

# **PLIN059 Proseminář z počítačové lingvistiky – úvod**

**Mgr. Dana Hlaváčková, Ph.D.**

[hlavacko@phil.muni.cz](mailto:hlavacko@phil.muni.cz)

**Ústav českého jazyka FF MU**

A. Nováka 1, budova D

# Formality

- účast povinná, možná 1 neomluvená absence
- jinak omluvenky do IS
- 1 vnitrosemestrální test
- 1 závěrečný test
- praktická seminární práce

# Co je to počítačová lingvistika?

- Matematická lingvistika, Kvantitativní lingvistika, Komputační lingvistika, Počítačové zpracování přirozeného jazyka, Jazykové inženýrství
- Computational Linguistics, Mathematical Linguistics, Natural Language Processing (NLP)
- průnik mezi lingvistikou, informatikou a matematikou
- součást oblasti Digital Humanities
- **přirozený jazyk + počítačové zpracování**
- **detailní analýza jazyka a jeho formální popis**
- **strojové učení a neuronové sítě**

# Co počítačová lingvistika dělá?

- výsledkem jsou denně používané **aplikace**, např.:
  - korektor překlepů
  - korektor gramatiky
  - vyhledávání na webu
  - prediktivní psaní
  - překladače jazyků
  - online slovníky
  - syntéza řeči
  - chatboty

# Co počítačová lingvistika poskytuje?

- urychlení a zefektivnění práce lingvisty
- ověřování existujících teorií a hypotéz
- objevení nového jazykového jevu, zákonitosti
- počítačový lingvista ví, co a jak může použít
- a co může a nemůže od nástrojů očekávat
- autorská práva a přístupy k nástrojům
  - veřejně dostupné (dostupné na MU)
  - hromadný přístup (společné heslo)
  - vlastní přístup (registrace)

# Pár obecných zásad

- **proč** to chceme? (cíl, účel, uživatel)
- **jak** toho dosáhneme? (efektivita)
- **uživatelská přívětivost**, uživatelská zkušenost, User Experience
- maximum **automatizace** – minimum ruční práce (při vytváření i používání)
- zpracování **velkého objemu** dat
- **univerzálnost** (široká množina vstupů, spojování více nástrojů do jednoho)
- **nezávislost** na jednotlivých lingvistických teoriích
- dříve desktopové, dnes **webové aplikace**
- při zpracování i používání je nutná **PŘESNOST**

# Obsah kurzu

- **počítačová lexikografie** – DEBDict, DEBWrite, lexikální databáze, Vokabulář webový a další
- **jazykové korpusy** – KonText, Sketch Engine
- **morfologická analýza** – Ajka, Majka, Morče, MorphoDiTa (atributivní a poziční systém)
- **derivační rozhraní** – Deriv, Morfio a další
- **syntaktická analýza** – Synt, Set, PDT (stromové banky)
- **sémantická analýza** – WordNet, FrameNet, VerbNet
- **valenční databáze** – Vallex, VerbaLex
- **rozpoznávání a syntéza řeči**
- seminární práce

# Příbuzná pracoviště

- **Centrum zpracování přirozeného jazyka** FI MU Brno – <http://nlp.fi.muni.cz/>
- **Ústav formální a aplikované lingvistiky** MFF UK Praha – <http://ufal.mff.cuni.cz>
- **Ústav teoretické a počítačové lingvistiky** FF UK Praha – <http://utkl.ff.cuni.cz>
- **Ústav Českého národního korpusu** FF UK Praha – <http://www.korpus.cz>
- **Ústav pro jazyk český** AV ČR – <http://www.ujc.cas.cz>



# Příbuzná pracoviště

- **Fakulta informačních technologií** VUT Brno –  
<http://www.fit.vutbr.cz>
- **Katedra informatiky a výpočetní techniky** –  
<http://www.kiv.zcu.cz>, **Katedra kybernetiky**  
<http://www.kky.zcu.cz> FAV ZČU Plzeň
- **Ústav informačních technologií a elektroniky** FM TU Liberec –  
<http://www.fm.tul.cz>
- **Slovenský národný korpus**, JÚLŠ SAV Bratislava –  
<http://korpus.juls.savba.sk/>

# Ukázky

- [Internetová jazyková příručka](http://prirucka.ujc.cas.cz) <http://prirucka.ujc.cas.cz>

- [WebMetaTrans](http://metatrans.fi.muni.cz)

<http://metatrans.fi.muni.cz>

© Jan Pomikálek (FI MU)