

# ÚVODNÍ SEMINÁŘ CJBB84

Úvod do korpusové lingvistiky  
seminář pro magisterské studium

# dnes

- náplň semináře
- požadavky k ukončení
- harmonogram
- studijní literatura

# Požadavky

**Podmínky – aktivní účast na semináři**  
**Požadavky ke kolokviu: zpracování**  
**zvoleného lingvistického problému**  
**formou výzkumné zprávy (5-10 s.).**

# Osnova výzkumné zprávy

- 1) zpracování příslušného jazykového jevu v literatuře
- 2) popis pracovního postupu při získávání jazykového materiálu z korpusu
- 3) analýzu korpusových dat
- 4) srovnání výsledků vlastní analýzy s údaji nalezenými v odborné literatuře
- 5) pokus o hodnocení přínosu korpusu pro zpracování daného problému

# Náplň semináře

- 1) Prakticky zaměřené cvičení s korpusovým manažerem BONITO
- 2) Konzultace k problémům zpracovávaným pro závěrečnou výzkumnou zprávu

# Harmonogram seminářů

- 1. Úvodní seminář (harmonogram, literatura) 25.9.
- 2. Bonito – I. (základní vyhledávání) 2.10.
- 3. Morfologické značkování 9.10.
- 4. Bonito – II. (třídění, frekvenční seznamy) 16.10.
- 5. Výběr tématu a konzultace 23.10.
- 6. Ukázka k tématům z formální morfologie 30.10.
- 7. Prezentace – host 6.11.
- 8. Ukázka k tématům se složitější konstrukcí dotazu 13.11.
- 9. Ukázka k tématům z teorie tagování 20.11.
- 10. Ukázka k tématům mapujícím slovní zásobu 27.11.
- 11. Konzultace k závěrečným zprávám 4.12./11.12.
- 12. Odevzdání závěrečných zpráv – hodnocení 18.12.

# Komentáře ke studijní literatuře

- přesnější informace během semináře při řešení jednotlivých problémů

# Úvody

- Šulc, M.: Korpusová lingvistika. První vstup. Praha : Karolinum. 1999
- Kocek, J. - Kopřivová, M. - Kučera, K. (eds.) (2000): Český národní korpus - úvod a příručka uživatele Praha : FF UK - ÚČNK 2000.



# Zajímavosti

- Čeština doma a ve světě 1 a 2, 2001.

# Cvičebnice

- Blatná, R. - Čermák, F. (eds.) (2005): Jak využívat Český národní korpus. Praha : Nakladatelství Lidové noviny.

# ČNK

- **František Čermák, Věra Schmiedtová (2004): Český národní korpus – základní charakteristika a širší souvislosti, Národní knihovna – knihovnická revue roč. 15, č. 3, s. 152-168.**
- **<http://ucnk.ff.cuni.cz>,**

# BONITO

- **Marie Kopřivová, Jan Kocek : Manuál korpusového manažeru *Bonito***
- ***<http://nlp.fi.muni.cz/projekty/bonito/>***
- Kocek, J. - Kopřivová, M. - Kučera, K. (eds.) (2000): Český národní korpus - úvod a příručka uživatele Praha : FF UK - ÚČNK 2000.

# Práce v seminářích

- spuštění programu BONITO
- přihlášení
- vyvolání www ČNK
- paralelní práce s návodem

# upozorňujeme

- download – články z KL ke stažení a studiu
- jak citovat korpus – citační normy pro jakoukoliv odbornou práci odvolávající se na korpus
- manuál a instalace – podrobný manuál zacházení s korpusovým manažerem

# BONITO

Autor :Pavel Rychlý, FI MU

Použití: České korpusy (ČNK, korpusy FI MU)

Korpusový manažer: program umožňující efektivní práci s korpusem.

# PŘÍSTUP- HESLO

- **Jméno: CJBB**
- **Heslo:leei7458**



# Mluvené korpusy a KSK

- Hlaváčková, D.: Brněnský mluvený korpus a jeho morfológická analýza. In: 3. mezinárodní setkání mladých lingvistů Olomouc. 2002, s. 167 – 173.
- Hlaváčková, D., Sedláček, R.: Morfológické značkování korpusu soukromé korespondence, XIV. kolokvium mladých jazykovedcov, 8. - 10. 12. 2004, Šintava pri Seredi, Slovenská republika. (V tisku)
- Hladká Z. (2005): Zkušnosti s tvorbou korpusů češtiny v ÚJČ FF MU v Brně, SPFFBU, A 53, s. 115-124.
- Osolsobě, K.: Hypokoristika v korpusu soukromé korespondence KSK, SP FF MU A, 53, Brno, 2005, s. 125-136.
- Osolsobě, K.: Korpus soukromé korespondence z hlediska morfológického značkování, SPFFBU A 54, s. 187-201, Brno.

