

## Vaje II

---

### Pregled korpusov FidaPLUS podrobneje

---

---

---

---

---

---

---

---

## Dogajanja

---

- <http://www.sdjt.si/> - Slovensko društvo za jezikovne tehnologije;
- <http://nl.ijs.si/is-ltc08/index-sl.html> - konferenca JEZIKOVNE TEHNOLOGIJE;
- <http://lojze.lugos.si/jota/> - Jezikovnotehnološki abonma;
- <http://videlectures.net/solomon/> - Solomonovi seminarji;
- <http://sdrv.fe.uni-lj.si/> - seminarji Slovenskega društva za razpoznavanje vzorcev;
- <http://groups.google.com/group/lingvisticnikrozek/>
- Lingvistični krožek na Filozofski fakulteti.

23.10.2008

Amanda Saksida Korpusno  
jezikoslovje

2

---

---

---

---

---

---

---

---

## Orodja

---

- iskanje po korpusih:
  - slovenski besedilni korpus [FidaPLUS](#)
  - slovenski besedilni korpus [Nova BESEDA](#)
  - angleško-slovenski [EVROKOPUS](#), [\\*-IJS KOPUSI](#)
- [RDR/CLOG](#): spletni vmesnik za lematizacijo besedil
- [ZRCola](#): vnašalni sistem za jezikoslovno rabo v programu WinWord (opis v [prispevku](#))

23.10.2008

Amanda Saksida Korpusno  
jezikoslovje

3

---

---

---

---

---

---

---

---

## Slovarji in leksikalni viri

- [BOS@ZRC](#): Slovanske in besedilne zbirke na Inštitut za slovenski jezik Frana Ramovša ZRC SAZU:
  - [Slovar slovenskega knjižnega jezika](#) (iskalnik)
  - [Spletni seznam besed slovenskega jezika](#) (iskalnik)
  - [Odzadnji slovar slovenskega jezika](#) (iskalnik)
  - [Besedišče slovenskega jezika](#) (iskalnik)
  - [Zbirka za določevanje osnovnih besednih oblik](#) (iskalnik)
- [Terminološki slovar Evroterm](#) (iskalnik)
- [Terminološki slovar Informatike](#) (iskalnik)
- [Angleško-slovenski računalniški slovar](#) (iskalnik)
- [Nemško-slovenski in slovensko-nemški slovar](#) (iskalnik)
- [Resjanica](#): portal z informacijami (slovarjem) o Rezijanščini
- [Dictionarium quatuor linguarum](#): slovar Hieronymusa Megiserja (iskalnik)
- Sezname spletnih slovarjev: [gov.si](#), [Lexicool](#)
- [sloWNet](#) (semantični slovar): za vse jezike glej [Global WordNet Association](#)

23.10.2008

Amanda Saksida Korpusno  
jezikoslovje

4

## Preverjanje in učenje jezika

### Govor

- [Amebis sklanjanje](#): računalnik vam sklanja vnešeni samostalnik ali pridevnik
- [Seznam pripomočkov za učenje slovenščine, Slovenian for Travelers](#)
- Sintetizatorji govora:
  - [Govorec \(Odsek za inteligentne sisteme, Institut "Jožef Stefan"\)](#)
- [SAMPA-sl](#): SAMPA fonetični simboli za slovenski jezik

23.10.2008

Amanda Saksida Korpusno  
jezikoslovje

5

## Dveri

- [Revija "Jezik in slovstvo"](#)
- [Tematska številka o jezikovnih tehnologijah \(3/4 2003\)](#)
- [ACL Anthology](#)
- [HLT Central](#): EU portal o jezikovnih tehnologijah
- [Joint European Website for Education in Language and Speech, JEWELS](#)
- [EisNet CEE Survey](#): Central and Eastern European Centres of Human Language Technologies
- [EisNet's Directory of Language and Speech Technology Experts and Organisations](#)
- [Language Technology World](#): Nemški virtualni informacijski center za jezikovne tehnologije
- [Bookmarks for Corpus-based Linguists](#)
- [Statistical NLP / corpus-based computational linguistics resources](#)
- [Hrvaški portal za jezikovne tehnologije](#)
- [Euralex 2000 Tutorial - Listing of selected links](#)

23.10.2008

Amanda Saksida Korpusno  
jezikoslovje

6

## Vrste korpusov

■ Korpusi na spletu: kako določimo, kakšni sploh so  
(glede na tipologijo: pisni oz. govorni korpusi, referenčni oz. korpusi podjezikov, celoviti oz. vzorčni korpusi, statični oz. spremljevalni korpusi, enojezični oz. večjezični, označeni oz. neoznačeni, + način zapisa, velikost, kakovost označevanja ipd.)

- Slovenski korpusi (Fida+, Nova beseda, Evrokorpus, korpusi na IJS, MultextEast)
- Hrvaški korpusi (<http://www.hnk.ffzg.hr/jthj/korpusi.htm>)
- Češki korpus (<http://ufal.mff.cuni.cz/rest/CAC/doc-cac10/cac-guide/eng/html/index-en.html>)
- CHILDES (<http://chilides.psy.cmu.edu/data/>)
- LDC (<http://www ldc.upenn.edu/Catalog/>)
- ELRA (LR database: <http://www.elra.info/Catalogue.html>)

23.10.2008

Amanda Saksida Korpusno  
jezikoslovje

7

## Fida+: delo s spletnim vmesnikom na [www.fidaplus.net](http://www.fidaplus.net)

- osnovno iskanje
  - konkretne pojavitve besed
  - nadomestni znaki
  - sopojavitve besed (fraze, iskanje po bližini)
  - notranja 'in' in 'ne'
- iskanje po lemah
- iskanje po kriterijih (napredno iskanje)
  - iskanje po platformi besedil (po zvrsteh, po vrstah, po jezikovni obliki - (ne)lektorirano)
- shranjevanje spisikov konkordanc in obdelanih seznamov

23.10.2008

Amanda Saksida Korpusno  
jezikoslovje

8

## Vaje:

1. Kaj se v korpusu pojavlja pogosteje: "otvoritev razstave" ali "odprtje razstave"?
2. Kolikokrat se v korpusu pojavi besedna oblika »obnoviti« in kolikokrat lema "obnoviti"?
3. Poiščite vse leme v korpusu, ki vsebujejo koren "trd".
4. Poiščite čimveč pridevnikov, ki se pojavljajo v bližini lem "ženska", "punca", "baba" in "dekle". So med njimi kakšne razlike? Primerjajte rezultate s pridevniki v bližini "moški", "fant" in "dedec".
5. Poiščite vsaj 4 različne pomene besede *zemlja* ter ugotovite, ali obstaja navezava med tipom besedila in pomenom.

23.10.2008

Amanda Saksida Korpusno  
jezikoslovje

9