

Mgr. Tomáš Jelínek, Ph.D.
Univerzita Karlova v Praze
Filozofická fakulta
Ústav teoretické a počítačnické lingvistiky
U kříže 8
158 00 Praha 5 – Jinonice
email: Tomas.Jelinek@ff.cuni.cz

Univerzita Karlova v Praze
Filozofická fakulta
k rukám spectabilis doc. PhDr. Michala Pullmanna, Ph.D.
nám. Jana Palacha 2
116 38 Praha 1


Věc: přihláška do Interního výběrového řízení na místo ředitele/ky Ústavu teoretické a počítačnické lingvistiky

V Praze 3. 5. 2018

Spectabilis,

vážený pane děkane,
přihlašuji se do výběrového řízení na místo ředitelky/ředitele Ústavu teoretické a počítačnické lingvistiky, jež bylo vypsáno ke dni 9. 4. 2018 po rezignaci doc. Vladimíra Petkeviče na místo ředitele ÚTKL. Přikládám svůj profesní životopis, bibliografii odborných prací a koncepci rozvoje ústavu na období 1. 7. 2018 – 1. 7. 2021 s výhledem na další roky.

S uctivým pozdravem
Tomáš Jelínek



Profesní životopis

Mgr. Tomáš Jelínek, Ph.D.

Od svého nástupu na Ústav teoretické a počítačnické lingvistiky v roce 2004 se věnuji především korpusové a počítačnické lingvistice s orientací na českou morfologii a syntax a na češtinu jako druhý/cizí jazyk. V rámci výzkumných projektů, především projektu Český národní korpus, vyvíjím softwarové nástroje pro automatické morfologické značkování (desambiguaci) a lemmatizaci českých textů a pro syntaktické značkování češtiny; vývoj počítačové analýzy češtiny zároveň přináší nové poznatky o jazykovém systému češtiny. Zkoumám především (standardní) psaný jazyk, kromě toho se ale zabývám i jazykem mluveným nebo texty tvořenými nerodilými mluvčími češtiny.

Po celou dobu svého akademického působení se soustředím především na vědecký výzkum a řešení výzkumných projektů (a příležitostně i projektů aplikovaného výzkumu), věnuji se také pedagogické činnosti v oboru korpusová lingvistika.

Ukončené vzdělání

2004–2012 Matematická lingvistika, doktorské studium na FF UK

1996–2004 Český jazyk a literatura – Francouzština, magisterské studium na FF UK

Zaměstnání

2004– dosud **Ústav teoretické a počítačnické lingvistiky**, Filozofická fakulta Univerzity Karlovy

2005–2007 **Ústav pro jazyk český Akademie věd ČR**, Oddělení gramatiky

2005–2006, 2008–2009 **Ústav formální a aplikované lingvistiky**, Matematicko-fyzikální fakulta Univerzity Karlovy

Podíl na řešení výzkumných projektů

2004–2007 Diagnostické a evaluační nástroje pro lingvistický software: Národní program výzkumu 1ET100610409; řešitel doc. RNDr. K. Oliva, Dr.

2005–2011 Český národní korpus a korpusy dalších jazyků: VZ RVV MSM 0021620823; řešitel prof. PhDr. F. Čermák, DrSc.

2007–2011 Právní e-slovník – PES: grant GAČR 407/07/0679; řešitel doc. JUDr. F. Cvrček, CSc.

2007–2008 Od jazyka ke znalostem a sémantickému webu: grant GAČR 1ET201120505; řešitel prof. RNDr. J. Hajič, Dr.

2009–2012 Inovace vzdělávání v oboru čeština jako druhý jazyk: grant ESF CZ.1.07/2.2.00/07.0259; řešitel prof. PhDr. K. Šebesta, CSc.

2010–2012 Syntaktická anotace českých korpusů: grant GAČR P406/10/0434; doc. RNDr. V. Petkevič, CSc.

2010–2013 Moderní mluvnice češtiny pro studenty magisterských a doktorských programů: grant ESF CZ.1.07/2.2.00/15.0275; řešitel prof. PhDr. O. Uličný, DrSc.

2012–2015 Český národní korpus: velká infrastruktura pro VaVal LM2011023; doc. PhDr. V. Cvrček, Ph.D.

2012–2016 Český národní korpus: Program rozvoje vědních oblastí na Univerzitě Karlově č. P11.; doc. PhDr. V. Cvrček, Ph.D.

