

# A Nyelvtechnológiai és Élőnyelvi Osztály kutatási célkitűzései a 2012. évre

- Kiemelt feladat a Magyar Nemzeti Szövegtár új, kibővített és átdolgozott annotációval ellátott változatának előállítás és korszerű technológiákat alkalmazó közzététele. Erre a munkára a CESAR projekt keretében kerül sor. (Forrás: CESAR)
- A helyesiras.hu Nyelvi Tanácsadó Portál moduljainak kifejlesztése, a portál üzembe helyezése, a felhasználói felület elkészítése. A kezdeményezés célja egy közszolgálati feladatokat ellátó internetes portál létrehozása és üzemeltetése. A portál nyelvtechnológia eszközök és interaktív módszerek segítségével a következő területeken adhat segítséget a felhasználók számára: tulajdonnevek helyesírása, egybeírás-különírás bizonyos részterületei, keltezés és számnevek helyesírása, elválasztás, abc-rendbe sorolás, valamint a Magyar Helyesírási Szabályzat böngészhető/kereshető formában közzététele. (Forrás: központi)
- A Magyar Generatív Történeti Szintaxis projektben az év során befejeződik az ómagyar anyagok betűhű szövegének előállítása. Folytatódik a morfológiai elemző ómagyarrá való adaptálása, valamint megtörténik a már normalizált szövegek elemzése és egyértelműsítése. Elkezdődik a középmagyar szövegek feldolgozása, valamint a projekt honlapjának megtervezése. (Forrás: OTKA)
- Megnyitni a BUSZI-2 itemizált és társalgási moduljait egyaránt kutathatóvá tevő weblapot.
- Beszélt gyermeknyelvi adatbázis hangfelvételeinek elkészítése és azok lejegyzése. Szociolingvisztikai elemzésük elkezdése.
- Cél sikeresen folytatni a megkezdett CESAR, AMICUS és INNET projektet. Munkálkodni a CLARIN ERIC-hez való felzárkózáson, illetve CLARIN központtá váláson. A magyar tudományos folyóiratok hivatkozási adatbázisát létrehozó HUN-ERIH projekt sikeres teljesítése. A *slova.ee* projekt elnyesése, illetve a munkálatok elkezdése. Kis- és középvállalati kapcsolatok megszilárdítása közös alap kutatás kialakítására. Nyelvtechnológia a kulturális örökség feldolgozásában (LaTeCH-2012) workshop szervezése és lebonyolítása az EACL konferencián (2012. április, Avignon). Az AMICUS projekt keretében workshop szervezése. (Forrás: a megfelelő pályázat)
- Korpusznyelvészeti, számítógépes nyelvészeti kutatások: Kommunikatív aktusok számítógépes felismerése multimodális jellemzők alapján. Magyar nyelvű dialógusrendszer fejlesztése. Szemantikus és többnyelvű erőforrások fejlesztése a NooJ eszközre, valamint szemantikus technológiákra alapozva, ezek alkalmazása és szöveges adatokra adaptálása a kulturális örökség és a humaniorák területén. A NooJ igei és névszói lexikonok továbbfejlesztése. A Wikipedia felhasználásával automatikusan tulajdonnév-annotált korpuszt építő módszer kutatása és publikálása. A *Magyar igei szerkezetek* mintájára egy névszói kollokációs szótár megtervezése. Korábban kifejlesztett nyelvtechnológiai algoritmusok közzététele, szabadon letölthetővé tétele. Az automatikusan generált online prototípus szótárak kibővítése több szavas kifejezésekkel. (Forrás: központi illetve EFNILEX) A GeoX Kft-vel együttműködésben ebben az évben elkészül az OpinHuBank nevű erőforrás (korpusz). Célja, hogy alapot adjon magyar nyelvű számítógépes véleményelemző rendszerek fejlesztéséhez és kiértékeléséhez. A korpusz 10.000, online médiumokból (hírportálok, blogok, közösségi hálózatok) kiválasztott mondatot tartalmaz. A mondatokban automatikus eszközökkel azonosított tulajdonnevekhez humán annotátorok polaritás-értéket rendelnek (pozitív, negatív vagy semleges megítélés a mondatban). Az elkészült erőforrást mély nyelvi elemzésen alapuló véleményelemző algoritmusok

fejlesztéséhez fogjuk felhasználni. (Forrás: központi) Az eredmények publikálása hazai és nemzetközi fórumokon.

- A folyamatban lévő PhD disszertációk elkészítése (témák: *Gépi tanulási algoritmusok a magyar morfoszintaxis elsajátítására* illetve *Tulajdonnévfelismerés*).
- Számítógépes nyelvészeti támogatás nyújtása nyelvészeti kutatásokhoz, nyelvi adatok szolgáltatása, kutatói kérések teljesítése a Magyar Nemzeti Szövegtár és az ebből származtatott nyelvi erőforrások tekintetében. Informatikai támogatás számítógépes nyelvészeti munkálatokhoz, intézeti kutatási projektekhez.
- Az *Utónév portál* megtervezése és megvalósításának megkezdése. A portál lényege egy többszemponú számítógépes keresőrendszer, mely segít eligazodni a nevek között, segíti a szülők névválasztását. A keresés és rendszerezés szempontjai a korábbi és jelenleg is folyó névtani kutatások alapján alakulnak ki.

### **Nyelvművelő és Nyelvi Tanácsadó Kutatócsoport:**

- A nyelvművelés, nyelvi normativitás, helyesírás, stilisztika, retorika, névtan egyes kérdéseinek további kutatása. Az eredmények közzététele konferenciákon, folyóiratokban, kötetekben.
- A nyelvi tanácsadói tevékenységi körök folytatólagos technológiai megújítása.
- Részvétel a helyesírási közönségszolgálat új webfelületének kialakításában.
- Az utónév-szakvéleményezés, a névkutatás, a nevek gondozása területén. A honlap és a névadással kapcsolatos információk folyamatos bővítése. Az adható nevek szócikkszerű leírása, adatainak számítógépes rögzítése. Névadással kapcsolatos ismeretterjesztés. Az adható nevek kiadványszerű internetes megjelentetése. A névadás trendjeinek, jellemző motívumainak további kutatása. A névadási szabályozás kutatása. A névadás és a jog - közös kutatás jogász és közgazdász kutatókkal.
- Hazai és nemzetközi konferencián való részvétel.
- Nyelvi és nyelvtudományi ismeretterjesztés.
- Kapcsolattartás a médiával közérdekű kérdésekben.