

otázky mgr státnice 09 2017

-1-

- a) (Nevěřilová) – jazykové modely a desambiguace, morfologické značkování – docela podrobně jak funguje morfologická analýza např. u Majky, zabralo nejvíce času
- b) (Osolsobě) – historie českých korpusů a korp. lingvistiky, SYN – chtěla slyšet o DIAKORP, mluvené korpusy ORAL i jiné, INTERKORP, o PDT (úrovně značkování), pracoviště (nezapomenout na Akademii věd), značkování (první SYN2000 byl jen automaticky), nějaké osobnosti: Tešitelová. Hajičová apod.
- c) (Navrátilová) – vývoj deklinace u substantiv, adjektiv a zájmen do dnešních dnů – velké překvapení, nechtěla ani vědět jaké byly dříve kmeny, jen tendence (typy skloňování, složená adjektiva si vyvinula ze sloučení s ukazovacím zájmenem, vývoj hlásek, přechody mezi kmeny, zánik konsonantických kmenů apod.)

-2-

- a) (Nevěřilová): množiny, množinové operace, zadání množin, přirozená čísla v teorii množiny atd.
- b) (Osolsobě): lingvistické značkování, tokenizace, lemmatizace, automatická morfologická analýza, disambiguace (statistické metody, pravidlová disambiguace, hybridní metody) – chtěla zmínit FRANTu, lingvistické problémy grafického zachycení korpusů mluveného jazyka (přepis a pravidla přepisu v mluvených korpusech budovaných v rámci ČNK), stručná charakteristika olomouckého mluveného korpusu (chtěla vědět rozdíly mezi PMK + BMK vs. OMK) – **Klára oblíbené téma je morfologická disambiguace**
- c) (Navrátilová): praslovanština – popis, časové zařazení atd. + nářeční štěpení prsl., na základě kterého se vydělily jednotlivé slovanské jazyky (3 změny – popsat, časově zařadit)

-3-

- a) (Nevěřilová): Přirozené vs. formální jazyky. Formální gramatiky – základní pojmy, typy g. dle Chomského, vymyslet přepisovací pravidla dle jednotlivých typů a něco málo vygenerovat. Syntaktická analýza – SYNT a SET, frázové vs. složkové stromy – rozdíly ukázat na modelové větě. CQL – popsat základní metaznaky a uvést pár příkladů + všechna slova, která ten výraz vyhledá.
- b) (Osolsobě): Reprezentativnost korpusů – kdy je vhodné použít reprezentativní korpus (zajímaly jí konkrétní jazykové jevy), výhody a nevýhody webových korpusů. Morfologické značky – pražský vs. brněnský tagset (rozdíly, výhody a nevýhody) – vše docela podrobně (např. že pražský měl původně o pozici míň, protože se v SYN2000 neznačkoval vid, dále jaké negramatické vlastnosti mohou tagy vyjadřovat – stylové příznaky atd.). Korpusové manažery – co dělá serverová a co klientská část, vyjmenovat ty nejpoužívanější, popsat jejich základní funkce (f. analýza, kolokace), charakterizovat word sketches.
- c) (Navrátilová): Stratifikace národního jazyka – hlavně co vše pod něj spadá a rozdělit jej na strukturní a nestrukturní útvary. Ke každému útvaru stačilo říct pár vět.