

23. září 2010, 14 hodin
MFF UK, Malostranské náměstí 25, učebna S1

Zápis z obhajoby doktorské disertace Mgr. Pavla Straňáka

Téma práce: Annotation of Multiword Expressions in the Prague Dependency Treebank
Školitel: prof. RNDr. Jan Hajič, Dr.
Oponenti: doc. PhDr. Karel Pala, CSc.
RNDr. Pavel Pecina, Ph.D.

Členové komise: prof. dr. Jarmila Panevová, DrSc. (předsedkyně)
prof. dr. Eva Hajičová, DrSc.,
Mgr. Barbora Vidová Hladká, Ph.D.
doc. dr. Vladimír Petkevič, CSc. (FF UK)
doc. dr. Karel Pala, CSc. (FI MU)

Hosté: viz prezenční listina

Předsedající prof. *Panevová* zahájila obhajobu za přítomnosti pěti členů komise a obou oponentů. Konstatovala, že všechny vyžadované podklady byly řádně předloženy a že k vystavené práci nedošly žádné připomínky. Představila disertanta Mgr. Straňáka a nechala kolovat disertační práci a seznam publikací.

Školitel prof. *Hajič*, zhodnotil průběh studia Mgr. Straňáka a charakterizoval jeho disertaci. Práce reflektuje splnění vytyčených cílů – vybudování slovníku víceslovných výrazů (VV) „zdola“, na základě dat s použitím anotace PDT 2.0 na tektogramatické rovině. Celkově práce přes svou formulační stručnost splňuje zadání.

Disertant Mgr. *Straňák* ve své presentaci představil výstižně svou práci. Poté byli oponenti vyzváni k přednesení svých posudků.

Doc. *Pala* vidí hlavní přínos práce v prozkoumání problematiky VV, což je aktuální téma. Další přínos spočívá ve výstupech, jako jsou anotační nástroj SemAnn, slovník SemLex a rozšíření editoru TrEd, dále v pravidlech pro anotaci a metodologii měření shody pomocí vážené kappy, která je uplatnitelná i jinde. Disertantovi vytkl nedostatky v citacích (chybí brněnský slovník VSV autorů Paly, Svobody a Šmerka, nedostatečný je odkaz k autorství TrEdu). Disertantovi položil následující otázky:

- (i) Proč se práce nevěnuje více automatickému rozpoznávání VV?
 - (ii) Proč se sémantická klasifikace používá jen pro pojmenované entity (PE) a ne i pro ostatní VV?
 - (iii) V jakém je vztahu slovníková „glossa“ v SemLexu k definicím známým z jiných slovníků?
- Celkově oponent považuje výsledky (například související s nástrojem SemAnn) za průměrné; více si cení měření shody.

Dr. *Pecina* ocenil, že se práce zabývá také identifikací VV, nikoli jen pouhou extrakcí běžnou v jiných projektech, a fakt, že návrh anotace byl od počátku promyšlený. Za nedostatky považuje stručnost a nedostatečnou preciznost práce (chybějící rešerše v úvodu, nedostatečné vymezení vágních

přetěžovaných termínů, chybějící anotační instrukce). Položil mu tyto doplňující otázky:

(i) Jaký je vztah VV a PE, které druhy byly v SemLexu obsaženy a které ne?

(ii) Jak pomohla předanotace? Nebyly znevýhodněny počáteční anotace?

(iii) Proč bylo potřeba modifikovat kappu, měřící anotátorskou shodu?

Na závěr doporučil získaná data, která jsou neobyčejně velká, nabídnout pro nějakou mezinárodní soutěž.

Disertant položené otázky zodpověděl a vyjádřil se ke konkrétním výtkám oponentů. Následovaly upřesňující dotazy a poté byla otevřena všeobecná diskuse. Doc. *Petkevič* se ptal, zda bylo výhodné používat tektogramatickou rovinu, když jsou VV dány také svou formou. Prof. *Hajičová* podotkla, že prezentace a disertantovy odpovědi v diskusi ukázaly, že práce bylo odvedeno víc, než je v disertační práci předloženo, a že by práce měla být po doplnění a úpravách publikována. Doc. *Petkevič* se dále ptal po důvodech pro stručnost disertační práce. Disertant průběžně na dotazy uspokojivě odpovídal. Předsedkyně shrnula průběh obhajoby a veřejnou rozpravu. Uvedla, že doktorand reagoval na připomínky a otázky věcně, přiznal některá svá opomenutí a stručnost, zejména v oblasti citací. Vyjádřil se k možnému přínosu slovníku VSV vypracovaného na FI MU v Brně, ke klasifikaci pojmenovaných entit a k použití metriky KAPPA. Zdůvodnil, proč použil pro tyto účely tektogramatickou rovinu.

Reakce doktoranda shledala komise jako vyhovující a ocenila jeho publikační aktivitu.

Po neveřejné diskusi proběhlo tajné hlasování. Bylo přítomno všech 6 členů oprávněných hlasovat; hlasovali s tímto výsledkem: 4 hlasy kladné – 1 hlas záporný – 1 hlas neplatný

Eduard Bejček
Zapsal: Eduard Bejček

Schválila: Jarmila Panevová

V Praze 23.září 2010