

CJBB75 Základy využití korpusu pro praxi

st. 10.50-12.25 G13

Výuka ZS 2012

- 19. 9. ÚKOL 1 : ,Vyplnit žádost, získat heslo.
- 3. 10. Organizační pokyny, základy – *Bonito*.
- 10. 10. ÚKOL 2
- 17.10. Různé korpusy a rozdíly v anotačních schématech (tokenizace, lemmatizace, tagging, disambiguace, tagset).
- 24.10. ÚKOL 3
- 31.10. Jak lze v korpusech hledat doklady pro výzkum hláskových alternací?
- 7. 11. ÚKOL 4
- 14. 11. Je možné získat z korpusů informace o slovtvorbě ?
- 21. 11. ÚKOL 5 (Jak využívat ČNK : 17/5, 19/15).
- 28. 11. Písemka
- 5. 12. ÚKOL 6
- 12.2. Zpt. test

19.9. Organizační pokyny

- a) Na [www http://ucnk.ff.cuni.cz/](http://ucnk.ff.cuni.cz/) - Dohody a registrace si opatřete uživatelská přístupová práva k programu Bonito.
- b) Na [www http://ucnk.ff.cuni.cz/](http://ucnk.ff.cuni.cz/) - Manuál a instalace - Instalace si stáhněte na svůj domácí počítač program Bonito.
- c) Na [www http://ucnk.ff.cuni.cz/](http://ucnk.ff.cuni.cz/) - Dostupné korpusy si nastudujte informace o korpusech ČNK.
- d) Na [www http://ucnk.ff.cuni.cz/](http://ucnk.ff.cuni.cz/) - Manuál a instalace si nastudujte informace o tom, jak zadávat dotazy v programu Bonito.

Bonito

Manuál korpusového manažeru Bonito

Marie Koprivová
Jan Koček

The screenshot shows the Bonito software interface with the following components and annotations:

- dotazový řádek**: The search bar at the top containing the query "[lemma="korpus"]".
- výběr korpusu**: The dropdown menu for selecting a corpus, currently set to "syn2000".
- pojmenování dotazu**: The dropdown menu for naming the query, currently set to "korpus (lemma)".
- konkordanční řádek**: The list of concordance results, each starting with a corpus code (e.g., opus=smrt) and a snippet of text.
- označený konkordanční řádek**: A specific row in the list is highlighted in blue, representing the selected concordance.
- vyhledaný výraz - KWIC (key word in context)**: The word "korpus" is highlighted in red within the concordance snippet.
- konkordanční seznam**: The entire list of concordance results.
- kód jednoznačně identifikující text**: The corpus code "opus=nylon" at the beginning of a row.
- rozšíření kontextu vyhledaného výrazu**: The expanded text snippet below the concordance list, showing the word "korpus" in context.
- stavový řádek**: The status bar at the bottom showing "Zobrazeno: 1+100/276 (36%) Řádek: 7 Vybráno: 1".

Menu Korpus

- Souhrnné informace
- Implicitní atribut

Základní vyhledávání v korpusu

- **Obsah:**

Vyhledávání tvaru slova nebo slovního spojení

Vyhledávání podle atributu **lc** (lowercase)

Vyhledávání podle atributu **lemma**

Vyhledávání podle atributu **tag** (morfologická značka)

Nastavení implicitního atributu

Hledání v rámci jedné věty

Menu Konkordance

- Zpět k předchozí
- Vpřed k následující
- Statistiky – Frekvenční distribuce

Menu Zobrazení

- Zdroje
- Atributy
- Kontext
- Rozsah

Zobrazení názvu díla a autora

The screenshot shows the Bonito software interface. At the top, there are menu options: Manažer, Korpus, Dotaz, Konkordance, Zobrazení, Výběr, and Nápověda. Below the menu is a search bar with a dropdown menu for 'jméno:' and a search button 'syn'. The search results are displayed in a list on the left and a detailed view on the right. The list on the left contains the following entries:

- opus.nazev=Ikaros, ročník 1999,opus.autor=Y
- opus.nazev=Naše řeč, ročník 1997,opus.autor=Y
- opus.nazev=Slovo a slovesnost, ročník 1997,opus.autor=Y
- opus.nazev=Slovo a slovesnost, ročník 1997,opus.autor=Y
- opus.nazev=Efeméra - ostatní,opus.autor=Y
- opus.nazev=Lidové noviny - Magazin, č. 49/1997,opus.autor=Y
- opus.nazev=Právo, 4. 12. 2003,opus.autor=Y
- opus.nazev=Deníky Bohemia, 24. 11. 2007,opus.autor=Y
- opus.nazev=Letohradský měsíčník, ročník 1998,opus.autor=Y

The detailed view on the right shows a snippet of text from the first result, with the word 'použití' highlighted in red. The text reads: 'dosud nevidaného jazykového materiálu bude konci věty ; v této pozici je ho v našem tutéž cenu . Malou cenu bude mít použití tomto jazyce pro vytvoření diversifikovaného dosud nevidaného jazykového materiálu bude přibližně od 6 do 25 korun) . K ovinutí včetně autobusového terminálu . Stavbě jako loni v lednu , " svěřila se keramička . Na adresou http : // UCNK . FF . CUNI . CZ/CNC .

At the bottom of the window, there is a status bar showing 'Zobrazeno: vše/9 Řádek: 1'. Above the status bar, there is a summary of the search results: 'Počet výskytů: 9', '> Query : "korpus"', and '> P-filter: ". *už[i]t. *"'. The interface also includes a scrollbar on the right and a status bar at the bottom.

Navrhněte postup, jak vyhledat v korpusu

SYN2010 tvary **maskulin v genitivu plurálu**

- a) pomocí tagu
- b) bez pomoci tagu.