

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

## **PRO JEDNOBOROVÉ STUDIUM A HLAVNÍ STUDIJNÍ PLÁN**

### **PREAMBULE**

Státní závěrečná zkouška (SZZ) se skládá z obhajoby bakalářské práce a ústní zkoušky.

### **OBHAJOBA BAKALÁŘSKÉ PRÁCE**

Obhajoba bakalářské práce probíhá 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 bakalářské SZZ. Vycházejí z látky probrané v povinných a profilových kurzech bakalářského studia programu Počítačová lingvistika na FF MU.

## **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.
- Wiki Českého národního korpusu dostupná zde: <https://wiki.korpus.cz/doku.php>.

**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 bakalářského studijního programu Počítačová lingvistika imatrikulované **od podzimu 2019** včetně (jednooborový a hlavní studijní plán).

### I. LINGVISTIKA (NA MATERIÁLU ČEŠTINY)

Kromě zvládnutí níže specifikovaného vědomostního penza se u bakalářské zkoušky předpokládá znalost elementární lingvistické terminologie (včetně propojení termínů s konkrétními jazykovými příklady) a orientace v základní literatuře oboru.

#### Fonetika a fonologie

1. Fonetika – vztah mezi grafickou a zvukovou stránkou (grafém – hláska), fonetický přepis (IPA, česká varianta).
2. Základní nástroje fonologického popisu – foném: identifikace fonémů (volná vs. vázaná distribuce, kontrastivní funkce, minimální páry), vztah foném – alofon, alternace fonémů, neutralizace kontrastu mezi fonémy; rozdíl mezi hloubkovou a povrchovou strukturou, role lexikální reprezentace fonologické struktury, fonologické pravidlo (X → Y / Z).

3. Segmentální fonologie (se zaměřením na češtinu) – vokály a jejich fonologické rysy, reprezentace vokalické délky; konsonanty a jejich fonologické rysy, škála sonority.
4. Fonologické procesy (se zaměřením na češtinu) – asimilační procesy: asimilace znělosti, asimilace nazál, degeminace; vliv morfologie: ztráta znělosti, palatalizace, dloužení vokálů, epenteze vokálu *e*.
5. Prozodická (suprasegmentální) fonologie – slabičná struktura: principy sylabifikace a sonorita, extraslabičné konsonanty, slabičné konsonanty; slabičná typologie: otevřené vs. zavřené slabiky, lehké vs. těžké slabiky; metrická stopa: prominence, trochejská stopa vs. jambická stopa; fonologické slovo (přízvukový takt): přízvuk, vztah k morfologickému slovu, proklitika, enklitika; fonologická fráze (promluvvový úsek, kólon): polokadence, intonace, pauza.

### **Morfologie derivační a flektivní**

1. Morfém, kombinatorika morfémů – kořeny vs. afixy, typy afixů; kompozicionalita, produktivita, rozdíl mezi kořenem a kmenem, alomorfie (supletivní alomorfy a jejich distribuce, fonologicky podmíněné alomorfy, morfonologické alternace).
2. Slovní druhy – kritéria klasifikace (morfologická, syntaktická, sémantická); číslovky a participia z hlediska slovního druhu / slovnědruhového přesahu.
3. Substantiva – gramatické kategorie substantiv; deklinační paradigmata/typy, pádový synkretismus / pádová homonymie (uvnitř paradigmát a mezi nimi); diachronní vývoj deklinačního systému, kmenový vs. rodový princip, vliv slovo tvorné utvářenosti substantiv na zachování deklinačních typů a podtypů.
4. Adjektiva – typy deklinace (složená, jmenná, smíšená deklinace posesivních adjektiv) a její diachronní vývoj; sémantické typy adjektiv a jejich vliv na stupňování.
5. Slovesa a konjugační systém – gramatické kategorie sloves, finitní a nefinitní tvary, syntetické a analytické formy; konjugační paradigmata / slovesné třídy, variantní/dubletní formy; vývojové tendence v konjugaci.

### **Syntax**

1. Větné členy – podmět, předmět, příslovečné určení, přísudek, doplněk, atribut (jak je lze poznat a jaké mají vlastnosti).
2. Konstituentové testy, struktura vět s pomocnými slovesy.

3. Hlava a její komplement – jak lze poznat hlavu: syntaktická, sémantická a distribuční kritéria, jak se hlava projevuje v různých typech frází: PP, VP, NP, AP, projevy funkčních hlav, zvláště T, C, Mod, Fáze, posun hlavy a hlava slova.
4. Frázový posun – wh-posun: tvorba otázek doplňovacích, vztažných vět, posun subjektu, plovoucí kvantifikátory, raisingová slovesa, vznik diskontinuitních větných členů.
5. Teorie vázání – princip A, B, C, ilustrace na příkladech.

## **Lexikologie**

1. Lexikální onomaziologie – pojmenování vs. usouvztažnění, pojmenovací proces; autosémantika a synsémantika; krajní případy onomaziologie: termíny, propria.
2. Lexikální sémaziologie – lexikální význam a jeho popis, významová vágnost; základní významové vztahy strukturující lexikon: synonymie, opozitnost, hyperonymie – hyponymie, kohyponymie, holonymie – meronymie; lexikografická reflexe uvedených významových vztahů.
3. Slovní zásoba – strukturace; historické proměny, současné vývojové tendence, neologismy.
4. 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.
5. Výkladový slovník – charakteristika, základní výkladové slovníky češtiny: PSJČ, SSJČ, SSČ; perspektivy české výkladové lexikografie.

