

Koncepce rozvoje *Ústavu teoretické a počítačnické lingvistiky FF UK* na období 1. 2. 2016 – 31. 1. 2019

1. Úvod

Ústav teoretické a počítačnické lingvistiky FF UK (dále ÚTKL) byl založen roku 1990 prof. PhDr. Petrem Sgallem, DrSc., jako sesterské pracoviště *Ústavu formální a aplikované lingvistiky MFF UK* (dále ÚFAL) a jeho obecné zaměření je zřejmé z názvu. Je to ústav zaměřený především vědecky, v oblasti pedagogické ústav zajišťuje (spolu s Ústavem Českého národního korpusu FF UK) výuku doktorského studia oboru *matematická lingvistika*. Ředitelem ústavu je od června roku 1994 doc. RNDr. Vladimír Petkevič, CSc., naposledy byl jmenován ředitelem dne 21. 1. 2013, a to na tříleté funkční období od 1. 2. 2013 do 31. 1. 2016.

2. Celkové obecné zaměření ÚTKL

2.1 VĚDECKÁ ČINNOST

Ústav pracuje v těchto lingvistických odvětvích a oborech:

A. Matematická lingvistika

A1. korpusová lingvistika:

Ústav se věnuje přípravě rozsáhlých korpusů současné češtiny, a to zejména v rámci projektu velké infrastruktury *Český národní korpus* (hlavní řešitel dříve prof. PhDr. František Čermák, DrSc., z Ústavu Českého národního korpusu FF UK, dále ÚČNK, nyní doc. PhDr. Václav Cvrček, Ph.D.) a v různých obdobích i projektů Grantové agentury České republiky (GAČR) a Evropského sociálního fondu (ESF). V projektu infrastruktury se ÚTKL konkrétně zabývá touto problematikou:

- (a) gramatické (morfologické a syntaktické) značkování korpusů (tokenizace vstupního textu, větná segmentace, morfologická analýza, morfologická disambiguace včetně disambiguace slovnědruhové a lematizace, syntaktická analýza, zpracování frazémů);
- (b) příprava různých typů korpusů, zejména cizojazyčných paralelních korpusů v rámci infrastrukturního podprojektu *InterCorp*; ÚTKL zde v součinnosti s ÚČNK konkrétně zajišťuje budování datové základny *InterCorpu*, automatické zarovnávání (alignment) a dále gramatické značkování a úkoly související; práce na podprojektu *InterCorp* navíc řídí ing. Alexandr Rosen, Ph.D. z ÚTKL. Dále se ÚTKL zabývá budováním syntakticky anotovaných korpusů (treebanks) umožňujícího mj. různé pohledy na táž jazyková data; a vytvářením žákovských korpusů češtiny nerodilých i rodilých mluvčích a jejich anotace (ve spolupráci s pracovníky Katedry českého jazyka a teorie komunikace FF UK, s pracovníky Technické univerzity Liberec a Matematicko-fyzikální fakulty UK)

A2. Formální lingvistika

Ústav pracuje v těchto oblastech:

- formální a teoretický popis přirozeného jazyka (zejména češtiny); implementace a vývoj formálních gramatik
- syntax přirozeného jazyka a její teoretické a počítačové zpracování (automatická syntaktická analýza češtiny);
- tvorba valenčního slovníku a slovníku kolokací;
- gramatické formalismy (Funkční generativní popis jazyka, Head-Driven Phrase-Structure Grammar aj.);
- příprava encyklopedických hesel oboru matematická lingvistika (včetně lingvistiky korpusové);

B. Obecná lingvistika

Ústav se podílí na vydávání (překladových) publikací souvisejících hlavně s Pražskou školou strukturní lingvistiky.

C. Teoretická lingvistika

Ústav se věnuje výzkumu v oblasti morfologie a syntaxe přirozeného jazyka, především češtiny.