2013–2015 Treebank češtiny na základě gramatiky: grant GAČR 13-27184S; doc. RNDr. V. Petkevič, CSc.

2016– dosud Český národní korpus: Projekty velkých infrastruktur pro VaVal, LM2015044; Mgr. M. Křen, Ph.D.

2016– dosud Počítačnická psycholingvistická analýza českého textu, hledáme mezi lidmi a slovy: grant GAČR 16-19087S; PhDr. D. Kučera, Ph.D.

2016– dosud Mezi slovníkem a gramatikou: grant GAČR 16-07473S; doc. RNDr. V. Petkevič, CSc.

2016– dosud Čeština nerodilých mluvčích z pohledu teoretického a počítačnického: grant GAČR 16-10185S; řešitel ing. A. Rosen, Ph.D.

2017– dosud Český národní korpus: Progres Q08 FF UK; Mgr. M. Křen, Ph.D.

2017– dosud Jazyková variabilita v CNC: OP VVV CZ.02.1.01/0.0/0.0/16_013/0001758; Mgr. M. Křen, Ph.D.

Aplikovaný výzkum a vývoj

vývoj komerčních lingvistických aplikací v rámci týmu pracovníků ÚJČ AV ČR (doc. RNDr. Karel Oliva, Dr. et al.):

2005–2006 Gramatický korektor češtiny I., II. pro Microsoft Office

2007 Rozpoznávání českého rukopisu na Tablet PC

Pedagogická činnost

2009–2012 Výuka korpusové lingvistiky na FP Technické univerzity v Liberci

2010–2013 Výuka morfologie a syntaxe češtiny ve vzdělávacím cyklu Moderní mluvnice češtiny

2012–2016 Výuka korpusové lingvistiky v seminářích ÚČNK FF UK

2017–2018 Výuka ve výběrovém semináři Korpus a syntax na FF UK

1998–2012 Výuka češtiny pro cizince jako OSVČ (firemní, soukromá)

Bibliografie odborných prací Mgr. Tomáš Jelínek, Ph.D.

Hnátková M., Jelínek T., Kopřivová M., Petkevič V., Rosen A., Skoumalová H., Vondříčka P.: Eye of a needle in a haystack. In Mitkov R.: *Computational and Corpus-Based Phraseology*. Berlin, Springer, 2017, s. 160-175.

Jelínek T.: Errors in Inflection in Czech as a Second Language and Their Automatic Classification. In Ekštejn K., Matoušek V.: *Text, Speech, and Dialogue*. Berlin, Springer International Publishing, 2017, s. 263-271.

Jelínek T.: FicTree: a Manually Annotated Treebank of Czech Fiction. In Hlaváčová J.: *ITAT 2017: Information Technologies – Applications and Theory*. Praha, Aachen & Charleston: Technical University & CreateSpace Independent Publishing Platform, 2017, s. 181-185.

Altman G., Andres J., van der Auwera J., Bachmannová J., Balhar J., Bičan A., Bičanová L., Bílková J., Biskup P., Bláha O., Blaszczyk I., Bojar O., Bořil T., Bořkvcová M., Jelínek T., et al.: *Nový encyklopedický slovník češtiny*. Praha, Czechia, Nakladatelství Lidové noviny, 2016.

Jelínek T.: Combining Dependency Parsers Using Error Rates. In Sojka P., Horák A., Kopeček I., Pala K.: *Text, Speech, and Dialogue*. Heidelberg, Springer Berlin Heidelberg, 2016, s. 82-92.

Jelínek T.: Partial Accuracy Rates and Agreements of Parsers: Two Experiments With Ensemble Parsing of Czech. In Brejová B.: *Proceedings of the 16th ITAT: Slovenskočeský NLP workshop (SloNLP 2016)*. Bratislava, CreateSpace Independent Publishing Platform, 2016, s. 42-47.

Křen M., Cvrček V., Čapka T., Čermáková A., Hnátková M., Chlumská L., Jelínek T., Kovářiková D., Petkevič V., Procházka P., Skoumalová H., Škrabal M., Truneček P., Vondříčka P.: SYN2015: Representative Corpus of Contemporary Written Czech. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, ELRA, 2016, s. 2522-2528.

