

Korpuszok a nyelvészeti kutatásban szeminárium

Oktató: Dr. Vincze Veronika

Időpont: péntekenként tömbösítve

Elérhetőség: <http://www.inf.u-szeged.hu/~vinczev/>

A kurzus célja, hogy a hallgatókkal megismertesse, miként használatosak a korpuszok és egyéb nyelvi adatbázisok a nyelvészeti kutatásban. A félév során áttekintjük a nyelvi adatgyűjtés módszertani követelményeit, sorra vesszük a korpuszépítés különböző lépéseit, illetve a korpuszból való adatgyűjtés módjait, továbbá az adatok statisztikai elemzésére is figyelmet fordítunk. A kurzus során minden hallgató elkészíti saját minikorpuszát, amely az elsajátított elméleti ismeretek gyakorlati alkalmazására is kiváló lehetőséget nyújt.

Tematika

Dátum	Témák
Szeptember 14.	Bevezetés, követelmények ismertetése Nyelvi adatok, az adatgyűjtés módszertana Korpuszok és adatbázisok Korpuszok létrehozása
Szeptember 28.	Annotáció, nyelvi elemzés Annotációs eszközök Statisztikai elemzések
Később egyeztetve	Saját korpuszépítő projektek megbeszélése Annotációs eszközök beüzemelése Java I. (alapok, Netbeans használata) Java II. (adatgyűjtés korpuszból)
Később egyeztetve	Projektbemutatók

Követelmények

Az órákon való részvétel kötelező (a tömbösítés miatt max. 1 hiányzás megengedett).

Teljesítés: a félév végére saját minikorpusz létrehozása és az abból kigyűjtött adatok rendszerezése, statisztikai elemzése. Rövid projektbemutató tartása a félév végén. Beadandó: a kész korpusz + az elemzés leírása.

Értékelés: ötfokozatú minősítés (1-5).

Ajánlott irodalom

Cor J. Koster: *Statisztika nyelvtanárok és nyelvszakos diákok számára*. Janus Pannonius Tudományegyetem, 1996.

Szirmai Mónika: *Bevezetés a korpusznyelvészetbe*. Tinta Kiadó, 2005.

http://www.korpusz.com/Monika/A_szerzorol_files/SzirmaiMonikaKorpusznyelviszeszet.pdf

Michael Hammond: *Programming for Linguists: Java Technology for Language Researchers*. Blackwell, 2002.

<http://sharif.ir/~bahram/cp4cl/Programming%20for%20Linguists.%20Java%20Technology%20for%20Language%20Researchers.pdf>