně obstojí.

⁹Pro úplnost zde uvádíme i homonymie s recipročním *se/si*.

1. zároveň se nemohou vyskytovat hodnoty AuxT a AuxR.¹⁰
2. nemůže být zároveň AuxT a Obj - vymezili jsme, že pro AuxT bude vždy zvláštní lemma.

B. Rozlišení *se* a *si* v zápisu. Problémy spojené s tvarem *si*.

Hodnota AuxR se váže jen k tvaru *se*. U AuxT bude částice *si* zahrnuta přímo v lemmatu, u Obj jsme výše navrhli rozlišení dvou možností této hodnoty: Objse, Obj*si*. Atribut bezděčná činnost se může týkat obou částic, jde však o několik sloves a myslíme si, že není vhodné popis dále komplikovat rozlišováním případů *se se* a *si si*.

Si může dále vyjadřovat, často ve spojení s prefixy *za-*, *po-*¹¹, "uspokojení z děje" ([PMČ], s. 213). Schopnost vyjádření tohoto příznaku má široká skupina sloves, do níž spadají jak slovesa, která považujeme za odvozená reflexiva (*zakouřit si*, *zalyžovat si*, *pospat si*) a která by měla mít samostatné lemma, tak i slovesa, u nichž je *si* nevalenční a viditelně plní jinou funkci - srov. *zanotovat si*, *zazpívat si*, *zatroubit si* atd. Pro tato užití *si* platí tvrzení v článku [Králíková,1981], s. 293:

"Hranice objektového a volného *si* je však velmi nejasná, užití objektového *si* je typické zejména pro projevy hovorové a je podmíněno stylisticky."

Jak dále autorka podotýká "si neobměňuje význam základních sloves", s nimiž pracuje, a tuto otázku dále neřeší. Máme za to, že ve valenčním slovníku je zásadní prvotně zachytit vazebné¹² *si*; pro druhou skupinu sloves není podle našeho názoru nutné možnost užití *si* uvádět. Uvědomujeme si však, že hranice mezi vazebným a volným dativem je někdy vágní a dosáhnout naprosté důslednosti proto bude velmi obtížné.

5.2 Reciprocita

Základní pojmy

Při zpracování reciprocit vycházíme ze studie Panevové ([Panevová,1999]), zde v krátkosti shrnujeme její závěry podstatné pro anotování jednotlivých sloves.

Reciprocita znamená možnost prvku valenčního rámce vstoupit do **vztahu symetrie s jiným prvkem valenčního rámce** (příp. s více prvky). Nutnou podmínkou je vzájemná zaměnitelnost příslušných prvků, tj. aby mohlo být sloveso v daném významu užito reciprocně, musí existovat takové lexikální obsazení, kdy jsou splněny následující podmínky:

- příslušné prvky jsou lexikálně homogenní (např. oba životné),
- prvky jsou v symetrickém vztahu
(*mluvili o sobě hezky* (*Jan mluvil o Marii hezky a Marie mluvila o Janovi hezky*), ale ne *nařídili si* (*Jan nařídil něco Marii ≠ Marie nařídila něco Janovi*).

¹⁰[Mluvnice II], s. 177 upozorňuje na možnost tvoření konstrukcí typu: *To se ti to směje!* i od refl. tantum. Jde vlastně o mediopasivum, v němž je elidováno adverbiale. Tuto problematiku však považujeme za okrajovou a nebudeme ji ve slovníku řešit.

¹¹Ovšem nejenom při prefixaci, srov. "náladové si" u Šmilauera, s. 42 a 222: *žiti si*, *jíti si*.

¹²Srov. i u [Skoumalová,2001], s. 21).

(Navíc doplnění musí mít stejnou hodnotu co do kontextové zapojenosti a sloveso nesmí být rozvíto doplněním, které se vztahuje pouze k jednomu z prvků vstupujících do recipročního vztahu.)

Přejímáme závěr Panevové, podle kterého jde při reciprocitě o speciální užití základního rámce potenciálně recipročního slovesa (*viděli se (navzájem)*) je jen speciálním případem vazby *viděli někoho*).

Morfematické vyjádření reciprocity. Pokud do vztahu reciprocity vstupuje aktor a jedno další doplnění, je toto druhé doplnění "zrovnoprávněno", tj. přesouvá se jako koordinovaná součást aktoru (případně jsou oba členy vyjádřeny syntaktickým jménem v plurálu) a reciprocity se obvykle vyjadřuje reflexivním zájmenem *se/si*, v krátké nebo dlouhé formě a v příslušném pádu (tj. Gen *sebe*, Dat *si/sobě*, Acc *se/sebe*, Loc *sobě*, Ins *sebou*), příp. ve formě předložkové skupiny (např. na+4 *na sebe*).

Někdy však reciprocity potřebuje explicitní vyjádření pomocí adverbii *spolu, navzájem* apod. Předložková skupina s+Instr se obvykle nahrazuje příslovcem *spolu* (*Jan a Petr spolu mluvili*, ale **Jan a Petr mluvili se sebou (navzájem)*).

Pokud se na reciprocitě aktor nepodílí, může plynout pouze z lexikálního významu (formálně může být reciprocity vyjádřena nulovým prostředkem: *seznámil je = seznámil někoho s někým*).

V případě reflexiv tantum se reflexivní částice se nezdvojuje.

Zachycení ve slovníku

Možnost recipročního užití se netýká celého slovesa, ale vždy jeho jednotlivých rámců. Proto ji zachycujeme u každého valenčního rámce, kterého se týká, a to pomocí atributu reciprocity, ve kterém se specifikují prvky rámce vstupující do vztahu reciprocity.

□ *představovat*₁ ACT(1;obl) ADDR(3;obl) PAT(4;obl) EFF(jako+4;opt)
-reciprocity: ACT-ADDR-PAT
představovali se (sobě navzájem) (=seznamovat se)

□ *představovat*₂ ACT(1;obl) PAT(4;obl) REG(v+6;typ)
představovat v něčem špičku, úspěchy; Robinson představuje nového člověka; plat představuje jediný příjem, to představuje zajímavou otázku (=symbolizovat)

Druhý z uvedených rámců nemá možnost recipročního užití.

Příklady

Nejrozšířenější typ reciprocity je reciproční užití aktoru a adresátu:

- ▷ *domlouvat se (spolu) na něčem*
- ▷ *slíbovat si (navzájem) něco*
- ▷ *hradit si (navzájem) cestovní náklady*
- ▷ *kupovat si dárky*
- ▷ *učit se (navzájem) násobilku/násobilce*
- ▷ *dovedli si pomoci navzájem, když bylo potřeba*
- ▷ *zaslat si dárky poštou*

Reciproční vztah aktoru a patientu se může uplatnit např. u sloves *bít, kreslit, vidět, potkat, akceptovat, chodit spolu, zapadat do sebe (mozaika do sebe přesně zapadá)* nebo u reflexiv tantum – *bát se, brát si/vzít si, hádat se, setkávat se, dívat se.*

Poměrně řídký je reciproční vztah aktoru a doplnění efektu nebo původu:

- ▷ *žádají od sebe navzájem omluvu*
- ▷ *dosáhli na sobě slibu*
- ▷ *mít od sebe něco na památku (ACT-ORIG)*
- ▷ *soustředit pozornost na sebe (ACT-EFF)*

Do vztahu reciprocity nemusí vstupovat aktor – může se týkat patientu a adresátu

▷ *obrátil je proti sobě*
patientu a původu (*oddělovat, odlišovat*)

- ▷ *dělili je od sebe dva roky*
- ▷ *lišila je povaha*

nebo patientu a efektu

- ▷ *přeložili klády (přes sebe)*
- ▷ *spojit lana (=jedno s druhým)*
- ▷ *srovnat různé způsoby*
- ▷ *střídat nálady (veselou se smutnou).*
- ▷ *navazovat provaz a lano*

V daleko širším měřítku než u [Panevová,1999] je zřejmě potřeba uvažovat o kombinaci aktantu a volného doplnění. Může jít o reciproční vztah aktoru a určení směru, resp. místa:

- ▷ *přepadnout přes sebe (ACT-DIR2)*
- ▷ *přiblížit se k sobě, docházet k sobě (ACT-DIR3)*
- ▷ *válet se po sobě, pořídili u sebe dobře (ACT-LOC)*

o vztah aktoru a benefaktoru:

- ▷ *budovali si (jeden druhému) šťastnou budoucnost*
- ▷ *čekají si (v případě potřeby) s dluhy*
- ▷ *udržovali si zahradu (navzájem jeden druhému, když byl někdo z nich nemocný)*

či aktoru a jiných volných doplnění:

- ▷ *hráli (sami) za sebe, mluví za sebe (navzájem)(ACT-SUBST)*
- ▷ *čekají spolu dítě, udržovali spolu krok (jako udržoval krok s dobou), žili spolu, hrají si spolu (ACT-ACMP)*

Objevil se i příklad, kde je nutno uvažovat o symetrickém vztahu mezi aktorem a doplněním normy

- ▷ *zařídili se podle sebe.NORM*

Reciprocita se může týkat různých dvojic doplnění jednoho slovesa:

| | |
|--|--|
| ACT-PAT <i>předvedli se přátelům, před přáteli popisovali se (navzájem)</i> | ACT-ADDR <i>předvedli si své vynálezy popisovali si zážitky</i> |
| ACT-PAT <i>hodili se domů (autem)</i> | ACT-BEN <i>hodili si domů věci autem</i> |
| ACT-PAT <i>očekávali se (s napětím) přijímat se</i> | ACT-ORIG <i>očekávat od sebe omluvu přijímat něco od sebe</i> |

Do vztahu reciprocity mohou vstupovat tři doplnění u sloves se třemi či čtyřmi aktanty.

ACT-ADDR-PAT *mluvit, seznamovat, jednat, četli si (jeden druhému) o sobě (navzájem)*

ACT-PAT-ORIG *dozvěděli se od sebe o sobě/na sebe všechno, ? udělali si ze sebe otroky*

ACT-ADDR-EFF *hráli spolu (jeden s druhým) karty o sebe (jeden o druhého)*

ACT-PAT-EFF *připojili se k sobě, dvojčata se vydávala za sebe*

Reciproční vztah může být i mezi dvěma aktanty a volným doplněním:

ACT-ADDR-BEN *dělali si pro sebe dárky*

ACT-PAT-DIR1 *posunout se od sebe*

ACT-PAT-DIR3 *obracet se k sobě, posunout se k sobě*

ACT-ADDR-DIR2 *posílali si po sobě zprávy*

ACT-PAT-CAUS *trápili se (navzájem) kvůli sobě*

Panevová popisuje zejména případy, kdy do recipročního vztahu vstupují aktanty, připouští však i volná doplnění (*přiblížit se k sobě*, ACT-DIR3). Na zkoumaných slovesech se ukázalo, že tento jev je nutno chápat širěji – symetrické vyjádření aktantu a volného doplnění je poměrně obvyklé. Další zkoumání zasluhuje reciproční vztah mezi několika členy rámce.

5.3 Kontrola

Základní pojmy

Problematika kontroly¹³ se týká tzv. **sloves kontroly** (verbs of control), tj. sloves, jejichž valenční člen může být vyjádřen slovesem v infinitivu.^{14,15} U takových sloves zachycujeme vztah mezi **kontrolovaným členem** rámce (controllee, tj. člen rámce, který by byl subjektem infinitivu, jeho povrchová realizace není možná) a příslušným **kontrolujícím** valenčním členem řídicího slovesa kontroly (controller).

Příklad:

▷ *Marie se pokouší zpívat.*

Marie je aktorem slovesa *pokoušet se* a zároveň je subjektem závislého slovesa *zpívat*, protože zpívá Marie; Marie je zde tedy člen kontrolující i kontrolovaný.

V odborné literatuře (zejména pro angličtinu) se setkáváme s jiným členěním sloves s infinitivem:

¹³V této sekci jsme vycházeli z následující literatury: [Hajičová et al.,2001], [Mluvnice III], [Panevová,1996], [Panevová,1998], [Skoumalová,2002].

¹⁴Obecně se za slovesa kontroly považují taková slovesa, která mají dva koindexované členy valenčního rámce, tj. kromě vyžadování valenčního členu ve formě infinitivu může mít např. i formu substantivní nominalizace *Petr_i od maminky_j požadoval pomoc /Sb_j/*. Takové případy ve slovníku zachyceny nejsou.

¹⁵Infinitiv může u sloves stát v subjektové pozici, může být vyjádřením PAT, EFF, INTT.

- Raising verbs: subjekt infinitivu se stává subjektem nebo objektem řídicího slovesa, zároveň ale nepatří do rámce řídicího slovesa: *Jirka se zdá být veselý*. Subjekt může být vyjádřen celou klauzí.
 - subject raising verbs - zejména modální a fázová slovesa, sloveso *zdát se*. Subjekty obou slovesných tvarů jsou koreferenční. *Petr smí odejít*.
 - object raising verbs - zejména slovesa vnímání. Subjekt infinitivu je objektem řídicího slovesa. *Slyšel Karla zpívat*.
- Equi verbs (slovesa s vlastností kontroly): jeden z aktantů řídicího slovesa je koreferenční s aktantem slovesa vyjádřeného infinitivem. *Jirka se snaží být veselý*.

