

Informace ke státní závěrečné zkoušce programu FF N-PLIN Počítačová lingvistika pro studenty imatrikulované od podzimu 2019 včetně

PRO JEDNOBOROVÉ STUDIUM A HLAVNÍ STUDIJNÍ PLÁN

PREAMBULE

Státní závěrečná zkouška se skládá ze dvou částí: obhajoby diplomové práce a ústní zkoušky.

OBHAJOBA DIPLOMOVÉ PRÁCE

Obhajoba diplomové práce probíhá zpravidla ve stejný den a před stejnou komisí jako státní závěrečná zkouška. Student si připraví prezentaci (v rozsahu cca 15 minut). Obhajoby jsou veřejné. Musí být přítomen vedoucí práce a/nebo její oponent (zejména liší-li se zásadně jejich hodnocení práce).

ÚSTNÍ ZKOUŠKA

Otázky jsou vybírány na základě tematických okruhů, které mají studentům sloužit jako opora pro přípravu k SZZ. Vycházejí z látky probrané v povinných a profilových kurzech navazujícího magisterského studijního programu Počítačová lingvistika na FF MU.

Ve srovnání s bakalářskou SZZ se u magisterské SZZ očekává vedle rozšířeného a odlišného penza znalostí (viz níže uvedené studijní okruhy) též schopnost pojednat problematiku v širších souvislostech a v elementárním teoretickém (či metodologickém) zakotvení. Předpokládá se, že student ovládá i látku, která byla součástí předchozího bakalářského studia programu Počítačová lingvistika. Samozřejmostí je znalost terminologie, orientace v základní odborné literatuře a přehled o hlavních oborových časopisech.

OBECNÉ PŘÍRUČKY DOPORUČENÉ PRO PŘÍPRAVU KE SZZ

K části orientované na **český jazyk**:

- Černý, J. *Dějiny lingvistiky*. Votobia: Olomouc, 1999.

- Karlík, P. – Nekula, M. – Rusínová, Z. (eds.). *Příruční mluvnice češtiny*. Nakladatelství Lidové noviny: Praha, 1995.
- Karlík, P. – Nekula, M. – Pleskalová, J. (eds.). *Nový encyklopedický slovník češtiny*. Nakladatelství Lidové noviny: Praha, 2016.
- Lamprecht, A. – Šlosar, D. – Bauer, J. *Historická mluvnice češtiny*. Státní pedagogické nakladatelství: Praha, 1986.
- *Mluvnice češtiny I-III*. Praha, 1986, 1987.

K části orientované na **počítačovou a korpusovou lingvistiku, informatiku**:

- Čermák, F. – Klímová, J. – Petkevič, V. (eds.) *Studie z korpusové lingvistiky*. Praha: Karolinum, 2000.
- Osolobě, K. Matematická lingvistika. In Krčmová, M. – Večerka, R. – Pleskalová, J. – Karlík, P. *Kapitoly z dějin české jazykovědné bohemistiky*. Praha: Academia, 2007, s. 447–466.
- Kocek, J. – Kopřivová, M. – Kučera, K. (eds.) *Český národní korpus – úvod a příručka uživatele*. Praha: ÚČNK FF UK, 2000.
- Osolobě, K. *Česká morfologie a korpusy*. 1. vyd. Praha: Karolinum, 2014.
- Hesla z NESČ, která se týkají korpusové lingvistiky dostupná zde: <https://www.czechency.org/slovník/search?search=korpus>.
- Wiki Českého národního korpusu dostupná zde: <https://wiki.korpus.cz/doku.php>.
- Sketch Engine User Guide dostupná zde: <https://www.sketchengine.eu/guide/>.

Literatura k jednotlivým okruhům je zadána v sylabech povinných a povinně volitelných kurzů bakalářského studia v IS MU.

TEMATICKÉ OKRUHY KE SZZ

Přehled okruhů z lingvistiky, matematiky, statistiky, informatiky, počítačového zpracování přirozeného jazyka a korpusové lingvistiky pro studenty navazujícího magisterského studia programu Počítačová lingvistika imatrikulované **od podzimu 2019** včetně (jednooborový a hlavní studijní plán).

I. LINGVISTIKA (NA MATERIÁLU ČEŠTINY)

Vývoj spisovné češtiny

1. Raná stará čeština – časové vymezení, prameny (bohemika, glosy, přípisky), nejstarší doložené české texty, základní jazyková charakteristika, rozvoj slovní zásoby, poměr ke konkurenčním jazykům (latina a němčina).

