

Studijní okruhy k magisterské státní závěrečné zkoušce z oboru Český jazyk se specializací počítačová lingvistika

Preambule

Tematické okruhy mají studentům sloužit jako opora pro přípravu k magisterské státní závěrečné zkoušce. Vycházejí z obsahu povinných a povinně volitelných kurzů **bakalářského i navazujícího magisterského** studia oboru Český jazyk se specializací počítačová lingvistika na FF MU.

Ve srovnání s bakalářskou SZk se u magisterské SZk očekává vedle rozšířeného penza znalostí (viz níže uvedené studijní okruhy) též schopnost pojednat problematiku v širších souvislostech a v elementárním teoretickém (či metodologickém) zakotvení. Samozřejmostí je znalost terminologie a orientace v základní odborné literatuře a přehled o hlavních oborových časopisech.

Obecné příručky doporučené pro přípravu ke SZk:

- k části orientované na **český jazyk**:

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

Helbig, G. *Vývoj jazykovědy po roce 1970*. Praha, 1991.

Hoffmannová, J. *Stylistika a...: současná situace stylistiky*. Praha: Trizonia, 1997.

Karlík, P. – Nekula, M. – Rusínová, Z. (eds.). *Příruční mluvnice češtiny*. Nakladatelství Lidové noviny: Praha, 1995.

Karlík, P. – Krčmová, M. – Pleskalová, J. – Večerka, R. (eds.). *Kapitoly z dějin české jazykovědné bohemistiky*. Academia: Praha, 2007.

Karlík, P. – Nekula, M. – Pleskalová, J. (eds.). *Encyklopedický slovník češtiny*. Nakladatelství Lidové noviny: Praha, 2002.

Lamprecht, A. – Šlosar, D. – Bauer, J. *Historická mluvnice češtiny*. Státní pedagogické nakladatelství: Praha, 1986.

Mluvnice češtiny. Academia: Praha, 1986.

- 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.

Osolsobě, K. *Česká morfologie a korpusy*. 1. vyd. Praha: Karolinum, 2014.

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

STUDIJNÍ OKRUHY

I. Část orientovaná na český jazyk a obecnou lingvistiku:

1) Fonetika a fonologie češtiny

1. Fonetika, její náplň a cíle práce, zvukový projev a jeho záznam.
2. Proces vzniku řeči, jeho fáze. Mluvní orgány a jejich funkce.
3. Fonologický aspekt zkoumání procesu komunikace.
4. Fonologický systém, foném, vztahy mezi fonémy (zejména korelační) a jejich fonetická manifestace.
5. Vokalické fonémy současné češtiny a jejich realizace.
6. Konsonantické fonémy současné češtiny a jejich realizace.
7. Zákonitosti spojování fonémů v češtině a výběru jejich alofonů.
8. Zvuková stavba souvislé řeči (takt, slabika ad.), fonické prostředky souvislé řeči.
9. Ortoepie, ortofonie. Zásady kultivované české výslovnosti.
10. Přínos české lingvistiky k rozvoji fonologie.

2) Morfologie a tvoření slov

1. Morfematická analýza – morfém/morf/alomorf; princip opakovatelnosti a princip kontrastu.
2. Vztah morfologie a fonologie – typy alomorfie a typologie morfonologických alternací.
3. Vztah morfologie a syntaxe – typy kompozit, klitika jako syntakticky vázaná slova ad.
4. Typy deklinace u jmen s ohledem na diachronní vývoj.
5. Typy deklinace u adjektiv s ohledem na diachronní vývoj.
6. Typy konjugace s ohledem na diachronní vývoj.

7. Nominalizace – typy deverbálních jmen, jejich syntaktické vlastnosti.
8. Verbální prefixy – vztah k vidu, valenci, předložkám.
9. Klasifikace slovních druhů – morfologická, syntaktická a sémantická kritéria.
10. Vztah mezi flexivní a derivační morfologií – produktivita, kompozicionalita.

3) Syntax

Tematické okruhy této oblasti byly po inovaci výuky aktualizovány a jsou platné od jarního semestru 2016 včetně (tj. pro státní závěrečnou magisterskou zkoušku z českého jazyka konanou od června 2016 vč. dále).

1. Lexikální kategorie
2. Semilexikální kategorie
3. Funkční kategorie
4. Hybridní kategorie
5. Prázdné kategorie
6. Elipsa – gapping, VP-elipsa, sluicing, stripping, comparative deletion
7. Implikatury – presupozice, entailment, invited inference, konverzační implikatury
8. Raising
9. Klitika
10. Templát
11. Linearizace

4) Lexikologie

1. Lexikologie, předmět zájmu, členění, hlavní metodologické postupy, interdisciplinární souvislosti.
2. Jednotky lexikonu v systému jazyka a v textové realizaci, problémovost jejich definice a delimitace.
3. Lexikální onomaziologie, pojmenování; slovní druhy z hlediska onomaziologického, krajní případy lexikální onomaziologie: termíny, propria.
4. Lexikální sémaziologie, lexikální význam, jeho složky a možnosti popisu; základní významové vztahy strukturující lexikon.