Zachycení ve slovníku

Do jednotlivých rámců je zaveden atribut control. Hodnotami tohoto atributu jsou členy valenčního rámce řídicího slovesa, plnící funkci kontrolujícího členu. U jistých sporných příkladů užíváme hodnoty ex (viz níže).

Příklady

Funkci kontrolujícího členu mohou plnit následující aktanty a volná doplnění, jež vyžaduje řídicí sloveso:

- | | |
|-----------------------|---|
| □ <i>dojít</i> | ACT(1;obl) INTT(Inf;obl) <i>došel sousedce nakoupit</i> -control: ACT |
| □ <i>dát</i> | ACT(1;obl) PAT(4;obl) EFF(Inf;obl) <i>dala děti spát</i> -control: PAT |
| □ <i>dovolovat</i> | ACT(1;obl) ADDR(3;obl) PAT(4,Inf,aby;obl) <i>každý víkend jim tatínek dovoloval jít do kina</i> -control: ADDR |
| □ <i>chtít</i> | ACT(1;obl) PAT(4,Inf,aby;obl) ORIG(po+6,na+6,od+2;opt) AIM(na+4,k+3;typ) BEN(pro+4,typ) <i>chce od rodičů zaplatit školu</i> -control: ORIG |
| □ <i>představovat</i> | ACT(1,Inf;obl) PAT(4,že;obl) BEN(pro+4;typ) <i>najít vhodnější řešení tohoto úkolu představovalo pro několik generací problém</i> -control: BEN |

Další příklady lze nalézt v článku [Panevová,1996].

U některých sloves může být kontrolujícím členem ACT, nebo ADDR, např.

- | | |
|--------------------|---|
| □ <i>nabídnout</i> | ACT(1;obl) ADDR(3;obl) PAT(4,Inf,že,aby,ať;obl) EFF(7,jako,za+4;opt) AIM(k+3,na+4;typ) EXT(za+4;typ) |
|--------------------|---|

přítel mi nabídl napsat ten dopis místo mne (ACT) × *přítel mi nabídl přespat ještě jednu noc u něj* (ADDR)
-control: ACT, ADDR

Kontrolující aktant může být všeobecný, srov. všeobecný ADDR ve větě:

- ▷ *Tento obchod umožňuje nakoupit zbrusu nový nábytek za nízké ceny.*

Jelikož se jedná o jednu z možností realizace konkrétního typu aktantu, zvláště tento fakt nezachycujeme.

Problémy, otevřené otázky

Jistý problém představují struktury typu: *dala (dětem) spravit boty*. Subjekt infinitivu nepovažujeme za všeobecný aktant, na jeho konkretizaci lze však usuzovat jen na základě pragmatických znalostí. Zpracování těchto konstrukcí je jistě komplikováno faktem, že se ve skutečnosti jedná o struktury sekundární diateze (srov. [Mluvnice II] III, s. 251-252); pro jejich marginalnost je však nezaznamenáváme do atributu diateze (viz kapitola 5.5). Do atributu control tak prozatím u příslušného rámce slovesa *dát, dávat, nechat, nechávat* zapisujeme zkratku ex.

Problémy může působit také určení kontrolujícího členu u konstrukcí s neosobním slovesem *patřit se*:

- ▷ *Pustit dámu první ve dveřích se patřilo.*

I u tohoto slovesa také ponecháváme zatím označení ex v atributu control.

5.4 Vidové dvojice

Vysvětlení základních pojmů

Slovesa dokonavá, nedokonavá. Vid¹⁶ představuje vztah slovesa k završenosti (ukončenosti) děje. Slovesa dokonavá vyjadřují, že děj byl ukončen, nebo bude završen; slovesa nedokonavá se k faktu ukončení děje nevyjadřují. Existují také slovesa obouvidová, která mohou v kontextu vystupovat jako dokonavá, nebo nedokonavá. ([PMČ], s. 318-319.)

Vidová dvojice. V našem pojetí se jedná o tzv. "čisté vidové dvojice" - tedy o dvě slovesa se stejným lexikálním významem, lišící se pouze videm. Ve zpracovávaném vzorku jsou velice málo uváděny prefigované dokonavé protějšky, předpona často mění význam. V souladu s přístupem mluvnice [Mluvnice II] jsou uvedeny některé vidové protějšky s tzv. prostě vidovými předponami (např. *na-* v *napsat - psát*). Statut tohoto pojmu je však, jak uznávají sami autoři Mluvnice, značně neurčitý.

Násobenost. U nedokonavých sloves je možné vymezit kategorii násobenosti. Slovesa násobená vyjadřují neaktuální, uzuální děje.¹⁷

Zachycení ve slovníku, příklady

Atribut vid je zpracováván pro každý slovesný rámec, neboť jednotlivé rámce jednoho lemmatu se mohou ve vidové charakteristice lišit. Zápis atributu obsahuje tyto informace: v závorce vidovou charakteristiku slovesa (hodnoty dok., nedok., dok. i nedok., nás.); vidovou dvojici a

¹⁶V této kapitole byla použita následující literatura: [Mluvnice II], [PMČ], [SSJČ], [SSČ].

¹⁷Význam násobenosti mají i některé jiné typy sloves, např. *chodil na houby*; tato slovesa však nutně nemusí mít rys neaktuálnosti. Srov. [Mluvnice II], s.185

za ní v závorce její vidovou charakteristiku a případně znak ”=”.

Příklad: rámce slovesa *odpovídat*:

- ACT(1;obl) ADDR(3;obl) PAT(na+4;opt)
EFF(4,7,že,zda,aby,ať;obl) MANN(;typ) MEANS(7;typ)

odpovídal mu na jeho dotaz pravdu / že ... / činem, smíchem, se smíchem; odpovídali si (navzájem) na dotazy
-vid: (nedok.); odpovědět (dok.) =
- ACT(1,že;obl) PAT(3;obl) REG(7;typ)
řešení odpovídá (svými vlastnostmi) požadavkům
-vid: (nedok.)
- ACT(1;obl) ADDR(3;opt) PAT(za+4;obl) MEANS(7;typ)
odpovídá za své děti, odpovídá za ztrátu svým majetkem; odpovídají (si) za sebe (navzájem)
-vid: (nedok.)

Znaménko = za vidovou dvojicí označuje totožnou valenci obou sloves (viz výše). Ve velké většině případů se rámce obou sloves neliší. Problematickým zůstává např. posouzení sloves *požádat někoho o něco* a *požadovat něco od někoho*, jejichž výklad se liší i ve slovnících češtiny.¹⁸

Obouvidová slovesa jsou označována ”dok. i nedok.”.

Násobenost. Za informací o dokonavosti, resp. nedokonavosti slovesa a za určením vidové dvojice následuje násobené sloveso, pokud existuje, s informací (nás.) a příp. znakem =, srov. jeden z rámců slovesa *rušit*:

- *rušit* ACT(1;obl) PAT(4;obl) MEANS(7;typ)
z technických důvodů rušíme plánovaný koncert, rušit federální fond, rušíme malé obchody
-fvid: (nedok.); rušivát (nás.) =

Násobená slovesa je možno tvořit jen od některých slovesných tříd (od sloves 1. tř. typu brát, mazat, 3. tř. kupovat, 4. a 5. slovesné třídy: prosit, trpět, sázet; dělat). Navrhujeme zachycovat násobené protějšky jen v případech, kdy je jejich vytvoření přirozené (např. *říkávat*). Násobená slovesa je jistě možné utvořit od velkého množství sloves tak, že je adresát textu bude schopen identifikovat, půjde však o aktualizaci (srov. např.

▷ *Ona ho o to kdysi žádávala, ale nevyhověl jí.*

Ve slovníku nejsou zachyceny tvary se zdvojenou předponou *-va-*, tento jev považujeme spíše za jev slovotvorný.

Dvě podoby jednoho slovesa jsou psány s lomítkem: *podepisovat/podpisovat*.

¹⁸Slovník spisovného jazyka českého považuje *požadovat* za nedok. protějšek pouze k významu slovesa *požádat*, který je označen jako ”poněkud zast.”: *vyslovit požadavek něčeho, projevit žádostivost něčeho, vyžádat si - bude ti dávat, cokoli požádáš* (Něm.). Naopak Slovník spisovné češtiny pro školu a veřejnost nerozlišuje více významů slovesa *požádat* a obě slovesa spojuje do jednoho hesla.

Kritéria pro rozhodování

Základním kritériem pro rozhodování byl jazykový cit zpracovatelů hesel. U většiny sloves byla provedena kontrola se Slovníkem spisovné češtiny pro školu a veřejnost, kde je vidová charakteristika zachycena. Ne vždy pro nás však byla informace ze Slovníku směrodatná: SSČ např. neuznává u slovesa *dostat* (=dopadnout) vidovou dvojici *dostávat*. Tato vidová dvojice je pro nás i v uvedeném významu přijatelná.

V odborné literatuře není jednotný názor na vidovou charakteristiku sloves pohybu. Neprefigovaná slovesa pohybu (*jít, jet, chodit*) jsou podle našeho názoru nedokonavá bez vidového protějšku. Vztahy mezi *jít* a *chodit*, *jet* a *jezdit* se nedají omezit kategorií vidu, popř. násobenosti; roli zde hraje tzv. determinovanost nebo indeterminovanost. U prefigovaných sloves pohybu jsou ve slovníku soustavně uváděny vidové protějšky (např. *projít - procházet*).

Problémy a otevřené otázky

Zpracovávání vidové problematiky je komplikováno faktem, že jsou nejednou do stejného slovníkového hesla zařazena homonyma: např. u hesla *dokázat* spolu s významem 1. =umět (bez vid. protějšku jako většina sloves s modálním význ.) 2. a 3. =provést důkaz, osvědčit (dokazovat) také 4. =skončit kázání; u hesla *dostávat* také význam: *dostával svým slibům*, s dokonavým protějškem *dostát*.

Při vytváření vidové charakteristiky je třeba zvláště pozorně přihlížet k určení vidových dvojic u ustálených spojení, frazeologismů apod. Často pro svou ustálenost nemohou být slovesa v takových spojeních nahrazena vidovým protějškem, srov. *přestat - přestávat* (x *to už přestává všechno*).

V SSČ jsou u některých sloves zvláště vymezeny významy pasivních tvarů (např. *být ztracen = být v beznadějném stavu*). Tato skutečnost není zatím zahrnuta v rámcích, proto jsme ji nezkoumali ani u určování vidového protějšku.

Praktické ověření

Prověření provázanosti sloves *dovést* a *dovádět* z hlediska vidové opozice bylo provedeno na vzorku vět z ČNK. 500 náhodně vybraných příkladů lemmatu *dovést* a 173 příkladů lemmatu *dovádět* potvrzuje přiřazení vidových protějšků k jednotlivým rámcům slovesa *dovádět*: pro význam primárního slovesa pohybu s PAT a DIR3 neexistuje čistá vidová dvojice, a to

- ▷ *Z turnovského náměstí nás například pěšky spolehlivě <dovede> přes Betlémské skály a Drábovnu červená , od roku 1929 zvaná Šimáková stezka.*

V případě užití tzv. historického presentu se však nedokonavý tvar výjimečně u slovesa pohybu objevil:

- ▷ *Zdejší " třiaosmdesátka " mě pak <dovádí> na nejsevernější bod Oahu a právě zde - na samém severu Oahu - pak konečně poznávám město Laie , které v mém havajském cestovním plánu stálo na jednom z nejpřednějších míst.*

Vidový protějšek také nemá *dovést* v platnosti modálního slovesa (=umět, být schopen).

Na rozdíl od charakteristiky ve Slovníku spisovné češtiny pro školu a veřejnost tedy můžeme konstatovat, že jedině přenesený význam tohoto slovesa má čistou vidovou dvojici - *dovádět*.

V materiálu se také mnohokrát vyskytlo homonymum *dováďet₂* (= skotačit). Toto sloveso je bez vidového protějšku.

5.5 Diateze

Základní pojmy

Diateze¹⁹ označuje vztah mezi sémantickými participanty věty a syntaktickými funkcemi (pozicemi), zvl. subjektem a objektem, vyjadřovaný především slovesným rodem ([Lotko,2000]). V popisu užíváme termínů diateze primární (výchozí, s verbem v aktivním rodu) a sekundární (od primární odvozená).

Zachycení ve slovníku

Podle našeho názoru by při slovníkovém zpracování bylo vhodné do atributu diateze zapisovat:

- zda a jaké jeho sekundární diateze jsou možné
- jak se změní rámeček.

Navrhujeme zachycovat možnost tvoření opisného pasiva (OP), diateze s *mít*+trpné participium (m+p) a *dostat*+trpné participium (d+p). Pro zápis by např. bylo možné vyjít od aktantu, nebo volného doplnění, které je na povrchové rovině v sekundární diatezi vyjádřeno jménem v nominativu, např. zápis opisného pasiva: diateze: OP - PAT.