## **Historie češtiny a její teritoriální diferenciac**

1. Čeština v rámci slovanských jazyků, vývojové fáze češtiny.
2. Vývoj fonologického systému češtiny od roku 1000 do roku 1400 – staročeské přehlásky, vývoj /r/ a /l/ sonans, vývoj pobočných slabik, pračeské depalatalizace, změna /dz' > z'/, změna /g > γ > h/, asibilace ve staré češtině, hlavní historická depalatalizace, výslovnost stč. /v/, hlásky v proudu řeči.
3. Vývoj fonologického systému češtiny od roku 1400 do současnosti – rozvinutí jotace a labializace, diftongizace a monoftongizace, úženi /é > í/, změna /aj > ej/, změny /šč > št/ atd.

4. Teritoriální stratifikace národního jazyka jako výsledek historického vývoje češtiny – nástin českých dialektů; nivelizační procesy, interdialekty a obecná čeština; teritoriální diference v užívání spisovné češtiny.
5. Jazykovědná pracoviště Ústavu pro jazyk český AV ČR v Brně – dialektologické a etymologické oddělení, jejich zaměření a hlavní práce.

### **Základy jazykovědy**

1. Klasifikace jazyků – podle genetického a typologického hlediska, zařazení češtiny.
2. Jazyk jako nástroj komunikace, jazykové funkce – komunikace lingvální a nelingvální, psaný a mluvený jazyk; komunikační model; základní jazykové funkce podle R. Jakobsona.
3. Norma, úzus, kodifikace – péče o spisovný jazyk; aktuální kodifikační příručky.
4. Základní obrysy historickosrovnávací jazykovědy, strukturalistické jazykovědy, pragmatického obratu v jazykovědě ad.
5. Jazykovědná bohemistika – hlavní osobnosti (J. Dobrovský, J. Gebauer, J. Zubatý, F. Trávníček, V. Mathesius, B. Havránek, F. Daneš, M. Komárek ad.); současná institucionální centra, jazykovědné časopisy.

## **II. MATEMATIKA, INFORMATIKA, POČÍTAČOVÉ ZPRACOVÁNÍ JAZYKA**

### **Základy matematiky a statistiky**

1. Definice, věty, důkazy.
2. Výroková logika, predikátová logika 1. řádu.
3. Množiny, uspořádané množiny a operace na nich.
4. Funkce a relace, jejich uzávěry.
5. Posloupnosti a řady.
6. Přirozená čísla a základní operace na nich.
7. Kombinatorika a pravděpodobnost, podmíněná pravděpodobnost.
8. Statistický soubor, charakteristiky polohy a variability.

9. Statistika a pravděpodobnost: pravděpodobnostní rozložení, pravděpodobnostní prostor.
10. Bayesův vzorec a jeho aplikace.
11. Zipfův zákon, zákon velkých čísel.

### **Základy informatiky**

1. Von Neumannův model počítače.
2. Reprezentace čísel v počítači.
3. Reprezentace znaků v počítači, znakové sady a jejich kódování.
4. Algoritmy a datové struktury – pole, spojový seznam, fronta, zásobník.
5. Grafy, stromy, vlastnosti grafů, základní grafové algoritmy, prohledávání stromu do hloubky, do šířky.
6. Základní pojmy z algoritmické složitosti – časová a prostorová složitost algoritmu, asymptotická složitost a třídy složitosti, složitost algoritmu a složitost problému.

### **Počítačové zpracování přirozeného jazyka**

1. Formální jazyky vs. přirozené jazyky.
2. Chomského hierarchie jazyků, gramatiky, automaty.
3. Regulární výrazy a jejich využití ve zpracování přirozeného jazyka.
4. Automatická morfologická analýza přirozeného jazyka.
5. Automatická syntaktická analýza přirozeného jazyka.
6. Formální gramatiky pro přirozené jazyky.
7. Jazykový model, n-gramy.
8. Využití statistiky ve zpracování přirozeného jazyka.
9. Automatická sémantická analýza přirozeného jazyka.
10. Komunikační agenty (chatbots).
11. Reprezentace znalostí (ontologie, WordNet, rámce).

12. Word sense disambiguation.
13. Textové korpusy.
14. Uložení korpusu v počítači.
15. Indexování a prohledávání korpusu.
16. Nástroje pro práci s korpusy.
17. CQL.

### **III. KORPUSOVÁ LINGVISTIKA**

1. Charakteristika jazykových korpusů a jejich typologie.
2. Základní pojmy korpusové lingvistiky.
3. Český národní korpus, jeho složení a vývoj.
4. Psané a mluvené korpusy, charakteristika a rozdíly.
5. Klasické a webové korpusy, charakteristika a rozdíly.
6. Vnitrotextové a vnětextové značkování korpusů.
7. Systémy morfologického značkování pro češtinu. Problematika desambiguace.
8. Znalost uživatelských aplikací z oblasti korpusové lingvistiky.