Jelínek T., Petkevič V., Rosen A., Skoumalová H., Vítovec P.: Taking care of orphans: ellipsis in dependency and constituency-based treebanks. In Dickinson M., Hinrichs E., Patejuk A., Przepiórkowski A.: *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT14)*. Warszawa, Polish Academy of Sciences, 2015, s. 119-133.

Jelínek T.: *Skladební funkce a pád v korpusu: Frekvenční analýza*. Praha, Nakladatelství Lidové noviny, 2015. 75 s.

Skoumalová H., Rosen A., Jelínek T., Petkevič V., Vítovec P., Znamenáček J.: A grammar-licensed treebank of Czech. In Henrich V., Hinrichs E., de Kok D., Osenova P., et al.: *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13)*. Tübingen, Department of Linguistics (SfS), University of Tübingen, 2014, s. 218-229.

Jelínek T.: Improvements to Dependency Parsing Using Automatic Simplification of Data. In Calzolari N.: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavík, Island, European Language Resources Association (ELRA), 2014, s. 73-77.

Petkevič V., Jelínek T.: Verbonominální přísudek v české větě. In *Kapitoly ze syntaktologie češtiny. Studie k moderní mluvnici češtiny*. Olomouc, Univerzita Palackého v Olomouci, 2013, s. 150-181.

Jelínek T.: Improving Dependency Parsing by Filtering Linguistic Noise. In Habernal I., Matoušek V.: *Text, Speech, and Dialogue*. Berlin, Springer, 2013, s. 288-294.

Jelínek T.: A System for Syntactic Annotation of Large Czech Corpora. In Zakharov V.: *Trudy meždunarodnoj konferencii Korpusnaja lingvistika - 2013*. St. Petersburg, St. Petersburg State University, 2013, s. 44-51.

Jelínek T.: Srovnání žánrů v korpusu na základě syntaktických funkcí substantiv. *Korpus - gramatika - axiologie*, 2013, s. 3-11.

Petkevič V., Rosen A., Štindlová B., Jelínek T., Hnátková M., Jäger P.: Anotace chybových textů v českém žákovském korpusu. In Šebesta K., Škodová S.: *Čeština – cílový jazyk a korpusy*. Liberec,

Technická univerzita v Liberci, 2012, s. 61-88.

Jelínek T., Štindlová B., Rosen A., Hana J.: Combining Manual and Automatic Annotation of a Learner Corpus. Lecture Notes in Computer Science, 2012, č. 7499, s. 127-134.

Jelínek T., Petkevič V., Rosen A., Skoumalová H.: Czech Treebanking Unlimited. In Hajič J., De Smedt K., Tadić M.: Proceedings of the META-RESEARCH Workshop on Advanced Treebanking 2012. European Language Resources Evaluation (ELRA). Istanbul, ELRA, 2012, s. 37-44.

Jelínek T.: Automatic Rule-Based Correction of Stochastic Syntactic Annotation of Czech. In Ziková M., Dočekal M.: Slavic Languages in Formal Grammar. Proceedings of FDSL 8.5. Frankfurt am Main, Peter Lang, 2012, s. 65-81. ISBN 978-3-631-63609-1.

Jelínek T., Petkevič V.: Systém jazykového značkování současné psané češtiny. In Čermák F.: Korpusová lingvistika Praha 2011, sv. 3: Gramatika a značkování korpusů. Praha, Nakladatelství Lidové noviny, 2011, s. 154-170.

Jelínek T.: Systém pro syntaktické značkování velkých korpusů. In Petkevič V., Rosen A.: Korpusová lingvistika Praha 2011. 3 Gramatika a značkování korpusů. Praha, Nakladatelství Lidové noviny, 2011, s. 123-142.

Bartoň T., Cvrček V., Čermák F., Jelínek T., Petkevič V.: Statistiky češtiny. Nakladatelství Lidové noviny 2009.

Jelínek T., Petkevič, V.: Automatic Identification of Prepositional and Non-prepositional Cases in Czech. In Dočekal M., Ziková M.: Czech in Formal Grammar. Studies in Slavic Linguistics 36. Lincom Europa 2009, s. 103-114.

Jelínek T.: Nové značkování v Českém národním korpusu. Naše řeč 91, 2008, s. 13-20.

Jelínek T.: Morfologické značkování a lemmatizace v korpusech ČNK. In Štícha F., Fried M.: Gramatika a korpus / Grammar & Corpora. Academia, Praha 2008, s. 169-179.