Příklady

- *prohlašovat* ACT(1;obl) PAT(4;obl) EFF(7,za+4;obl) [A]
prohlašovali ho za blázna, prohlašoval ho ze svého nástupce
-diateze: OP-PAT
- *povolit* ACT(1;obl) ADDR(3;opt) PAT(4,Inf,aby;obl) [B]
povolili mu přestup na jinou školu
-diateze: OP-PAT, m+p-ADDR, d+p-ADDR
- *ověřit* ACT(1;obl) PAT(4,že,zda,jak;obl) MEANS(7;typ) BEN(3;typ)
ověřit (si) správnost hypotézy pokusy; ověřit platnost podpisu
-diateze: OP-PAT, m+p-ACT,BEN

Kritéria pro rozhodování

V této fázi přípravy slovníkového zpracování bylo důležité prozkoumat, při užití kterých slovesných tvarů může dojít ke změnám v rámci. Výsledky takového zkoumání vedly k dalším závěrům o typech struktur, jež je nutno v tomto atributu zaznamenávat. U některých struktur podáváme na základě dostupné literatury podrobnější výklad o podmínkách tvoření apod.

¹⁹V této sekci jsme vycházeli z následující literatury: [Grepš,Karlík,1998], [Lotko,2000], [Mluvnice II], [Mluvnice III], [PMČ], [Skoumalová,2001], [Skoumalová,2002], [Šmilauer,1966].

Příčestí minulé a aktivním užití.

Rámec beze změny.

Infinitiv aktivní.

Tato problematika je řešena v kapitole o kontrole (5.3).

Imperativ.

Rámec se zachovává, v bezpříznakových větách na povrchové rovině explicitně nevyjádřen ACT.

Přechodník přítomný aktivní.

Rámec nezměněn, ACT povrchově může být vyjádřen coby ACT vyžadovaný predikátem věty, ACT je tak vyjádřen jen jednou.

Přechodník minulý aktivní.

Platí to, co u přechodníku přítomného aktivního

Reflexivní forma slovesná.

Pod tento pojem řadíme jak zvratný tvar u sloves přechodných, který má pak pasivní význam (*Dopis se píše hodinu.*²⁰), tak zvratný tvar sloves předmětových nepřechodných i sloves nepředmětových, kde je význam aktivní (*Hlasovalo se o tom v parlamentu. Jelo se tam celou noc.*) (srov. [Mluvnice II], s. 176, [PMČ], s. 324).

Změny ve ztvárnění rámce: nelze povrchově vyjádřit ACT:

- ▷ *Úředník píše dopis hodinu. Dopis se píše hodinu.*
- ▷ **Dopis se píše úředníkem hodinu.*

Pro užívání reflexivní slovesné formy (se zkratkou RFS, jindy nazývané také reflexivním, zvratným pasivem) existují určitá pravidla, vymezovaná v jednotlivých příručkách, která se nyní pokusíme shrnout:

Pravidla pro užívání reflexivní formy slovesné (reflexivního pasiva) (podle [Grepl,Karlík,1998], str. 135):

- RFS nelze užít, je-li patientem mluvčí nebo adresát (RFS je omezena na 3. osobu)²¹, a většinou ani tehdy, je-li patient specifikován jmény označujícími osoby:
 - ▷ *Maminka umývá malou Janičku.*
 - ▷ **Malá Janička se umývá.*- možná záměna s primární diatezí²².
- RFS je možná také ve větách typu
 - ▷ *Léčím se u dr. Pažourka už dva roky.*

²⁰[Mluvnice III] (s. 235) v souvislosti s tím upozorňuje na fakt, že pokud je ACT v primární diatezi vyjádřen názvem instituce, je možné jej v diatezi sekundární vyjádřit pomocí místního předložkového pádu: *Výbor o tom rozhodl - Ve výboru se o tom rozhodlo.*

²¹Srov. [Mluvnice II], s. 174-5 a [Mluvnice III], s.240, kde je připouštěno i užití v 2. os. v administrativním stylu: *Vyzýváte se, abyste... , Odsuzujete se k 5 letům vězení.* K dané problematice viz též [Šmilauer,1966], s. 150-1.

²²Srov. však v [Mluvnice III], s. 239 výklad o tom, že dané pravidlo se vztahuje k dějům, které konatel může vykonávat sám na sobě. U dějů, které konatel sám na sobě vykonávat nemůže, nedochází k uvedené homonymii a užití refl. formy je možné: *Děti se očekují v pátek dopoledne.*

- tzv. podnícené pasivum, možné ve všech osobách, jde o jiný způsob vyjádření hierarchizace²³

- RFS lze užít u většiny predikátorů implikujících agens nebo procesor (kromě reflexiv).
 - ▷ *Spali jsme až do rána. Spalo se až do rána.*
- RFS dáváme přednost u sloves nedokonavých.

O tvoření tvarů opisných a zvratných tvarů pasiva podle vidové charakteristiky pojednává také Šmilauer ([Šmilauer,1966], (s. 151)), když říká: "Opisné pasivum je statictější. Naproti tomu se zvratného pasiva užívá zvláště o činnosti opakované, pravidelné, typické, hlavně u sloves nedokonavých. Je dynamičtější."

Mluvnice [Mluvnice II], s. 178-8 k těmto faktům dodává další: "reflexivní tvar obvykle předpokládá osobní agens", reflexivní tvar tak netvoří slovesa, která nemají osobní agens, např. *pršet, tát, téct*. Jako nespisovné se pocituje použití refl. formy u statických predikátů, např. *Když se má auto, to se to jezdí*. ([Mluvnice III], s. 239).

Opisné pasivum.

Tvar slovesa *být* + trpné participium. [GrepI,Karlík,1998], s. 133-4 uvádí tento transformační vzorec:²⁴

$$\square V1_{nom} - \text{PREDvf akt} - V2_{ak} \implies V2_{nom} - \text{PREDvf op} - V1_{instr}$$

Úředník napsal dopis. \implies Dopis byl napsán úředníkem.

resp.

$$\square V1_{nom} - \text{PREDvf akt} - V2 \text{ jiná než ak} \implies \text{PREDvf op} - V1_{instr} \\ - V2 \text{ jiná než ak}$$

Poslanci o zákonu hlasovali dvakrát. \implies O zákonu bylo (poslanci) hlasováno dvakrát.

Pravidla pro užívání opisného pasiva (podle Skladba češtiny, s. 135):

- OP - u akčních sloves, která vyžadují předmět (jak vyplývá z dřív. výkladu, nemusí jít o předm. v Ak; v našem pojetí půjde nejčastěji o vyjádření aktantů: PAT, EFF)
- OP dáváme přednost oprotirefl. pasivu u sloves dokonavých (*Ta záležitost už byla projednána.*)²⁵

[Mluvnice II] (s. 173) udává další podmínky pro tvoření opisného pasiva:

- nelze jej tvořit od přechodných sloves vyjadřujících tělesné nebo duševní stavy
 - ▷ *Zebou ho nohy.*

²³Při zpracování valenčního slovníku necháváme otázku podníceného pasiva stranou.

²⁴Podobně jako u refl. formy i zde je možné ACT v sekundární diatezi u názvů institucí vyjádřit pomocí předložkového pádu: *Dráha zavedla nové typy místenek. - Na dráze byly zavedeny nové typy místenek.* mluvnice [Mluvnice III], s. 235. Otázkou zůstává, zda v tomto případě nedochází kromě změny hierarchizace propozice oproti primární diatezi i k dalším změnám.

²⁵Konstrukce s OP tvořeným od sloves nedokonavých se podle autorů Skladby češtiny, s. 136, pocítují jako knižní, často až strojené. Srov. obdobnou charakteristiku v [Mluvnice III], s. 241.

- nelze jej tvořit od nedokonavých sloves vyjadřujících statické vztahy různého druhu
 - ▷ *Mám chalupu.*
 - ▷ *Věž převyšuje okolní budovy.*
- nelze jej tvořit od procesuálních přechodných sloves vyjadřujících pohyby částí těla (*Sklopil oči.*) Taková dokonavá slovesa pak však mívají particip. konstrukce se stavovým, rezultativním významem, nebo složený tvar participia v atributu:
 - ▷ *Jeho ruce byly sevřeny v pěst. Seděl nohu přehozenou přes nohu.*
- nelze jej tvořit od refl. sloves s morfémem *se* a s volným morfémem *si* (*smát se, dát si sklenici vína*), lze jej však vytvořit např. od refl. tantum *tázat se*; původní PAT, který v sekundární diatezi získává formu nominativu, je vyjádřen genitivem ([Skoumalová,2001], s. 43).
- netvoří se od sloves nepředmětových (*křičet*)
- agens nemusí být, na rozdíl od agentu reflexivních konstrukcí, živý:
 - ▷ *Byl zabit padajícím stromem.*
 Jak uvádí ve své disertaci H. Skoumalová ([Skoumalová,2001],s. 43, 45), jistou výjimku ve tvoření opisného pasiva představují slovesa, která vyžadují dva akuzativy - *stát koho co* - nelze pasivizovat vůbec, *učit koho co/čemu*. V případě pasivizace druhého slovesa musí být jeden z původních akuzativů vypuštěn:
 - ▷ *Děti jsou učeny. . .*
 - ▷ *Matematika je učena. . . .*

Mít + trpné participium.

Tuto konstrukci se sekundární diatezí lze tvořit od sloves, která kromě ACT v nominativu a aktantu v akuzativu (většinou PAT) vyžadují ještě aktant, popř. volné doplnění vyjádřené formou dativu nebo formou pro+Ak.

- ▷ *Její bývalý spolužák měl kancelář přidělenou správcem objektu. (tzn. Správce objektu přidělil jejímu bývalému spolužáku kancelář.)*

Otázka ACT.

- ACT a doplnění vyjádřené formou dativu v primární diatezi jsou referenčně identické²⁶:
 - ▷ *Jakmile bude vyjasněna situace, vydáme definitivní barevnou verzi, kterou máme připravenou.*
 ACT není třeba znovu vyjadřovat.
- ACT v primární diatezi není referenčně identický s doplněním vyjádřeným formou dativu - v sekundární diatezi pak může být vyjádřen předl. vazbou od+G nebo Instr., tedy
 - $V1_{nom} - \text{PRED}v_f \text{ akt} - V2_{ak,abySent,zeSent,inf.} - V3_{dat}$
 $\implies V3_{nom} - \text{mít+part.} - V2_{ak,abySent,zeSent,inf.} -$
 $V1_{instr./od+V1_{gen}}$

²⁶Existují predikáty, které vždy předpokládají jejich referenční identičnost: *Mám ten problém už promyšlený.*, a naopak také predikáty, které vylučují jejich referenční identičnost: *Mám doporučeno léčit se v lázních.* [Mluvnice III], s. 248

- ▷ *V kuloárech se však znovu mluvilo o sporu kolem dvou divokých karet, které nakonec připadly právě Sukové a Wiesnerové, přestože jednu měla od Čs. tenisového svazu slíbenou Jana Pospíšilová za vítězství v letošním halovém mistrovství ČSFR. (tzn. Čs. tenisový svaz slíbil Janě Pospíšilové jednu kartu.)*

Jiná doplnění a jejich formy se v rámci zachovávají.

Dostat + trpné participium.

Také slovesa, která jsou schopna vytvářet tuto konstrukci, vyžadují doplnění v dativu, popř. ve formě pro+Ak.

Pro typ diateze "dostat + trpné participium" platí v zásadě obdobné podmínky jako pro "mít + trpné participium", s těmito rozdíly:

- ACT v sekundární diatezi může a nemusí být stejně jako u a) povrchově vyjádřen. ACT a doplnění vyjádřené formou dativu však nejsou nikdy referenčně identické.
- užití této konstrukce je možné i u nepřechodných sloves typu: *namlátit, nařezat, nafackovat* a také *vyhubovat, vynadat, vyčinit, vyspílat*

[Mluvnice III] kromě této charakteristiky vymezuje ještě sémantické třídy predikátů, u nichž lze tento typ diateze vytvořit. Jedná se o predikáty s prvkem, který [Mluvnice III] nazývá recipientem (příjemcem); nejde zde tedy o vlastníka, beneficienta, proživatele v termínech [Mluvnice III]:

- predikáty s významem: "způsobit, aby někdo něco měl" (kromě *dát*) - *nahradit, přidat*
- predikáty s významem: "způsobit, aby někdo měl instrukci k nějaké činnosti" - *poručit, přikázat, doporučit*
- predikáty s významem: "způsobit, aby někdo měl nějakou informaci" - *oznámit*²⁷
- výše uvedené predikáty typů *nafackovat* a *vynadat*. ([Mluvnice III], s. 251)

Konstrukce dát/nechat + infinitiv.

Jedná se o sekundární diatezi k primárním větám, v nichž je obsažen tzv. iniciátor činnosti. Věta s primární diatezí tak obsahuje spojení typu *z podnětu/z popudu/na popud+2*.

- ▷ *Lékařka na podnět primáře udělala pacientu kontrolní CT. - Primář dal pacientu udělat kontrolní CT*

Tuto problematiku však považujeme za okrajovou a nebude ve slovníku specifikována.

Konstrukce s tzv. mediopasivem.