5. Konotační složky lexikálního významu z pohledu kognitivní sémantiky.
6. Významová vágnost; od aristotelovské teorie kategorizace (nutné a dostačující podmínky) k wittgensteinovskému modelu rodinných podobností a prototypové sémantice.
7. Syntagmatika slovníku, ustálené a neustálené pravidelné a nepravidelné lexikální kombinace; frazeologie a frazeografie.
8. Slovní zásoba češtiny, její strukturace, historické proměny, aktuální vývojové tendence.
9. Lexikografie, stavba slovníku, typologie slovníků; přehled současných specializovaných slovníků češtiny.
10. Historie, současnost a perspektivy výkladové lexikografie v českém prostředí; základní výkladové slovníky češtiny 20. století: PSJČ, SSJČ, SSČ.

5) Teorie národního jazyka (při aplikaci na češtinu)

1. Stratifikace národního jazyka.
2. Spisovný jazyk a jeho funkce.
3. Sociálně a teritoriálně vymezené variety národního jazyka a jejich funkce.
4. Přehled hlavních skupin tradičních teritoriálních dialektů (+ diferenciací teritoriálních nářečí jako výsledek jazykového vývoje) a situace v pohraničních oblastech českého jazykového území; základní znaky vybrané nářeční skupiny.
5. Český jazykový atlas.
6. Obecná čeština – různé způsoby vymezení; interdialekty; specifické rysy běžně mluveného jazyka.
7. Jazyk města Brna; sociolingvistické aspekty studia městské mluvy.
8. Úzus, norma, kodifikace; standard, substandard; charakteristika kodifikačních příruček pro psanou a mluvenou spisovnou češtinu.
9. Jazykový purismus v českém prostředí, jeho fáze.
10. Ústav pro jazyk český AV ČR – mimouniverzitní centrum studia národního jazyka.

6) Stylistika

1. Postavení stylistiky mezi dalšími odbornými disciplínami.
2. Základní definice stylu; různé způsoby klasifikace stylů, včetně příslušných kritérií

a průnikových/přechodových pásem relevantních pro dané stylové oblasti.

3. Stylotvorné faktory (slohotvorní/stylové činitele – stylový faktor obecně, objektivní s. f., subjektivní s. f., rozdíly mezi nimi, příklady).

4. Kompozice textu z hlediska stylistiky (horizontální a vertikální členění textu, základní slohové postupy a jejich modifikace, koheze, koherence, intertextovost).

5. Prostědělovací funkční styl.

6. Odborný funkční styl.

7. Administrativní funkční styl.

8. Publicistický funkční styl.

9. Rétorický funkční styl.

10. Umělecký funkční styl.

7) Historie jazyka od praslovanštiny dodnes; vývoj spisovné češtiny

1. Klasifikace slovanských jazyků jako výsledek jazykové divergence.

2. Praslovanština – její obecná charakteristika a prameny jejího poznání; praslovanský fonologický systém a jeho vývoj; nářeční štěpení praslovanštiny jako proces diferenciací slovanských jazyků.

3. Původ, vznik a vývoj nejstaršího spisovného slovanského jazyka; počátek a rozvoj slovanského písma.

4. Vývoj fonologického systému češtiny od roku 1000 do současnosti.

5. Vývoj deklinace substantiv, adjektiv a zájmen od pračeštiny do současnosti.

6. Vývoj konjugace od pračeštiny do současnosti.

K vývoji spisovné češtiny:

7. Počátky písemné vzdělanosti v českých zemích – latina, staroslověnština moravské a české redakce.

8. Rozmach psané („spisovné“) češtiny ve 14. století; doba husitská a pohusitská; vospělý jazyk humanistické epochy.

9. Barokní regrese, národní obrození, 19. století.

10. Vývoj spisovného jazyka ve 20. století (včetně vývoje pravopisu).

8) Obecné otázky a lingvistické teorie

1. Typologická klasifikace jazyků; typologie V. Skaličky, typologické vlastnosti češtiny.
2. Genetická klasifikace jazyků; indoevropské jazyky, slovanské jazyky, čeština; nostratická teorie.
3. Základní metodologické přístupy diachronní a synchronní lingvistiky.
4. Mladogramatická škola ve světě a u nás.
5. Lingvistický strukturalismus: F. de Saussure, pražská škola, kodaňská a ženevská škola, americký deskriptivismus.
6. Pragmatický obrat v jazykovědě: Wittgenstein, teorie mluvních aktů, teorie konverzačních implikatur, zdvořilost ad.
7. Sociolingvistika – základní obrysy.
8. Generativní gramatika – univerzální gramatika; modularita; teorie principů a parametrů.
9. Jazykovědná bohemistika: hlavní osobnosti (J. Dobrovský, J. Gebauer, J. Zubatý, F. Trávniček, V. Mathesius, B. Havránek, F. Daneš, M. Komárek ad.).
10. Jazykovědná bohemistika – současný stav, institucionální centra, jazykovědné časopisy.

II. Část orientovaná na počítačovou a korpusovou lingvistiku, informatiku:

1) Matematika a statistika

1. Logika a její aplikace na přirozený jazyk.
2. Matematické objekty – množiny, relace, uspořádání, posloupnosti, ...
3. Statistika a pravděpodobnost – statistický soubor a jeho charakteristiky, pravděpodobnostní prostor.
4. Aplikace statistiky na zpracování jazyka.
5. Statistický jazykový (n-gramový) model – konstrukce, využití.
6. Vyhodnocování aplikací zpracování jazyka – přesnost, pokrytí, F-míry a zlatý standard, křížová validace.

2) Informatika a programování

1. Programování – neimperativní, imperativní, praktické programování.

2. Algoritmy a datové struktury (reprezentace objektů, serializace, techniky algoritmizace) – pole, seznamy, objekty, algoritmy typu rozděl a panuj, dynamické programování...
3. Grafy a stromy, grafové a stromové algoritmy.
4. Reprezentace čísel a znaků v počítači.
5. Databáze, jazyk SQL.
6. Značkovací jazyky – (X)HTML, XML, DTD, ...
7. Algoritmická složitost – lineární, polynomiální a exponenciální algoritmy, složitost algoritmu a složitost problému, třídy složitosti.
8. Formální jazyky a automaty – regulární jazyk, bezkontextový jazyk, konečný automat, formální gramatika, typy gramatik podle Chomského.
9. Regulární výrazy, CQL.
10. Umělá inteligence (jaké problémy řeší a jaké algoritmy používá).

3) Počítačové zpracování přirozeného jazyka

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

12. Automatické rozpoznání emocí – techniky, jazykové zdroje.

4) Počítačová a korpusová lingvistika

1. Počítačová lingvistika (komputační – formální – kvantitativní lingvistika, NLP, language engineering). Vývoj ve světě a v ČR.
2. Počítačové porozumění přirozenému jazyku a počítačové zpracování jednotlivých rovin přirozeného jazyka (morfologická, syntaktická, sémantická, slovotvorná).
3. Korpusová lingvistika – empirická věda. Historie korpusové lingvistiky. Definice korpusu v moderním slova smyslu. První korpusy. Chomského kritika korpusové lingvistiky.
4. Typy korpusů (kritéria – časové hledisko, druh zachycené komunikace, rozsah, žánrové hledisko, účel, jazyk). Specializované korpusy. Žákovské korpusy.
5. Budování korpusů. Získávání textů. Přepis, OCR metody, konverze dat. Značkování korpusů (strukturní značky, atributy). Metatextové anotace. Sociolingvistické značkování korpusů mluvených a korpusu soukromé korespondence.
6. Lingvistické značkování. Tokenizace, lemmatizace, automatická morfologická analýza, disambiguace (statistické metody, pravidlová disambiguace, hybridní metody). Lingvistické problémy grafického zachycení korpusů mluveného jazyka (přepis a pravidla přepisu v mluvených korpusech budovaných v rámci ČNK). Stručná charakteristika olomouckého mluveného korpusu.
7. Korpusy v českém prostředí. ČNK – historie a současnost. Charakteristika jednotlivých korpusů pod ČNK (korpusy řady SYN, projekt DIAKORP a další). Charakteristika korpusů budovaných na FI MU (webové korpusy).
8. Korpusové manažery a další nástroje pro práci s jazykovými korpusy. Volba korpusu vhodného pro řešení lingvistického problému. Morfologické značkování českých korpusů (dva analyzátoři – pražský a brněnský, poziční a atributový systém, shody a odlišnosti, porovnání tagsetů).
9. Značkování syntaktické roviny jazyka. Pražský závislostní korpus. Projekty VALLEX a VERBALEX.
10. Lexikální významy – slova a kolokace z hlediska automatických nástrojů, využití korpusů, elektronické slovníky (obsah, budování), lexikální databáze (WordNet, EuroWordNet, thesauri).