

## 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** doc. RNDr. Karel Oliva, Dr.  
**Pracoviště** ÚJČ AV ČR, v.v.i.

### *Oponentský posudek práce*

Na úvod svého posudku bych chtěl zdůraznit, že titul „docent“ označuje (alespoň dle mého mínění) hodnost vědecko-pedagogickou. Účelem habilitační práce je tedy mj. prakticky předvést, že uchazeč je schopen jak samostatné vědecké práce na vysoké úrovni, tak její pedagogicky přesvědčivé (tj. především jasné a srozumitelné) prezentace. K nezanedbatelným požadavkům patří samozřejmě i kultivované formální vypracování textu. Při hodnocení práce jsem se tedy soustředil na tato právě uvedená kritéria, přičemž jsem ovšem na kritéria obsahová kladl váhu větší než na kritéria formální.

Práce *Computer Processing of Czech Syntax*, kterou k habilitačnímu řízení na Fakultě informatiky Masarykovy univerzity v Brně předložil RNDr. Aleš Horák, Ph.D., se fakticky skládá ze čtyř víceméně samostatných studií (*New Language Resources and Tools*, *Czech Syntax Analyzer*, *Transparent Intensional Logic as a Way to Semantics* a *Application in Dialogues – the Electrical Power Systems Simulation*), doplněných o kapitoly takřikajíc podpůrné – předmluvu (*Preface*), obsah (*Contents*), úvod (*Introduction*), závěr (*Conclusions and Future Directions*), seznam použité literatury (*Bibliography*) a stručnou anotaci (*Annotation*). Návaznost „hlavních“ kapitol (výše uvedených samostatných studií) je dosti volná, jde spíše o samostatné studie k jednotlivým tématům, habilitační spis byl tedy zřejmě pojat v zásadě jako kumulativní, tj. jako souhrn témat, kterým se uchazeč za dobu svého vědeckého působení věnoval, nemá však tento charakter plně – zejména nejde o soubor prací již publikovaných. Všechny tyto kapitoly jsou pojaty jako popis projektů, které uchazeč o habilitaci vedl nebo ve kterých alespoň působil jako vedoucí vědecká osobnost. Pro recenzenta z toho vyplývá poněkud těžký úkol rozpoznat, kolik z popsanych výsledků vzniklo bezprostřední a samostatnou vědeckou prací uchazeče, na kolika z nich se podílel ve spoluautorství a na kolika z nich se podílel víceméně „pouze“ ideově a manažersky: snažil jsem se toho – snad úspěšně – dobrat jak na základě jistého přehledu o uchazečově dosavadní

vědecké dráze, tak srovnáním textu s autory bibliografických odkazů. V dalším textu posudku se soustředím pouze na ty části (kapitoly či podkapitoly), kde jsem dohledal autentický osobní vědecký přínos.

V kapitole 2. *New Language Resources and Tools*, která popisuje jazykové zdroje a nástroje *VerbaLex* (slovník valenčních rámců českých sloves), *VisDic* (editor lexikálních databází typu *WordNet*) a platformy založené na technologii *DEB* (Dictionary Editor and Browser, tedy platformy na kompilaci a prohlížení slovníků) je originálním přínosem uchazeče především implementace nástroje *VisDic*, jehož první původní verzi vytvořil sice jiný autor (T. Pavelek), ale uchazeč tento nástroj zcela přepracoval a výrazně všestranně zlepšil jeho parametry, a to tak, že jeho verze *VisDic-u* byla následně použita v projektu *Balkanet* financovaném Evropskou unií a jehož verzi *DebVisDic* lze v současné době považovat za standardní nástroj pro tvorbu a údržbu databází typu *WordNet* v celosvětovém měřítku (používá se pro budování *WordNet-ů* pro řadu jazyků: pro slovenštinu a ruštinu, ale také pro 10 bantuských jazyků a pro afrikaans v JAR, pro nepálštinu, čínštinu a i další jazyky).

Pro diskusi při prezenčním habilitačním řízení bych navrhl, aby uchazeč podrobněji vysvětlil topologii syntaktické struktury na obr. 2.5 v této kapitole, konkrétně proč je zde tak nápadný rozdíl mezi hierarchičností jmenných frází (np) a plochostí fráze verbální (clause): čím je tento rozdíl motivován?

Zásadně originální příspěvek uchazeče lze nalézt v kapitole 3. *Czech Syntax Analyzer*, kde jsou popsány návrh, implementace a empirické výsledky tabulkového analyzátoru pro češtinu, jenž je schopen na výstupu poskytovat jak složkové, tak závislostní struktury (z tohoto pohledu jde o jediný nástroj takového druhu pro češtinu). Na rozdíl od výsledků popsaných výše (v textu o kapitole *New Language Resources and Tools*), je zde autorství uchazeče již od počátku prací jedinečné a nesdílené; jeho kvalitu (resp. kvalitu práce uchazeče) pak ukazují závěrečné pasáže, v nichž je provedeno srovnání s dalšími analyzátory češtiny, které vyznívá veskrze pozitivně.