Možnost tvoření konstrukcí, které obsahují subjekt v dativu, zvrtný tvar slovesa a hodnotící výraz typu *snadno, lehce, dobře*, např.

- ▷ *Chemie se mi studuje špatně*
- ▷ *Večer se mi dobře přemýšlí*

apod. bude řešena již zachycením schopnosti slovesa tvořit reflexivní formu slovesnou. Jak totiž tvrdí H. Skoumalová ve své disertaci ([Skoumalová,2001], s. 47), tuto konstrukci může

²⁷ Konstrukce s tímto sémantickým typem predikátu jsou však pro nás jen stěží přijatelné. *?dostal oznámeno, ?dostal nahlášeno.*

tvořit jakékoli sloveso, které je schopno tvořit reflexivní pasivum.

Transponované užití členů kategorie slovesné osoby.

- konstrukce s transponovaným užití formy 3. os. pl.:
Jedná se o konstrukce s osobním ACT, který však nelze povrchově vyjádřit.
 - ▷ *Hlásili to v rádiu.*
 - ▷ *Psali o tom v posledním čísle tohoto časopisu.*

Tyto konstrukce jsou homonymní s konstrukcí s realizovatelným vyjádřením ACT. Vzhledem k tomu, že kromě nemožnosti povrchově realizovat ACT se zápis rámce nemění, nepovažujeme za nutné zachycovat ve slovníku tento typ diateze u jednotlivých lemmat.

- konstrukce s transponovaným užití forem 1. a 2. os. sg. a pl.:
 - ▷ *S poctivostí nejdál dojdeš.*
 - ▷ *Když něco slíbím, tak to splním.*

Podobně jako v předchozím se jedná o konstrukce s osobním ACT a ani zde jej nelze explicitně povrchově vyjádřit. Ani možnost tvoření těchto konstrukcí nebudeme uvádět u jednotlivých lemmat.

Závěr: Navrhujeme zapisovat možnost tvoření opisného pasiva (OP), diatezi s mít + participium (m+p) a dostat + participium (d+p) s uvedením doplnění, které v sekundární diatezi bude mít formu nominativu.

Za zbytečné považujeme zachycovat možnost tvoření reflexivního formy slovesné, neboť tato možnost bude zaznamenána již u atributu reflexivita (refl.) v podobě AuxR.

Problémy a otevřené otázky

- Poněkud složitější postup při navrhovaném zachycování změn v rámci vyžadují věty se strukturou mít+trpné participium, a to v případech, kde ACT může být v primární diatezi referenčně identický s doplněním v dativu, nedochází (viz výše) tedy ke změně v rámci, viz citovaný př.
 - ▷ *Jakmile bude vyjasněna situace, vydáme definitivní barevnou verzi, kterou máme připravenou.*V jiných kontextech však ACT a doplnění v dativu identické být nemusí:
 - ▷ *Veškeré materiály pro dnešní jednání mám připraveny od své sekretářky.*Takové chování daných sloves bude nutno zaznamenat: m+p-ACT,BEN.

- U sloves je nutné do zvláštního rámce zachycovat pasiva, která nesou jiný - další význam, např. *být ztracen (nemít žádnou možnost k záchraně)*, *být nasazen (k nuceným pracím)*. Podle našeho názoru by bylo žádoucí zachycovat tyto významy přímo při prvotním zpracování lemmatu.

5.6 Způsob užití slovesa

Základní pojmy

Při teoretickém popisu valenčních rámců se zatím zkoumala slovesa především v jejich základním, bezpříznakovém významu. Snaha o komplexní a úplné zpracování sloves s sebou nese

potřebu zpracovávat a zachycovat nejen tyto základní významy, ale i jejich další, příznakové významy.²⁸

Způsob užití slovesa úzce souvisí s vyčleněním jednotlivých valenčních rámců slovesa, tedy s rozpadem lexikální jednotky, lexému do jednotlivých lexií. Způsob užití slovesa se váže k jednotlivým valenčním rámcům.

Víceslovné lexémy (idiomy a frazémy) řadíme v hesle ke slovesnému lemmatu, ke kterému se vztahují, nevyčleňují se tedy jako samostatná lemmata.

Primární užití.

Jednotlivé lexie nejsou na stejné úrovni. Jedna z lexií bývá chápána jako základní, primární; slovníky ji obvykle uvádějí na prvním místě (viz [SSJČ] a [SSČ])²⁹. Výjimkou jsou lexémy homonymní, u kterých se vyskytuje více primárních významů: např. *žít – on žije* (prim) a *on žne trávu* (prim).

Posunuté užití.

Ostatní lexie jsou od lexie primární derivované. Z toho vyplývá i určitý významový posun, ke kterému při derivaci dochází a který se projevuje v rozdílných významových rysech kognitivní složky významu. Rozdíly ve významové složce se mohou, ale nemusí odrazit v podobě valenčního rámce:

- ▷ *překvapit někoho.ADDR zprávou.PAT* (use: prim)
- ▷ *překvapit pachatele.PAT při činu*(use: posun)
- ▷ *absolvovat školu.PAT (= zakončit)* (use: prim)
- ▷ *absolvovat operaci.PAT (= zažít)* (use: posun)

U derivované lexie se může proměňovat životnost doplnění:

- ▷ *seznámit přítele s rodiči* (živ.) (use: prim)
- ▷ *seznámit žáky s rozvrhem* (neživ.) (use: posun)

Derivace lexií může probíhat různými způsoby: postupným a souřadným odvozením (viz [PMČ]). (Genetické vazby mezi jednotlivými lexemi valenční slovník nezachycuje. Stejně tak rozdíly v pragmatických rysech významu lexií valenční slovník zachycuje v atributu note pouze jako okrajovou poznámku.)

Okrajové užití.

Při okrajovém užití dochází k výraznému posunu významu; na rozdíl od primárních a posunutých užití, které chápeme jako centrum, okrajové užití reprezentuje periferii: idiomy a frazémy. Pojetí frazému je postaveno na vyšší míře anomálnosti různého druhu i stupně: formální, sémantické a kolokační. Za idiom nebo frazém považujeme lexie, v nichž sloveso se svým doplněním tvoří jedinou významovou jednotku. U idiomu toto spojení není vázáno na konkrétní lexém, popř. omezenou skupinu závislých členů (např. *chodit s někým*). Frazém je naopak charakterizován tím, že spojení je vázáno na konkrétní lexém, popř. velmi omezenou skupinu závislých členů: *držet hubu, mít něčeho dost/tak akorát/(až) po krk/pocud/do alelujá/doboha/ažaz*.

Výchozím přístupem je snaha "syntakticky rozebrat, co rozebrat lze", tj. přiřadit i idiomům a frazémům v co nejširší míře 'standardní' funktoři. Funktor DPHR využíváme pouze

²⁸V této kapitole jsme vycházeli z následující literatury: [Čechová et al., 1996], [PMČ], [Filipec, Čermák, 1985], [SČFI], [SSČ], [SSJČ].

²⁹Za základní, na 1. místě zaznamenaný, význam SSJČ většinou považuje význam v současném jazyce nejběžnější; u některých slov pak význam původní

u frazémů, které lze syntakticky rozebrat jen stěží (např. *brát roha*.DPHR). Informace, že jde o frazém nebo idiom, je zachycena v atributu způsobu užití.

Pojem frazém a idiom se tak liší od tradičního chápání těchto pojmů v české lexikologii, kde "frazém je ustálená kombinace alespoň dvou slovních forem, která má celistvý význam (zpravidla nerozložitelný na významy jednotlivých složek) a jejímž charakteristickým rysem je, že minimálně jeden z jejích komponentů je v dané funkci omezen pouze na tuto kombinaci (popřípadě na několik málo dalších)" ([SČFI]). Frazém označuje formální stránku jednotky, zatímco idiom stránku významovou. Pojem frazém a idiom se tedy vztahuje ke stejné jazykové jednotce, avšak pojmenovává ji z různých hledisek.

Pomocná slovesa.

Pomocná, modální a fázová slovesa mají blízko ke slovům synsémantickým, jsou vázána na spojení se slovy plnovýznamovými – jejich význam "se vyprazdňuje". V PDT se pomocná a modální slovesa zachycují jako významového slovesa. Ve valenčním slovníku mají v atributu use přiřazenu hodnotu pomocná.

Primární a posunuté užití slovesa chápeme jako centrální, neokrajové užití. Cílem valenčního slovníku je pokrýt především tuto oblast užití.

Zachycení ve slovníku

Způsob užití je zachycen v atributu use, který má pro primární užití hodnotu prim, pro posunuté užití hodnotu posun a okrajová užití jsou označena jako idiom nebo frazém. Pomocná, modální a fázová slovesa mají hodnotu pomocná.

Příklady

- ▷ *přijít domů*.DIR3 (use: prim)
- ▷ *přišel na nákup/nakoupit*.INTT (use: prim)
- ▷ *přišel na skvělou myšlenku*.PAT (use: posun)
- ▷ *přišli mu*.BEN *na podvod*.PAT, (use: posun)
- ▷ *konečně v jeho jednání přišel obrat*.ACT, (use: posun)
- ▷ *slunce*.PAT *mi*.ACT *dnes přijde jako umrlec*.EFF (use: posun)
- ▷ *sukně přišla na 1000 Kč*.PAT, (use: posun)
- ▷ *přišel o hodinky*.PAT (use: posun)
- ▷ *přišel k penězům*.PAT (use: posun)
- ▷ *přišlo (mu) to vhod*.DPHR (use: frazém)
- ▷ *přijít na řadu/na přetřes/na pořad*.DPHR (use: idiom)
- ▷ *kdo s tím*.PAT *přišel*, .(use: idiom)
- ▷ *přišla do jiného stavu*.DPHR (use: frazém)
- ▷ *přišel k sobě*.DPHR (use: frazém)

Kritéria

Prvním kritériem je rozlišení lexii určitého lexému na užití "centrální, neperiferní, neokrajová": prim, posun, a "okrajová": idiom, frazém. Při "okrajovém užití" by sloveso se svým doplněním mělo tvořit jedinou významovou jednotku, jakési relativně ustálené spojení. U "centrálních užití", která svým počtem ve valenčním slovníku převažují, tomu tak není. Lze je vůči "okrajovým užitím" definovat negativně, tedy jako užití, při němž sloveso se svým doplněním netvoří jedinou významovou jednotku. Rámce s DPHR patří jednoznačně k okrajovým užitím. Při rozlišení "neokrajového" a "okrajového užití" může pomoci [SČFI], avšak pouze informací o těsnosti spojení slovesa a jeho doplnění. (Pojetí frazému v [SČFI] je velmi široké, proto se ne vždy shoduje s valenčním slovníkem, který uplatňuje i syntaktické hledisko.)

Kritérii pro označování způsobu "centrálního užití" lexii jsou rysy: neodvozenost (use: prim; vývojově často primární) či odvozenost (use: posun) významu lexie. Hranice mezi lexemi neodvozenými (primární) a odvozenými (posunuté, přenesené) je těžko definovatelná, řešení je obvykle intuitivní. Jako pomůcka může posloužit SSJČ a SSČ. Při nejistotě, zda se jedná o lexii derivovanou či základní, valenční slovník preferuje spíš označení primární.

Kritériem pro rozlišení "okrajových užití" na idiom a frazém je skutečnost, zda spojení slovesa a jeho doplnění je vázáno na konkrétní lexém, popř. omezenou skupinu závislých členů (frazém), či nikoli (idiom). Hranice mezi idiomy a frazémy je pozvolná. Pomůckou může být SČFI, který sice nerozlišuje mezi idiomy a frazémem, ale může podat informaci o tom, nakolik je spojení slovesa vázáno na konkrétní lexém.

Problémy a otevřené otázky

Základním problémem při klasifikaci valenčních rámců z hlediska jejich užití se stává vzhledem k dynamičnosti slovní zásoby samo vymezení kritérií "neokrajového" (primární, posunutá užití) a "okrajového užití" (frazém, idiom). Asi nejsnáze lze určit primární, základní užití, avšak již mezi posunutým a "okrajovým užitím" je hranice velmi povlná. Česká lexikologie tradičně chápe oblast frazeologie a idiomatiky velmi široce [Filipec, Čermák, 1985]. Neostrost hranic frazeologie, stejně jako splývání s pravidelným jazykem je vnímáno jako přirozený jev. Zatím chybí kritéria, která by přesně stanovila rozdíl mezi "neokrajovým" a "okrajovým užitím".

Další otevřenou otázkou zůstává, které z idiomů a frazémů by měl valenční slovník obsahovat. Nutno podotknout, že každé sloveso se jako frazém či idiom chová individuálně, těžko tak lze chování sloves po této stránce zobecnit a následně vytvořit určitá pravidla, která by platila pro všechna slovesa. Kritériem pro zařazení určitého idiomu či frazému do slovníku prozatím zůstává jeho vyšší frekvence. (Lze vycházet ČNK, viz kap. 5.9)

Empirická zjištění

Zhruba 80% valenčního slovníku tvoří "neokrajová" užití (41% primární užití a 39% posunutá užití) a 20% "okrajová" užití.