2.2 PEDAGOGICKÁ ČINNOST

Pracovníci ústavu vyučují v seminářích oboru matematická lingvistika v těchto 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
- gramatika češtiny
- korpusová lingvistika.

Pracovníci ÚTKL mimoto vyučují v oborech obecná lingvistika a jazykovědná bohemistika předmět: základy jazykovědy a úvod do obecné jazykovědy

I v letech 2016 až 2019 se ÚTKL soustředí na vědeckovýzkumnou a pedagogickou činnost ve výše uvedených odvětvích s tím, že bude zejména plnit úkoly stanovené ve zmíněném projektu velké infrastruktury *Český národní korpus* (2012–2016 a i v návazném období) a podle možností v projektu *PRVOUK*, a to ve spolupráci s ÚČNK. Mimoto bude řešit úkoly stanovené v jiných projektech (podrobněji v odst. 4).

3. Oblast personálního rozvoje

V ÚTKL pracuje v současnosti šest pracovníků, z toho pět na plný úvazek:

- **doc. RNDr. Vladimír Petkevič, CSc.**, 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.

V ústavu je tedy pět vědeckých pracovníků na plný úvazek a sekretářka na úvazek poloviční. Vědečtí pracovníci sice mají smlouvy až do roku 2019, ty jsou však bohužel podmíněny prostředky získávanými z grantů. Pracovníci jsou v současnosti, tj. na konci roku 2015, placeni jednak z prostředků přidělených přímo ústavu v podobě tzv. mzdového „balíčku“, jednak z projektu velké infrastruktury *Český národní korpus*, a též z projektu GAČR *Treebank češtiny na základě gramatiky* (od 1. 2. 2013). Podrobněji o těchto projektech viz níže v odst. 4. „Balíček“ ÚTKL je ovšem nepřiměřeně nízký ve srovnání s tím, co ústav fakultě přináší granty a svou vědeckou produkcí.

Práci na uvedených projektech považuje ústav za svou vědeckovýzkumnou prioritu. Na projektu infrastruktury a grantů GAČR se podílejí a budou podílet též (zejména) mladí externisté, kteří jsou zaměstnáváni na dohody o provedení práce a na dohody o pracovní činnosti. Zvládnout vědeckou práci na uvedených projektech je ovšem velmi náročné, a v tomto směru proto plánujeme (pochopitelně v závislosti na finančních možnostech) takovéto personální posílení ústavu k řešení grantových úkolů:

- o programátora-počítačového lingvistu
- nejméně o jednoho lingvistu schopného exaktního, algoritmického uvažování.

Ústav úzce spolupracuje především s ÚČNK a dále pak s Ústavem formální a aplikované lingvistiky MFF UK, zčásti i s Ústavem pro jazyk český AV ČR, Filozofickou fakultou a Fakultou informatiky Masarykovy univerzity v Brně a Filozofickou fakultou Univerzity Palackého v Olomouci a dalšími obdobně zaměřenými pracovišti.

Za jednu z priorit považuje ústav *kvalifikační postupy pracovníků ÚTKL*, tj. profesury, docentury a úspěšné absolvování doktorského studia doktorandů ať už oboru matematická lingvistika, tak oborů jiných (pracovníci ústavu jsou školiteli i doktorandů jiných jazykových oborů).

4. Rozvoj vědecké činnosti

Hlavní náplní činnosti ÚTKL je vědeckovýzkumná práce v oboru *matematická lingvistika*. Ústav se specializuje a v nejbližších třech letech také bude specializovat na tyto výzkumné oblasti:

- morfologie a syntax přirozených jazyků, zejména češtiny, za použití matematických, formálních a počítačových metod; vytváření vhodných formalismů pro popis přirozených jazyků
- vytváření jazykových korpusů češtiny, paralelních korpusů a žákovských korpusů, zejména jejich gramatického popisu, a to po stránce lingvistické i informatické
- péče o dědictví Pražské školy strukturální lingvistiky

