**Bc. státnice – září 2015**

Komise: Hladká, Hlaváčková, Nevěřilová

1)

1. 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 ze své podstaty; když se vás na to bude ptát a budete si chtít šplhnout, řekněte, že v Drobnostech předjímali, jak se budou slova přejatá psát a většinou se totálně sekli)
2. token, tokenizace, lemma, lematizace, tag, tagset, KWIC, konkordance; automatická morfologická analýza a desambiguace
3. jazykový model (jsou to v podstatě n-gramy, to ze mne tahala Zuzka asi 5 minut)

2)

1. Synonymie, opozitnost, hyperonymie a hyponymie, holonymie a meronymie, onomaziologie, semaziologie
2. deaambiguace, znackovani korpusu, typy znackovani
3. cql, desambiguace, textové korpusy, prohledávání korpusu, indexování, nástroje pro práci s korpusy

3)

1. teritorialni a socialni stratifikace cestiny, ceske dialekty podrobne, o obecne cestine, cesky jazykovy atlas
2. formalni vs prirozene jazyky, chomskeho hierarchie, gramatiky s priklady (napsat gramatiku ktera generuje PSČ), priklad na regexpy, CQL, syntakticka analyza
3. struktura korpusu, reprezentativnost, metainformace, strukturni a pozicni znacky

4)

1. slovní druhy, Naše řeč
2. funkce, relace a jejich vlastnosti; přirozená čísla v PC; posloupnost, řady
3. korpus jako zdroj dat pro lingvistický výzkum (jaký vybrat korpus apod.)

**Státnice červen 2015**

Komise: Klára Osolsobě, Michaela Boháčová, Zuzana Nevěřilová

Otázky:

1.

1. Typologie slovníků + příklady
2. Automatická syntaktická analýza
3. Textové korpusy – uložení v PC, indexace a vyhledávání, korpusové nástroje, CQL

2.

1. Kombinatorika, pravděpodobnost, podmíněná pravděpodobnost, pravděpodobnostní rozložení; Zipfův zákon, náhodná veličina, zákon velkých čísel
2. Fonetická transkripce a její typy
3. Automatická morfologická analýza

3.

1. Relace a funkce, přirozená čísla, posloupnosti, množiny
2. Slabika, stavba slabiky, vlastnosti slabiky
3. Korpusové manažery

4.

1. Množiny – způsob zápisu, definice, operace nad nimi; přirozená čísla – operace nad nimi, definice, zápis pomocí následníka
2. Hláskové změny v současné češtině – asimilace, disimilace, hiát, proteze (změny byly v zadání vyjmenované)
3. Typy korpusů

5.

1. Definice – věta – důkaz, výroková a predikátová logika 1. Řádu
2. Historický vývoj deklinace v čj
3. CQL prakticky

6.

1. Formální a přirozené jazyky, formální gramatiky, regulární výrazy, CQL, automatická syntaktická analýza
2. Změny ve slovní zásobě
3. Jak se převádí texty do korpusu, transkripce a transliterace

**Leden 2015**

**Osolsobě, Boháčová, Nevěřilová. Otázky zněly přibližně takto:**

1) Korpusy, jejich obecná definice a značkování a použití pro vědeckou práci (měla jsem mluvit hlavně o metadatech), historické korpusy (byla jsem tázána i na ty "předpočítačové" a na půlmilionový český korpus, který vznikl na UK ještě před ČNK).

2) Úzus, norma a kodifikace (definovat, srovnat, uvést příklady); vývoj české deklinace (jmenná, složená a smíšená deklinace).

3) Formální a přirozené jazyky, různé typy gramatik (měla jsem napsat příklad nějakého formálního jazyka, z abecedy a pravidel odvodit pár slov a podobně), regulární gramatika a výrazy.

**Hladká - Dana - Vojta.**
1a) Typy deklinace
1b) Historicko-srovnávací jazykověda - základní obrysy: Chtěla časový vymezení, vůbec podstatu té metody a něco k mladogramatikům.
2) Von Neumannův model, reprezentace čísel, reprezentace znaků, znakový sady.
3) Značkování korpusů, značkovací systémy v ČR a problematika desambiguace.

a) automatická morfologická analýza a desambiguace - že se desambiguátory učí z dat a frekvencí, pořád jsem uváděl příklady (slovo jak), brněnský a pražský značky;

b) korpus a internet

c) hláskové změny 1400-1600

d) teritoriální vymezení dialektů - stačilo říct "Dej mouku ze mlejna na vozejk" čtyřikrát jinak a vyjmenovat ty dialektový oblasti

**Září 2013**

* vývoj deklinace, relace
* typy deklinace a historickosrovnávací jazykověda, matika (korelace, statistika), korpus a programy

**Mgr. státnice**

**Červen (Hlaváčková, Osolsobě, Kovář)**

* první člověk: vývoj slovanských písem, strojový překlad
* druhý člověk: vývoj korpusové lingvistiky

**Září (Boháčová, Osolsobě, Kovář)**

* skloňování adjektiv
* z té poč. lingvistické morfologickou, syntaktickou, sémantickou analýzu, značkování, desambiguaci
* z informatiky zase morfolog. analýzu, rozpoznávání emocí, dialogové systémy

**Leden 2016 (Hladká, Hlaváčková, Kovář)**

* typologie jazyků podle Skaličky
	+ papírek – přiřaďte jednotlivé příklady k typům: dobr-ý (flexivní), přítel – přátel (interflexivní), maloměsto (polysyntetický), orám – vyorám – přeorám (aglutinační), budu nosit (analytický/izolační)
* obecná čeština
* statistika, pravděpodobnost a její aplikace na počítačovou lingvistiku (hlavně korpusy), strojové učení a jeho metody, strojový překlad (ptal se jen na statistický strojový překlad)
* počítačové porozumění přirozenému jazyku a počítačové zpracování jednotlivých rovin přirozeného jazyka (morfologická, syntaktická, sémantická, slovotvorná) + jednotlivé nástroje (majka, ajka – morfologická, syntaktická – synt apod.) – jenom vyjmenovat nástroje pro jednotlivé roviny
* okrajově disambiguace (chtěla slyšet jen statistický x pravidlový)

2)

* formální jazyky, značkovací jazyky, CQL
* korpusy: historie korpusu v ČR i ve světě, paralelní korpusy, diacorp, webové korpusy, rozdíl pražské vs. brněnské
* lexikální sémaziologie, mladogramatici