5.7 Syntakticko-sémantické třídy sloves

Základní pojmy

Vytváření syntakticko-sémantických tříd v rámci slovníku VALLEX je zatím v samých počátcích. Protože se nám nepodařilo jednoduchým způsobem adaptovat žádnou z existujících klasifikací, přistoupili jsme k budování slovesných tříd 'zdola', od jednotlivých zpracovávaných sloves. Slovesa jsou zatím anotována pouze intuitivně, rozdělení je jen pracovní a zdaleka nepokrývá všechna slovesa, přesto je i v této fázi využitelné pro konzistentní zpracování i pro další výzkum.

Slovesa stejné syntakticko-sémantické třídy (termín je zatím pouze pracovní) by mělo spojovat **podobné syntaktické chování** a dále by slovesa jedné třídy měla vykazovat **podobné sémantické rysy**.

Nutno zdůraznit, že třídy jsou tvořeny nikoli lexémy, ale jednotlivými **lexemi**. Jeden lexém se proto může vyskytovat v několika třídách, srov.:

- *přistoupit*₁ ACT(1;obl) DIR3(;obl)
 přistoupit do autobusu
 -class: pohyb

- *přistoupit*₂ ACT(1;obl) PAT(k+3;obl) MANN(;typ)
 přistoupit k názornému výkladu
 -class: fázové

Domníváme se, že pro vytvoření syntakticko-sémantických tříd by měl být relevantní zejména **valenční rámec**; předpokladem je, že valenční doplnění (obligatorní a fakultativní aktanty a obligatorní volná doplnění) by se typicky měla shodovat. Existují však slovesné třídy výrazně nehomogenní, vnitřně strukturované. Např. verba dicendi:³⁰

- *říkat* ACT(1;obl) ADDR(3;opt) PAT(o+6,na+4;opt)
 EFF(4,že,aby,ať,zda,jak;obl)
 říkal jí o její dceři jen pravdu

- *upozorňovat* ACT(1;obl) ADDR(4;obl) PAT(na+4,že,aby;obl) ME-
 ANS(7;typ)
 upozorňoval jsem tě, abys to nedělal

Vzhledem k obdobným sémantickým rysům těchto sloves je chápeme jako jednu syntakticko-sémantickou třídu. Toto pozorování vede k úvahám o hlubší strukturační syntakticko-sémantických tříd.

Na druhou stranu samotná shoda valenčního rámce nemůže zaručit, že daná lexie bude patřit ke stejné třídě:

- *absolvovat* ACT(1;obl) PAT(4;obl)
 absolvoval školu (zakončit)

³⁰O verbech dicendi viz podrobněji v kapitole 4.4.5

- *charakterizovat* ACT(1;obl) PAT(4;obl)
ta vlastnost ho dobře charakterizuje
- *kontrolovat* ACT(1;obl) PAT(4;obl)
kontrolovat celou zemi
- *obejít* ACT(1;obl) PAT(4;obl)
úředník vás obejde s dotazníkem
- *porušit*₁ ACT(1;obl) PAT(4;obl)
porušit sliby
- *porušit*₂ ACT(1;obl) PAT(4;obl)
porušit zámek

Domníváme se, že shoda tzv. obohaceného valenčního rámce - rámce s typickými volnými doplněními a s doplněními kvazivalenčními nebude pro vytvoření syntakticko-sémantických tříd primárně relevantní. Předpokladem je, že vytvoření slovesných tříd zpětně přispěje k dopracování této problematiky.

Morfematické realizace valenčních doplnění sloves jedné syntakticko-sémantické třídy se nemusí plně shodovat:

- ▷ *obrátit se na někoho*.ADDR *se žádostí*.PAT
- ▷ *uchýlit se k někomu*.ADDR *se žádostí*.PAT.

Domníváme se, že propojením sloves do syntakticko-sémantických tříd vykrytalizují kombinace možných morfematických realizací určitých valenčních členů rámce. Morfematická realizace valenčních doplnění tedy nebude určující pro vytváření těchto tříd, spíš přispěje k popisu morfematických realizací.

Při vytváření syntakticko-sémantických tříd by dále měly napomoci informace o **reciprocitě**, **kontrole** a o možnostech tvoření sekundárních **diatezí** (zvl. reflexivní formy slovesné, opisného pasiva, konstrukci s *mít*, *dostat* + trpné participium (podobně jako v [Levin,1993])).

Vytvoření syntakticko-sémantických tříd by mělo významně přispět ke kompaktnímu dopracování valenčního slovníku; cílem je jednotně propracovávat celé skupiny sloves, nikoli řešit problémy u jednotlivých lexii. Vzniklé problémy by po vytvoření tříd bylo možné řešit systémově.

Jak se daný jev zachycuje ve slovníku

Informace o syntakticko-sémantické třídě slovesa je zachycována v atributu *class*. Zatím pracovně vyčleňujeme následující třídy: pohyb, vnímání, děláni, výměna, dicendi, fázové, umístění v prostoru, modální, změna, změna stavu, transport a šíření.

Příklady

- *přijít* ACT(1;obl) DIR3(;obl)
přijít domů

- class: pohyb
- *vidět* ACT(1;obl) PAT(4,že,jak;obl)
 vidí Karla vcházet
 -class: vnímání

 - *vytvořit* ACT(1;obl) PAT(4;obl) ORIG(z+2;opt)
 vytvořila pro děti.BEN z hadrů panenku
 -class: děláni

 - *vyměnit* ACT(1;obl) ADDR(s+7;opt) PAT(4;obl) EFF(za+7;obl)
 vyměnit s bratrem staré auto za nové
 class: výměna

 - *řici* ACT(1;obl) ADDR(3;opt) PAT(o+6,na+4;opt)
 EFF(4,že,aby,ať,zda,jak;obl)
 řekl mu o své ženě, že je velmi hezká
 class: dicendi

 - *pustit se* ACT(1;obl) PAT(do+2;obl)
 pustil se do práce
 class: fázové

 - *dát* ACT(1;obl) PAT(4;obl) DIR3(;obl)
 dát věci do skříně
 class: umístění v prostoru

 - *smět* ACT(1;obl) PAT(Inf;obl)
 smí přijít
 class: modální

 - *vyvinout se* ACT(1;obl)
 vyvinul se v něm pocit bezpečí
 class: změna

 - *vyrůst* ACT(1;obl) PAT(v+4;opt) ORIG(z+2;opt)
 vyrůstl z chlapce v muže
 class: změna stavu

 - *převést* ACT(1;obl) PAT(4;obl) DIR2(;obl)
 převést chodce přes přechod
 class: transport

 - *proniknout* ACT(1;obl) DIR3(;obl)
 nemoc pronikla do sousedního kraje
 class: šíření

Otevřené otázky

Zatím otevřenou otázkou zůstává využití EuroWordNetu při vytváření syntakticko-sémantických tříd, stejně jako využití dalších existujících klasifikací sémantických tříd (pro češtinu např. [Daneš,Hlavsa,1987]). Inspirativním zdrojem jsou i rozpracované klasifikace sloves pro angličtinu - klasifikace podle Levinové (viz kapitola 2.6), FrameNet (viz kapitola 2.4) a LCS (viz kapitola 2.5).

Další otevřenou otázkou je, které významové rysy budou pro konstituování sémantických tříd sloves relevantní a které nikoli. K řešení této problematiky mohou významnou měrou přispívat výše zmíněné klasifikace.

Zatím neprozkoumanou oblastí také zůstává vztah vidu a sémantických tříd.

Empirická zjištění

Hodnota atributu class byla vyplněna prozatím jen u méně než čtvrtiny anotovaných rámců (ve slovníku je nyní asi tři tisíce rámců, jen 665 má hodnotu tohoto atributu vyplněnou). V nejvyšším počtu jsou ve slovníku zachyceny výše jmenované třídy sloves: dicendi 151, pohyb 126, výměna 63, fázová 30, vnímání 27, změna stavu 17, děláni 17, transport 16, změna 14, umístění v prostoru, modální 6, šíření 5. Uvedená zjištění však uvádíme pouze pro orientaci. Dosud nedošlo k soustavnému zpracování atributu class a hodnoty zde uvedené se zakládají pouze na intuitivním anotování tohoto atributu. Lze očekávat, že po zpracování této oblasti valenčního slovníku se uvedené rozložení změní.

5.8 Navázání rámců na EuroWordNet

Základní pojmy, způsob zachycení ve slovníku

Zhruba polovina sloves dosud zpracovaných ve slovníku VALLEX je obsažena i v české části EuroWordNetu. Každé sloveso se může vyskytovat v několika synsetech (pojem synset viz kapitola 3.5.3), tyto jeho jednotlivé výskyty jsou v EuroWordNetu označeny indexem s přirozeným číslem. Do atributu ewn valenčního rámce zapisujeme index (nebo množinu indexů oddělených čárkou) takového synsetu (resp. množiny synsetů), který svým významem nejlépe odpovídá významu příslušného rámce.

Příklad

Sloveso *odpovídat* je v EuroWordNetu uvedeno ve dvou synsetech:

- český synset: *odpovídat-1, odepsat-1, zodpovědět-1*
anglický ekvivalent: *answer-1, reply-1, respond-1*
anglická glosa: *reply or respond to; "She didn't want to answer"; "answer the question"; "We answered that we would accept the invitation"*
- český synset: *odpovídat-3, shodovat se-2, souhlasit-2, korespondovat-2*
anglický ekvivalent: *match-1, fit-5, correspond-1, jibe-1, gibe-1, tally-1, agree-1*
anglická glosa: *be similar or consisten; conincide in their characteristics; "The two stories don't agree in many details"*

Navázání valenčních rámců na odpovídající synsety pak vypadá takto:

- ACT(1,že;obl) PAT(3;obl) REG(7;typ)
řešení odpovídá (svými vlastnostmi) požadavkům
-ewn: 3
- ACT(1;obl) ADDR(3;obl) PAT(na+4;opt)
EFF(4,7,že,zda,aby,ať;obl) MANN(;typ) MEANS(7;typ)

odpovídal mu na jeho dotaz smíchem/že ...
-ewn: 1
- ACT(1;obl) ADDR(3;opt) PAT(za+4;obl) MEANS(7;typ)
odpovídá za své děti (má zodpovědnost)
-ewn:

Problémové případy

Praktické problémy s přiřazováním synsetů rámcům lze rozdělit do několika skupin:

- Chybějící synset. V EWN u daného slovesa chybí synset, který by dostatečně odpovídal některému z významů slovesa.
- Přebývající synset. V EWN je u některého českého slovesa synset s takovým anglickým ekvivalentem a anglickou glosou (česká glosa není v EWN obsažena), že si použití toho českého slovesa ve významu odpovídajícím anglickému ekvivalentu nedokážeme představit.
- Nezřetelné (příliš jemné, subjektivní) vymezení synsetu. Z českého synsetu ani z anglického ekvivalentu nebo glosy není jasné, kterému z rámců synset odpovídá.

Ideální situace, kdy bylo možné mezi množinou valenčních rámců daného slovesa a množinou jeho synsetů nalézt vztah jedna k jedné, nastávala velice zřídka.

5.9 Frekvence rámce

Základní pojmy

Atribut frekvence obsahuje záznam o testování valenčních rámců. Každé sloveso je testováno na dvou souborech po třiceti větách (s daným slovesem) náhodně vybraných z ČNK. První kolo testování rámců probíhá již při anotování. Každé sloveso je tak už v první fázi zpracování konfrontováno s konkrétním jazykovým materiálem. Anotátorům se tím usnadňuje práce s disambiguací významu, mohou v hesle využít konkrétních příkladů z ČNK, při konfrontaci hesla s ČNK mají možnost rychle odhalit absenci některého významu či povrchové formy jednotlivých prvků rámce apod. Cílem této konfrontace je i snížit počet chyb z nepozornosti, popř. odhalit systémové problémy. Podruhé je sloveso konfrontováno s příklady z ČNK s cílem odstranit možné nedostatky v rámcích a dále má tato kontrola i ověřit, zda se frekvence rámců v náhodně generovaném vzorku 30 vět přibližně shoduje. (Tento předpoklad se nemusí vždy potvrdit následkem toho, že se jedná o skutečně náhodný výběr vět.) Kontroly potvrzují, že anotování valenčního slovníku se bez konkrétního jazykového materiálu neobejde.

Způsob zachycení ve slovníku

V atributu *freq* jsou dvě čísla oddělená středníkem, která odpovídají počtu vět z prvního a druhého testovacího souboru, ve kterých byl daný rámeček použit. Místo nulové hodnoty je použit prázdný řetězec, jsou-li obě hodnoty nulové, není atribut vyplněn vůbec.

