

## Oponentský posudek doktorské disertační práce

### **PhDr. Zdeňka Uřešová: Valence sloves v Pražském závislostním korpusu**

Oponentka: Markéta Lopatková

V předkládané disertační práci autorka představuje valenční slovník českých sloves, vybraných substantiv, adjektiv a adverbii PDT-Vallex, který vznikl jako součást Pražského závislostního korpusu (PDT) a jehož je hlavní autorkou. Soustřeďuje se přitom na popis koncepce slovníku, na propojení slovníku s korpusovými daty a na popis formálních prostředků pro zachycení diatezí.

#### **Obsah a členění práce**

Práce sestává z šesti kapitol – po krátkých úvodních kapitolách (Předmluva a Úvod), ve kterých autorka stručně představuje cíl a obsah práce, se ve třetí kapitole věnuje valenci a existujícím valenčním slovníkům pro češtinu (BRIEF, VerbaLex, Slovesa pro praxi, Český syntaktický lexikon a VALLEX) a představuje teoretický rámec daný Funkčním generativním popisem (FGD) a jeho valenční teorií (rozebírá kritéria pro stanovení hranice mezi aktanty a volnými doplněními, mezi obligatorními a fakultativními doplněními a představuje princip posouvání).

Jádro práce tvoří čtvrtá a pátá kapitola.

Ve čtvrté kapitole (str. 39-94) je představena koncepce slovníku PDT-Vallex a jeho propojení s Pražským závislostním korpusem – autorka popisuje strukturu slovníkového hesla a jeho jednotlivé součásti: tektogramatické lema, valenční rámec, vztah valenčního rámce a významu, a dále charakteristiky jednotlivých členů valenčního rámce, zejména možné povrchové realizace jednotlivých valenčních doplnění a jejich zápis ve slovníkovém hesle.

Sedmá kapitola (str. 95-191) je věnována zachycení diatezí – autorka popisuje transformační pravidla, která umožňují vygenerovat příslušné povrchové vyjádření jednotlivých valenčních doplnění pro sekundární diateze (pasivní, deagentní, rezultativní, recipientní, dispoziční). Zvláštní pozornost věnuje reciprocitě, kterou pro účely slovníkového zachycení přiřazuje k diatezím.

Práce je zakončena krátkým závěrem upozorňujícím na další možná rozšiřování a obohacování slovníku.

Dále práce obsahuje seznamy obrázků, transformačních pravidel a tabulek, bohatou literaturu a dvě přílohy – pravidla pro kontroly anotace využitá při finálních úpravách PDT, verze 2.0 a ukázkou slovníku PDT-Vallex. Nedílnou součástí práce je celý slovník PDT-Vallex, a to jednak v podobě připravené pro tisk (CD-ROM přiložený k práci), jednak ve formě webových stránek umožňujících pohodlné abecední vyhledávání, <http://ufal.mff.cuni.cz/lindat/PDT-Vallex.html>.

#### **Hodnocení práce**

Předkládaná práce si neklade ambice přinášet nová převratná řešení otevřených teoretických otázek – její přínos je potřeba hledat v rozsáhlé práci s daty: zaměřuje se na aplikaci obecně přijímaných teoretických řešení (konkrétně valenční teorie v rámci FGD). Taková práce, sama o sobě velmi náročná a poněkud nevděčná, je nesmírně důležitá pro verifikaci každé teorie a její další zjemňování.

Autorka přistupuje k problematice valence „zdola“, jejím úkolem je popsat slovníkovou formou valenční chování všech výskytů sloves v korpusu reálných dat (novinových textů) – každý výskyt slovesa (a též ty výskyty substantiv a adjektiv v PDT, které vykazují valenční charakteristiky) má přiřazen jeden z valenčních rámců uložených v postupně budovaném slovníku. Přitom musí mít na paměti praktické problémy, které nutně vznikají při koordinaci anotátorských prací.

Zdůrazněme zde zejména tři oblasti, kterým autorka věnuje zvýšenou pozornost. První oblastí je problematika vymezení významu slovesa, která je klíčovým problémem pro udržení konzistence anotací – v lexikologii ani lexikografii neexistují nesporná a všeobecně přijímaná kritéria pro odlišení jednotlivých významů slov, proto je pro autora jakéhokoliv slovníku nezbytné se vůči této otázce vymezit (alespoň formou anotačních zásad; viz též má otázka níže). Za podnětné považují rovněž zavedení příznaku stavu.

Dále se autorka podrobně věnuje formálnímu zachycení povrchové podoby valenčních doplnění, což úzce souvisí s propojením slovníkových hesel a korpusových instancí jednotlivých sloves – bez takového formalizovaného popisu by slovník nebylo možné využít pro kontrolu konzistence anotací PDT. Tento zápis je pro češtinu unikátní a je nesporně přínosný také pro další nástroje automatického zpracování jazyka, např. pro generování povrchových podob českých vět.

Konečně, podstatná část práce se zabývá problematikou diatezí z pohledu jejich možného zachycení ve slovníku. Jednotlivá hesla PDT-Vallexu explicitně popisují povrchové realizace jednotlivých valenčních doplnění pouze pro primární členy diatezí, tj. zachycují jejich základní realizace v aktivní větě (v podobě tzv. kanonického rámce). U sekundárních členů diatezí neodpovídá povrchová realizace valenčních doplnění jejich kanonické podobě. Autorka zde volí cestu formálních transformačních pravidel, která umožňují generovat povrchové realizace valenčních doplnění i pro sekundární členy diatezí. Odpovídá tak – po mém soudu velice adekvátně – na (implicitní) otázku vztahu gramatiky a lexikonu, na otázku, jaké informace zachycovat v podobě slovníkových hesel a jaké změny jsou natolik pravidelné, že je vhodné a/nebo ekonomické zachycovat je v gramatikalizované podobě. Komplexnost navržených pravidel podává též dobrou představu o složitosti těchto pravidelných změn v povrchové struktuře valenčních rámců (alespoň pokud je chceme zachytit vyčerpávajícím způsobem).