Během své existence obdržel ústav řadu českých a mezinárodních grantů, v nichž figuroval jako hlavní řešitelské nebo spoluřešitelské pracoviště, většinou s doc. Petkevičem jako hlavním odpovědným řešitelem. V posledních patnácti letech to byly zejména tyto granty:

Granty Grantové agentury České republiky (GAČR):

- **Syntaktická analýza českých textů** (2010–2012) (ÚTKL byl hlavním řešitelským pracovištěm)
- **Slovní poklad češtiny v informační společnosti** (2003–2005) (hlavním řešitelským pracovištěm byl Ústav pro jazyk český AV ČR, spoluřešitelským pracovištěm pak FF UK, konkrétně ÚTKL)
- **Velké jazykové korpusy a jejich automatická analýza** (2003–2005) (hlavním řešitelským pracovištěm byl Ústav formální a aplikované lingvistiky MFF UK, ÚTKL se podílel na plnění úkolů projektu)
- **Možnosti a meze gramatiky češtiny ve světle Českého národního korpusu** (2003–2005) (hlavním řešitelským pracovištěm byl Ústav pro jazyk český AV ČR, spoluřešitel pak FF UK, konkrétně ÚTKL)
- **Elektronizace postupů diachronní lexikografie** (1999–2001) (hlavním řešitelským pracovištěm byl Ústav pro jazyk český AV ČR, spoluřešitelským pracovištěm pak FF UK, konkrétně ÚTKL)
- Komplexní projekt GAČR **Čeština ve věku počítačů** (1996–2001) (hlavním řešitelským pracovištěm byl Ústav formální a aplikované lingvistiky MFF UK, ÚTKL byl spoluřešitelským pracovištěm)

Grant MŠMT:

Příprava testovacích dat a nástrojů pro testování lingvistického software (2005–2007) (hlavním řešitelským pracovištěm byl Ústav pro jazyk český AV ČR, ÚTKL se podílel na plnění úkolů projektu).

Výzkumný záměr:

- **Český národní korpus a korpusy jiných jazyků** (2005–2011, reg. č. MSM0021620823, vedoucí záměru prof. PhDr. František Čermák, DrSc., ÚČNK) ve spolupráci s ÚČNK

Projekt typu ESF (Evropský sociální fond):

- **Inovace vzdělávání v oboru čeština jako druhý jazyk** (červen 2009 – květen 2012, reg. č. CZ.1.07/2.2.00/07.0259, *OP Vzdělávání pro konkurenceschopnost*, hlavní koordinátor prof. PhDr. Karel Šebesta, CSc.) ve spolupráci s Katedrou českého jazyka a teorie komunikace FF UK a ve spolupráci s Technickou univerzitou Liberec.

Pracovníci ÚTKL se též jako externí spolupracovníci podíleli na řešení projektů:

- **Nový encyklopedický slovník češtiny** (ESF, 2012–2014).
- **Moderní mluvnice češtiny** (grant ESF, 2010 – 30. 6. 2013)

Z mezinárodních grantových projektů uvádíme tyto (za celé období existence ústavu):

- **Shared formal grammar of Czech and Polish** (program vědecko-technické spolupráce MŠMT KONTAKT 2004/23). ÚTKL byl spoluřešitelským pracovištěm.
- Mezinárodní projekt **CONCEDE. Consortium for Central European Dictionary Encoding** (1998–2000) (PL-1142). ÚTKL byl hlavním řešitelským pracovištěm za českou stranu.

- Mezinárodní projekt **TELRI (Trans–European Language Resources Initiative)** (1995–1997). ÚTKL se mimo další pracoviště v České republice podílel na plnění úkolů tohoto grantu za českou stranu.
- Mezinárodní projekt **MULTEXT–EAST. Multilingual Text Tools and Corpora for Central and Eastern European Languages** (1995–1997) (COP106). ÚTKL byl hlavním řešitelským pracovištěm za českou stranu.
- Mezinárodní projekt **Language Technologies for Slavic Languages (LATESLAV) – PECO 2824** (1993–1995). ÚTKL se mimo Ústav formální a aplikované lingvistiky MFF UK podílel na plnění úkolů tohoto grantu za českou stranu.