Kritéria

Při konfrontaci hesla valenčního slovníku s materiálem načerpaným z ČNK je cílem ke každé větě nalézt odpovídající valenční rámeček. Nejprve je tedy nutné rozlišit, jaká lexie daného slovesného lemmatu byla ve větě použita, a poté zkontrolovat, zda se valenční rámeček užitý ve větě shoduje s podobou valenčního rámce ve slovníku. V případě, že anotátor nedokáže určit, o jakou lexii se v dané větě jedná, lze vyjít od formy valenčního rámce a teprve následně srovnávat významy lexie. Při nedostatku kontextu či v eliptických konstrukcích je někdy možné ke konkrétnímu užití slovesa přiřadit několik rámečků. V takových případech, kdy má anotátor k dispozici pouze formu valenčního rámce, nikoli význam jemu odpovídající lexie, preferujeme lexie primárního užití (atribut *use* s hodnotou *prim*). Např. sloveso *dát* ve větě *Vy mi ty klíče nedáte?* může odpovídat dvěma lexii s různými rámečky:

- ACT(1;obl) ADDR (3;opt) PAT (4;obl)
dát někomu něco (=podat)

- ACT(1;obl) ADDR(3;obl) PAT (4;obl) EFF (za+4;opt)
dát někomu něco (=předat)

Preferujeme však první rámeček primárního užití slovesa.

Problémy a otevřené otázky

Největším problémem při anotování atributu *freq* se stávají věty eliptického charakteru. V takových případech dáváme přednost primárnímu užití slovesa.

Kapitola 6

Budování slovníku – postup a nástroje

V této kapitole jsou nastíněny některé praktické otázky, které bylo nutné řešit při budování slovníku VALLEX. Nejprve je hrubě načrtnut pracovní postup, způsob rozdělení sloves na jednotlivé dávky a distribuce práce na slovníku. Dále jsou popsány použité softwarové nástroje: (i) editor pro ruční anotování slovníku WinEdt a datový formát určený pro ruční anotaci, (ii) WWW rozhraní pro vyhledávání ve slovnících, (iii) editor XSH pro práci s XML soubory a ukázka jeho použití na slovníková data.

6.1 Pracovní postup

Jako první testovací vzorek bylo pro zpracování ve slovníku vybráno přibližně 170 nejčastějších českých sloves (přesněji řečeno šlo o slovesa, jejichž lemmata po případném odtržení reflexivní částice byla nejčastější v ČNK). Po započetí práce na tomto vzorku byl velmi brzo překonán počáteční nepodložený optimismus a ukázalo se, že vytvoření jakkoli zjednodušeného, ale konzistentního valenčního slovníku je velmi náročná úloha. Jedná se především o časové nároky kladené na ruční anotaci, nároky na velkou pečlivost a dostatečnou lingvistickou erudici anotátorů, kterému je také třeba připravit technické zázemí. Jelikož nejde o úlohu, která by byla v rozumném časovém horizontu řešitelná jedním člověkem, je třeba přičíst i čas potřebný pro distribuci práce, koordinování činnosti, kontrolu kvality atd. Dále je samozřejmě nutné vytvářet dokumentaci.

Z praktických důvodů nebylo možné, aby anotátoři sdíleli data celého slovníku a anotovali je všichni najednou (museli by být po celou dobu práce připojeni k Internetu). Proto bylo nutné zpracovávaná slovesa rozdělit do dávek. Po zkušenostech z první fáze se ukázalo, že 170 sloves v jedné dávce je z psychologických důvodů příliš mnoho. Velikost dávky tedy byla snížena na 100 sloves. Dalších 1000 nejčastějších sloves bylo rozděleno podle své četnosti do deseti dávek po sto slovesech, uvnitř dávek byla slovesa seřazena podle abecedy.

V současnosti je tedy ve slovníku zpracováno jedenáct dávek. Důvod, proč se velikost současného slovníku liší od součtu velikostí všech jedenácti dávek, spočívá v tom, že máme oproti zdrojovým slovníkům změnil jiná kritéria pro zpracování lemmat s reflexivní částicí.

Co se týká zpracování jednotlivých dávek sloves, postup lze velmi hrubě načrtnout takto:

1. Automatické "předgenerování" rámců pro všechna slovesa dávky a uložení dat do formátu pro anotaci (viz kapitola 6.2). Rámce jsou přežaty ze slovníku BRIEF a jsou v nich

automaticky předvyplněny funktory¹.

2. První ruční průchod. Slévání rámců, doplňování chybějících rámců, forem, valenčních prvků atd., s využitím příkladových vět z ČNK. Doplnění příkladové věty ke každému rámcu.
3. Konfrontace rámců s druhou dávkou příkladů z ČNK s daným slovesem, s daty z EuroWordNetu a dalšími zdroji (tištěné slovníky, pracovní poznámky anotátorů PDT atd.). Opravy rámců.
4. Doplnění dalších atributů rámců. Opravy rámců.

Každá dávka projde rukama několika anotátorů a je jimi kontrolována, což by mělo eliminovat některé chyby způsobené nepozorností apod. Některé další typy chyb lze nalézt technickými nástroji (překlep v označení funktoru apod.).

Zdá se, že je výhodnější v jednom průchodu dávkou doplňovat vždy jen jeden typ informace (např. jenom reciprocitu, jenom vidové protějšky), než psát celé heslo najednou.

Jak bylo řečeno, anotátoři zpracovávají jednotlivé dávky odděleně. Může se ale stát, že anotátor odhalí nějakou systematickou chybu a má podezření, že by se podobná chyba mohla vyskytnout třeba u několika dalších sloves, na kterých ale zrovna pracuje někdo jiný. Z tohoto důvodu byla slovesa nedávno přerozdělena do dávek založených na sémantických třídách a budou dále zpracovávána tímto způsobem. Udržování konzistence by tak mělo být snadnější.

6.2 Prostředí pro ruční anotaci slovníku

Pro první pokusy s anotováním slovníku byla vytvořena databáze v MS Access. Už během několika měsíců ale bylo zřejmé, že editování takto reprezentovaného slovníku je příliš těžkopádné a málo přehledné, navíc jakákoli změna anotačního schématu nebo rozdělení dat mezi více anotátorů a jejich následné slévání vyžaduje prakticky permanentní programátorskou podporu.² Jedinou výhodou takového přístupu bylo to, že zabránil anotátorovi v některých typech chyb (například použití neexistujícího funktoru).

V další fázi byla data z databáze převedena do formátu XML. Ruční anotace spočívala v editování XML v editoru WinEdt. Výhodou tohoto přístupu byla velmi snadná uživatelská manipulace s daty: rámce bylo možné jakkoli upravovat, kopírovat, mazat a vyhledávat stejným způsobem jako při práci s obyčejným textem. Navíc přestala být problémem současná práce několika anotátorů (XML soubor bylo možné velmi jednoduchým způsobem rozdělovat a slévat). Nevýhodou byla velká "vizuální náročnost" práce s takovými daty. Data v této podobě byla naprosto nevhodná pro tisk, do tisknutelné podoby byla převáděna pomocí XSL transformace.

Poslední formát, který byl pro potřeby valenčního slovníku navržen a který se pro ruční práci se slovníkem dlouhodobě velmi osvědčil, je jednoduchý, řádkově orientovaný textový formát. Pravidla notace pro tento formát byla zavedena takto:

¹Původně byly jako vstup použity valenční rámce ze slovníku Hany Skoumalové, které vznikly automatickým spojováním rámců ze slovníku BRIEF. Spojování rámců ale bylo podle našeho názoru v mnoha případech nadměrné a ukázalo se, že je méně pracné nadbytečné rámce vygenerované přímo ze slovníku BRIEF umazávat než chybně sloučené rámce rozdělovat.

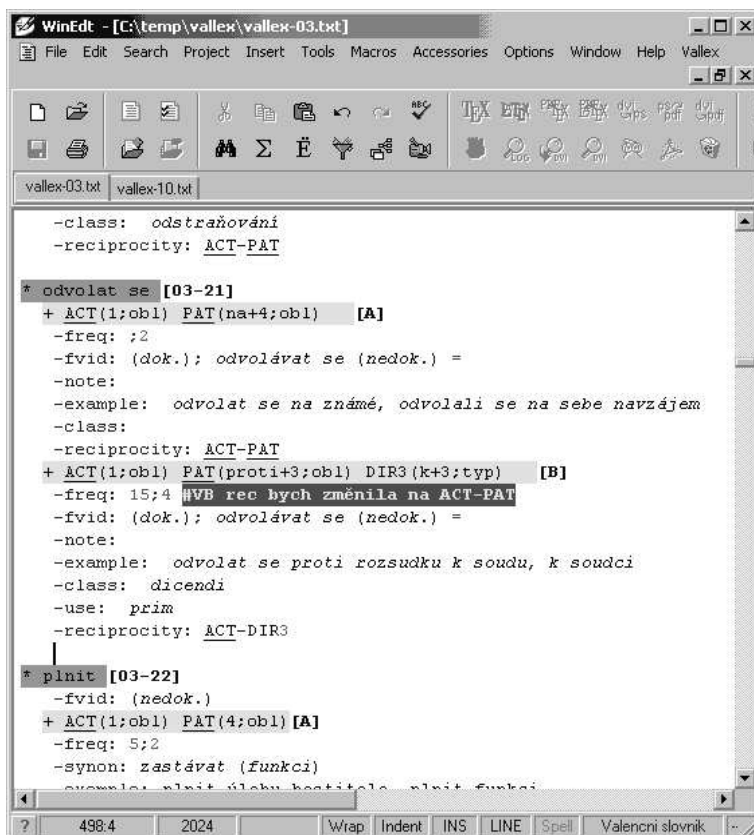
²Techniky pro distribuování databáze by bylo účelné použít jen tehdy, pokud by byli všichni anotátoři neustále on-line. To byla v našem případě nesplnitelná podmínka.

- Opakovaná mezera je považovaná za jedinou mezeru. Prázdný řádek je ignorován.
- Řádek začínající znakem "*" označuje začátek hesla, za tímto znakem následuje lemma slovesa.
- Řádek začínající mezerou a znakem "+" obsahuje valenční rámec, na dalších řádcích jsou pak atributy tohoto rámce.
- Valenční rámec je zapsán jako posloupnost valenčních prvků oddělených mezerou.
- Valenční prvek je zapsán jako funktor, za kterým v kulatých závorkách následuje povrchová forma (nebo posloupnost povrchových forem oddělených čárkou), znak ";" a typ prvku (je-li místo povrchové formy uveden prázdný řetězec, znamená to, že může být použita jakákoli forma vyjadřující daný funktor).
- Povrchovou formu zapisujeme jako číslo pádu (přímý pád), lemma předložky následované znakem "+" a číslem pádu, zkratku inf (infinitiv), lemma podřadící spojky.
- Řádek s atributem rámce začíná mezerou, znakem "-" následovaným názvem atributu, znakem ":" a hodnotou atributu.
- Kdekoli může být vložen dočasný komentář. Dočasný komentář začíná znakem "#" a končí na konci řádku.
- Na konec řádku s lemmatem slovesa a s rámcem může být vložen jejich identifikátor v hranatých závorkách.
- Řádky s atributy rámce mohou být uváděny v jakémkoli pořadí. Jediný povinný atribut je příklad užití (atribut example).
- Kvůli vyšší přehlednosti je každý řádek s rámcem odsazen, každý řádek s atributem je odsazen ještě o další mezeru.

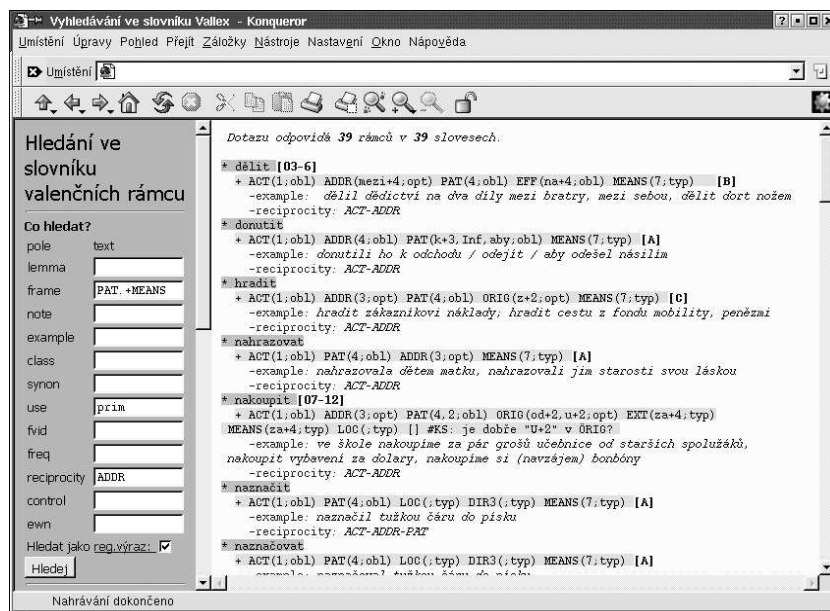
K editaci těchto dat je použit editor WinEdt³ (viz obrázek 6.1). V editoru WinEdt lze zavádět pro různé druhy dat různé nové editační módy. Pro potřeby valenčního slovníku byl vyroben speciální mód, který mj. velmi usnadňuje přehlednost dat díky zvýrazňování jejich syntaxe: řádek s lemmatem je zvýrazněn žlutým pozadím, řádek s rámcem je zvýrazněn zeleným pozadím, názvy atributů mají jiný vzhled než hodnoty atributů, komentář je na kontrastním červeném pozadí atd. Zvýrazňování probíhá automaticky na základě výše uvedených konvencí.

