

Doktorský studijní program Filologie, oboru Korpusová a teoretická lingvistika (P0232D090021)

[Ústav Českého národního korpusu](#) (ÚČNK) přijímá doktorandy do studijního programu Filologie, obor **Korpusová a teoretická lingvistika** (P0232D090021, dříve Matematická lingvistika). Předmětem tohoto oboru je počítačové zpracování jazyka za pomoci lingvistických a matematických metod, a to zvláště na bázi elektronických jazykových korpusů. Jsou to metody statistické a/nebo symbolické (algebraické, formální), typicky se oba typy metod kombinují. Obecné informace o studijním oboru jsou na [této stránce](#).

ÚČNK se zabývá se touto problematikou:

- rozvoj a budování korpusů češtiny (synchronní i diachronní, psaného a mluveného jazyka), paralelních (vícejazyčných) a žákovských korpusů různých typů, jejich gramatická (morfologická a syntaktická) anotace
- vývoj nástrojů pro strukturní a specializovanou anotaci jazykových korpusů
- veřejná služba poskytující internetový uživatelský přístup ke korpusům pomocí specializovaných korpusových nástrojů
- jazyková variabilita
- mluvená čeština, její syntax a vlastnosti v komunikaci nerodilých a rodilých mluvčích
- závislostní a složková syntaktická analýza přirozených jazyků
- gramatické formalismy (HPSG) a formální gramatiky
- matematické metody užívané v lingvistice (zejména algebraická teorie formálních gramatik, jazyků a automatů)
- dějiny jazykovědy (zvláště Pražská škola strukturní lingvistiky)

Specifické údaje o studijním oboru

Doktorandův studijní plán je rozčleněn do tří oblastí:

- A. povinnosti spojené s disertací a kontrolou práce na disertaci
- B. základní studijní povinnosti
- C. odborné vědecko-výzkumné či pedagogické aktivity.

Povinnosti student plní formou zápočtu, není-li uvedeno jinak. Jsou zahrnuty do studentova [individuálního studijního plánu](#) (ISP). Plán je navržen tak, aby student všechny své základní studijní povinnosti včetně státní doktorské zkoušky splnil již ve 3. roce studia, a ve 4. roce studia se tak mohl plně věnovat psaní disertace.

A. Disertace

Postup práce na disertaci je ve studentově ISP kontrolován pomocí následujících studijních povinností:

A.1 Doktorandský seminář

- a. Student je povinen se doktorandského semináře účastnit **po celou dobu studia**. Seminář slouží k průběžné prezentaci výsledků, jichž student dosáhl v doktorském

studiu; jeho náplní je především společná prezentace částí disertační práce a dalších odborných aktivit (např. předběžných verzí článků / konferenčních příspěvků). Student tak kolegy informuje o stavu rozpracovanosti disertace v jednotlivých letech studia. Ve třech studijních letech skládá student celkem tři atestace, každý studijní rok jednu.

- b. V případě výjezdu na stáž je student omluven z účasti na semináři, musí však vhodným způsobem prezentovat svou práci.
- c. Povinnost Doktorandský seminář atd. je studentovi automaticky zapisována v každém roce studia.
- d. Seminář se koná jednou za 4 týdny; po 2 týdnech střídá s teoreticko-metodologickým seminářem.

A.2 Teze disertační práce

- a. Rozsah předložených tezí, jež jsou hlavním nástrojem kontroly postupu studentovy práce na disertaci, odpovídá rozsahu a teoretickému záběru disertace. Teze musí obsahovat veškeré podstatné součásti disertační práce: tematické vymezení, hlavní otázky, hypotézy, předpokládané/očekávané závěry, náležitá zdůvodnění, pojmy, strukturu, metody, prameny, používanou literaturu, příp. další údaje, např. přílohy, a časový rozvrh dalšího postupu. Teze prokazují, zda student již splnil převážnou část úkolů spojených s přípravou disertace. Rozsah rozpracování je 7–15 normostran.
- b. Teze se prezentují na doktorandském semináři.
- c. Teze musí být prezentovány nejpozději ve 2. roce studia.

A.3 První kapitola nebo jiný ucelený výstup z disertační práce

- a. Minimální rozsah je 25 normostran.
- b. Kapitola (první nikoli ve smyslu pořadí v disertační práci) se prezentuje na doktorandském semináři.
- c. Kapitola musí být prezentována nejpozději ve 2. roce studia.

A.4 Další kapitola disertační práce

- a. Minimální rozsah je 25 normostran.
- b. Kapitola se prezentuje na doktorandském semináři.
- c. Kapitola musí být prezentována nejpozději ve 3. a každém dalším roce studia.

B. Základní studijní povinnosti

B.1 Cizí jazyk

- a. Angličtina, případně němčina / francouzština.
- b. Zkouška z jazyka musí být složena v prvních 3 letech studia.