V **současnosti** pracuje ÚTKL na těchto projektech:

- Velká infrastruktura **Český národní korpus** (2012–2016, reg. č. LM2011023, vedoucí doc. PhDr. Václav Cvrček, Ph.D., ÚČNK) ve spolupráci s ÚČNK.
- Grant GAČR **Treebank češtiny na základě gramatiky** (1. 2. 2013 – 31. 12. 2015, reg. č. P406/13-27184S). Hlavním řešitelem tohoto právě končícího grantu je doc. RNDr. Vladimír Petkevič, CSc., hlavním řešitelským pracovištěm je ÚTKL.

Vědecká činnost ÚTKL v nejbližším období bude probíhat především v rámci těchto projektů:

- Probíhající velká infrastruktura **Český národní korpus** (2012–2016), ÚTKL zde spolupracuje a bude spolupracovat s ÚČNK podle stanoveného rozvrhu zatím do konce roku 2016 včetně, ve výhledu je pokračování infrastruktury do roku 2022;
- Nově udělený grant GAČR **Mezi slovníkem a gramatikou** (reg. č. 16-07473S, hlavní řešitel Vladimír Petkevič, 2016–2018)
- Grant GAČR **Čeština nerodilých mluvčích z pohledu teoretického a počítačového** (reg. č. 16-10185S, hlavní řešitel Alexandr Rosen, 2016–2018).

a také následujícího projektu, bude-li ústavu přidělen:

- Grant Ministerstva kultury ČR v rámci programu na podporu aplikovaného výzkumu a experimentálního vývoje národní a kulturní identity (NAKI II):
Pravopis a gramatika češtiny v éře počítačové gramotnosti (2016–2019)

Mimo uvedené oblasti činnosti pracovníci ústavu příležitostně přednášejí/vyučují v zahraničí. Zúčastňují se vědeckých konferencí doma i v zahraničí (z posledních jmenujme účast s referáty na konferencích ve Varšavě, Řezně, Sankt-Petěrburku, Dubrovniku, Istanbulu, Göttingen, Tübingen).

Ústav spolupracuje s obdobně zaměřenými zahraničními lingvistickými pracovišti, například s Univerzitou v Řezně (spolupráce v oblasti syntaxe slovanských jazyků a paralelních korpusů) s Univerzitou v Torontu (dlouhodobá spolupráce na vývoji topologického parseru a formalismu pro adekvátní popis slovosledu), s univerzitou ve Varšavě (gramatické formalismy). Ke kontaktům se zahraničím patří i návštěvy významných zahraničních odborníků, kteří ústav navštívili v poslední době: byli to například prof. Zygmunt Saloni z Varšavské univerzity, dr. Ruprecht von Waldenfels z Univerzity v Bernu, prof. Leonid Iomdin z Ruské akademie věd, prof. Gerald Penn z Univerzity v Torontu, doc. Viktor Zacharov z Univerzity v Petrohradě.

Pracovníci ústavu odborně působí rovněž v oblasti obecné lingvistiky, zejména pečují o myšlenkové dědictví Pražské lingvistické školy. Roku 2005 vyšla česká verze publikace *Lingvistického slovníku Pražské školy* od Josefa Vachka (knihu připravil Vladimír Petkevič spolu s Jaromírem Tláskalem), v říjnu roku 2011 byla 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.

4.1 Zhodnocení realizace cílů dosavadní koncepce vědecké činnosti

V oblasti vědecké činnosti byly v posledním období, tj. v letech 2013–2015 (tedy v předcházejícím funkčním období ředitele ústavu Vladimíra Petkeviče) veškeré naplánované úkoly v rámci hlavního výzkumného projektu – výzkumného záměru **Český národní korpus a korpusy jiných jazyků** – splněny. Dokončuje se projekt GAČR **Treebank češtiny na základě gramatiky** (1. 2. 2013 – 31. 12. 2015, reg. č. P406/13-27184S).

