

# ÚSTAV TEORETICKÉ A KOMPUTAČNÍ LINGVISTIKY

Celetná 13, 110 00 Praha 1

<http://utkl.ff.cuni.cz>

telefon: +420 221 619 752

e-mail: [utkl@ff.cuni.cz](mailto:utkl@ff.cuni.cz)

## Hlavní oblasti výzkumu

Ústav teoretické a počítačové lingvistiky (ÚTKL) FF UK se zabývá vědecko-výzkumnou činností v následujících lingvistických odvětvích:

- počítačové zpracování přirozeného jazyka
- teoretická jazykověda
- obecná lingvistika

V oblasti počítačového zpracování přirozeného jazyka včetně korpusové lingvistiky ústav vyvíjí počítačové nástroje pro automatické zpracování (především) českého jazyka, zvláště software pro morfologické a syntaktické značkování češtiny, syntaktickou analýzu češtiny a paralelní korpusy. Vytváří disambiguační gramatiku češtiny založenou na automatické morfologické disambiguaci a syntaktické analýze češtiny, spoluvytváří a zpracovává počítačový valenční slovník češtiny a slovník kolokací a gramatickým značkováním se podílí na vytváření morfologicky a syntakticky anotovaných korpusů.

V oblasti teoretické jazykovědy se ústav zabývá též výzkumem gramatických formalismů a systémů.

V oblasti obecné lingvistiky a teoretické jazykovědy se ústav zabývá formálním a teoretickým popisem přirozeného jazyka, zejména češtiny. Ústav publikuje původní odborné články a věnuje se překladům publikací významných děl souvisejících s tematikou Pražského lingvistického kroužku.

## Profilové projekty

ÚTKL spolupracoval s Ústavem Českého národního korpusu FF UK na sedmiletém výzkumném záměru *Český národní korpus a korpusy dalších jazyků*, který byl financován MŠMT (2005–2011). Řešil také úkoly tříletého grantu GA ČR *Syntaktická anotace českých korpusů* (2010–2012), jehož cílem bylo vytvořit a uživateli (odborníkovi i lingvisticky vzdělanému laikovi) zpřístupnit syntakticky anotovaný korpus SYNT současné češtiny.

Ústav se dále podílel na řešení úkolů grantu ESF *Inovace vzdělávání v oboru čeština jako druhý jazyk* (červen 2009 – květen 2012). V tomto projektu se podařilo vytvořit korpus CzeSl obsahující texty vytvořené nerodilými mluvčími češtiny a romskými mluvčími češtiny. Projekt byl připravován ve spolupráci s Technickou univerzitou Liberec.

V současné době (listopad 2013) ÚTKL pracuje na grantu GA ČR *Treebank na základě gramatiky*, který navazuje na projekt *Syntaktická anotace českých korpusů*. Ústav je také zapojen do projektu velké infrastruktury *Český národní korpus*, kde připravuje zejména nové morfologické značkování současné češtiny, morfologicky značkuje nové korpusy současné češtiny (SYN2013PUB) a zpracovává paralelní korpus InterCorp.

## Nejdůležitější publikační projekty

ÚTKL se ve spolupráci s Ústavem Českého národního korpusu FF UK podílel na publikacích v edici *Studie z korpusové lingvistiky: Korpusová lingvistika: Stav a modelové přístupy* (2006), *Kolokace* (2006), *Mnohojazyčný korpus InterCorp: Možnosti studia* (2010), *InterCorp: Exploring a Multilingual Corpus* (2010).

V edici *Korpusová lexikografie* ústav spolupracoval na těchto publikacích: *Slovník Karla Čapka* (2007), *Slovník Bohumila Hrabala* (2009), *Slovník komunistické totality* (2010).

Dále se podílel na přípravě publikací *Frekvenční slovník češtiny* (2004) a *Statistiky češtiny* (2009).

Ústav se též spolu s Ústavem Českého národního korpusu FF UK podílí a podílí na tvorbě jazykových korpusů současné češtiny: SYN2005, SYN2006PUB, SYN2009PUB, SYN2010, SYN, SYN2013PUB, a to gramatickým značkováním a technickou přípravou; účastní se rovněž budování paralelního korpusu InterCorp.

Ústav též spolupracoval na přípravě českého překladu *Lingvistického slovníku Pražské školy Josefa Vachky* (2005) a připravil překlad knihy Jindřicha Tomana *Příběh jednoho moderního projektu. Pražský lingvistický kroužek 1926–1948* (2011).

## Spolupracující domácí a zahraniční instituce

ÚTKL spolupracuje s Ústavem Českého národního korpusu FF UK na tvorbě a zpracování jazykových korpusů současné češtiny i na korpusech paralelních. S Ústavem formální a aplikované lingvistiky MFF UK spolupracuje na slovnědruhovém a morfologickém značkování korpusů češtiny a vývoji nového zachycení morfologie češtiny. S Technickou univerzitou Liberec se podílí na tvorbě žakovského korpusu češtiny CzeSl.

K zahraničním institucím, se kterými je ústav v kontaktu, patří: Universität Regensburg – Institut für Slavistik (Německo); Polska Akademia Nauk – Instytut Podstaw Informatyki, Warszawa (Polsko); University of Toronto – Department of Computer Science (Kanada) – spolupráce v oblasti morfologie a syntaxe slovanských jazyků a gramatických formalismů.

Další spolupracující instituce: Jazykovedný ústav Ľ. Štúra SAV Bratislava; Uniwersytet Warszawski: Wydział Polonistyki, Instytut Badań Interdyscyplinarnych; Sankt-Petěrburgskij gosudarstvennyj universitet, Filosofskij fakultět; Russkaja Akademia Nauk, Institut Problem Peredači Informacij, Moskva; Universität Tübingen, Seminar für Sprachwissenschaft; Universität Bern, Institut für slavische Sprachen und Literaturen; Freie Universität Berlin, Institut für Deutsche und Niederländische Philologie.