Jelínek T., Doležalová D.: Pravidlová disambiguace a získávání informací o povrchové valenci sloves a adjektiv z ČNK. In Štícha F., Šimandl J.: Gramatika a korpus / Grammar & corpora. ÚJČ AV ČR, Praha 2007, s. 42-48.

Jelínek T.: Neshoda adjektiva s následujícím substantivem. In Čermák F., Blatná R.: Korpusová lingvistika: Stav a modelové přístupy. Nakladatelství Lidové noviny, Praha 2006, s. 153-179.

Mgr. Tomáš Jelínek, Ph.D.

Koncepce rozvoje Ústavu teoretické a počítační lingvistiky FF UK na období 1. 7. 2018 – 1. 7. 2021, s výhledem na další roky

Ústav teoretické a počítační lingvistiky (dále ÚTKL) působí na Filozofické fakultě Univerzity Karlovy od svého založení v roce 1990 v oblasti počítačového zpracování přirozeného jazyka (především češtiny) a formálního popisu jazyka. Ústav je zaměřen vědecky, pracovní náplní ústavu je z větší části řešení výzkumných projektů. V pedagogické oblasti pak zajišťuje spolu s Ústavem Českého národního korpusu (dále ÚČNK) výuku doktorského studia v oboru *matematická lingvistika*. Ředitelem ústavu je od roku 1994 doc. Vladimír Petkevič, CSc., který na svou funkci rezignoval ke 30. červnu 2018.

ÚTKL se podílí spolu s ÚČNK na plnění mnoha společných úkolů, především na vytváření a správě Českého národního korpusu. Od založení ústavu úzce spolupracuje s Ústavem formální a aplikované lingvistiky MFF UK (ÚFAL) a s Ústavem pro jazyk český AV ČR. V rámci FF UK má ÚTKL (kromě ÚČNK) nejvíce společných zájmů s Ústavem českého jazyka a teorie komunikace (ÚČJTK) a s Ústavem bohemistických studií (ÚBS).

Koncepce rozvoje ÚTKL představí současný stav, plánovaný rozvoj v nejbližších třech letech a výhled na další roky ve třech oblastech: ve vědeckém výzkumu, v pedagogické činnosti a v oblasti personální.

I. Vědecký výzkum

1. Vědecké zaměření ústavu

ÚTKL se zaměřuje na počítačové zpracování češtiny, lingvistický výzkum češtiny a na formální popis jazyka a obecnou lingvistiku. Ve všech těchto oblastech je nutné sladit vlastní výzkumný zájem s možnostmi financování výzkumu vzhledem k tomu, že příspěvek („balíček“) na vědu poskytovaný fakultou je v rozpočtu ÚTKL zanedbatelný.

1.1 Počítačové zpracování češtiny

V oblasti počítačového zpracování češtiny se ÚTKL v současné době věnuje především automatické morfologické anotaci a lemmatizaci češtiny, pro tyto účely dlouhodobě vyvíjí desambiguační systém LanGr založený na lingvistických pravidlech, věnuje se úpravám morfologické analýzy a tokenizace.

Dále se zabývá syntaktickou anotací češtiny i dalších jazyků (ve standardu Pražského závislostního korpusu, nově i ve standardu Universal Dependencies), automatickou anotací frazémů a jiných víceslovných lexikálních jednotek, chybovou anotací nestandardního jazyka (čeština nerodilých mluvčích) aj. Po několik let také ÚTKL vede kolektiv lingvistů, kteří spolu vytvářejí návrh nového systému morfologické anotace češtiny NovaMorf.

V příštích třech letech by se v oblasti počítačového zpracování češtiny měl ÚTKL soustředit na zdokonalování a rozvoj počítačových nástrojů, které již vyvíjí, tj. především na zkvalitňování automatické morfologické a syntaktické anotace textů, dále na práci s víceslovnými lexikálními jednotkami a s nestandardním jazykem. ÚTKL nebude rozšiřovat záběr na zcela nové výzkumné oblasti. ÚTKL se tedy zaměří na vylepšení automatické morfologické anotace pomocí systému LanGr, především na vyhledání a odstranění desambiguačních chyb, které LanGr v určitých kontextech způsobuje. Více než dosud se bude ÚTKL zabývat lemmatizací a morfologickou anotací nestandardních textů (textů mluvených, webových, textů nerodilých mluvčích), což dosud nástroje vyvinuté v ÚTKL zvládají jen v omezené míře. Požadavky na kvalitní (v rámci možností) značkování nestandardních textů přibývají jak pro účely infrastruktury ČNK (jako hlavního zadavatele výzkumné práce pro ÚTKL), tak pro další uživatele.

