

Komise: Hladká, Hlaváčková, Nevěřilová

- 1)
 - a) finitní a infinitní tvary sloves, přechodníky; současné kodifikační slovníky (chtěla hlavně slyšet, kam bych koukala na pravopis přejatých slov, mluvily jsme o slovnících neologismů -- žádný z nich nemůže být kodifikační u ze své podstaty; když se vás na to bude ptát a budete si chtít šplhnout, řekněte, že v Drobnostech předjímal, jak se budou slova přejatá psát a většinou se totálně sekli)
 - b) token, tokenizace, lemma, lematizace, tag, tagset, KWIC, konkordance; automatická morfologická analýza a desambiguace
 - c) jazykový model (jsou to v podstatě n-gramy, to ze mne tahala Zuzka asi 5 minut)
- 2)
 - a) Synonymie, opozitnost, hyperonymie a hyponymie, holonymie a meronymie, onomaziologie, semaziologie
 - b) deambiguace, znakování korpusu, typy znakování
 - c) cql, desambiguace, textové korpusy, prohledávání korpusu, indexování, nástroje pro práci s korpusy
- 3)
 - a) teritoriální a sociální stratifikace cestiny, české dialekty podrobně, o obecně cestině, český jazykový atlas
 - b) formální vs. přirozené jazyky, chomského hierarchie, gramatiky s příklady (napsat gramatiku která generuje PSČ), příklad na regexpy, CQL, syntaktická analýza
 - c) struktura korpusu, reprezentativnost, metainformace, strukturní a pozicní značky
- 4)
 - a) slovní druhy, Naše řeč
 - b) funkce, relace a jejich vlastnosti; přirozená čísla v PC; posloupnost, řady
 - c) korpus jako zdroj dat pro lingvistický výzkum (jaký vybrat korpus apod.)