5. Rozvoj pedagogické činnosti

ÚTKL zajišťuje (ve spolupráci s ÚČNK) na FF UK studijní obor: *Filologie – matematická lingvistika* (doktorské studium), nabízí výuku v kursech *počítačové a formální lingvistiky, základů jazykovědy a obecné lingvistiky* v podobě povinně volitelných a výběrových přednášek a seminářů pro pregraduální i postgraduální studenty. Takto se také podílí na výuce pro tyto obory: *logika, srovnávací jazykověda, lingvistika a fonetika, český jazyk a literatura* na FF UK a konečně *počítačová a formální lingvistika* na MFF UK. Na téže fakultě vyučuje také zahraniční studenty v rámci magisterského oboru *jazyk a informační technologie* programu Erasmus Mundus. Mimo vlastní přednášky a semináře se ústav rovněž podílí na výuce v semináři *korpusové lingvistiky*, který je organizován ÚČNK. Přednášky a semináře navštěvují nejen studenti (hlavně doktorandi) FF UK, ale i studenti z MFF UK a dalších fakult UK. Pracovníci ústavu vedli a vedou také doktorandy v oborech *matematická lingvistika, český jazyk a literatura, obecná lingvistika a germanistika*. Členové ústavu dr. Skoumalová a dr. Rosen jsou členy oborové rady doktorského oboru *matematická lingvistika*, doc. Petkevič je jejím předsedou. Doc. Petkevič je rovněž členem oborové rady oboru *logika a český jazyk a literatura* na FF UK, *obecný a indoevropský jazykozpyt* na Filozofické fakultě Masarykovy univerzity v Brně (FF MU), oboru *český jazyk* na Filozofické fakultě Univerzity Palackého v Olomouci (FF UP) a oboru *anglický jazyk* na téže fakultě. Doc. Petkevič a dr. Rosen jsou též členy oborové rady oboru *matematická lingvistika* na MFF UK. Pracovníci ústavu bývají oponenty disertačních prací v oboru *matematická lingvistika* a dalších oborech (například *český jazyk a literatura, germanistika*), zasedají v komisích pro státní doktorské zkoušky v uvedených oborech na FF UK, MFF UK, Fakultě informatiky Masarykovy univerzity (FI MU), FF MU, FF UP. Píší také posudky na projekty podávané u Grantové agentury Univerzity Karlovy. Rovněž příležitostně přednášejí v semináři formální lingvistiky pořádaném Ústavem formální a aplikované lingvistiky MFF UK.

Doc. Petkevič a dr. Rosen se podílejí na výchově vědeckých pracovníků v rámci doktorského studia oboru *matematická lingvistika*, i oborů jiných (český jazyk a literatura, germanistika).

5.1 Zhodnocení realizace cílů dosavadní koncepce pedagogické činnosti

Cíle dosavadní koncepce byly splněny ve výše uvedeném smyslu.

6. Výhled dalších oblastí rozvoje základní součásti

Mimo plány uvedené výše se ÚTKL bude v příštích letech věnovat obecně těmto aktivitám:

(a) širší spolupráci s dalšími oborovými pracovišti, zejména pak s těmito: s ÚČNK a dalšími lingvisticky zaměřenými ústavy a katedrami FF UK, katedrou logiky FF UK, ÚFAL MFF UK, ÚJČ AV ČR, FI MU, FF MU a FF UP;

(b) prohlubování spolupráce se zahraničními oborovými pracovišti, s nimiž už má ÚTKL dlouhodobé kontakty, a navazování odborných kontaktů s novými pracovišti.