Pokud projekt NovaMorf postoupí do fáze implementace (což mj. záleží na vedení infrastruktury ČNK), bude nutné nemalou část kapacit ústavu věnovat úpravě desambiguačního systému LanGr a dalším souvisejícím činnostem.

I po dokončení grantu zaměřeného na značkování víceslovných lexikálních jednotek bude ÚTKL (v rámci infrastruktury ČNK) pracovat na vývoji databáze těchto jednotek a na automatickém značkování textů s využitím této databáze.

Pracovníci ÚTKL budou nadále ve spolupráci s ÚČNK pracovat na vytváření paralelních korpusů a zlepšování jejich automatické lemmatizace, morfologické (a výhledově i syntaktické) anotace.

V oblasti automatické syntaktické anotace bude ÚTKL pracovat na zlepšení syntaktické anotace češtiny podle Pražského závislostního korpusu (mimo jiné s cílem lepší syntaktické anotace korpusu syn2020) a na syntaktické anotaci paralelních korpusů ve formátu Universal Dependencies.

ÚTKL bude také pokračovat ve vývoji metod a nástrojů pro zpracování a chybové značkování češtiny nerodilých mluvčích, a to jednak v návaznosti na dříve řešené projekty, jednak podle potřeb nových projektů ÚBS a ÚČJTK.

Z dlouhodobého hlediska je třeba zaměření výzkumu v oblasti počítačového zpracování přirozeného jazyka přizpůsobovat jednak možnostem financování daného výzkumu, jednak širšímu vývoji v oboru, kvůli němuž některé v současnosti používané metody mohou být za několik let již nekonkurenceschopné. ÚTKL se při zvažování účasti v nových projektech nevzdá své výzkumné priority, již je studium české morfologie a syntaxe s využitím počítačů a vývoj počítačových nástrojů pro tyto účely.

ÚTKL se bude nadále věnovat především automatické morfologické a syntaktické anotaci češtiny, ovšem jaký systém bude používat, nelze v delším časovém horizontu rozhodnout. Financování ÚTKL po roce 2022 závisí z větší části na pokračování infrastruktury ČNK, ÚTKL by tedy měl vyvinout maximální úsilí při spolupráci na předpokládaném podání přihlášky do nové výzvy (v roce 2020 či 2021).

1.2 Výzkum českého jazykového systému

Zároveň s vývojem nástrojů pro automatické zpracování češtiny probíhá na ÚTKL i výzkum českého jazykového systému, který se opírá o rozsáhlé textové korpusy češtiny, korpusy paralelní, žakovské aj. Údaje zjištěné pro účely například morfologické desambiguace jsou většinou zároveň poznatky o české morfologii, syntaxi nebo lexikologii. Tyto poznatky ÚTKL systematicky shromažďuje a (podle možností) publikuje.

V této vědecké činnosti budou pracovníci ÚTKL pokračovat i v příštích letech. Pokud to dovolí jiné úkoly, mohou publikační činnost v této oblasti ještě zintenzivnit. Jedním z dlouhodobých cílů ÚTKL (jehož plnění ovšem nebylo ještě započato) je vydat publikaci o českém jazykovém systému z pohledu automatické morfologické desambiguace.

1.3 Formální popis jazyka a obecná lingvistika

Pracovníci ústavu se věnují i formálnímu a teoretickému popisu přirozených jazyků (zvláště češtiny), rozvíjejí formální gramatiky a gramatické formalismy, zvláště HPSG (Head-Driven Phrase-Structure Grammar). Kromě teoretického výzkumu a vývoje formálních gramatik po teoretické stránce se věnují i jejich počítačové implementaci, např. v prostředí Trale.