2. Čeština 14. stol. – časové vymezení, prameny (žánrová skladba, významné texty), základní jazyková charakteristika, rozvoj slovní zásoby, stylové prostředky, pravopis, poměr ke konkurenčním jazykům (latina a němčina), první jazykovědné reflexe češtiny.
3. Čeština doby husitské – časové vymezení, prameny (žánrová skladba, významné texty / významní autoři), základní jazyková charakteristika, rozvoj slovní zásoby, stylové prostředky, pravopis, Husův jazyk a vliv na vývoj češtiny, jazykovědné reflexe češtiny, proměny psaného jazyka – „demokratizace“ jazyka (?), poměr ke konkurenčním jazykům (latina a němčina).
4. Humanistická čeština – časové vymezení, diskuze o pojmenování epochy (alternativní pojmenování), prameny (žánrová skladba, významné texty / významní autoři), základní jazyková charakteristika, rozvoj slovní zásoby, stylové prostředky, stabilizace psaného jazyka na konci 16. stol., rozvoj pravopisu (vliv knihtisku), jazykovědné reflexe češtiny (dobové gramatiky a slovníky), poměr ke konkurenčním jazykům (latina a němčina).
5. Barokní čeština – časové vymezení, stereotypy v hodnocení této epochy – koncept jazykového úpadků, prameny (žánrová skladba, významné texty / významní autoři), základní jazyková charakteristika, rozvoj slovní zásoby, stylové prostředky, jazyk českého exilu, jazykovědné reflexe češtiny (dobové gramatiky a slovníky), poměr ke konkurenčním jazykům (latina a němčina).
6. Čeština 19. stol. – a) obrozenská fáze, časové vymezení, počátky kodifikace a její autoři, rozvoj funkcí a sociální báze spisovné češtiny, vývoj výrazových prostředků, jazykovědné reflexe češtiny (dobové gramatiky a slovníky), vývoj pravopisu, poměr k cizím jazykům (vztah k němčině a jiným slovanským jazykům), b) poobrozenská fáze, časové vymezení, vývoj výrazových prostředků, vývoj funkcí spisovné češtiny, poměr k cizím jazykům (vztah k němčině a jiným slovanským jazykům), péče o spisovný jazyk (brusiči), vývoj dialektů.
7. Čeština 20. stol. – časové vymezení, rozvoj výrazových prostředků, jazyková politika meziválečného Československa, funkční pestrost spisovného jazyka, jazyková kultura a diskuze o ní, institucionální báze spisovné češtiny, strukturní a nestrukturní útvary češtiny, vlivy cizích jazyků.
8. Jazykový kontakt – definice, podtypy (kolektivní × individuální, bilingvismus × diglosie), interference, výpůjčka (borrowing), replika jazykového materiálu a repliky jazykového vzoru (kalk), důvody přejímek, procesy adaptace přejímek, hierarchie výpůjček (borrowing scales), internacionalismy, standardní průměrná evropština (Standard Average

European), internacionalismy; čeština v jazykovém kontaktu v historické perspektivě (vliv němčiny, latiny, polštiny, ruštiny, francouzštiny a angličtiny).

9. Teritoriální stratifikace češtiny – přehled hlavních skupin tradičních teritoriálních dialektů, diferenciací nářečí jako výsledek jazykového vývoje; nivelizační procesy ve vývoji českých nářečí, česko-moravské diference v rámci těchto procesů, obecná čeština, interdialekty; základní znaky vybrané nářeční skupiny; jazyk města Brna.
10. *Český jazykový atlas* – celková charakteristika (obsah, metody zpracování jazykového materiálu), zařazení do kontextu odborné reflexe českých teritoriálních dialektů; další aktivity dialektologického oddělení ÚJČ AV ČR v Brně.

Čeština v komparativní perspektivě

1. Subjekt, komplement, adjunkt.
2. Pro-drop parametr.
3. Konstrukce s reflexivem a jejich analýza.
4. Tranzitiva, ditranzitiva, neakuzativa, neergativa.
5. Slabičná struktura – slabičné konsonanty, extraslabičné konsonanty, alternace vokálů s nulou.

