

UNIVERZITA KARLOVA

Filozofická fakulta

Zápis o obhajobě disertační práce

Akademický rok: 2019/2020

Jméno a příjmení studenta: Mgr. Petr Plecháč, Ph.D.

Identifikační číslo studenta: 87605770

Typ studijního programu: doktorský

Studijní program: Filologie

Studijní obor: Matematická lingvistika

Identifikační čísla studia: 472264

Název práce: Atribuce autorství básnických textů

Pracoviště práce: Ústav českého národního korpusu (21-UCNK)

Jazyk práce: čeština

Jazyk obhajoby: čeština

Školitel: doc. Mgr. Václav Cvrček, Ph.D.

Oponent(i): PhDr. Jiří Milička, Ph.D.

Ján Mačutek

Datum obhajoby : 21.11.2019 **Místo obhajoby :** Praha

Hlasování komise: prospěl/a: 5 neprospěl/a: 0

Průběh obhajoby: Zápis z obhajoby disertační práce

Student: Mgr. Petr Plecháč, Ph.D.

Identifikační číslo studenta: 87605770

Studijní program: Filologie

Studijní obor: Matematická lingvistika

Název disertační práce: „Atribuce autorství básnických textů“

Jazyk práce: čeština

Jazyk obhajoby: čeština

Školitel: doc. Mgr. Václav Cvrček, Ph.D.

Oponenti: PhDr. Jiří Milička, Ph.D., doc. Mgr. Ján Mačutek, Ph.D.

Datum obhajoby: 21. 11. 2019

Místo obhajoby: FF UK, nám. Jana Palacha 2, Praha 1, S214, týmová studovna

Termín: řádný

Předseda komise: doc. RNDr. Vladimír Petkevič, CSc.

Členové komise: prof. PhDr. Jarmila Panevová, DrSc., prof. PhDr.

Karel Kučera, CSc., Mgr. Michal Křen, Ph.D., doc. Mgr. Ján

Mačutek, Ph.D., doc. RNDr. Vladimír Petkevič, CSc.

Přítomní: prof. PhDr. Jarmila Panevová, DrSc., prof. PhDr. Karel

Kučera, CSc., Mgr. Michal Křen, Ph.D., doc. Mgr. Ján Mačutek,

Ph.D., doc. RNDr. Vladimír Petkevič, CSc.

11:00 Předseda komise doc. Vladimír Petkevič zahájil obhajobu, představil sebe a přítomné členy komise a přítomné oponenty.

Přítomným pak představil studenta.

11:03 Školitel seznámil přítomné stručně se svým hodnocením

studentova studia a jeho disertační práce, které uvedl ve svém písemném vyjádření.

11:07 Student seznámil přítomné s tezemi své disertační práce.

Představil zejména hlavní kapitoly své práce, kterými jsou:

(1) Kvantitativní metody určování autorství (historie vývoje stylometrie, především pak propojení versologických a lexikálních rysů, míra Delta, Support Vector Machine)

(2) Versologické rysy (rytmický profil, rytmické typy, kompromisní cesta navržená samotným autorem disertační práce: rytmické bigramy)

(3) Experimenty (výsledky konkrétních atribučních experimentů)

(4) Aplikace testovaných přístupů na určení sporného autorství dvou konkrétních textů. Zaprvé veršovaného dramatu *The Famous History Of The Life Of King Henry VIII*, jehož autorství je v průběhu let (1850–2018) na základě různých studií přisuzováno Shakespeareovi, Fletcherovi, Massingerovi (doktorand natrénovával svůj testovací model a při křížové validaci tohoto modelu vyšlo najevo, že autorská účast Massingera je velmi nepravděpodobná, naopak validace potvrdila autorství Shakespeara a Fletchera).

Podobně zaměřený výzkum provedl student zadruhé na *Básních* (z let 1858–1862) připisovaných Josefu Barákovi, u nichž existuje hypotéza, že skutečným autorem byl Jan Neruda.

Student také zdůraznil hlavní závěr svého výzkumu i záměr celé své disertační práce: versologické rysy představují relevantní stylometrický indikátor a (přestože jejich extrahování z textu je obvykle znatelně komplikovanější než u tradičně využívaných rysů, jako jsou četnosti slovních tvarů, lemmat, znakových či slovních n-gramů) mohou a měly by být při atribuci autorství básnických textů využívány.

11:23 Oponent doc. Ján Mačutek seznámil přítomné s hlavními body svého posudku se závěrem, že doporučuje předloženou disertační práci k obhajobě a považuje ji za vysoce nadprůměrnou. Položil studentovi jednu otázku, která se týkala vztahu rozpoznatelnosti autora a objemu jeho díla, zmínil fakt, že se má obecně za to, že plodnější autoři jsou snáze rozpoznatelní hlavně díky větší variabilitě stylu. Je tedy možné, že se styl autora více mění s časem než s objemem? Mělo by smysl uvažovat také o intenzitě tvorby? I velké dílo v krátkém období může být relativně objemné a naopak.