Pracovníci ÚTKL se také věnují výzkumu v oblasti obecné lingvistiky, navazují zejména na myšlenky Pražské lingvistické školy. V roce 2011 byla například vydána publikace *Jindřich Toman: Příběh jednoho moderního projektu. Pražský lingvistický kroužek 1926–1948* (knihu přeložil Vladimír Petkevič), na počátku roku 2015 byla vydána publikace *Pražská škola v korespondenci. Dopisy z let 1924–1989* (knihu připravila Marie Havránková z Ústavu pro českou literaturu AV ČR spolu s Vladimírem Petkevičem). Připravuje se též sborník českých překladů klíčových statí významných protagonistů Pražské školy *Prague School Reader in Linguistics*, jehož redaktorem byl prof. Josef Vachek.

V této činnosti budou pracovníci ÚTKL pokračovat i v následujících letech, nakořik to jejich pracovní povinnosti v ostatních výzkumných směrech ústavu umožní (na tuto vědeckou činnost v současnosti nemá ÚTKL zajištěn žádný zdroj financování a není ani pravděpodobné, že by takový zdroj získal v dohledné době).

2. Vědecké projekty

ÚTKL se v současné době podílí na následujících vědeckých projektech:

- Český národní korpus, projekt velké infrastruktury pro VaVal, LM2015044
- Český národní korpus, program Progres Q08 FF UK
- Jazyková variabilita v CNC, projekt OP VVV CZ.02.1.01/0.0/0.0/16_013/0001758
- Mezi slovníkem a gramatikou, grant GAČR 16-07473S
- Čeština nerodilých mluvčích z pohledu teoretického a počítačového, grant GAČR 16-10185S

V rámci infrastruktury **Český národní korpus** pracuje ÚTKL na automatické lemmatizaci, morfologické a syntaktické anotaci korpusů, na anotaci frazémů a jiných víceslovných lexikálních jednotek. Spolupracuje také na tvorbě a anotaci paralelních (tj. vícejazyčných) korpusů.

Infrastruktura ČNK je schválena na roky 2016–2019, ale díky úspěšné průběžné evaluaci z roku 2017 se dostala na tzv. cestovní mapu velkých infrastruktur (Roadmap of Large Infrastructures for Research, Experimental Development and Innovation of the Czech Republic for the years 2016–2022), jež by měla zajišťovat financování do konce roku 2022.

V programu Progres **Český národní korpus** se pracovníci ÚTKL věnují obecnějším, teoretičtějším otázkám a publikační činnosti v oblasti lingvistiky (obecné, počítačové, korpusové), které nelze realizovat v rámci infrastruktury, jež je zaměřena na aplikovaný výzkum. Program je schválen na období 2017–2018.

Cílem projektu OP VVV **Jazyková variabilita v CNC** je shromáždění a zpřístupnění informací o variabilitě jazykových jednotek (slov, víceslovných lexikálních jednotek) v češtině, primárně na základě synchronních korpusů ČNK (psaných i mluvených). Období realizace: 7/2017 – 6/2020. Po ukončení tohoto projektu OP VVV lze předpokládat vyhlášení dalšího projektu OP VVV pro doplnění financování velkých infrastruktur.

V grantu GAČR **Mezi slovníkem a gramatikou** zkoumá ÚTKL ve spolupráci s některými pracovníky ÚČNK a externisty víceslovné lexikální jednotky (frazémy, idiomy, přísloví a úsloví aj.) a buduje jejich databázi. Grant byl přidělen na roky 2016–2018.

Grant GAČR **Čeština nerodilých mluvčích z pohledu teoretického a počítačového**, na němž ÚTKL spolupracuje s pracovníky ÚBS, ÚJOP UK a ÚFAL MFF UK, je zaměřen na vývoj nového systému chybové anotace češtiny nerodilých mluvčích a na vývoj nových metod automatické jazykové analýzy nestandardního jazyka. Grant byl přidělen na roky 2016–2018.

V letech 2019-2021 by měla být spolupráce s ÚČNK na velké infrastruktuře **Český národní korpus** a na společném grantu OP VVV **Jazyková variabilita v CNC** hlavní činností ústavu. Tento projekt ústavu zajišťuje finanční stabilitu a zároveň umožňuje věnovat se dlouhodobým výzkumným zájmům ÚTKL.

V roce 2018 ÚTKL nové granty (které by nahradily současné dva granty **Mezi slovníkem a gramatikou** a **Čeština nerodilých mluvčích z pohledu teoretického a počítačového** končící v roce 2018) na GAČR nepodává, především kvůli nutnosti většího zapojení do řešení úkolů v rámci infrastruktury ČNK. Jednotliví pracovníci ÚTKL budou (v případě úspěšného podání) zapojeni do nově podávaných grantů vedených jinými ústavami (ÚČJTK, ÚBS) a do projektu **KREAS**. V dalších letech bude ÚTKL samostatně nové granty podávat, pouze bude-li jasné, že stávající granty nevyužijí plně kapacitu pracovníků.