7. Stručný profil kandidáta na ředitele doc. RNDr. Vladimíra Petkeviče, CSc.

Stručné odborné curriculum vitae

- Narodil se 2. 3. 1954 v Praze
- Vystudoval Matematicko-fyzikální fakultu Univerzity Karlovy (MFF UK), obor *Matematické zabezpečení výpočetní techniky* (1974–1979)
- Doktorát z přírodních věd (RNDr.), obor: *matematická informatika a teoretická kybernetika* (1985)
- Kandidát věd (CSc.), obor: *matematická informatika a teoretická kybernetika* (1992)
- Zaměstnán ve *Výzkumném ústavu matematických strojů (VÚMS)*. Podílel se na vývoji operačních systémů pro sálové (mainframe) počítače, dále na vývoji překladačů programovacích jazyků a na vývoji databázových programů (1979–1992).
- Dne 1. ledna 1993 přijat jako samostatný vědecký pracovník do *Ústavu teoretické a komputační lingvistiky* FF UK v Praze.
- Od 10. 6. 1994 dosud ředitelem ÚTKL FF UK.
- Habilitace v oboru *matematická lingvistika* (1996), název práce: *Underlying Structure of Sentence Based on Dependency*
- Od roku 1996: vedoucí lingvistické sekce (pro synchronní jazyk) Ústavu Českého národního korpusu FF UK, a to do roku 2006
- Přednášky na zahraničních univerzitách: Düsseldorf, Heidelberg, Erlangen, Tübingen (SRN), Bratislava; výuka na letní škole (Sozopol 2002)
- Účast na mezinárodních konferencích (nejvýznamnější: COLING'88, ACL'93, EURALEX 1994, COLING–ACL 1998, EURALEX 2000, Corpus Linguistics 2005 Birmingham, Corpus Linguistics 2011 Sankt-Petěrburg, SlaviCorp 2011 Dubrovník, Treebanks and Linguistic Theories Tübingen, Grammar and Corpora Varšava 2014)
- člen Jazykovědného sdružení České republiky
- člen Pražského lingvistického kroužku
- člen těchto oborových rad: pro obor *matematická lingvistika* (FF UK, předseda), *logika* (FF UK), *český jazyk a literatura* (FF UK), *matematická lingvistika* (MFF UK), *obecný a indoevropský jazykozpyt* (FF MU), *český jazyk* (FF UP).

Odborná specializace:

matematická lingvistika a lingvistická bohemistika:

- korpusová lingvistika – lingvistické značkování jazykových korpusů (zejména korpusů synchronní češtiny v rámci projektu *Český národní korpus*)
- syntax a morfologie přirozeného jazyka a její formální a počítačové zpracování, automatická morfologická a syntaktická analýza češtiny, gramatické formalismy

obecná jazykověda

- pražský strukturalismus

Konkrétní pedagogická a odborná činnost v posledních pěti letech:

- výukav oboru matematická lingvistika, konkrétně: základy oboru (formální popis přirozeného jazyka, korpusová lingvistika, matematické metody v lingvistice, gramatické formalismy); výuka základů jazykovědy a úvodu do obecné jazykovědy; občasné přednášky na semináři pořádaném na MFF UK a na jiných fakultách (např. FF UP, FF MU, FI MU, Západočeská univerzita v Plzni)
- vedení devíti doktorandů v doktorském oboru *matematická lingvistika*, z nichž čtyři doktorandi v období prosinec 2007 – prosinec 2012 úspěšně absolvovali doktorské studium.
- gramatické (morfologické a syntaktické) značkování jazykových korpusů češtiny
- příprava žákovského korpusu nerodilých mluvčích
- výzkum (morfo)syntaxe české věty na základě korpusů současné češtiny
- překlady v oboru lingvistika a matematická lingvistika
- redakční činnost při přípravě publikací v uvedených oborech.

V Praze dne 1. prosince 2015

doc. RNDr. Vladimír Petkevič, CSc., ředitel ÚTKL FF UK