11:26 Oponent dr. Jiří Milička seznámil přítomné s hlavními body svého posudku se závěrem, že doporučuje předloženou disertační práci k obhajobě. Položil studentovi tyto dvě otázky:

Zda student sám četl Barákovy básně a text o Jindřichovi VIII.

Zda by v případě, že by student vzal své použité metody a k tomu zcela nový korpus, který by při trénování dat neviděl, byla úspěšnost nižší. A zda je výběr metody ovlivněn už tím, jaká data jsou k dispozici.

11:32 Student reagoval na posudky oponentů a jejich položené otázky.

Uvedl, že co se týče data publikace jednotlivých básní i knih, tak je k dispozici anotovaný jen český korpus, anglický je tvořen většinou sebranými spisy, proto se student díval pouze na datum narození, nikoli na datum publikace (úspěšnost autorství student ukázal podrobně na grafech). Dále student zkoumal, zda je možné rozpoznat dobu vzniku díla od jednoho autora, k výzkumu vzal Vrchlického díla ze sedmdesátých a osmdesátých let proti devadesátým letům 19. století a nultým letům století dvacátého, dále Heyduka, Sládka, Klášterského a prokázal, že skutečně u některých autorů tomu tak je. Závěr tedy zní, že určitě takoví autoři jsou, ale nebývá to pravidlem. Co se týče aplikace na jiný korpus, úspěšnost je vždycky funkcí korpusu, což dokázaly i jiné velké stylometrické studie. Dále student potvrdil, že četl pouze jeden z textů, a to Barákovy básně.

Oponenti byli spokojeni s odpověďmi.

11:44 Předseda komise zahájil diskusi.

V následné diskusi vystoupili:

Prof. Karel Kučera s otázkou, zda studentovi vyšel Vrchlický jako autor, u kterého byly časové rozdíly nejvýraznější. Student odpověděl, že nikoliv, spíše naopak.

Prof. Jarmila Panevová se zeptala na to, jakým způsobem se prokazuje hypotéza v případě, že máme k dispozici vstupní text. Je to Neruda? Je to Barák? Kdo je to? S jakými typy otázek se nejvíc pracuje a jaké modely se na tento výzkum uplatňují? Další otázka byla: když víme, že je to Neruda, jak se liší vstupy lyrických a epických básní; pozná se, že je to Neruda?

Student odpověděl, že ve versologickém výzkumu nenarazil na to, že by se epické a lyrické texty od sebe tolik odlišovaly, rozpoznatelnost Nerudy v datech je. Trénování lyrických a testování epických básní u Nerudy může být problém v případě, že bereme v potaz lemmata nebo slova; čím více lemmat, tím se snižuje rozpoznání mezi lyrickými a epickými texty.

Student se dále vyjádřil k jednotlivým metodám a jejich východiskům. Zdaleka nejrozpracovanější je ve stylometrii metoda klasifikační: existuje omezený počet kandidátů, k čemuž se používají různé metody (např. metoda nejbližšího souseda), další metoda spočívá v tom, že se vezmou sporné texty a jejich potenciální kandidát, základní hypotéza pak je: když bude úspěšnost klasifikátoru nízká a na jiných textech bude vysoká, jde o indikátor toho, že by to mohl být ten daný autor. Poslední možnost je, že známe text a nevíme jinak vůbec nic, proto chceme zjistit aspoň něco (pohlaví atd.), v tom případě se uplatňuje tradiční klasifikační metoda, při níž se vezmou texty a natrénuje se rozpoznávání, třeba na staletí nebo na pohlaví (muži, ženy), příp. geolokace atp.

Dr. Michal Křen se zeptal, zda doktorand nezkoušel použít eufonii (Altmannovu metodu). Student odpověděl, že nezkoušel, protože má k metodě teoretické i empirické námítky, zejména se prokázalo to, že distribuce jazykových prostředků je stejná napříč básnickým i prozaickým korpusem.

Doc. Vladimír Petkevič se zeptal, co je optimální z hlediska slovních druhů: synsémantika, či autosémantika? Student odpověděl, že důležitá jsou právě synsémantika, která jsou velmi výrazným ukazatelem.

11:55 Předseda komise ukončil samotnou obhajobu a komise zahájila neveřejné zasedání o klasifikaci obhajoby disertační práce.

Předseda komise seznámil studenta a přítomné s výsledkem obhajoby: komise hlasovala zdvižením ruky, počet členů komise 5 – přítomno členů komise 5 – kladných hlasů 5, záporných hlasů 0. Obhajoba disertační práce byla klasifikována jako „prospěl“.

Zapisovatel: Mgr. Lucie Benešová

Jméno a podpis předsedy komise:

doc. RNDr. Vladimír Petkevič, CSc.

Jméno a podpis dalšího člena komise:

Mgr. Michal Křen, Ph.D.

Klasifikace obhajoby:	prospěl/a (P)	
Předseda komise:	doc. RNDr. Vladimír Petkevič, CSc. (přítomen)
Členové komise:	prof. PhDr. Karel Kučera, CSc. (přítomen)
	prof. PhDr. Jarmila Panevová, DrSc. (přítomen)
	Mgr. Michal Křen, Ph.D. (přítomen)
	Ján Mačutek (přítomen)