Čeština v teorii a praxi

1. Slovní druhy – kritéria klasifikace (morfologická, syntaktická, sémantická).
2. Substantiva – gramatické kategorie substantiv; deklinační paradigmaty.
3. Adjektiva – typy deklinace (složená, jmenná, smíšená deklinace posesivních adjektiv).
4. Slovesa – gramatické kategorie sloves; finitní a nefinitní tvary, syntetické a analytické formy; konjugační paradigmaty/slovesné třídy.
5. Větné členy – podmět, předmět, příslovecné určení, přísudek, doplněk, atribut (jak je lze poznat a jaké mají vlastnosti).
6. Slovní zásoba – strukturace; vývojové tendence, neologismy.
7. Lexikografie – předmět zájmu; stavba slovníku, prezentace makrostruktury a mikrostruktury na vybraném slovníkovém díle; typologie slovníků, zastoupení jednotlivých typů v české lexikografii od počátku 20. století;

výkladový slovník – charakteristika, základní výkladové slovníky češtiny: PSJČ, SSJČ, SSČ; perspektivy české výkladové lexikografie.

8. Čeština v rámci slovanských jazyků, vývojové fáze češtiny.
9. Teritoriální stratifikace národního jazyka – nástin českých dialektů; nivelizační procesy, interdialekty a obecná čeština; teritoriální diference v užívání spisovné češtiny.
10. Norma, úzus, kodifikace – péče o spisovný jazyk; aktuální kodifikační příručky.

II. MATEMATIKA, INFORMATIKA, POČÍTAČOVÉ ZPRACOVÁNÍ JAZYKA

Matematika a statistika

1. Logika a její aplikace na přirozený jazyk.
2. Aplikace statistiky na zpracování jazyka.
3. Statistický jazykový (n-gramový) model – konstrukce, využití.
4. Vyhodnocování aplikací zpracování jazyka – přesnost, pokrytí, F-míry a zlatý standard, křížová validace.

Informatika a programování

1. Algoritmy a datové struktury (reprezentace objektů, serializace, techniky algoritmizace) – pole, seznamy, objekty, rozhodnutelnost, složitost.
2. Reprezentace jazykových informací v počítači – kódování znaků, XML, morfologická, syntaktická anotace aj.
3. Značkovací jazyky – (X)HTML, XML, DTD, ...
4. Algoritmická složitost – lineární, polynomiální a exponenciální algoritmy, složitost algoritmu a složitost problému, třídy složitosti.
5. Formální jazyky a automaty – regulární jazyk, bezkontextový jazyk, konečný automat, formální gramatika, typy gramatik podle Chomského.
6. Regulární jazyky a regulární výrazy, CQL.
7. Strojové učení a umělá inteligence – jaké problémy řeší a jaké algoritmy používá.

Počítačové zpracování přirozeného jazyka

1. Techniky word sense disambiguation a algoritmy.
2. Strojový překlad – pravidlové a statistické přístupy, existující systémy, vyhodnocování.
3. Morfologická analýza přirozeného jazyka a desambiguace, přístupy.
4. Syntaktická analýza přirozeného jazyka, syntaktické stromy závislostní a složkové, algoritmy analýzy.
5. Budování a indexace korpusů, webové korpusy, statistické nástroje pro korpusy, jazykové modelování.
6. Počítačová lexikografie – systémy pro editaci slovníků, značkování slovníkového hesla.
7. Paralelní korpusy, zarovnání.
8. Strojové učení – princip, techniky, vyhodnocování, křížová validace.
9. Reprezentace znalostí, jazyková znalost, znalost o světě, encyklopedická znalost, common-sense.
10. Dialogové systémy – dialogové strategie, typy dialogů, značkovací jazyk pro popis dialogu, využití dialogových systémů, komunikační agenty (chatbots).

III. POČÍTAČOVÁ A KORPUSOVÁ LINGVISTIKA

1. Přehled vývoje počítačové lingvistiky ve světě a v ČR.
2. Počítačové zpracování jednotlivých rovin přirozeného jazyka z pohledu lingvistiky – morfologická, syntaktická, sémantická, slovtvorná rovina.
3. Historie korpusové lingvistiky – raná korpusová lingvistika, Chomského kritika korpusové lingvistiky, budování prvních korpusů.
4. Vývoj korpusové lingvistiky v ČR.
5. Morfologický slovník jakožto součást automatického analyzátoru – zachycení gramatických významů v morfologickém slovníku, zachycení standardních a substandardních tvarů v morfologickém slovníku.

6. Automatické nástroje pro studium gramatiky budované nad jazykovými korpusy – konkrétní aplikace, využití složitějších dotazů v jazyce CQL pro studium gramatického systému jazyka.
7. Výběr vhodného korpusu pro řešení lingvistického problému – volně dostupné korpusy a jejich charakteristika, DIY korpusy, lingvistické příručky založené na korpusech.