

**Univerzita Karlova**  
**Matematicko-fyzikální fakulta**

**ZÁZNAM O PRŮBĚHU OBHAJOBY**  
**DISERTAČNÍ PRÁCE**

**Název práce:** Coreference from the Cross-lingual Perspective

**Jazyk práce:** Angličtina

**Jméno studenta/studentky:** Mgr. Michal Novák

**Studijní program:** Informatika

**Studijní obor:** 4I3 Matematická lingvistika

**Školitel:** doc. Ing. Zdeněk Žabokrtský, Ph.D.

**Oponenti:**

prof. Manfred Stede  
Ing. Alexandr Rosen, Ph.D.

**Členové komise:**

prof. RNDr. Jan Hajič, Dr., předseda (přítomen)  
doc. RNDr. Markéta Lopatková, Ph.D. (přítomna)  
RNDr. Pavel Pecina, Ph.D. (přítomen)  
prof. PhDr. Eva Hajičová, DrSc. (přítomna)  
prof. PhDr. Jarmila Panevová, DrSc. (přítomna)  
doc. RNDr. Vladimír Petkevič, CSc. (přítomen)  
Ing. Alexandr Rosen, Ph.D. (přítomen)

**Datum obhajoby:** 19.9. 2018

**Průběh obhajoby:**

Obhajoba proběhla ve čtvrtek 19. 9. 2018 od 10:45 hodin v budově MFF UK v místnosti S1, Malostranské nám. 25, Praha 1.

Předseda komise, prof. RNDr. Jan Hajič, Dr. přivítal přítomné, představil členy komise a konstatoval, že komise je usnášeníschopná. Kromě komise byl přítomen i školitel, oponent Ing. Alexandr Rosen, Ph.D. a řada hostů. Předseda představil studenta a jeho dosavadní studium. Konstatoval, že byly splněny všechny formální podmínky pro konání obhajoby a že nedošly žádné písemné připomínky mimo posudků oponentů. Student absolvoval zkoušku z angličtiny, doktorandskou státní zkoušku a odevzdal disertační práci, životopis i seznam publikací. Přítomní dostali možnost nahlédnout do výtisků obhajované disertační práce a ostatních materiálů.

---

Pokyny pro předsedy nebo místopředsedy komisí:

Práce v elektronické podobě musí být studentem vložena do SIS. Formulář vyplňte ve všech bodech v elektronické podobě. V bodě Členové komise se uvedou všichni členové komise a za jejich jména se uvede „(přítomen)“ nebo „(nepřítomen)“. Předseda nebo místopředseda komise je jejím členem. V bodě Průběh obhajoby by měly být uvedeny alespoň čtyři věty vystihující průběh obhajoby. Po vyplnění formuláře ho vytiskněte, dole formulář ještě vlastnoručně podepište a přiložte k zápisu o státní závěrečné zkoušce. Současně vložte formulář v elektronické podobě (bez vlastnoručního podpisu) do SIS.

Školitel doc. Ing. Zdeněk Žabokrtský, Ph.D. shrnul studentovo studium a výzkum. Samotnou práci hodnotil jako originální a precizně provedenou. Vyzdvihl především aktuálnost řešeného tématu a kvalitu experimentálních výsledků, kterých uchazeč ve své práci dosáhl.

Uchazeč dostal slovo, aby představil svou práci, která se zabývá automatickou analýzou koreference v angličtině a češtině a jejich rozdíly, které jsou dány především odlišnými typologickými vlastnostmi obou jazyků. Představil reprezentaci jazyka (která vychází z tzv. tektogramatické roviny používané v Pražském závislostním korpusu) a data, se kterými v práci pracoval: paralelní treebank anotovaný na tektogramatické úrovni, český treebank a anglická data z dřívější mezinárodní soutěže v rámci CoNLL. Dále představil vlastní evaluační metriku, kterou používá k evaluaci experimentů. Ta vychází z běžně používaných jazykově nezávislých metod, ale liší se tak, aby lépe zachytila jevy, kterými se uchazeč ve své práci zabýval.

V další části prezentace uchazeč představil vlastní metodu pro řešení koreference v rámci jednoho jazyka založené na strojovém učení. Výsledky této metody porovnal se systémy, které jsou vyvíjeny na Stanford University a pracují přímo s textem bez tektogramatické analýzy a ukázal, že jsou srovnatelné. Dále představil rozšíření této metody, která využívá informaci z obou jazyků současně. Experimenty ukázaly výrazné zlepšení v obou jazycích. Překvapivým výsledkem je, že informace z anglického jazyka výrazně pomáhá při řešení koreference v českém jazyce. Další experimenty s projekcí koreference mezi paralelními potvrzují, že kvůli typologickým rozdílům mezi angličtinou a češtinou není možné projekci jednoduše provést. Experimenty navíc potvrdily překvapivý závěr, že projekce z angličtiny do češtiny je informativnější než opačným směrem, především kvůli tomu, že je možné z angličtiny získat informaci o životnosti referencovaných jmen a informace o nevyjádřených podmínkách.