B.2 Teoreticko-metodologický seminář I–II nebo Filozofie

B.2.1 Teoreticko-metodologický seminář I–II

- a. Student navštěvuje seminář alespoň po dobu 2 let. Seminář se koná jednou za 4 týdny; po 2 týdnech se střídá s doktorandským seminářem.
- b. Náplň semináře:

1. Přednášky odborníků na různá oborová témata
 2. Četba a rozbor teoreticko-metodologických textů
 3. Diskuse o teoreticko-metodologických otázkách vyplývajících z disertačních projektů
 4. Diskuse s odborníky z jiných pracovišť, případně včetně zahraničních odborníků.
- c. Povinnost Teoreticko-metodologický seminář se studentovi automaticky zapisuje v prvních 2 letech studia (není-li ovšem nahrazena atestací z filozofie).
- d. Seminář musí student absolvovat po dobu 2 let v prvních 3 letech studia.

Sylabus teoreticko-metodologického semináře

Náplň semináře tvoří přednášky odborníků na různá oborová témata, rozdělená do těchto tematických okruhů:

a. vymezení korpusu a metodologie korpusové lingvistiky

- korpusová lingvistika jako odvětví lingvistiky a jako metoda zkoumání jazyka
- terminologie užívaná v korpusové lingvistice

b. výstavba, složení, technické aspekty a anotace korpusu

- postupy při tvorbě korpusu
- složení korpusu a jeho aspekty, klasifikace korpusových textů: reprezentativnost a vyváženost korpusových textů, centrální versus periferní v jazyce
- technické aspekty tvorby korpusu: kódování, filtry textových duplicit a netextového vstupu, standardy (např. Text Encoding Initiative – TEI), značkovací jazyky (SGML, XML), struktura dat v korpusu a jejich počítačová implementace
- gramatika (morfologie, syntax) v korpusu, syntagmatika v korpusu (včetně statistických měr, technik a nástrojů k jejímu zkoumání)
- anotace korpusu: strukturní, fonetická, fonologická, morfologická (včetně lemmatizace), syntaktická, sémantická, pragmatická, obsahová a další

c. využívání korpusu a jeho přínos pro explicitní popis přirozeného jazyka

- korpus jako odraz a vzorek jazykové reality
- konkordance a vyhledávání v korpusu
- přístup typu corpus-driven a corpus-based
- variabilita v jazyce a její odraz v korpusu, jazykové variety
- syntagmatika a paradigmatica v korpusu
- homonymie a synonymie v jazyce a jejich odraz v korpusu a v jeho značkování
- statistické vlastnosti textu, kvantifikace dat a výsledků analýzy korpusu (pravděpodobnost, frekvence, disperze)

d. druhy korpusů z různých hledisek

- korpusy psaného a mluveného jazyka
- synchronní, diachronní, monitorovací korpusy
- paralelní a srovnatelné korpusy
- akviziční korpusy a korpusy nerodilých mluvčích
- autorské korpusy

- web jako korpus
- (ne)značkové korpusy
- přehled zahraničních korpusů

e. aplikace korpusů a korpusové lingvistiky

- korpus a tezaurus
- korpus a zpracování přirozeného jazyka (natural language processing)
- korpusová lexikografie, lexikologie založená na korpusovém materiálu
- pedagogické využití korpusů, jazykové učebnice na základě korpusu
- přínos paralelních korpusů pro překladatele
- korpusová lingvistika a „vnější“ aplikace (sociolingvistické, psycholingvistické, pedagogické, též parole/diskurs)

f. matematické a počítačové metody užívané v jazykovědě a specificky při vytváření a využívání korpusů

- statistické metody zpracování jazyka, statistické míry
- strojové učení, neuronové sítě
- Word Sketches a textové a gramatické struktury
- nástroje Treq, SyD, KWords, konkordancery

g. formální lingvistika

- strukturní lingvistika a její zdroje
- formální popis přirozeného jazyka
- závislostní přístupy
- funkční generativní popis
- model „smysl-text“
- vývoj Chomského teorie gramatiky
- další základní gramatické formalismy (HPSG, LFG, kategoriální gramatiky, (L)TAG)
- fonetika, fonologie, morfologie, syntax, sémantika z pohledu formální lingvistiky
- počítačová lexikografie
- aktuální členění věty, informační struktura, diskurs
- koreference
- programovací jazyky pro popis a analýzu jazyka (lingware)
- formální jazyky a automaty, regulární výrazy

B.2.2 Filozofie

- a. Zkouška z filozofie je alternativou teoreticko-metodologického semináře. Student si ji volí ve výjimečných případech nebo je-li to z hlediska tématu disertace žádoucí. Zkoušku z filozofie místo teoreticko-metodologického semináře skládá student po dohodě se školitelem.
- b. Zkouška má podobu písemné práce na filozofické téma vztažené k projektu disertace; téma si student volí po dohodě se školitelem, konzultantem z Ústavu filozofie a religionistiky FFUK a v souladu se zaměřením disertační práce.
- c. Zkoušku musí student složit v 1.–2. roce studia.