Uvedený formát v kombinaci s editorem WinEdt se v současné době jeví jako optimální prostředí pro ruční anotaci slovníku. Data jsou velmi přehledná a současně poměrně kompaktní (při vizuálním vjemu nepřekázejí např. koncové značky jako v XML), lze s nimi na uživatelské úrovni snadno manipulovat, anotátor prakticky nepotřebuje žádné zručení, protože způsob editování dat se téměř neliší od běžné práce s textem v obvyklých editorech. Drobné změny anotačního schématu (např. zavedení nového atributu) jsou proveditelné prakticky na uživatelské úrovni. Některé typy chyb jsou okamžitě odhaleny pomocí zvýrazňování syntaxe (např. chyba v názvu atributu). Na druhou stranu cenou za velkou uživatelskou volnost jsou některé jiné typy možných chyb, kterým by v rigidnějším anotačním prostředí bylo možné zabránit (např. použití nesprávných oddělovačů v neatomických hodnotách).

³<http://www.winedt.com>



Obrázek 6.1: Editor WinEdt v módu Valenční slovník, vzorek rozpracované dávky sloves.



Obrázek 6.2: WWW rozhraní pro vyhledávání ve VALLEXu.

6.3 Vyhledávací rozhraní pro WWW

Pro usnadnění některých činností při budování slovníku byla v programovacím jazyku Perl vyvinuta dvě vyhledávací rozhraní, ke kterým lze z internetu přistupovat pomocí jakéhokoli WWW prohlížeče:⁴

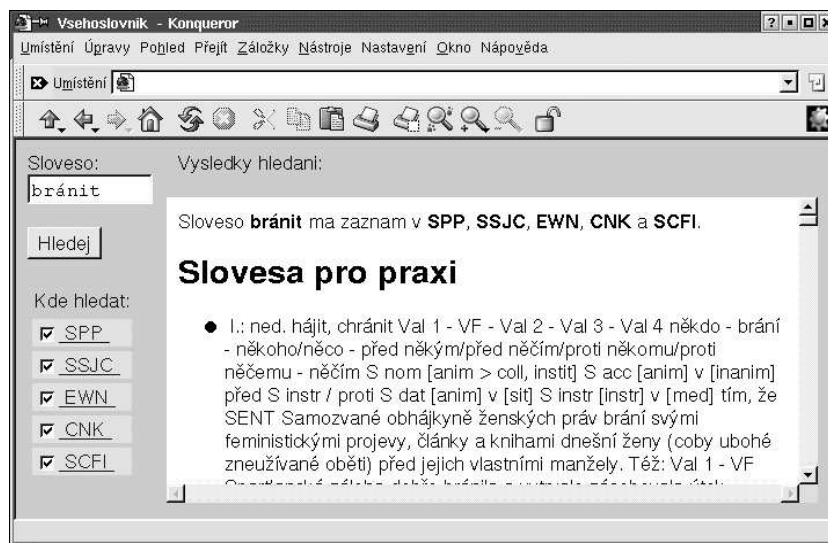
- **Vyhledávání ve slovníku VALLEX** (část rozhraní viz obrázek 6.2). Umožňuje vyhledávat rámce podle toho, jestli daný rámec nebo jeho vybraný atribut obsahují určité podřetězce nebo odpovídají regulárnímu výrazu (většinu dotazů na obsah rámce je možné přepsat na regulární výraz). Na uživatelské úrovni tak lze velmi snadno klást dotazy typu "najdi všechna slovesa kontroly", "najdi všechna slovesa začínající předponou ..." atd. Podmínky zadané v jednotlivých polích musí platit současně.
- **Vyhledávání v některých dalších slovnících** (viz obrázek 6.3). Umožňuje rychle nahlédnout, jak zadané sloveso zpracovali lexikografové v některých tištěných slovnících (SSJČ, SPP, SČFI) nebo jak je zpracováno v EuroWordNetu.

Vývoj obou rozhraní byl pro tuto chvíli zastaven. Ve VALLEXu lze nyní vyhledávat daleko silnějším dotazovacím jazykem XPATH v prostředí XSH (viz následující kapitola). Co se týče obecného vyhledávacího v různých slovnících, jde o nástroj potřebný všem současným lexikografům. Je tedy velmi pravděpodobné, že už byl někde vyvinut nějaký dokonalejší systém a bude rozumnější jej převzít.

6.4 XSH

Valenční slovník byl od úplného začátku budován s představou, že konečná datová reprezentace bude založena na XML (*Extensible Markup Language*). Hlavním důvodem k tomuto

⁴Přístup k oběma vyhledávacím rozhraním není veřejný, proto zde neuvádíme jejich adresy.



Obrázek 6.3: WWW rozhraní pro vyhledávání v různých slovnících.

rozhodnutí je skutečnost, že pro práci s XML daty dnes existuje celá řada technologických standardů a softwarových nástrojů.

Jedním z nástrojů pro práci s XML je editor XSH⁵ (*XML Editing Shell*) Petra Pajase. V XSH je možné se pomocí příkazové řádky pohybovat uvnitř XML dokumentu podobným způsobem jako v souborovém systému, klást dotazy na XML dokument pomocí standardu XPATH, transformovat jej podle standardu XSL nebo s ním jinak manipulovat. XSH má vlastní skriptovací jazyk, v rámci kterého lze používat také příkazy operačního systému a programovacího jazyka Perl.

V následujícím textu se pokusíme na několika velmi jednoduchých příkladech předvést možnosti využití XSH pro práci s valenčním slovníkem. Předpokladem použití XSH je převedení slovníku do XML. Poté můžeme valenční slovník otevřít v XSH a podívat se, jak je v XML reprezentováno heslo např. pro sloveso *odhalit*:⁶

```
sup:...zabokrtsky/ xsh -i
-----
xsh - XML Editing Shell version 1.6/0.9 (Revision: 1.15)
-----
Copyright (c) 2002 Petr Pajas.
This is free software, you may use it and distribute it under
either the GNU GPL Version 2, or under the Perl Artistic License.
Using terminal type: Term::ReadLine::Perl
Hint: Type 'help' or 'help | less' to get more help.
xsh scratch:/> quiet
xsh scratch:/> open v=vallex.xml
xsh v:/> cd /vallex/verb[@lemma="odhalit"]
xsh v:/vallex/verb[246]> ls .
<verb lemma="odhalit">
```

⁵<http://xsh.sourceforge.net>

⁶XML soubor, který je zde použit, má pouze ilustrativní funkci. Skutečná XML reprezentace valenčního slovníku se od něj může co do obsahu i formy značně lišit.

```

<frame>
  <elm func="ACT" type="obl"> <form>1</form> </elm>
  <elm func="PAT" type="obl"> <form>4</form> </elm>
  <elm func="BEN" type="typ"> <form>3</form> </elm>
  <example>odhalit si tvář, odhalit se</example>
  <synon>odkrýt</synon>
  <use>prim</use>
  <fvid>(dok.); odhalovat (nedok.) =</fvid>
  <freq>;1</freq>
  <reciprocity>ACT-BEN-PAT</reciprocity>
</frame>
<frame>
  <elm func="ACT" type="obl"> <form>1</form> </elm>
  <elm func="PAT" type="obl"> <form>4</form> </elm>
  <elm func="BEN" type="typ"> <form>3</form> </elm>
  <example>odhalili mu pomník</example>
  <synon>odevzdat slavnostně</synon>
  <use>idiom</use>
  <fvid>(dok.); odhalovat (nedok.) =</fvid>
  <freq>2;2</freq>
  <reciprocity>ACT-BEN</reciprocity>
</frame>
<frame>
  <elm func="ACT" type="obl"> <form>1</form> </elm>
  <elm func="PAT" type="obl"> <form>4</form> <form>jak</form> <form>že</form> </elm>
  <elm func="LOC" type="typ"> </elm>
  <example>odhalit chyby v úkolu, odhalit tajemství</example>
  <synon>nalézt, vypátrat</synon>
  <use>posun</use>
  <fvid>(dok.); odhalovat (nedok.) =</fvid>
  <freq>28;26</freq>
</frame>
</verb>
xsh v:/vallex/verb[246]> cd /
xsh v:/>

```

V XSH lze velmi snadno zjistit základní kvantitativní vlastnosti slovníku, jako je počet sloves, rámců a prvků v rámcích

```

xsh v:/> count //verb
1017
xsh v:/> count //frame
2935
xsh v:/> count //elm
8639
xsh v:/>

```

nebo rozložení počtu rámců na jedno sloveso (v prvním sloupci je počet sloves, která měla tolik rámců, kolik je uvedeno v druhém sloupci)

```

xsh v:/> foreach //verb count ./frame | sort | uniq -c | sort -n -k2
356 1
246 2

```

```
141 3
96 4
64 5
32 6
26 7
22 8
9 9
7 10
3 11
3 12
3 13
1 14
3 15
1 16
1 17
1 18
1 19
1 22
```

```
xsh v:/>
```

najít slovesa s více než 15 rámci,

```
xsh v:/> ls //verb[count(./frame)>15]/@lemma
lemma="brát"
lemma="přijít"
lemma="táhnout"
lemma="vyjít"
lemma="vzít"
xsh v:/>
```

nebo najít rozložení hodnot nějakého atributu nebo elementu,

```
xsh v:/> ls //form | sort | uniq -c | sort -nr | head -n10
2906 <form>1</form>
1664 <form>4</form>
724 <form>3</form>
418 <form>7</form>
349 <form>na+4</form>
339 <form>že</form>
183 <form>s+7</form>
168 <form>k+3</form>
162 <form>pro+4</form>
147 <form>z+2</form>
xsh v:/> ls //reciprocity/text() | sort | uniq -c | sort -nr | head -n10
573 ACT-PAT
286 ACT-ADDR
110 ACT-BEN
60 ACT-ADDR-PAT
21 ACT-DIR3
17 ACT-ORIG
14 ACT-ACMP
13 ACT-PAT-BEN
12 PAT-EFF
```

```

10 ACT-LOC
xsh v: /> ls //elm/@func | sort | uniq -c | sort -nr | head -n10
2926 func="ACT"
2332 func="PAT"
447 func="ADDR"
374 func="DIR3"
334 func="MEANS"
326 func="EFF"
307 func="LOC"
306 func="MANN"
305 func="BEN"
154 func="DIR1"
xsh v: />

```

případně rozložení podmíněné obsahem jiného elementu nebo atributu, např. jak často se používají které formy pro vyjádření funkce ADDR:

```

xsh v: /> ls //elm[@func="ADDR"]/form/text() | sort | uniq -c | sort -nr
322 3
46 s+7
35 4
18 na+4
14 na+6
7 k+3
4 před+7
4 proti+3
3 2
3 u+2
3 mezi+4
2 v+6
1 vůči+3
1 u+4
1 nad+7
xsh v: />

```

Lze samozřejmě klást i jiné než kvantitavní dotazy, například dotaz na nalezení sloves, která mají rámec s adresátem ADDR i rámec se směrovým doplněním DIR3 (následující příkaz⁷ vypíše jen prvních 10 z nich).

```

xsh v: /> foreach //elm[@func="DIR3"] if (
../../frame/elm[@func="ADDR"] ) ls ../../@lemma | uniq -c | head -n10
lemma="brát"
lemma="dát"
lemma="dávat"
lemma="dělat"
lemma="dodat"
lemma="dopustit"
lemma="dovážet"
lemma="dovést"
lemma="hlásit se"
lemma="hodit"

```

⁷Tento příkaz zde uvádíme spíše jako ilustraci užití `foreach` a `if`; lze jej nahradit jediným XPATH dotazem.

```
xsh v:/>
```

Nalezení lemmat s částicí *se*, která nemají nereflexivní protějšek (uvnitř příkazů je použit programovací jazyk Perl): :

```
xsh v:/> foreach //verb {$lemma=string(@lemma); eval {$L{$lemma}=1}
xsh v:/> eval {print join ", ", grep{/(.+)\s[ie]/ && !L{$1}}sort keys %L}
blížit se, bát se, dařit se, divit se, domnívat se, dostavit se,
dotknout se, dozvědět se, dívat se, líbit se, narodit se, obávat se,
ocitnout se, podařit se, podobat se, podílet se, podívat se, postarat
se, povést se, počíst si, ptát se, setkat se, setkávat se, shodnout
se, snažit se, spokojit se, starat se, stěžovat si, týkat se, vyhnout
se, vyhýbat se, vyskytnout se, vyskytovat se, vyžádat si, všimnout si,
zabývat se, zamyslet se, zdát se, zeptat se, zúčastnit se, účastnit se
```

HTML verzi slovníku lze z XML vytvořit pomocí XSL šablony (podobným způsobem je možné připravit předlohu pro sazbu slovníku typografickým systémem \LaTeX):

```
xsh v:/> xslt v vallex.xsl h
xsh v:/> save h vallex.html
xsh v:/>
```

Kapitola 7

Shrnutí a práce do budoucna

V kapitole 1.3 byl uveden seznam požadavků na ideální valenční slovník. Hodnocení dosažených výsledků a práce do budoucna přirozeně vyplynou ze srovnání ideálu a současného stavu slovníku.