# PDT

- Hajičová Eva, Panevová Jarmila, Sgall Petr (2002): K nové úrovni bohemistické práce, Využití anotovaného korpusu. Část I. Slovo a slovesnost, 63, s. 161-177.
- Eva Hajičová , Jarmila Panevová, Petr Sgall (2002): K nové úrovni bohemistické práce: Využití anotovaného korpusu. Část II. Slovo a slovesnost, 63, s. 241-262. ???

též: <http://ufal.mff.cuni.cz/pdt2.0/doc/pdt-guide/cz/html/ch06.html>

# MORFOLOGICKÉ ZNAČKOVÁNÍ

- Hajič J., Hladká B. (1997): Morfologické značkování korpusu českých textů stochastickou metodou. Slovo a slovesnost 4/1997, s. 288-304.
- Petkevič V.(2001): Neprojektivní konstrukce v češtině z hlediska automatické morfologické disambiguace. In: Hladká Z., Karlík P. (eds.) Čeština – univerzália a specifika 3. Brno : Masarykova univerzita, s. 197-206.

# MORFOLOGICKÉ ZNAČKOVÁNÍ

- Bartůšková, D., Hlaváčková, D., Ungermannová, M.: Manuál pro značkování a desambiguaci slovních tvarů v jazykových korpusech, rkp. 58 s. Brno : FI MU, 2004. (pdf verze: <http://nlp.fi.muni.cz/projekty/desman/>)

# DIACHRONNÍ KORPUSY

- Kučera K. (1998): Diachronní složka Českého národního korpusu : obecné zásady, kontext a současný stav. Listy filologické, 121, s. 303-313.

# KL v širších souvislostech

- ČERMÁK, F., KLÍMOVÁ, J., PETKEVIČ, V (eds.): Úvod do korpusové lingvistiky, Karolinum, 2000.

# Tzv. vytěžování (mining) korpusu

- Renata Blatná – Vladimír Petkevič (eds.): Jazyky a jazykověda: Sborník k 65. narozeninám prof. PhDr. Františka Čermáka, DrSc. Praha: Filozofická fakulta Univerzity Karlovy, Ústav Českého národního korpusu, 2005. (kap. Studie z korpusové lingvistiky)
- Karlík, P. (Ed.) (2004): Korpus jako zdroj dat o češtině, Brno : FF MU.

# Poznámky ke cvičebnici

## Jak využívat Český národní korpus

- Příručce by podle našeho názoru prospělo více jasně formulovaných návodů, kde získat znalosti, které si je třeba osvojit k tomu, aby hlubší zamyšlení se nad různými problémy jazyka (češtiny), k němuž mají dát podnět jednotlivá cvičení, mohlo být plodné. (Máme na mysli případy, kdy nelze u celého širokého spektra adresátů, jimž je příručka určena, předpokládat patřičné znalosti jak lingvistické, tak technické.)



# První kapitola

**„Práce s Českým národním korpusem krok za krokem“** obsahuje sedm tematických oddílů (A. *Pravopis*, B. *Tvoření slov / slovotvorba (morfologie širší)*, C. *Tvarosloví / morfologie užší*, D. *Slovní zásoba / lexikologie*, E. *Kolokace (slovní spojení)*, F. *Syntax*, G. *Kombinovaná zadání*) a zaměřuje se na jednodušší úkoly sledujíc přitom jednotlivé roviny jazyka.

# Druhá kapitola

**„Co říká o různých slovech korpus a co slovníky“** sestává z jazykového kvízu zaměřeného na určování významů cizích slov.

# Třetí kapitola

## **„*Význam slova prozrazuje kontext*“**

zahrnuje cvičení založená na příkladech vět vybraných z korpusu SYN2000. Jsou zaměřena na odhad výskytu konkrétního slova podle kontextu a řazena do tematických oddílů (*A. Slovní tvary, B. Lemmata, C. Kolokace, D. Části slov, E. Formálně podobná slova, F. Význam slov*).

# Čtvrtá kapitola

**„Úkoly pro náročnější“** - její náročnost spočívá v tom, že u zájemce o procvičování nabízených úkolů se předpokládá u některých cvičení základní znalost programování a u všech podrobné studium kapitoly *Popis morfologických značek v Manuálu korpusového manažeru Bonito*

( <http://ucnk.ff.cuni.cz/bonito/index.html> ).

# Pátá kapitola

- **„Práce se subkorpusy“** zahrnuje dva úkoly a je tedy jen úvodem do této problematiky.

# Klíč

- Velmi pozitivní je fakt, že učebnice má klíč.
- Bohužel řada chyb

# Hodnocení

*Jak využívat Český národní korpus* je textem zaměřeným primárně k **popularizaci** výsledků práce projektu ČNK. I přes některé nedostatky lze uvítat, že v době, kdy se korpusový přístup k výzkumu jazyka prosadil jako jedna z nejdůležitějších metodologií lingvistického výzkumu, vychází prakticky orientovaná příručka pro studenty, badatele i širší veřejnost případných zájemců.

# Návod

Cennou pomůckou se tato cvičebnice může stát v rukou těch z řad uživatelů ČNK, kteří nelitovali, popřípadě nebudou litovat námahy věnované studiu lingvistické literatury nejen korpusově zaměřené (lakoničnost úvodů jednotlivých kapitol a oddílů předpokládá uživatele, který má již nějakou předchozí zkušenost jak s prací s jazykovými korpusy, tak s prací v materiálově orientovaném lingvistickém výzkumu).



# Poučení

Ti ostatní si mohou „pohrát“. (Je ovšem třeba upozornit na to, že ne každá hra je tím, co měl Jan Amos Komenský na mysli, když psal svoje dílo „Schola ludus“ a že řečeno s učitelem národů „student musí mít zadek z olova“, a to nejen k tomu, aby na něm seděl jsa „připojen“ k ČNK. )