

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

## PRO VEDLEJŠÍ STUDIJNÍ PLÁN

### PREAMBULE

Vedlejší studijní plán neobsahuje obhajobu bakalářské 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**:

- 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.

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

- Čermák, F. – Klímová, J. – Petkevič, V. (eds.) *Studie z korpusové lingvistiky*. Praha: Karolinum, 2000.
- Osolsobě, 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.
- Osolsobě, 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 imatrikované **od podzimu 2019** včetně (vedlejší studijní plán).

### I. ELEMENTÁRNÍ POPIS JAZYKA (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 znalost základních oborových příruček.

1. Slovní druhy – kritéria klasifikace (morfologická, syntaktická, sémantická).
2. Substantiva – gramatické kategorie substantiv; deklinační paradigmata.
3. Adjektiva – typy deklinace (složená, jmenná, smíšená deklinace posesivních adjektiv).
4. Slovesa – gramatické kategorie sloves; finitní a nefinitní tvary, syntetické a analytické formy; konjugační paradigmata/slovesné třídy.
5. Větné členy – podmět, přísudek, přívlastek, předmět, příslovečné určení, doplněk (jak je lze poznat a jaké mají vlastnosti). Souvětí – souřadné, podřadné.
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 difference v užívání spisovné češtiny.

10. Norma, úzus, kodifikace – péče o spisovný jazyk; aktuální kodifikační příručky.

## **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 vněttextové značkování korpusů.

7. Systémy morfologického značkování pro češtinu. Problematika desambiguace.
8. Znalost uživatelských aplikací z oblasti korpusové lingvistiky.