- **Zachycení hloubkové i povrchové roviny.** Každý člen valenčního rámce je na hloubkové rovině popsán funktorem, morfematická realizace (nebo množina alternativních realizací) členu je buď uvedena explicitně, nebo vyplývá z funktoru. Slovník zachycuje několik dalších jevů z hloubkové roviny (např. kontrola) i povrchové roviny (např. reflexivita).
- **Explicitnost.** Jako podkladová teorie byl převzat Funkční generativní popis, koncepce valenčních rámců byla popsáným způsobem rozšířena. Pro anotaci slovníku byla převzata nebo zavedena různá kritéria. Řada z nich však přes veškerou snahu zůstává vágní a anotování slovníku se tak neobjede bez intuice.
- **Úplnost.** Slovník by měl zachycovat všechny rámce slovesa, od primárních významů po frazeologismy.
- **Maximální pokrytí.** Slovník obsahuje přibližně tisíc nejčastějších českých sloves.
- **Minimální redundance.** Slovník v tuto chvíli obsahuje několik typů redundancí. Na odstranění některých z nich v tuto chvíli pracujeme (např. vidové protějšky často sdílejí zcela identické rámce včetně většiny jejich vlastností). Úroveň poznání některých zákonitostí ale není dostatečná k tomu, abychom odstranili všechny typy redundancí, aniž by došlo ke ztrátě části informace (např. valenční rámec by měl být v některých případech odvoditelný ze syntakticko-sémantické třídy, nebo naopak?).
- **Kvalita.** Slovník prošel řadou ručních (heslo každého slovesa procházeli nejméně tři lidé) i automatických kontrol, byla využita informace z jiných slovníků, ve dvou kolech bylo provedeno testování na příkladových větách z Českého národního korpusu. Kvalifikovaný odhad možného množství chyb nebo odhad míry mezianotátorské shody ani metodika jejich měření v tuto chvíli nejsou k dispozici.
- **Užitná hodnota.** Některé možnosti využití valenčního slovníku v NLP byly zmíněny už v úvodní kapitole. V tuto chvíli je část slovníku reálně využívána pouze anotátory Pražského závislostního korpusu (přijatou verzi slovníku ovšem upravují a rozšiřují podle svých potřeb, z mnoha důvodů ji nebylo možné zatím synchronizovat s verzí popsané v této zprávě, proto se nyní oba slovníky ve většině aspektů liší).

- **Machine readability.** Valenční rámec je reprezentován jako posloupnost prvků rámce, všechny vlastnosti prvků rámce a většina vlastností rámce jako celku mají konečnou (většinou velmi malou) množinu hodnot. Pokud není hodnota atomická, je zavedena konvence pro zápis její struktury (např. hodnota reciprocity obsahuje posloupnost možných recipročních užití oddělených čárkou, reciproční užití je posloupnost funktorů oddělených pomlčkou). Slovník lze převést do XML formátu.
- **Human readability.** Editor pro práci se slovníkem umožňuje snadnou orientaci v jednotlivých heslech a jejich vytváření a upravování. Je použit jednoduchý, snadno srozumitelný řádkově orientovaný textový formát, jehož přehlednost je zvýšena odsazováním a barevným zvýrazněním struktury hesla. K usnadnění orientace mezi jednotlivými rámci slovesa je u každého rámce příklad užití. Vzhledem k nutnosti rozložit slovesa do jednotlivých dávek kvůli distribuovanému zpracování zůstává problémem prohlížení slovníku jako celku.

V blízké budoucnosti se budeme soustřeďovat zejména na následující oblasti:

- Zpřesňování kritérií, která by vedla ke zlepšení konzistence zpracování slovesných rámců i jejich atributů (konkrétní problémy byly uvedeny v příslušných kapitolách).
- Doplnování chybějících významů zpracovaných sloves.
- Systémové zpracování vidových protějšků (odstranění redundance plynoucí z toho, že některé vidové protějšky sdílí stejný rámec).
- Sjednání slovníku VALLEX s valenčními rámci vznikajícími při anotování PDT.

Literatura

- [Balabanova,Ivanova,2002] Elisaveta Balabanova and Krasimira Ivanova. Creating a machine-readable version of bulgarian valence dictionary (a case study of clark system application). In **Proceedings of The First Workshop on Treebanks and Linguistic Theories**, Sozopol, Bulgaria, 2002.
- [Bond,Shirai,1997] Francis Bond and Satoshi Shirai. Practical and efficient organization of a large valency dictionary. In **Proceedings of the 4th Natural Language Processing Pacific**, Phuket, Thailand, 1997.
- [Daneš,Hlavsa,1987] František Daneš and Zdeněk Hlavsa. **Větné vzorce v češtině**. Academia, Praha, 1987.
- [Silnickij et al.,1999] G.G. Silnickij et al. **Korreljacionnaja tipologija glagolnych sistem indoevropejskich i inostrukturnych jazykov**. Smolensk, Russia, 1999.
- [Čermák,Blatná,1995] František Čermák and Renata Blatná, editors. **Manuál lexikografie**. H&H, Jinočany, 1995.
- [Čechová et al.,1996] Marie Čechová et al. **ČEŠTINA – Řeč a jazyk**. ISV nakladatelství, Praha, 1996.
- [Fillmore,2002b] Charles J. Fillmore, Collin Baker, and Hiroaki Sato. Seeing arguments through transparent structures. In M. González Rodríguez and C. Paz Suárez Araujo, editors, **Proceedings of LREC 2002**, pages 787–791. ELRA, 2002.
- [Filipec,Čermák,1985] Josef Filipec and František Čermák. **Česká lexikologie**. Academia, Praha, 1985.
- [Fillmore,1968] Charles J. Fillmore. The case for case. In E. Bach and R. Harms, editors, **Universals in Linguistic Theory**, pages 1–90. New York, 1968.
- [Filipec,1973] Josef Filipec. K otázce sémantického popisu lexikálních jednotek. **Slovo a slovesnost**, 34, 1973.
- [Fillmore,1977] Charles J. Fillmore. The case for case reopened. In J.M. Sadock P. Cole, editor, **Syntax and Semantics 8**, pages 59–81. 1977.
- [Fillmore,2002a] Charles J. Fillmore. Framenet and the linking between semantic and syntactic relations. In Shu-Cuan Tseng, editor, **Proceedings of COLING 2002**, pages xxviii–xxxvi. Howard International House, 2002.

- [Grep1,Karlík,1998] Miroslav Grepl and Petr Karlík. **Skladba češtiny**. Votobia, Olomouc, 1998.
- [Horák,1998] Aleš Horák. Verb valency and semantic classification of verbs. In Petr Sojka, Václav Matoušek, Karel Pala, and Ivan Kopeček, editors, **Text, Speech and Dialogue - TSD 98**, Brno, 1998.
- [Hajičová et al.,2001] Eva Hajičová, Jarmila Panevová, and Petr Sgall. **Manuál pro tektonogramatické značkování**. ÚFAL, 2001.
- [Hajičová et al.,2002] Eva Hajičová, Jarmila Panevová, and Petr Sgall. K nové úrovni bohemistické práce: Využití anotovaného korpusu. In **Slovo a slovesnost 63**, pages 161–177. 2002.
- [Helbig,Schenkel,1969] Gerhard Helbig and Wolfgang Schenkel. **Wörterbuch zur Valenz und Distribution deutscher Verben**. VEB BIBLIOGRAPHISCHES INSTITUT, Leipzig, Germany, 1969.
- [Jackendoff,1983] Ray Jackendoff. **Semantics and Cognition**. The MIT Press, 1983.
- [Jackendoff,1990] Ray Jackendoff. **Semantic Structures**. The MIT Press, 1990.
- [Karlík et al.,2002] Petr Karlík, Marek Nekula, and Jana Pleskalová, editors. **Encyklopedický slovník češtiny**. Nakladatelství Lidové noviny, 2002.
- [Komárek,2001] Miroslav Komárek. Několik poznámek k reflexi reflexivity reflexiv. **Slovo a slovesnost**, pages 207–209, 2001.
- [Kingsbury,Palmer,2002] Paul Kingsbury and Martha Palmer. From treebank to propbank. In **Proceedings of the 3rd International Conference on Language Resources and Evaluation**, Las Palmas, Spain, 2002.
- [Kingsbury et al.,2002] Paul Kingsbury, Martha Palmer, and Mitch Marcus. Adding semantic annotation to the penn treebank. In **Proceedings of the Human Language Technology Conference**, San Diego, California, 2002.
- [Králíková,1981] Květa Kráčíková. Reflexivnost sloves z hlediska automatické analýzy češtiny. **Slovo a slovesnost**, pages 291–298, 1981.
- [Levin,1993] Beth C. Levin. **English Verb Classes and Alternations: A Preliminary Investigation**. University of Chicago Press, Chicago, IL, 1993.
- [Lotko,2000] Eduard Lotko. **Slovník lingvistických termínů pro filology**. Olomouc, 2000.
- [Šmilauer,1966] Vladimír Šmilauer. **Novočeská skladba**. SPN, 1966.
- [Mluvnice II] **Mluvnice češtiny II**. Academia, Praha, 1986.
- [Mluvnice III] **Mluvnice češtiny III**. Academia, Praha, 1987.
- [Macháčková,Svozilová,1990] Eva Macháčková and Naďa Svozilová. Slovesná valence ve slovnících. **Slovo a slovesnost**, 51, 1990.

- [Nižníková,Sokolová,1998] Jolana Nižníková and Miloslava Sokolová. **Valenčný slovník slovenských slovies**. Filozofická fakulta Prešovskej univerzity, 1998.
- [Oliva,2001] Karel Oliva. Reflexe reflexivity reflexiv. **Slovo a slovesnost**, pages 200–2007, 2001.
- [Panevová,1974] Jarmila Panevová. On verbal frames in functional generative description. In **The Prague Bulletin of Mathematical Linguistics 22**, pages 3–40. 1974.
- [Panevová,1980] Jarmila Panevová. **Formy a funkce ve stavbě české věty**. Academia, Praha, 1980.
- [Panevová,1992] Jarmila Panevová. O nekotoryx tipax obobschennyx aktantov. In T. Reuther, editor, **Festschrift für V. J. Rozencvejg, Wiener Slawistischer Almanach 33**, pages 73–78. Wien, 1992.
- [Panevová,1996] Jarmila Panevová. More remarks on control. **Prague Linguistic Circle Papers, John Benjamins**, 2:101–120, 1996.
- [Panevová,1998] Jarmila Panevová. Koreference v gramatice a textu (nutnost strukturního popisu a jeho hranice, 1998.
- [Panevová,1999] Jarmila Panevová. Česká reciproční zájmena a slovesná valence. **Slovo a slovesnost**, 4, 1999.
- [Pala,Ševeček,1997] Karel Pala and Pavel Ševeček. Valence českých sloves. In **Sborník prací FFBU**, pages 41–54, Brno, 1997.
- [Panevová,Řezníčková,2001] Jarmila Panevová and Veronika Řezníčková. K možnému pojetí všeobecnosti aktantu. In Zdeňka Hladká and Petr Karlík, editors, **sborník Čeština - univerzália a specifika 3**, pages 139–146. 2001.
- [PMČ] **Příruční mluvnice češtiny**. Nakladatelství LN, Praha, 1996.
- [Polański,1992] Kazimierz Polański, editor. **Słownik syntaktyczno-generatywny czasowników polskich**. Wydawnictwo Polskiej Akademii Nauk, Wrocław, 1980-1992.
- [Popova,1987] Maria Popova. **Kratak valenten rechnik na glagolite v savremennia bulgarski knizoven ezik**. Bulgarian Academy of Sciences Publishing House, 1987.
- [SČFI] **Slovník české frazeologie a idiomatiky**. Academia, Praha, 1983.
- [Sgall,1967] Petr Sgall. **Generativní popis jazyka a česká deklinace**. Academia, 1967.
- [Sgall,1992] Petr Sgall. Underlying structure of sentences and its relations to semantics. In T. Reuther, editor, **Festschrift für Viktor Jul'evič Rozencvejg zum 80. Geburtstag, Wiener Slawistischer Almanach 33**, pages 273–282. Wien, 1992.
- [Sgall et al.,1986] Petr Sgall, Eva Hajičová, and Jarmila Panevová. **The Meaning of the Sentence in Its Semantic and Pragmatic Aspects**. D. Reidel Publishing Company, Dordrecht, 1986.

- [Skoumalová,2001] Hana Skoumalová. **Czech syntactic lexicon**. PhD thesis, Univerzita Karlova, Filozofická fakulta, 2001.
- [Skoumalová,2002] Hana Skoumalová. Verb frames extracted from dictionaries. **The Prague Bulletin of Mathematical Linguistics** 77, 2002.
- [SPP] Naďa Svozilová, Hana Prouzová, and Anna Jirsová. **Slovesa pro praxi**. Academia, Praha, 1997.
- [SSČ] **Slovník spisovného jazyka českého**. Praha, 1964.
- [SSJČ] **Slovník spisovné češtiny pro školu a veřejnost**. Academia, Praha, 1978.
- [Tesnière,1959] Lucien Tesnière. **Eléments de syntaxe structurale**. Paris, 1959.