Podobně originální práce uchazeče jsou popsány v kapitole 4. *Transparent Intensional Logic as a Way to Semantics*, kde se po dosti stručném úvodu do transparentní intenzionální logiky P. Tichého (případně po rozpracování P. Materny) popisuje užití transparentní intenzionální logiky pro formální vyjádření významu (vět) přirozeného jazyka, konkrétně češtiny. Pro tyto účely uchazeč vytvořil experimentální systém analýzy vět češtiny, který je v podstatě rozšířením syntaktického analyzátoru popisovaného v předchozí kapitole o překlad výsledku syntaktické analýzy do transparentní intenzionální logiky. Systém přitom využívá i nástrojů popsaných v kapitole 2 (zejména *VerbaLex* k analýze slovesných skupin na základě valence sloves). Jde o velmi komplexní problematiku, takže jak návrh, tak implementace

systemu byly jistě velmi náročné a výsledek, byť jde o systém experimentální, který zdaleka nepokrývá celou češtinu, lze považovat za velmi cenný, a to i v mezinárodním srovnání.

Kapitola 5 *Application in Dialogues – The Electrical Power Systems Simulation* představuje motivaci, postupy i výsledky, které jsou zaměřeny poněkud jiným směrem než předchozí kapitoly, a ze svého posudku jsem se její hodnocení proto rozhodl vynechat.

V závěru podrobnějšího přehledu práce ještě zmíním sekci *Bibliography*, která je na jedné straně přiměřeně rozsáhlá a reprezentativní, na druhé straně se mi ale zdá být na některých místech zpracována bez patřičné formální péče – formát jednotlivých bibliografických záznamů není jednotný (např. autoři citace [HS00] jsou uváděni jako *Aleš Horák and Pavel Smrž*, autoři [HS03] jako *A. Horák and P. Smrž* apod.) a obsahuje i chyby věcné, některé zřejmě vzniklé nedostatkem formální pozornosti (např. v záznamu [NM04] se zřejmě nejedná o *Kluwerovu akademickou hospodu*, ale o známé vědecké nakladatelství), jiné ale opravdu zarážející – zde se mi jedná zejména o připsání hlavního autorství (a vlastně jediného – spoluautoři nejsou jmenovitě uvedeni) *Slovníku spisovného jazyka českého*, základního díla české lexikografie, Janu Petrovi, zmatený je ovšem celý záznam [P+02] (vydavatel, pořadí vydání, rok, ...). Nejasný je i údaj na str. 56 práce (tj. v kap. 2), kde se počet hesel tohoto slovníku udává číslem 180 000, což je v rozporu s údaji uvedenými ve slovníku samém (kde se hovoří o 192 000 hesel); na druhé straně velmi oceňuji „narovnání“ (na téže straně) údajů o počtu hesel *Příručního slovníku jazyka českého*, kde je v práci uvedeno dosti přesné číslo 200 000 hesel, oproti údajům ze samotného slovníku, kde se v doslovu uvádí 250 000 hesel.

Jako komentář k habilitační práci RNDr. Aleše Horáka, Ph.D., bych nejdříve dodal, že si z vědeckého hlediska nejvíce cením podrobné zpracování tématu automatického převodu (podmnožiny) českých vět do formalismu transparentní intenzionální logiky (tj. obsah kap. 4) a že v podstatě již podrobné zpracování této tematiky bych považoval za dobré a vhodně habilitační téma, jistým způsobem bych dokonce možná osobně dal přednost tomu, aby práce měla jedno takto vyhraněné a výrazné zaměření a téma, navíc v jasné a nedvojznačné podobě výhradního a do hloubky podaného zpracování uchazečem. Ani v předložené podobě však nemám k předložené práci námitky, podle mého názoru ji lze plně uznat za habilitační spis.

Kromě posouzení vlastní práce jsem si vzhledem k tomu, že (jak jsem uvedl na začátku posudku) považuji titul „docent“ za titul vědecko-pedagogický, u pracovníků Fakulty informatiky MU vyžádal i informace o dosavadní pedagogické činnosti A. Horáka, z nichž vyplynulo, že se takovou činností zabývá dlouhodobě a úspěšně: vyučuje na FI MU dva předměty, úspěšně vedl a vede bakalářské a magisterské práce a v posledních několika letech působí i jako školitel-konzultant při studiu postgraduálním (doktorském).

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

1. Můžete podrobněji vysvětlit topologii syntaktické struktury na obr. 2.5 v této kapitole, konkrétně proč je zde tak nápadný rozdíl mezi hierarchičností jmenných frází (np) a plochostí fráze verbální (clause): čím je tento rozdíl motivován?

### **Závěr**

Jako své celkové o stanovisko tedy předkládám habilitační komisi konstatování, že uchazeč o habilitaci na Fakultě informatiky Masarykovy univerzity RNDr. Aleš Horák, Ph.D., předloženou prací dokázal svou vědeckou kvalifikaci a svou dosavadní činností na Fakultě informatiky Masarykovy univerzity také své pedagogické schopnosti, jinými slovy, habilitační práce Aleše Horáka „*Computer Processing of Czech Syntax and Semantics*“ **splňuje** požadavky standardně kladené na habilitační práce v oboru Informatika.

Praha 31. ledna 2012

A handwritten signature in blue ink, appearing to be 'Aleš Horák', with a long, sweeping tail extending downwards and to the right.