

Nyelvészinformaticus-képzés terve az ELTE Bölcsészettudományi Karán

Kis Balázs
MorphoLogic
kis@morphologic.hu

Kis Ádám
SZAK Kiadó
adam.kis@szak.hu

Az előadás az ELTE Bölcsészettudományi Karán kialakítás alatt álló informatikai tanszék egyik tervezett programját, a nyelvészinformaticus vagy számítógépes nyelvész programot mutatja be. Ennek létrehozását az indokolja, hogy az országban már számos helyen folyik nyelvtechnológiai képzés a felsőfokú informatikaoktatás keretében (ennek legteljesebb példája a Pázmány Péter Katolikus Egyetem nyelvtechnológiai képzési programja), míg hazánkban meglehetősen nehéz olyan *filológust* találni, aki kellően fel van készítve a nyelvészeti kutatás számítógépes eszközeinek alkalmazására, illetve szükség esetén létrehozására.

A tervezett nyelvészinformaticus program a bölcsészkarai képzés utolsó két (negyedik és ötödik) évében történik majd, MA (Master of Arts) programként. Ezt olyan hároméves BA (Bachelor of Arts) program előzi meg, amely tudományos kutatási asszisztensek (bölcsészinformaticusok) képzésére irányul – erről Kis Ádám egy másik előadásában esik szó részletesen.

A nyelvészinformaticus-képzés programjának kialakításában fontos szempont, hogy a tanmenet mind a meglévő hazai, mind pedig az ismert európai képzési programokkal kompatibilis legyen. A képzés kialakításakor ezért két, jelenleg is folyó képzési program struktúrájára támaszkodunk: az egyik a Magyar Tudományos Akadémia Nyelvtudományi Intézete által működtetett elméleti nyelvészet szak számítógépes nyelvészeti programja, a másik pedig a groningeni egyetem (Rijksuniversiteit Groningen, Groningen, Hollandia) bölcsészinformaticai tanszékén (Alfa-Informatica vagy Humanities Computing) működő számítógépes nyelvészeti MA (Master of Arts), illetve PhD-program.

A számítógépes nyelvészeti képzéshez szükséges alapozó tantárgyak – a nyelvészeti alapismeretek (fonológia, szintaxis), a matematikai apparátus (halmazelmélet, statisztika, formális nyelvek), illetve a programozásoktatás bevezető kurzusai – nem képezik részét az MA-programnak; az elméleti nyelvészet szak alapozó tárgyai egyfelől a megelőző hároméves BA-programba kerülnek, másfelől egy részük oktatása elvárható a többi nyelv- és nyelvészet szaktól. A groningeni képzési program eleve ezt a kettős felosztást (BA-, illetve MA-program) követi.

Az MA-képzés főbb tervezett témái (nem fontossági sorrendben, a teljesség igénye nélkül): Számítógépes nyelvmodellek; számítógépes morfológiai és szintaktikai elemzés (algoritmusok és adatstruktúrák); korpusznyelvészet; nyelvészeti erőforrások létrehozása és kezelése (lexikonok, adatbázisok, korpuszok); szemantikai adatbázisok és ontológiák; tudásábrázolás, keresés és információkivonatolás; webes alkalmazások fejlesztése; gépi fordítás és fordítástámogatás. Az előadás során bemutatjuk a képzési program egy lehetséges tanmenetét is.

Az előkészítő munkában felelős szerepe van Lóth Lászlónak, az ELTE Informatikai és Könyvtártudományi Intézete megbízott igazgatójának, illetve Bánréti Zoltánnak, az MTA Nyelvtudományi Intézete igazgatóhelyettesének, akiknek – ötleteikért, gondolataikért – ezúttal mondunk köszönetet.