Na závěr uchazeč shrnul přínosy své práce. Vyzdvihl experimentální výsledky, které ukazují, že angličtina je z hlediska koreference pro češtinu informativnější než čeština pro angličtinu. Práce výrazně přispěla k vydání verze Pražského česko-anglického treebanku opatřeného anotací koreference. Výsledkem práce je také vylepšení slovního zarovnání mezi angličtinou a češtinou a softwarové nástroje pro řešení koreference v angličtině a češtině.

Předseda požádal oponenta Ing. Alexandra Rosena, Ph.D. o přednesení posudku. Ten práci hodnotil kladně. Chválil precizní metodologii, která využívá strojové učení a přínos provedených experimentům k lingvistické teorii. K práci podotkl, že přestože se vlastně věnuje translátologickému tématu, nepoužívá translátologickou terminologii a nezkoumá detailněji vlastnosti Pražského závislostního česko-anglického korpusu, který vznikl překladem anglických žurnalistických textů do češtiny.

Za nepřítomného oponenta prof. Manfreda Stedeho z univerzity v Postupimi přednesl posudek předseda komise. Prof. Stede považuje předloženou dizertační za velmi detailní studii, která vyniká především precizní metodologií, která je na vyšší úrovni než v oboru běžné. Kladně hodnotí také teoretické dopady experimentálních výsledků prezentovaných v práci. Zmiňuje se ovšem o tom, že v části, která se věnuje evaluační metrice, chybí systematictější porovnání s jinými metodami pro evaluaci koreference. Oponent v práci také postrádá detailnější vysvětlení vztahu mezi řešením koreference na tektogramatické úrovni, její anotací v prostém textu nebo v jiných anotačních schématech.

Uchazeč odpověděl na dotazy obou oponentů, které všechny vyžadovaly dovysvětlení detailů. Komise souhlasila, že dotazy nepřítomného oponenta byly uspokojivě odpovězeny.

Po zodpovězení dotazů oponentů předseda komise otevřel všeobecnou diskuzi.

První se do diskuse přihlásil doc. Petkevič s dotazem, jakými jazykovými jevy si uchazeč vysvětluje experimentální výsledky. Kromě již zmíněné životnosti jmen a nevyjádřených podmětů uchazeč zdůraznil přínos párování reflexivních zájmen, která v angličtině neexistují a v češtině se vždy vztahují k podmětu.

---

Pokyny pro předsedy nebo místopředsedy komisí:

Práce v elektronické podobě musí být studentem vložena do SIS. Formulář vyplňte ve všech bodech v elektronické podobě. V bodě Členové komise se uvedou všichni členové komise a za jejich jména se uvede „(přítomen)“ nebo „(nepřítomen)“. Předseda nebo místopředseda komise je jejím členem. V bodě Průběh obhajoby by měly být uvedeny alespoň čtyři věty vystihující průběh obhajoby. Po vyplnění formuláře ho vytiskněte, dole formulář ještě vlastnoručně podepište a přiložte k zápisu o státní závěrečné zkoušce. Současně vložte formulář v elektronické podobě (bez vlastnoručního podpisu) do SIS.

Doc. Pecina vznesl další dotaz k porovnání použité evaluační metriky. Uchazeč zopakoval nevýhody běžně používaných metrik.

Dr. Rosa položil otázku, jakým způsobem se liší výstupy nástrojů vyvinutých na Stanford University od nástrojů, které vyvinul uchazeč ve své dizertační práci. Uchazeč poukázal zejména na hlavní rozdíl, který spočívá v tom, že systémy ze Stanford University pracují navíc s nominální koreferencí.

Předseda ukončil veřejnou debatu, shrnul dosavadní průběh obhajoby, konstatoval, že byly provedeny všechny potřebné kroky a formální požadavky byly splněny. Následně vyzval všechny přítomné mimo členů komise, oponentů a školitele, aby opustili místnost a mohlo proběhnout hlasování. Následovala neveřejná diskuse členů komise a hlasování.

Komise schválila udělení titulu Ph.D.

Zápis provedl 19. 9. 2018 Mgr. Jindřich Libovický

**Počet publikací:** 17

**Výsledek hlasování:**

Počet členů s právem hlasovacím: 7

Počet přítomných členů: 7

Odevzdáno hlasů kladných: 7

Odevzdáno hlasů neplatných: 0

Odevzdáno hlasů záporných: 0

**Výsledek obhajoby:**  prospěl     neprospěl

**Podpis předsedy nebo místopředsedy komise:**

**Podpis jednoho dalšího člena komise:**

---

Pokyny pro předsedy nebo místopředsedy komise:

Práce v elektronické podobě musí být studentem vložena do SIS. Formulář vyplňte ve všech bodech v elektronické podobě. V bodě Členové komise se uvedou všichni členové komise a za jejich jména se uvede „(přítomen)“ nebo „(nepřítomen)“. Předseda nebo místopředseda komise je jejím členem. V bodě Průběh obhajoby by měly být uvedeny alespoň čtyři věty vystihující průběh obhajoby. Po vyplnění formuláře ho vytiskněte, dole formulář ještě vlastnoručně podepište a přiložte k zápisu o státní závěrečné zkoušce. Současně vložte formulář v elektronické podobě (bez vlastnoručního podpisu) do SIS.