Detailní popis koncepce slovníku je nesporně přínosný pro uživatele slovníku i celého PDT, a to jak pro uživatele-lingvistu, který tato data využívá pro teoretickou práci, tak pro informatika, který je využívá pro vývoj NLP nástrojů. Práce vyčerpávajícím způsobem popisuje způsob zachycení jazykových dat ve slovníku (v případech transformačních pravidel až na úrovni dokumentace).

Po formální stránce je práce velmi dobře strukturována, je přehledně a logicky členěna a je psána čitelným a srozumitelným jazykem. Obsahuje pouze malé množství nepřesných formulací (autorka je již dostala k dispozici). Nezaznamenala jsem žádné formální chyby a jen minimum neobratností stylistických (způsobených zřejmě přesuny v textu).

### **Připomínky, dotazy, návrhy**

První dotaz se týká vyčleňování jednotlivých významů daného slovesa. Nejprve terminologická poznámka: autorka používá termín „valenční rámec“ v (nejméně) dvou významech, (i) jednak jím označuje základní valenční charakteristiku daného slovesa v daném významu (vyjádřenou jako sekvence valenčních pozic s příslušnými informacemi o každé pozici, tj. funktorem, obligatorností a možnými povrchovými formami); (ii) stejným termínem ovšem označuje i dané sloveso v daném významu, tedy to, co je obvykle označováno jako (základní) lexikální jednotka. Vzhledem k existenci zavedenému termínu by bylo vhodné se této dvojznačnosti vyhnout.

K vlastnímu dotazu: PDT-Vallex vyčleňuje zvláštní rámec pro ta užití sloves, která vyjadřují ‚schopnost‘, např. *číst-5 ... ACT(1)* (umět číst), s příklady *dítě už čte; Alenka už čte, ale počítat ještě neumí; č. anglicky.MANN* (obdobně pro sloveso *psát*, rámec 8; ovšem např. u slovesa *počítat* PDT-Vallex zvláštní rámec pro toto užití neuvádí). Podobně vyčleňuje PDT-Vallex pro sloveso *jíst* jednočlenný rámec *jíst-2 ... ACT(1)* (žít) *jíst přes den; zdravě.MANN*.<sup>1</sup> Ve valenčních studiích bývají obdobné příklady klasifikovány jako příklady vět se všeobecným patientem (Daneš, 1971; Panevová, 1980). Zajímaly by mě důvody pro rozhodnutí vyčleňovat zvláštní rámec pro tato užití (se všemi nepříjemnými důsledky).

---

1 Počet užití daných rámců v PDT a PCEDT: *číst-5 ... 1+1, psát-8 ... 5+2, jíst-2 ... 1+9*.

Druhý dotaz se týká pravidel pro zachycení diateze – je zřejmé, že vyčerpávající zachycení povrchových podob jednotlivých valenčních doplnění vyžaduje poměrně komplikovaný formální zápis. Ráda bych se proto zeptala, zda byla tato pravidla nějakým způsobem validována, tj. zda byly podle (části z) nich automaticky vygenerovány možné povrchové realizace a zda byl tento výstup vyhodnocen?

Závěrem opět terminologickou poznámku – diateze bývají chápány jako permutace / změny vzájemného přiřazení sémantických participantů a větných členů formálně syntaktické struktury (přičemž vyjadřovaná reálná situace zůstává stejná, podle Daneše, 1985), či jako „relace mezi paradigmaticky tvořenými příbuznými větnými konstrukcemi, které se navzájem odlišují různým uspořádáním participantů v povrchové struktuře a navíc se vyjadřují zvláštními gramatickými prostředky“ (Panevová, rukopis). Autorka termínu diateze používá pro označení jednoho členu této relace, a to (zejména) členu příznakového (sekundární diateze), (viz s. 16 předložené práce), což by si zasloužilo alespoň poznámku.

### **Shrnutí a závěr**

Předložená práce podrobně popisuje koncepci valenčního slovníku PDT-Vallex, který je jednoznačně jejím nejdůležitějším přínosem. Autorka prokazuje koncepční přístup k řešené problematice a schopnost samostatné práce s rozsáhlými jazykovými zdroji.

Práce splňuje nároky kladené na disertační práci, navrhuji tedy přijmout ji jako práci požadovanou pro udělení titulu Ph.D.

Edinburk, 2. prosince 2011

doc. RNDr. Markéta Lopatková, Ph.D.  
Ústav formální a aplikované lingvistiky  
MFF UK

### Literatura:

- Daneš, Fr. (1971) Větné členy obligatorní, potenciální a fakultativní. *Miscellanea Linguistica*, s. 131–138.
- Daneš, Fr. (1985) *Věta a text*. Praha, Academia
- Panevová, J. (1980) *Formy a funkce ve stavbě české věty*. Praha, Academia.
- Panevová, J. et al. (rukopis) *Syntax současné češtiny (na základě anotovaného korpusu)*