

Příloha 6: Posudek oponenta habilitační práce

Masarykova univerzita

Fakulta Fakulta informatiky
Habilitační obor Informatika

Uchazeč RNDr. Aleš Horák, Ph.D.
Pracoviště Fakulta informatiky
Habilitační práce Computer Processing of Czech Syntax and Semantics

Oponent prof. PhDr. Eva Hajičová, DrSc.
Pracoviště ÚFAL MFF UK v Praze

Text posudku (rozsah dle zvážení oponenta)

Monografie, kterou autor předkládá jako svůj habilitační spis (tiskem vyšla v roce 2008 a byla vydána Masarykovou univerzitou v Brně), je koncipována velmi široce a v podstatě představuje popis tří modulů nutných při automatickém zpracování přirozeného jazyka, totiž lexikálního, syntaktického a sémantického. Vedle toho propojuje autor lexikální komponent v užším smyslu se sítí WordNet a věnuje pozornost i otázkám počítačové implementace; v závěru pak ukazuje i možnost jedné aplikace. Všechny uvedené složky jsou pro adekvátní počítačové zpracování přirozeného jazyka neodmyslitelné a je jedním z kladných rysů předložené práce, že je autor vidí ve vzájemném propojení.

V oddílu o lexikální složce věnuje autor velkou pozornost zachycení tzv. slovesných rámců. Pro lepší porozumění uváděně koncepce rámců by bylo vhodné již na tomto místě uvést několik slovníkových hesel s navrhovanými rámci (několik příkladů lze nalézt až v části o logické analýze, str. 136n., ale to už je ilustrace jiných aspektů). Ve stručném popisu (v podstatě jen str. 10 až 18) nenajdeme principy, podle kterých se přiřazení rámců řídí, ani kritéria, na jejichž základě se rozhoduje. Autor zmiňuje dvě úrovně (level) rámců, ale rozdíl mezi nimi není zcela jasný, alespoň rozbor příkladů ilustrující rozdíl mezi těmito úrovněmi by byl potřebný; měl by ukázat na nutnost takového rozlišování. Také nikde není zmínka o tom, že podobně jako o slovesných rámcích je při vývoji lexikálního komponentu třeba uvažovat o přiřazení rámců dalším druhům slov, což je dnes již neoddiskutovatelná skutečnost a v literatuře (i české) dobře popsána.

Slovesné rámce pak hrají významnou úlohu při popisu syntaktické analýzy: rámce předpokládají, že sloveso vystupuje jako centrální element celé struktury a v podstatě podobu této struktury určuje. Tady se zcela přirozeně nabízí jedna zásadní otázka: proč jako formální základ pro syntaktickou reprezentaci byla zvolena koncepce složková a nikoli závislostní. Vývoj zahraniční teoretické lingvistiky v podstatě od začátku devadesátých let ukazuje, že i teorie původně založené výhradně na analýze podle bezprostředních složek (frázové gramatiky) počítají při nejmenším s vyznačením řídicích či hlavních členů složek, tedy přibližují se ke koncepci závislostní. Pokud jde o parsery, v poslední době jsou častěji využívány parsery pracující s koncepcí závislostní. Podobně je tomu i v korpusové lingvistice při značkování jazykových korpusů, přičemž v některých případech jsou formulovány procedury, kterými lze ze složkové struktury vytvořit strukturu závislostní, a s ní se pak pracuje.

Jako jedna z řady motivací pro preferenci závislostní struktury se uvádí přechod od popisu povrchové stavby věty k její sémantické interpretaci: právě u sémanticky založených přístupů se jeví závislostní struktura jako výhodné východisko (srov. např. tzv. funkční strukturu u lexikálně funkční gramatiky i jinde). Habilitant pracuje při popisu sémantiky transparentní intenzionální logiku P. Tichého. Příklady, s nimiž se v habilitační práci pracuje, jsou v podstatě velmi jednoduché konstrukce (alespoň z jazykového hlediska); toho si je autor vědom (str. 156n.) ovšem bohužel nenaznačuje, které otázky, v lingvistické analýze již podrobně zpracované, by bylo třeba postupem v rámci TILu zpracovat, a jak. Jedním z pokračovatelů Tichého u nás je Pavel Materna, který v některých svých pracích na tyto otázky upozorňoval. Mám tu na mysli problematiku dosahu kvantifikátorů (i negace) a její souvislost s tzv. aktuálním členěním věty, a také relevanci aktuálního členění věty pro různé druhy vyvozování (entailment). Habilitant dvě Maternovy práce cituje, ale o této problematice se nezmiňuje. Stejně tak by bylo zajímavé, jak by při popisu sémantiky v rámci TILu (nebo nad tento rámec) postupoval při přechodu od věty jako základu analýzy k nadvětným celkům; ve své práci odkazuje k Pražskému závislostnímu korpusu, v němž je dnes již vyznačena nejen informační struktura věty (její aktuální členění), ale i gramatická koreference a základní koreferenční vztahy textové i některé základní vztahy diskursní. Jazykový materiál k automatickému zpracování, testování i porovnávání výsledků tedy pro češtinu již k dispozici je.

Předložený spis přináší řadu podnětných myšlenek a postupů, přičemž nejvýše hodnotím jeho snahu vidět jednotlivé komponenty propojené a v interakci. Vzhledem k šíři monografie se však nabízí otázka, které části jsou originálním přínosem habilitanta a které jsou odvozeny z prací kolegů či diplomantů. Autor bohatě odkazuje na statě vlastní i ty, v nichž je jen spoluautorem, to je naprosto korektní přístup, ale v textu tyto zdroje někdy přímo odlišeny nejsou (např. v oddíle o tvorbě lexikální databáze, odd.2.3). Na druhé straně je třeba kladně ocenit zřejmou schopnost habilitanta pracovat v týmu a vést jej.

Práce je psaná velmi přehledně, formulace jsou dobře strukturované, jasné.

Dotazy oponenta k obhajobě habilitační práce (počet dotazů dle zvážení oponenta)

1. Autor si je vědom důležitosti práce se slovesnými rámci. To přirozeně navozovalo, že reprezentace v syntaxi budou založeny na závislostní koncepci se slovesem jako centrem věty; habilitant však pracuje se syntaxí složkovou (viz už sama obálka knížky se strukturou opakovanou na str. 183). Proč?

2. Předložená práce vyšla tiskem v roce 2008; od té doby jistě habilitant ve svém výzkumu pokročil, ostatně možné směry naznačuje v závěrečné části spisu, stejně tak však pokročil výzkum i aplikace na jiných pracovištích, ať již českých nebo zahraničních. Bude zajímavé dovědět se, jak z tohoto hlediska habilitant ve své práci na tento vývoj i současné podněty reaguje

Závěr

Habilitační práce Aleše Horáka „*Computer Processing of Czech Syntax and Semantics*“ **splňuje** požadavky standardně kladené na habilitační práce v oboru Informatika.

Praha, 24. ledne 2012



xx (podpis)