B.3 Oborová zkouška

Zkouška prověřuje znalosti studenta z oboru blízkého tématu disertační práce, v němž student vzhledem k interdisciplinaritě oboru nemusí disponovat patřičnými znalostmi nabytými předchozím studiem. U studenta přicházejícího se zkušeností matematicko-informatickou jsou obsahem atestace lingvistické disciplíny vážicí se k tématu disertační práce (např. lexikologie, morfologie); naopak u studenta s formací převážně lingvistickou atestace prověřuje jeho dovednosti v oblasti matematiky/informatiky/logiky (např. matematické metody v lingvistickém popisu, statistika, základy programování atp.). Předměty atestace navrhuje školitel při sestavování ISP a schvaluje je oborová rada. Zkoušku musí student složit v 1.–3. roce studia.

C. ODBORNÉ AKTIVITY

Požadavky na tvůrčí činnost

Splnění odborných aktivit není podmínkou pro přihlášení k SDZK, nýbrž až pro obhajobu disertace. Předpokládá se, že si studenti po dohodě se školitelem rozloží jejich plnění podle osobních preferencí a též podle konkrétních možností a nabídek (možnosti stáží, publikační činnost) do doby celého studia. Povinné jsou tyto odborné aktivity:

C.1 Odborná studie

- a. Studie musí být původní (příp. přehledový článek, nepřipouští se recenze) publikovaná v časopise/sborníku se standardním recenzním řízením, nebo (minimálně) kapitola v kolektivní monografii.
- b. Studie musí být napsána v 1.–4. roce studia.

C.2 Aktivní účast na konferenci

- a. Připouští se účast na domácí i zahraniční konferenci/workshopu/kolokviu, a to jak formou přednášky, tak posteru.
- b. Konferenční účast je povinná v 1.–4. roce studia.

C.3 Povinná volitelná odborná aktivita

- a. Připouští se minimálně jedna z následujících odborných aktivit:
 1. podíl na řešení grantu nebo podání grantového návrhu
 2. krátkodobá stáž/studijní pobyt na zahraničním nebo na jiném domácím specializovaném pracovišti
 3. odborná editorská činnost
 4. odborný překlad
 5. akademická výuka (1 jednosemestrální kurs)
 6. odborná studie
- b. Termín: 1.–4. rok studia

Další odborné aktivity nad rámec výše vymezených povinností jsou vítány, a mají-li vztah k tématu disertace, student je vykazuje v pravidelném ročním hodnocení.

Požadavky na absolvování stáží

C.4 Zahraniční stáž

- a. Student je povinen absolvovat během studia zahraniční stáž/stáže, v celkovém rozsahu minimálně 1 měsíc.
- b. Rok plnění není předepsán, neboť výjezd na stáž je závislý na stipendijních nabídkách a nelze jej zcela předvídat předem. Povinnost lze plnit i po složení státní doktorské zkoušky, její splnění je až podmínkou pro přihlášení k obhajobě disertace.
- c. V závažných důvodech (péče o děti apod.) lze na základě žádosti schválené děkanem zahraniční stáž nahradit i některou jinou z následujících odborných aktivit se zahraničním přesahem: jako je např. aktivní účast na zahraniční konferenci, odborná ediční činnost, odborný překlad, akademická výuka, odborná studie.

Další studijní povinnosti

U všech doktorandů se požadují pravidelné konzultace s vedoucím práce, případně spolupráce se členy příslušných kateder na grantech. Předpokládá se, že doktorandi navazují zahraniční a přeshraniční styky.

Státní doktorská zkouška

Průběh státní doktorské zkoušky má charakter vědecké rozpravy. V rozpravě jde nejen o rozhovor o samostatné studijní aktivitě doktorandů, ale i o zjištění, zda se doktorand celkově orientuje ve své specializaci, resp. zda je schopen samostatné vědecké činnosti ve vlastní specializaci, na niž se zaměřuje jeho doktorské studium. Hlavní obsahovou náplní zkoušky je diskuse o široce pojatých odborných okruzích souvisejících s tématem disertační práce.

Pro studenta jsou stanoveny tyto okruhy:

1. Tvorba jazykových korpusů
2. Využívání jazykových korpusů
3. Formální lingvistika
4. Statistické zpracování jazykových dat
5. Specifika korpusového přístupu v popisu jazyka

Od studenta se vyžaduje znalost problematiky všech okruhů. Těsně před zkouškou stanoví zkušební komise dva okruhy, z nichž bude studenta zkoušet.