II. Pedagogická činnost

Ústav teoretické a počítačové lingvistiky zajišťuje spolu s ÚČNK výuku doktorského studia v oboru matematická lingvistika. Pracovníci ÚTKL vedou semináře v následujících oblastech:

- formální zpracování přirozeného jazyka včetně matematické teorie formálních jazyků a automatů
- teoretická lingvistika, s důrazem na deklarativní (netransformační) teorie
- gramatické formalismy a jejich aplikace na popis přirozeného jazyka
- korpusová lingvistika.

Kromě toho vyučují pracovníci ústavu v oborech obecná lingvistika a jazykovědná bohemistika předmět základy jazykovědy a úvod do obecné jazykovědy. V rámci volitelných seminářů ve filologických oborech magisterského studia vedou také semináře v oboru korpusové lingvistiky.

V příštích třech letech by výuka v doktorském studiu měla pokračovat ve stejné míře jako dosud. V bakalářském a magisterském studiu je třeba vzhledem k plánovanému zařazení korpusové lingvistiky do společného základu filologických oborů počítat s požadavkem na navýšení počtu hodin v seminářích korpusové lingvistiky pro bakaláře a magistry. Lze také očekávat nárůst zájmu o volitelné (specializovanější) semináře v tomto oboru. Na tento zájem bude ÚTKL muset adekvátně reagovat (i s cílem vyvolat zájem o studium matematické / korpusové lingvistiky v doktorském studiu u vyššího počtu studentů ve studiu magisterském a bakalářském).

III. Personální oblast

Na Ústavu teoretické a počítačové lingvistiky pracuje k 1. květnu 2018 pět akademických pracovníků na plný úvazek a jedna pracovnice neakademická (administrativní) na poloviční úvazek:

- **Doc. RNDr. Vladimír Petkevič, CSc.**, současný ředitel – netermínovaná smlouva
- **Ing. Alexandr Rosen, Ph.D.**, zástupce ředitele – smlouva do 30. 9. 2019
- **RNDr. Hana Skoumalová, Ph.D.**, tajemnice – smlouva do 30. 9. 2019
- **RNDr. Milena Hnátková, CSc.** – smlouva do 30. 9. 2019
- **Mgr. Tomáš Jelínek, Ph.D.** – smlouva do 31. 12. 2019
- **Lenka Horčíčková**, sekretářka – smlouva celkově na plný úvazek (půl úvazku v ÚTKL, půl úvazku v Ústavu Blízkého východu a Afriky FF UK) na dobu neurčitou.

Většina vědeckých pracovníků má smlouvy jen do roku 2019, nicméně vzhledem k jejich zapojení do infrastruktury ČNK a do projektu OP VVV (viz výše) je financování jejich mezd zaručeno i v delším časovém horizontu, smlouvy tedy bude možné prodloužit.

Věkový průměr ústavu je poměrně vysoký, 56 let, je tedy důležité včas zajistit generační obnovu ústavu tak, aby bylo možné znalosti a dovednosti starších kolegů předat novým pracovníkům. První krok k této obnově byl nejspíš učiněn vyhlášením výběrového řízení (VR) na nového pracovníka ÚTKL (s nástupem od září 2018), jehož práce bude financována z infrastruktury ČNK (ve výběrovém řízení ovšem věk nebude hrát žádnou roli, rozhodne jen kvalifikace uchazeče).

V následujících třech letech bude mít tedy ústav šest akademických pracovníků na plný úvazek. Pokud se v těchto třech letech nebo v letech následujících některý z kolegů rozhodne snížit svůj úvazek, bude vyhlášeno VR, aby byly kapacity ústavu odpovídajícím způsobem doplněny. ÚTKL však nebude usilovat o rozšíření nad šest AP na plný úvazek vzhledem k prostorovým i finančním omezením. Zároveň se ÚTKL bude více než dosud snažit zapojovat do své práce studenty v doktorském studiu tak, aby mohli postupně přebírat úkoly i zkušenosti od starších kolegů, kteří budou v následujících letech snižovat své úvazky kvůli důchodovému věku.